

Providing Confidentiality and Integrity on Data Stored in Cloud Storage by Hash And Meta-Data Approach

Sadashiv Mulawad¹, Rabin Samanta², Prof. Krishanchandra³

¹Master of Computer Applications, New Horizon College Of Engineering, Karnataka, India

²Master of Computer Applications, New Horizon College Of Engineering, Karnataka, India

³Master of Computer Applications, Assistant Professor, New Horizon College Of Engineering, Karnataka, India

ABSTRACT

Cloud computing is a new computing model which is widely emerging technology in the recent years is adopted by most of the IT companies and other organizations. Cloud computing enables individuals and organizations to gain access to huge computing resources without capital investment. It does mean that users can utilize computing resources in pay per use fashion. The cost of storing large amount of data in the local storage is higher than cloud storage. However, the cloud environment is considered untrusted as it is accessed through Internet. Therefore people have security concerns on data stored in cloud environment. We proposed a new approach for securely storing our data in cloud and integrity checking mechanism by which we can check whether data integrity is preserved or not at the time of data retrieval.

Keywords : Cloud Computing, Storage, Integrity, Metadata, Encryption.

I. INTRODUCTION

1.1 Cloud Computing

Cloud computing is a recent technology that uses the Internet, central servers to organize the data and applications, which the user can access. Cloud computing allows individual users and other business peoples to use application without the necessity to install in their computer. They can access their files, which is located in other computer using Internet. This technology allows for more inefficient computing by centralizing storage, processing memory, and bandwidth.

Cloud computing comes in three categories such as Software as a Service (SaaS), Infrastructure as a service (IaaS), Platform as a Service (PaaS). The SaaS provides application software which the user can use. The PaaS provides the platform for the user to do his operation. The IaaS provide physical or virtual

devices for user. And each provides different services to the user. The cloud is available in four-deployment model namely.

1. Public Cloud
2. Private Cloud
3. Community Cloud
4. Hybrid Cloud

Public Cloud: If the cloud computing resides outside an organization and any one access it is called public cloud. Third party hosts the files. Example: Amazon Elastic Compute Cloud (EC2), Google App Engine, Windows Azure Services Platform.

Private Cloud: If the cloud computing resides inside an organization and file or application accessed through a secure network is called private cloud.

Community Cloud: Different organization with same policy and requirement share a same cloud computing and this is called community cloud

Hybrid Cloud: Combination of public, private and community cloud is called hybrid cloud.

1.2 Cloud Storage

As cloud computing is popular and in demand similarly cloud storage technology has greater demand. Cloud storage is a virtualized storage areas over a network basis .It provides services on the basis of QoS assured. Cloud storage consist of many resources but yet act as single system. It has greater fault tolerance by redundancy. As the data generated by IT sectors are dramatically growing we can't just update our hardware frequently instead we can adopt for cloud storage which is a better choice. Cloud storage can we just for different purpose just backing up our home desktop data into cloud storage or as an archive to maintain data for regulatory. Cloud storage allows user to access broad range of application and resources immediately, which are hosted by others.

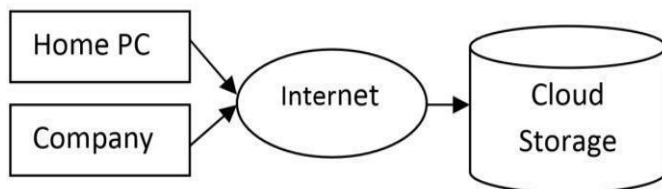


Figure 1. Sample Cloud Storage

Advantages:

- ✓ Cloud storage avoids the need to buy storage equipment.
- ✓ We have to just pay for the amount of storage we are using.
- ✓ Cloud storage allows user to access broad range of application and resources immediately, which are hosted by others.

Disadvantages:

- ✓ As data is redundant it leads to be hacked by unauthorized users.

- ✓ Cloud storage is costly for day users. Security is not guaranteed completely for our data.

1.2.1 Cloud Storage Architecture

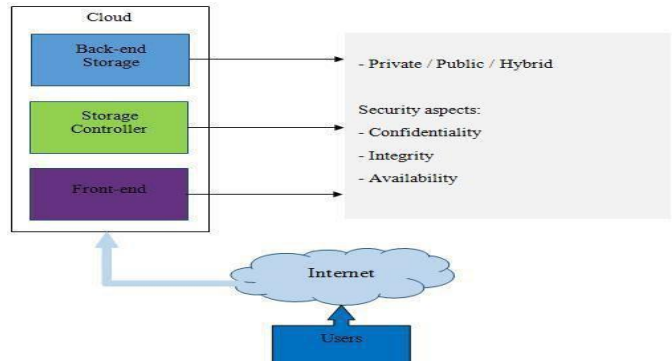


Figure 2. Generic Cloud Storage Architecture

In traditional storage systems, this API is the SCSI protocol; nonetheless in the cloud, these are evolving protocols. At this layer, there are web service, file-based Internet SCSI or iSCSI front ends. This layer is the first communication point between the user and the service provider. Users access the services using their credentials.

The midpoint component layer is called storage controller that interconnects and communicates from the front API to the backend storages.

This layer has a variety Cloud storage architectures are mainly about delivering storage on demand in a highly scalable and multi-tenant way. Basically, cloud storage architectures contain of a front end that exports an API to communicate with the backend storage of features such as replication, traditional data placement algorithms with geographical location.

Finally, the back-end consists of physical storage for data. This may be a central protocol that runs dedicated programs or a traditional back-end to the physical disks.

Advantages: Following are the advantages of cloud storage.

Usability: All cloud storage services reviewed in this topic have desktop folders for Mac's and PC's. This allows users to drag and drop files between the cloud storage and their local storage.

Bandwidth: You can avoid emailing files to individuals and instead send a web link to recipients through your email.

Accessibility: Stored files can be accessed from anywhere and anytime via Internet connection.

Disaster Recovery: It is highly recommended that businesses have an emergency backup plan ready in the case of an emergency. Cloud storage can be used as a backup plan by businesses by providing a second copy of important files. These files are stored at a remote location and can be accessed through an internet connection.

Cost Savings: Businesses and organizations can often reduce annual operating costs by using cloud storage; cloud storage costs about 3 cents per gigabyte to store data internally. Users can see additional cost savings because it does not require internal power to store information remotely.

Disadvantages: Following are the disadvantages of cloud storage.

Usability: Be careful when using drag/drop to move a document into the cloud storage folder. This will permanently move your document from its original folder to the cloud storage location. One can do a copy and paste instead of drag/drop, if you want to retain the document's original location in addition to moving a copy onto the cloud storage folder.

Bandwidth: Several cloud storage services have a specific bandwidth allowance. If an organization surpasses the given allowance, the additional charges could be significant. However, some providers allow unlimited bandwidth. This is a factor that companies should consider when looking at a cloud storage provider.

Accessibility: If you have no internet connection, you have no access to your data.

Software: If you want to be able to manipulate your files locally through multiple devices, you'll need to download the service on all devices.

Data Security: There are concerns with the safety and privacy of important data stored remotely. The possibility of private data commingling with other organizations makes some businesses uneasy.

1.3 Security Requirements in Cloud Storage

Security is the protection of information assets through the use of technology, processes, and training. Cloud storage is a service that includes inherent vulnerabilities, but these have never dissuaded users from taking advantage of its economies and flexibilities.

With adoption of a cloud model, users lose control over physical security. Users raised concerns whether their data are accessed by unauthorized person since there are many user sharing the resources over the cloud.

Sharing the cloud with other users possesses risks and concerns over security. Security overall covers mainly three aspects: Confidentiality, Integrity and Availability (CIA). These aspects are the topmost considerations in designing a security measure to ensure maximum protection.

Confidentiality: Protecting data and information from disclosure to unauthorized person.

Integrity: Protecting data and information from being modified by unauthorized person. **Availability:** Authorized people are able to access and use data and information whenever require.

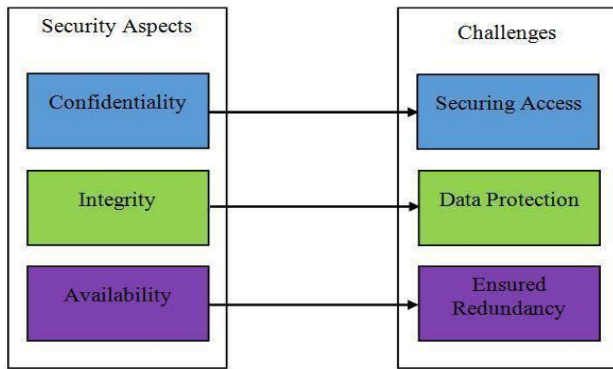


Figure 3. Cloud Storage Security Aspects and Challenges

1.3.1 Securing Access

Securing access to protected data and information is restricted to certain level of user authorized to access it. This requires mechanisms to be in place to control the access of protected data.

The sophistication of the access control mechanisms should be in parity with the value of the information being protected; the more sensitive or valuable the information the stronger the control mechanisms need to be. The foundation on which access control mechanisms are built starts with authentication, authorization and encryption.

1.3.2 Data Protection

Cloud storage that holds data and information on the cloud is obligated on data integrity. Data integrity depends on the assurance pursued by the user that data are unaltered on the provider infrastructure. Data integrity threats involve both malicious third party occurrences and hosting infrastructure weaknesses.

Protecting data from loss and leakage involves integrity of many parties involved in providing the resources. Some schemes and mechanism are needed to ensure the data and information kept on the cloud is unaltered or removed. It is suggested to practice auditing techniques such as proof-of-irretrievability and proof-of-data possession to enable verification.

Another issue was also raised as users found that even if they accidentally deleted their data, the provider can restore a backup file. This means the data is still kept by the provider. In FADE [7], a secure overlay with file assured deletion is presented. It is a policy-based scheme that reliably removes files and withdraws file-access policies on it. Thus, even if a data is restored by a provider, the file is restricted from read/write as the file-access policies are revoked.

1.3.3 Ensured Redundancy

Data availability is critical. Cloud storage providers must guarantee that the data will always be available autonomously regardless of hardware failures, corrupted physical disks or downtime. Hardware failures can happen at any time. This includes failures caused by environmental failures such as a natural disaster, flood or even fire.

A hardware design should be built on a basis of having redundancy and minimum single points of failure. At the design phase, the analyst creates a physical hardware map that shows all the connection points for server, storage, network and software.

1.4 Meta-data

Metadata is “data [information] that provides information about other data”. Three distinct types of metadata exist:

Descriptive metadata, Structural metadata, and Administrative metadata.

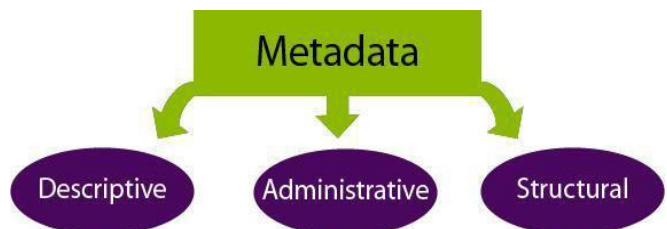


Figure 4. Metadata and types

1.4.1 Descriptive metadata

It describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.

1.4.2 Structural metadata

It is a metadata about containers of metadata and indicates how compound objects are put together, for example, how pages are ordered to form chapters.

1.4.3 Administrative metadata

It provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it.

II. LITERATURE SURVEY

2.1 Survey on Confidentiality and Integrity Providing Methods in Cloud Storage

1) Title: Security Issues for Cloud Computing

Abstract:

Security issues for the cloud computing is the vast accept to come through, but there are many issue like Integrity, Confidentiality, Availability, Map reduce. Here the discussion is related to the trending issues and how to build the trusted application from untrusted component of secure cloud computing

Overview:

There are numerous security issues for cloud computing as it encompasses many technologies including networks, database, operating system, virtualization, load balancing, concurrency control, Memory management. Therefore, security issues for many of these systems in a cloud have to be secure. Furthermore, virtualization paradigm in cloud computing results in several security concerns. For example, mapping the virtual machines to the physician machines has to be carried out securely. Data security involves for data sharing.

Conclusion:

The main goal is to securely store and manage the data that is not controlled by the owner of the data. Building trust application from the untrusted components will be the major aspects with respect to cloud security.

2) Title: HAIL: A High-Availability and Integrity Layer for Cloud Storage

Abstract:

HAIL (High-Availability and Integrity Layer), a distributed cryptographic system that allows a set of servers to prove to a client that a stored file is intact and retrievable. HAIL is efficiently computable by servers and highly compact typically tens or hundreds of bytes, irrespective of file size. HAIL improves on the security and efficiency of existing tools, like Proofs of Retrievability (PORs) deployed on individual servers. We propose a strong, formal adversarial model for HAIL, and rigorous analysis and parameter choices. We show how HAIL improves on the security and efficiency of existing tools, like Proofs of Retrievability (PORs) deployed on individual servers.

Overview:

In HAIL, a client distributes a file F with redundancy across n servers and keeps some small(constant) state locally. The goal of HAIL is to ensure resilience against a mobile adversary. This kind of powerful adversary can potentially corrupt all servers across the full system lifetime. There is one important restriction on a mobile adversary, though: It can control only b out of the n servers within any given time step. We refer to a time step in this context as an epoch.

In each epoch, the client that owns F (or potentially some other entity on the client's behalf) performs a number of checks to assess the integrity of F in the system. If corruptions are detected on some servers, then F can be reconstituted from redundancy in

intact servers and known faulty servers replaced. Such periodic integrity checks and remediation are an essential part of guaranteeing data availability against a mobile adversary: Without integrity checks, the adversary can corrupt all servers in turn across $[n/b]$ epochs and modify or purge F at will.

Conclusion:

Author have proposed HAIL, a high-availability and integrity layer that extends the basic principles of RAID into the adversarial setting of the Cloud. HAIL is a remote file integrity checking protocol that offers efficiency, security, and modelling improvements over straightforward multi-server application of POR protocols and over previously proposed, distributed file availability proposals. Through a careful interleaving of different types of error-correcting layers, and inspired by proactive cryptographic models, HAIL ensures file availability against a strong, mobile adversary.

Title: Identity Based Approach for Cloud Data Integrity in Multi-Cloud Environment

Abstract:

Cloud computing is a new computing model which gain access to huge computing resources without capital investment. With virtualization technology the commoditization of computing resources has become a reality. However, the cloud environment is considered in trusted as it is accessed through Internet. Many techniques have been proposed in the literature for ensuring data storage security in cloud computing. There propose a model for cloud Data integrity in the distributed multi-cloud environment. The proposed method is testing using a prototype application which demonstrates the proof of concept.

Conclusion:

The main focus is on data integrity in cloud computing environment. By distributing the block-tag pairs to the various cloud servers and when the

combiner get the retrieval request it get the challenge and that is distributed among the servers and the server responses are aggregated prior to sending response back to client. The client and cloud servers do their respective job while the proposed model is capable of ensuring data integrity in distributed environment.

Title: PORs: Proofs of Retrievability for Large Files

Abstract:

A POR scheme enables an archive or back-up service (prover) to produce a concise proof that a user (verifier) can retrieve a target file F . Explore POR protocols here in which the communication costs, number of memory accesses for the prover, and storage requirements of the user (verifier) are small parameters essentially independent of the length of F . Existing cryptographic techniques help users ensure the privacy and integrity of files they retrieve. The goal of a POR is to accomplish these checks without users having to download the files themselves. A POR can also provide quality-of-service guarantees, i.e., show that a file is retrievable within a certain time bound.

Conclusion:

The POR protocol is design to protect a static archived file F . Any natively performed, partial updates to F would undermine the security guarantees of our protocol. For example, if the verifier were to modify a few data blocks (and accompanying error correcting blocks), the archive could subsequently change or delete the set of modified blocks with (at least temporary) impunity, having learned that they are not sentinels.

1. Title: Providing Security and Integrity for Data Stored In Cloud Storage

Abstract:

A scheme by which there provided a secure saving of confidential data in cloud storage in an efficient manner which requires low computational power and time and disallowing hacker from penetrating into private data storage. So there provided a simple and easy integrity checking mechanism when compared to other already present one by which verification can be done whether data is not corrupted and deleted or modified ours is an efficient. Integrity checking mechanism is simple that it does not take more computational power. This mechanism even prevents the TPA who maintains our data in cloud storage from editing our file.

Overview:

In the proposed system encrypting the full file is eliminated. Instead we encrypt only some bits of each data block and thus client computation overhead is reduced. In the model irrespective of data file size only one cryptographic key is used. The client does not store any data at his side. While uploading a file it pre-processes the user file and create meta data later which is appended at the back of the file. The Meta data is later used for checking if integrity is preserved or not.

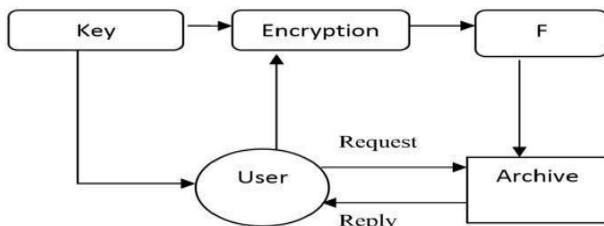


Figure 5. Overall system architecture

The proposal consists of two modules.

1. Providing Security

The data that the user stores in the cloud storage should be secure so that it prevents intruders from accessing our private data. To provide security we use a security key which is automatically generated

for each unique user and we use RSA encryption algorithm to encrypt the file and store it. It is a public key encryption algorithm, which eliminates the need to send our secret key over the network.

The public key is shared but the private key is not shared. The sender encrypts the file or data that is to be stored in the cloud storage using the Third party auditor public key. So the receiver with the particular private key can only decrypt the file.

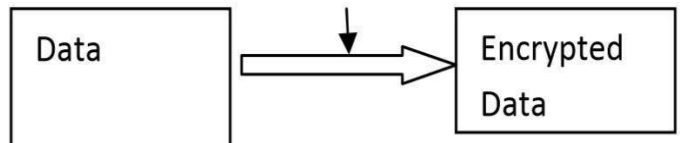


Figure 6. User Side Encryption

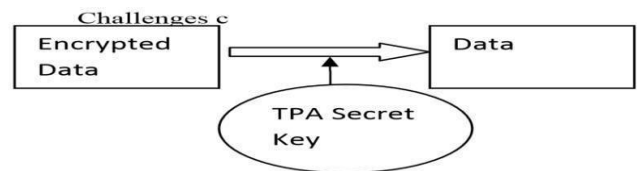


Figure 7. TPA side Decryption

2. Integrity Check

Integrity modules check the correctness of the user data by verifying the Meta data appended at back of the file. These are the following steps involved in finding the integrity of the user data

- a) Generation of Meta data.
- b) Encrypting the Meta data.
- c) Appending the Meta data.
- d) Verification Phase.

Conclusion:

Method to save our data in the cloud storage secure and provide an integrity check for our data to verify if integrity is preserved or not while we retrieve our data is been proposed and by creating meta-data an integrity of the stored data is been checked. It uses less computational power and processing time.

Problem Finding

Problem Identification

Cloud computing is becoming more popular similarly cloud storage has greater demand and cloud storage becomes a tool of choice for data storage. Information available in cloud storage only has value if it is correct. Information that has been tampered with could prove costly. Protecting data and information from being modified by unauthorized person is more important in cloud storage.

By considering literature review, we found that there are different techniques available for checking the integrity of data stored in cloud. We also found that above techniques or mechanism are only checking integrity of data stored in cloud. So, we think to provide both security as well as integrity checking mechanism simultaneously using security algorithm.

We proposed a new approach for securely storing our data in cloud storage and integrity checking mechanism by which we can check whether data integrity is preserved or not.

III. PROPOSED SOLUTION

We have proposed an approach to provide data security and data integrity in an efficient manner for data stored in cloud. In this approach, following process is carried out for storing data and checking integrity of data stored in cloud. The process is depicted in the figure 8 and figure 4.

Storing Encrypted file, hash file and Meta-data in cloud

Figure 8, shows the process of storing data securely in the cloud storage.

- ✓ In our proposed approach to store the local file securely in the cloud, first we encrypt the local file using AES-256

encryption algorithm and create the meta-data of that file.

- ✓ After encryption, the encrypted file and the Meta-data are stored in the Cloud.
- ✓ To provide integrity checking we generate the hash of the local file using SHA256 hashing algorithm.
- ✓ Then the hashed file is also stored in cloud.

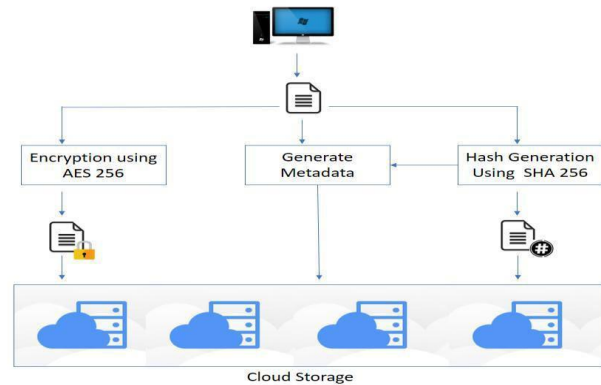


Figure 8. Storing Encrypted file, hash file and Meta-data in cloud

Integrity Checking

Figure 9, shows the process of integrity checking of the data stored in cloud.

- ✓ While retrieving the file we check the Meta-data of the stored file.
- ✓ Then we retrieve the encrypted file from the storage and decrypt the file using AES-256
- ✓ After decrypting the file we generate the hash of decrypted file using SHA256 hashing algorithm.
- ✓ Now, we download the hash file which is available in cloud and compare both hash file.
- ✓ If both the files hash are same and the meta-data value matches then integrity is preserved otherwise integrity is not preserved.

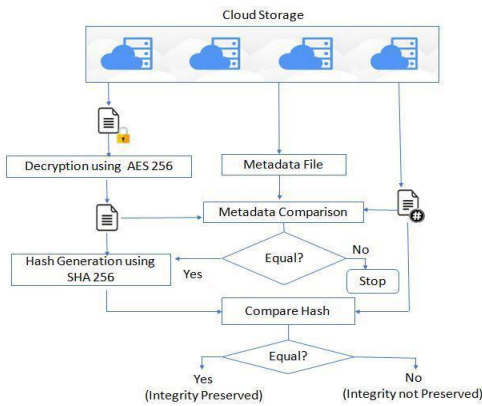


Figure 9. Integrity checking

IV. EXPERIMENTAL SETUP AND RESULTS

Experimental Setup

We implement our proposed approach using Amazon S3 live cloud. The many more experiments are carried out in the machine that have 4 GB RAM, 500 GB HDD, 2.50GHz CPU and Intel(R) Core(TM) i5-3210M processor. The machine is equipped with the Windows 10 Pro 64-bit operating system. The software which is used for the implementation is NetBeans IDE 8.0.2, jdk1.8.0_31. We have implemented our proposed approach using JAVA programming language.

Amazon S3 is a Simple Storage Service provided by Amazon where we can store our data over the internet. Amazon S3 has a humble web services interface that we can use to store and regain our data, at any time and from anywhere using internet. Amazon provides highly scalable, reliable, fast and inexpensive data storage infrastructure to the users. It provides one year free trial base access to the developers, after that they can renew it.

Key Concept of Amazon S3

Bucket: A bucket is a object of container which is stored in Amazon S3. E.g. Object named images/imagename.jpeg is stored in the bucket name bucket, then it is

addressable using the URL :http://bucketname.s3.amazonaws.com/images/imagename.jpeg

Object: Objects are the entities which are stored in Amazon S3. It consists of object data and metadata. The metadata is a set of name-value pair that describes the object.

Key: A key for an object has a unique identifier within a bucket. Every object has one key in the bucket. Because the combination of a bucket, key, and version ID each has uniquely identification for each object.

Regions: To store the Bucket we can choose the different geographical region of Amazon S3. We might choose a region to

Results

In this section, we present the experimental results of our proposed approach. The result of the proposed approach is depicted in the figure 10, figure 11, figure 12 and figure 13.

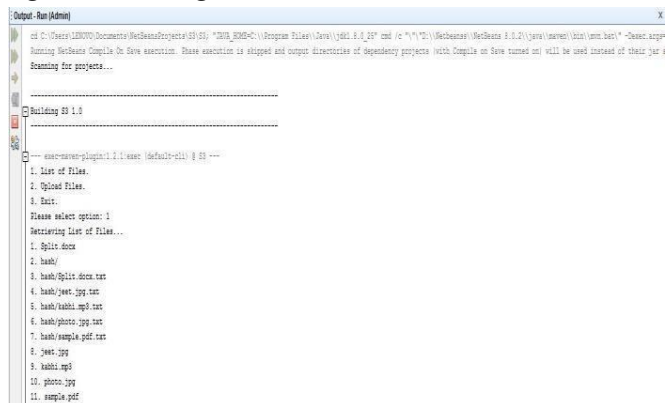
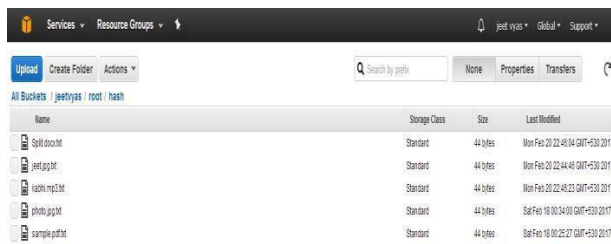


Figure 10. Screenshot for List of file uploaded

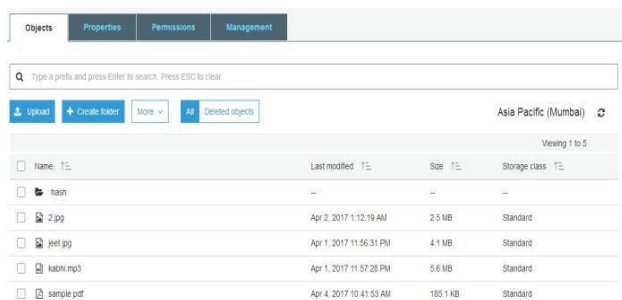
Figure 10, shows the process of uploading file from local storage to Amazon S3. For uploading a file, first we have to specify the file path which has to be uploaded. Then, our proposed approach create the hash of that file, metadata of the file and also encrypt that file. After that files are uploaded in cloud.



Name	Storage Class	Size	Last Modified
Split.docx	Standard	44 bytes	Mon Feb 20 22:42:44 GMT+05:30 2017
jeet.jpg	Standard	44 bytes	Mon Feb 20 22:44:48 GMT+05:30 2017
kabhi.mp3	Standard	44 bytes	Mon Feb 20 22:49:23 GMT+05:30 2017
photo.jpg	Standard	44 bytes	Sat Feb 18 00:34:00 GMT+05:30 2017
sample.pdf	Standard	44 bytes	Sat Feb 18 00:25:27 GMT+05:30 2017

Figure 11. Screenshot of Uploaded file in AWS S3

Figure 11, shows the list of files which are stored in cloud. It also shows the path of the files stored in the Amazon S3.



Name	Last modified	Size	Storage class
hash
2.jpg	Apr 2, 2017 1:12:19 AM	2.5 MB	Standard
jeet.jpg	Apr 1, 2017 11:56:31 PM	4.1 MB	Standard
kabhi.mp3	Apr 1, 2017 11:57:28 PM	5.6 MB	Standard
sample.pdf	Apr 4, 2017 10:41:53 AM	185.1 KB	Standard

Figure 12. Screenshot of Uploaded hash file in AWS S3

Figure 12, shows the list of hash files which are stored in cloud. It also shows the path of the hash files stored in the Amazon S3.

Figure 12, shows the metadata of the stored file, here the metadata object are.

- ✓ Length of file (bytes)
- ✓ Hash of file (32 byte)
- ✓ Last modified date (milliseconds)

V. FUTURE WORK AND CONCLUSION

Conclusion

In this report, there have presented an approach for securely storing our data in cloud. We have check the integrity of the file stored in cloud by integrity checking mechanism. In this integrity checking mechanism, we check whether data integrity is preserved or not through hashing technique and metadata approach. We also encrypt the file and then store in cloud for providing more security to the data. Thus we can say that, this approach will enhance the security of the data which is stored in

cloud. As, we store the encrypted file in cloud and also provide integrity checking mechanism.

Future Work

In future work, we may make approach that will generate a log file, which can record the information about the file access in cloud. The log file will store the information about the access time of file. So, that we can know whether file is accessed by some unauthorized person or not.

VI. REFERENCES

- [1]. W.O. Quine Word and object. Cambridge, M.I.T. Press, Mass., 1960.
- [2]. A. Thayse and P. Gribomont etc. Approche logique de l'intelligence artificielle. 1 De la logique classique a la programmation logique. Bordas, Paris, 1988.
- [3]. J.-L. Lauriere Intelligence artificielle. resolution de problemes par l'homme et la machine. Eyrolles, Paris, 1987.
- [4]. J. Barwise (Ed) Handbook of Mathematical Logic, North-Holland, Amsterdam, 1977.
- [5]. B.N. Pyatnitsyn Towards a Problem of induction-deduction relations. "Methods of Logical Analysis". Nauka, Moscow, 1977.