Volume 4, No. 4, March-April 2013

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

REVIEW ARTICLE

Available Online at www.ijarcs.info

A Review On A Mechanism Of Ensuring Data Storage Security In Cloud Computing

Gayatri V. Shastri
Department of information technology
JDIET Yavatmal, India
shastri.gayatri13@gmail.com

Charul D.Akhade
Department of information technology
JDIET Yavatmal, India
charul.akhade@yahoo.in

Dnyaneshwar P.Shirame
Department of information technology
JDIET Yavatmal, India
d2shrirame@gmail.com

Prof.Sandip T.Dhagdi
Department of information technology
JDIET Yavatmal, India
sandip.yml@gmail.com

Abstract: We know that Cloud Computing has been visualizing as the next generation architecture of IT Enterprise. The Cloud make a user to store the data remotely in cloud storage and remove the burden of storage and maintenance. In contrast to traditional solutions, where the information technology services are under proper physical, logical and personnel controls, Cloud Computing moves the application software and databases to the large data centers, where the management of the data and services may not be fully reliable.

Keywords: Cloud storage security, Cloud computing, Network security, Encryption, Decryption

I. INTRODUCTION

This paper examines data security in cloud computing along with data protection methods and approaches. Cloud data security involves far more than simply data encryption. The requirements for data security vary depending on your tolerance for risk as well as three service models (SaaS, PaaS, and IaaS), as in Figure 1.

A. Infrastructure as a Service (IaaS):

IaaS is the lowest layer and refers to the operating system and its virtualization. Different users will be allocated with dedicated CPU and memory virtually depending upon their accountability [3][4].

B. Platform as a Service (PaaS):

PaaS is the middle layer and refers to the programming models & environment, execution method, database, and web server.

C. Software as a Service (SaaS):

SaaS is the topmost layer and is the most important from user's perspective. It features a complete application offered as service on demand. This where the cloud providers install and operate application software in the cloud. The cloud users can access these software from cloud clients and are prevented from directly accessing infrastructure of the cloud and platform on which the application is running i.e., the users here would only be accessing the software online and storing the data back in the cloud eradicating the need of locally installing application on user's machine, thus lessening burden for software maintenance on the customer. This attribute offers ease of maintenance and support for various levels of user accountability [5].

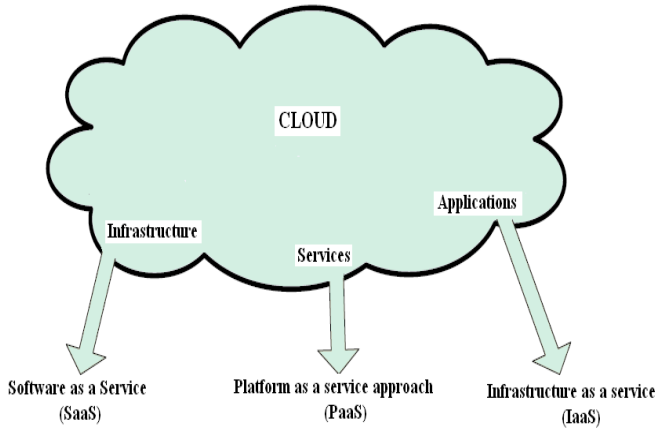

Figure 1: Cloud Layered Model

Meeting the requirements for cloud data security entails applying existing security techniques and following sound security practices. To be effective, cloud data security depends on more than simply applying appropriate countermeasures. Taken collectively, countermeasures must comprise a resilient mosaic that protects data at rest as well as data in motion. While the use of encryption is a key component for cloud security, even the most robust encryption is pointless if the keys are exposed or if encryption endpoints are insecure.

Rest of paper is organized as follows:

Part1 provides an overview of cloud data security issues, including data control and data types. **Part2** considers cryptographic techniques and common mistakes using data encryption for data stored on the Internet. **Part3** briefly reviews data protection methods and any unique aspects that may apply when they are deployed in a cloud. **Part4**describes cloud data storage which includes Cloud Lock-in (The Roach Motel Syndrome), Avoiding Cloud Lock-in (The Roach Motel Syndrome).

II. PART 1 - DATA SECURITY IN CLOUD COMPUTING: OVERVIEW

It is understandable that prospective cloud adopters would have security concerns around storing and processing sensitive data in a public or hybrid or even in a community cloud. Compared to a private data center, these concerns usually center on two areas:

- **a.** Decreased control by the owning organization when data is no longer managed within an organization's premises Securing the Cloud.
- **b.** Concern by the owning organization that multitenancy clouds inherently pose risks to sensitive data

In both cases, the potential risk of data exposure is real but not fundamentally new. This is not to say that cloud computing does not bring unique challenges to data security.

A. Control over Data and Public Cloud Economics:

In contrast to use of a public cloud, maintaining organizational physical control over stored data or data as it traverses internal networks and is processed by on-premises computers does offer potential advantages for security. But the fact is that while many organizations may enforce strict on-premises-only data policies, few organizations actually follow through and implement the broad controls and the disciplined practices that are necessary to achieve full and effective control.

So, additional risks may be present when data doesn't physically exist within the confines of an organization's controlled facility—this is not necessarily the security issue that it may appear to be. To begin, achieving the potential advantages with on-premises data requires that your security strategy and implementation deliver on the promise of better security. This is illustrating in Figure 2. Note that this situation is a function of generally available and anticipated offerings in the public cloud space. Quite likely, this will change as security becomes more of a competitive discriminator in cloud computing.

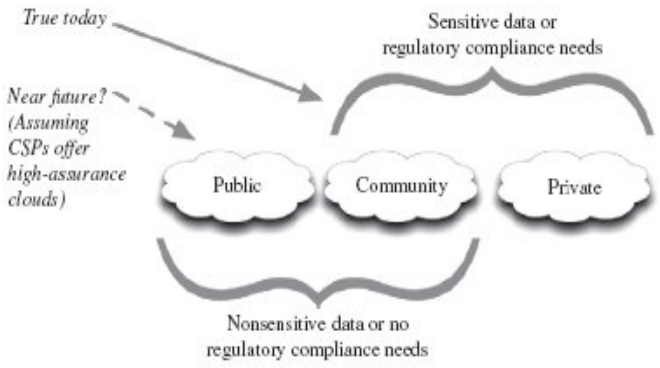

Figure 2: Meeting security needs

One can easily imagine future high-assurance public clouds that charge more for their service than lower-assurance public clouds do today. We might also expect that some higher-assurance clouds would limit access by selective screening of customers based on entry requirements or regulation. Limiting access to such a cloud would reduce risk—not eliminate it—by limiting access if screening is effective.

B. Data at Rest and in Motion:

Data at rest refers to any data in computer storage, including files on an employee's computer, corporate files on a server, or copies of these files on off-site tape backup. Protecting data at rest in a cloud is not radically different than protecting it outside a cloud. Generally speaking, the same principles apply. As discussed in the previous section, there does the potential for added risk as the data own enterprise does not physically control the data. But as also noted in that discussion, the trick to achieving actual security advantage with on-premises data is following through with effective security.

Data in motion refers to data as it is moved from a stored state as a file or database entry to another form in the same or to a different location. Any time you upload data to be stored in the cloud, the time at which the data is being uploaded data is considered to be data in transit. Data in motion can also apply to data that is in transition and not necessarily permanently stored. Your username and password for accessing a Web site or authenticating yourself to the cloud would be considered sensitive pieces of data in motion that are not actually stored in unencrypted form. Because data in motion only exists as it is in transition between points—such as in memory (RAM) or between end points—securing this data focuses on preventing the data from being tampered with as well as making sure that it remains confidential. One risk has to do with a third party observing the data while it was in motion. But funny things happen when data is transmitted between distant end points, to begin with packets may be cached on intermediate systems, or temporary files may be created at either end point. There is no better protection strategy for data in motion than encryption [2].

III. PART 2 - DATA SECURITY IN CLOUD COMPUTING: DATA ENCRYPTION APPLICATIONS AND LIMITS

In a recent article of Bruce Schneier discussed how the information age practice of encrypting data at rest deviates from the historical use of cryptography for protecting data while it is communicated or in transit. One of Schneier's key points is that for data in motion, encryption keys can be ephemeral, whereas for data at rest, keys must be retained for as long as the stored data is kept encrypted. As Schneier points out, this does not reduce the number of things that must be stored secretly; it just makes those things smaller (the size of a key is far smaller than a typical data file).

A. Overview of Cryptographic Techniques:

Cryptography is a complex and esoteric field. In modern times, cryptography has expanded from protecting the confidentiality of private communications to including techniques for assuring content integrity, identity authentication, and digital signatures along with a range of secure computing techniques. Going one level deeper in our background treatment of cryptography, for the purpose of this book, there are four basic uses of cryptography:

a. Block Ciphers:

These take as input a key along with a block of plaintext and output a block of cyphertext. Because messages are generally larger than a defined block, this method requires

some method to associate or knit together successive cyphertext blocks.

b. Stream Ciphers:

These operate against an arbitrarily long stream of input data, which is converted to an equivalent output stream of cyphertext.

c. Cryptographic Hash Functions:

Hash functions take an arbitrarily long input message and output a short, fixed length hash. A hash can serve various purposes, including as a digital signature or as a means to verify the integrity of the message.

d. Authentication Cryptography:

It is also widely used within authentication and identity management systems.

B. Common Mistakes or Errors with Data Encryption:

The most common mistakes or errors include:

a. Failing to use cryptography when cryptographic security is a viable option. Most likely, all payloads should be encrypted by default.
b. Thinking you can implement an existing cryptographic algorithm (when you shouldn't). Instead of reinventing the wheel, use a proven implementation.
c. Storing keys with data. This error is so profoundly egregious, one would expect not to need mentioning it except (sadly) there are reports that it happens time and time again
d. Sending sensitive data in unencrypted e-mail. Sending passwords, PINs, or other account data in unencrypted e-mail exposes that data in multiple places.
e. The bus test. If critical keys for the organization are kept by only one or a few individuals, how will your organization recover if these individuals suffer a disaster such as being hit by a bus?

IV. PART 3 - CLOUD DATA PROTECTION METHODS

When it comes to cloud data protection methods, no particularly new technique is required. Protecting data in the cloud can be similar to protecting data within a traditional data center. Authentication and identity, access control, encryption, secure deletion, integrity checking, and data masking are all data protection methods that have applicability in cloud computing. This section will briefly review these methods and will note anything that is particularly unique to when these are deployed in a cloud.

A. Authentication and Identity:

a. Authentication of users takes several forms, but all are based on a combination of authentication factors: something an individual knows (such as a password), something they possess (such as a security token), or some measurable quality that is intrinsic to them (such as a fingerprint). Single factor authentication is based on only one authentication factor. Stronger authentication requires additional factors; for instance, two factor authentications is based on two authentication factors (such as a pin and a fingerprint).
b. Maintaining confidentiality, integrity, and availability for data security is a function of the correct application and configuration of familiar network, system, and application security mechanisms at various levels in the cloud infrastructure. Among these mechanisms are a broad range of components that implement authentication and access control. Authentication of users and even of communicating systems is performed by various means, but underlying each of these is cryptography [2].

B. Access Control Techniques:

Access controls are generally described as either unrestricted or non-discretionary, and the most common access control models are:

a. Discretionary Access Control (DAC) In a system, every object has an owner. With DAC, access control is determined by the owner of the object who decides who will have access and what privileges they will have. Permission management in DAC can be very difficult to maintain; furthermore, DAC does not scale well beyond small sets of users.
b. Role Based Access Control (RBAC) Access policy is determined by the system. Where with MAC access is based on subject trust or clearance, with RBAC access is based on the role of the subject. A subject can access an object or execute a function only if their set of permissions—or role—allows it.
c. Mandatory Access Control (MAC) Access policy is determined by the system and is implemented by sensitivity labels, which are assigned to each subject and object. A subject's label specifies its level of trust, and an object's label specifies the level of trust that is required to access it. If a subject is to gain access to an object, the subject label must dominate—be at least as high as—the object label.

Following Figure3 depicts this point by contrasting MAC with discretionary access controls (DAC) and role-based access controls (RBAC).

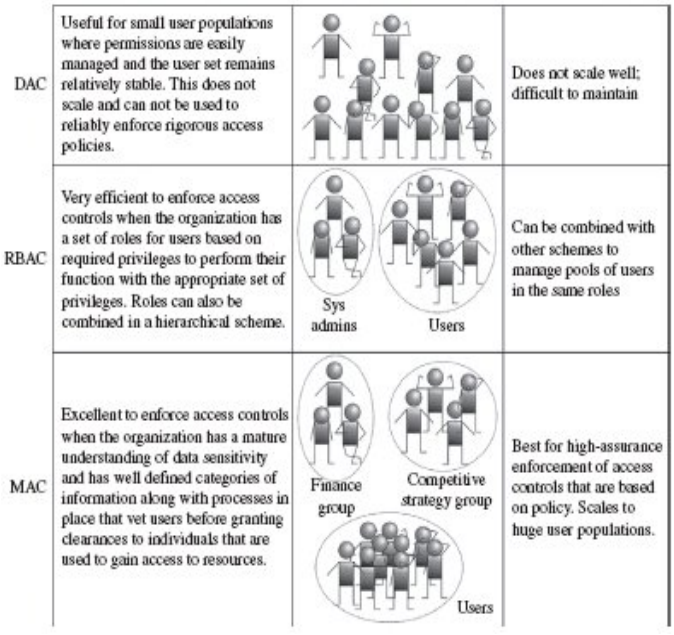

Figure 3: MAC scales better for data security than other schemes do.

C. Encryption:

According to Vic (J.R.) Winkler Encryption for Data at rest and in motion is described as follows:

a. Application of Encryption for Data at Rest:

Encryption is a key component to protect data at rest in the cloud. Employing appropriate strength encryption is important: Strong encryption is preferable when data at rest has continuing value for an extended time period. If such long-term value encrypted data is obtained by a third party and if they have an extensive period of time to break or *crack* the encryption, then the reward can be well worth the effort.

There are multiple ways of encrypting data at rest. Following is an outline of various forms of encryption that serve as protection methods for securing data at rest in the cloud [6].

(a). Full Disk: Encryption of data at the disk level—the operating system, the applications in it, and the data the applications use are all encrypted simply by existing on a disk that is encrypted. This is a brute-force approach to encrypt data since everything is encrypted, but this also entails performance and reliability concerns. If encryption is not done at the drive hardware level, then it can be very taxing on a system in terms of performance. Another consideration is that even minor disk corruption can be fatal as the OS, applications, and data.

(b). Directory Level (or File system): In this use of encryption, entire data directories are encrypted or decrypted as a *container*. Access to files requires use of encryption keys. This approach can also be used to segregate data of identical sensitivity or categorization into directories that are individually encrypted with different keys.

(c). File Level: Rather than encrypting an entire hard drive or even a directory, it can be more efficient to encrypt individual files.

(d). Application Level: The application manages encryption and decryption of application-managed data.

b. Application of Encryption for Data in Motion:

The two goals of securing data in motion are preventing data from being tampered with (integrity) and ensuring that data remains confidential while it is in motion. Other than the sender and the receiver, no other party observing the data should be able to either make sense of the data or alter it. The most common way to protect data in motion is to utilize encryption combined with authentication to create a conduit in which to safely pass data to or from the cloud.

Encryption is used to assure that if there was a breach of communication integrity between the two parties that the data remains confidential. Authentication is used to assure that the parties communicating data are who they say they are. Common means of authentication themselves employ cryptography in various ways. Transferring data via programmatic means, via manual file transfer, or via a browser using HTTPS, TLS, or SSL are the typical security protocols used for this purpose. A PKI is used to authenticate the transaction (trusted root CAs), and encryption algorithms are used to protect the payload.

D. Secure Deletion:

When it is time to delete sensitive or valuable data in a cloud, it is important to understand how that data is deleted. The U.S. Department of Defense has an excellent and well accepted definition illustrating the two key aspects of data deletion, as stated in National Industrial Security Program Operating Manual:

a. Clearing. Clearing is the process of eradicating the data on media before reusing the media in an environment that provides an acceptable level of protection for the data that was on the media before clearing. All internal memory, buffer, or other reusable memory shall be cleared to effectively deny access to previously stored information.
b. Sanitization. Sanitization is the process of removing the data from media before reusing the media in an environment that does not provide an acceptable level of protection for the data that was in the media before sanitizing. IS resources shall be sanitized before they are released from classified information controls or released for use at a lower classification level [7]-[10].

E. Data Masking:

Data masking is a technique that is intended to remove all identifiable and distinguishing characteristics from data in order to render it anonymous and yet still be operable. This technique is aimed at reducing the risk of exposing sensitive information. Data masking has also been known by such names as data obfuscation, de-identification, or depersonalization. These techniques are intended to preserve the privacy of records by changing the data so that actual values cannot be determined or re-engineered.

A common data masking technique involves substitution of actual data values with keys to an external lookup table that holds the actual data values. In operation, such resulting masked data values can be processed with lesser controls than if the original data was still unmasked.

But data masking must be performed carefully, or the resulting masked data can still reveal sensitive data. By example, if you mask salary data in an HR database by tokenizing what originally were employee names with name look up keys, the highest salary will probably be the CEOs. By using simple analysis techniques and methodically cross-referencing partially masked records with other employee information, more may be inferred by a process of elimination than should be.

V. PART 4 - CLOUD DATA STORAGE

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper.

Among other advances, cloud computing has brought advantages in the form of online storage. In this section, we are referring to Storage-as-a-Service. The range of service offerings in this space is remarkable, and they are continuing to grow.

Data security for such a cloud service encompasses several aspects including secure channels, access controls, and encryption. And, when we consider the security of data in a

cloud, we must consider the security triad: confidentiality, integrity, and availability. In the cloud storage model, data is stored on multiple virtualized servers. Physically the resources will span multiple servers and can even span storage sites. A common aspect of many cloud-based storage offerings is the reliability and availability of the service. Following figure 4. depicts an abstracted view of how many individual disks in many aggregated storage devices are composed into a virtualized unit of storage.

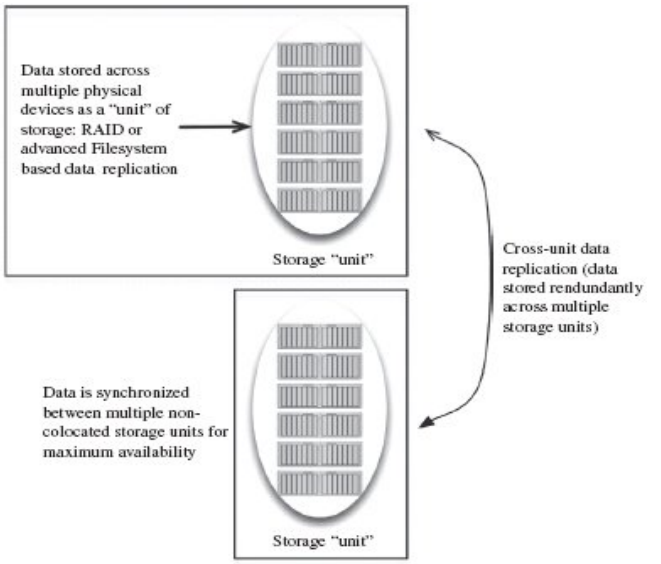

Figure 4: Cloud storage: replication and availability.

Replication of data is performed at a low level by such mechanisms as RAID or by a file system. One such file system is ZFS, which was designed by Sun Microsystems as both a file system and a volume manager. ZFS supports high storage capacities and performs numerous security relevant functions including copy-on-write cloning and continuous integrity checking along with automatic repair.

One of the more recent trends in online cloud-based storage is the cloud storage gateway. Several vendors offer such solutions that are generally implemented as an appliance that resides onsite at the customer premises. These appliances can provide multiple features, including:

a. Translation of client-used APIs and protocols (such as REST or SOAP) to those that are used by cloud-based storage services (such as NFS, ISCSI, or Fibre Channel). The goal is to enable integration with existing applications over standard network protocols.
b. Backup and recovery capabilities that work with in-cloud storage.
c. Onsite encryption of data that keeps keys local to the onsite appliance.
d. The vendors and products in this space include Gladnet, Nasuni Cloud Storage Gateway, StorSimple, and Emulex. The product and solutions that are available are seeing rapid changes and new functionality.

Following Figure 5. Depicts a typical cloud storage gateway application as it is used to augment local storage by acting as an onsite secondary copy and as an intermediary to the CSP storage service [8].

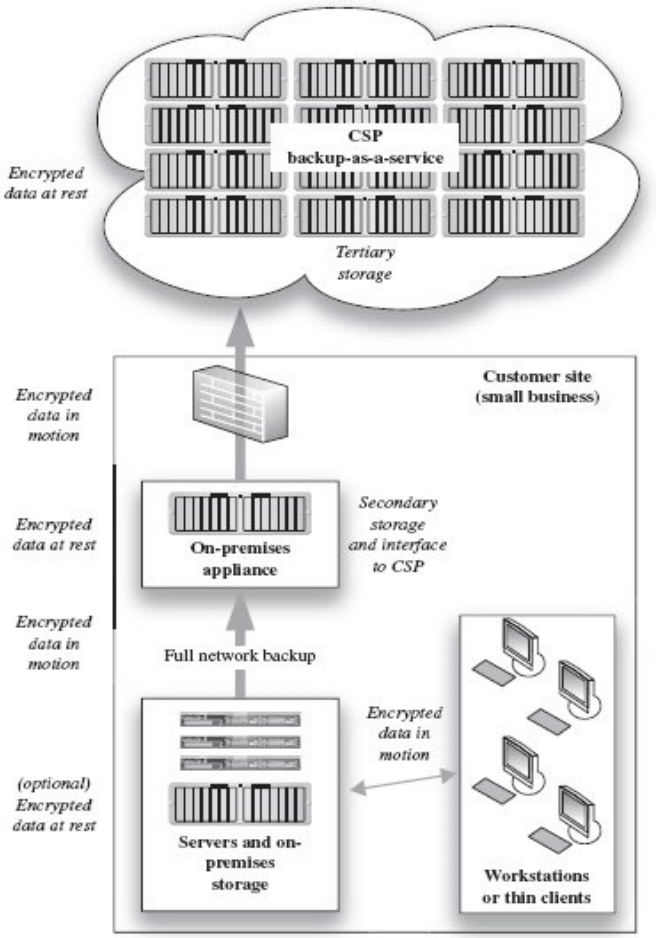

Figure 5: Cloud storage gateway appliance.

A. Cloud lock-in (The roach motel syndrome):

A number of questions about adopting public clouds have to do with what might happen when an external cloud becomes business-critical for the organization. One of these questions involves concern over cloud lock-in. Once you become dependent on the services of a cloud provider; you may find it extremely difficult to switch providers due to any number of technical reasons.

In one lock-in example, a company may subscribe to a specific public CSP service as their customer relationship management tool. This service may consequently end up being used to house all of the company's data relating to their customers. The company may invest significant effort in customizing rules or reporting routines in their use of this service. The service may also become the primary reporting engine that provides management insight to the health of the business.

B. Avoiding Cloud Lock-in (the Roach Motel Syndrome):

Fortunately, many of the large public cloud services organizations that exist today provide the ability to export not only data but also metadata generated by its subscribers. Any enterprise should seriously consider this as a vital feature to have before adopting any cloud service that could become critical to their business. It could be unrealistic to assume that

you will always maintain a service with a particular cloud provider. If there is no mechanism to retrieve your data, then the resulting situation can present a dilemma of costly proportions.

The presence of such a mass export feature isn't the only such requirement. How accessible and usable the data is after it has been exported is also important. If the data is exported in a proprietary file format, then that format might not be able to be intelligibly parsed. If it is exported in a plaintext format, it will have to be imported into the new system (or provider) in some intelligible way as well. As a result, one needs a real understanding of such challenges if you choose to leave a cloud.

Below are some examples of the cloud providers leading the industry in helping to avoid these lock-in concerns:

a. Salesforce.com offers its subscribers the ability to generate a complete export of all data within a subscribers instance on an on-demand basis. While some subscription levels include this export feature as a part of the package and others at an additional fee, it is available as an option. The exported data is available in a ZIP file containing plaintext CSV files, which have the raw data for each Salesforce object. This can also be setup in an automated task as well always archiving the data. If you are a subscriber, this feature is accessible under their Web Interface underSetup | Data Management | Data Export | Schedule Export. Also worth mentioning is that at the time of publishing, there are several other alternatives to Salesforce that are able to intelligibly and automatically parse this data, proving that it is indeed useful and not just satisfying a feature checkbox.

b. Another example of a public cloud provider helping to lead the way of addressing the lock-in problem is Amazon's Web Services and more specifically their Elastic Compute (EC2) service[3]. The same is also true for their surrounding cloud services for data storage, database computing, and several other services. Their approach to the problem is to offer an import/ export feature that accommodates amounts of data that are not feasible to transfer via a file download on the Internet. Subscribers can prepare a portable hard drive and submit a job to Amazon to perform a data import or export. At that point, the subscribers can physically mail their portable hard drive to an Amazon provided address, and the data migration occurs[4].

VI. ACKNOWLEDGMENT

We grateful to numerous local and global peers who have contributed towards shaping this seminar. At the outset, we would like to express my sincere thanks to **Prof. Sandip T. Dhagdi** for his advice. As our guide, he has constantly encouraged us to remain focused on achieving our goal. His observation and comments helped us to establish the overall direction of the review paper and to move forward with investigation in depth. He has helped us greatly and been a source of knowledge.

We highly indebted to**, Prof. Dr. R.M.Tugnayat, HOD IT Dept, Principal Prof. Dr. A.W.Kolhatkar,** for his continuous encouragement and support, as they has always been eager to help. We also thankful to all the professors of the department for their support.

We must acknowledge the academic resources that we have acquired from JDIET YAVATMAL. We would like to thank the non-teaching staff members of the department who have been kind enough to advice and help in their respective roles.

Last but not the least we would thankful to all our friends. Our sincere thanks to everyone who has provided us with inspirational words, a welcome eara, new ideas, constructive criticism, and their invaluable time, we truly indebted.

VII. CONCLUSION

We discuss the different ways of data security in cloud data storage, which is essentially a distributed storage system. Among other advances, cloud computing has brought advantages in the form of online storage.

We believe that data storage security in Cloud Computing, an area full of challenges and of paramount importance, is still in its infancy now, and many research problems are yet to be identified. We envision several possible directions for future research on this area.

VIII.REFERENCES

[1] Draft, National Institute of Standards and Technology, Information Technology Laboratory, January 2011.

[2] Journal of Computer Science and Network Security, VOL.10 No.6, June 2010.

[3] Amazon.com, “Amazon Web Services (AWS),” Online at http://aws.amazon.com, 2008.

[4] N. Gohring, “Amazon’s S3 down for several hours,” Online at http://www.pcworld.com/businesscenter/ article/142549/amazons s3 down for several hours.html, 2008.

[5] Ali M. Alakeel, A Guide to Dynamic Load Balancing in Distributed Computer Systems, IJCSNS International

[6] Cloud Computing Definition Gartner. http://www.gartner.com/it/page.jsp?id=1035013

[7] Peter Mell and Timothy Grance. The NIST Definition of Cloud Computing. Technical Report SP 800-145

[8] Martin Litoiu, Murray Woodside, Johnny Wong, Joanna Ng, Gabriel Iszlai, “A Buisness Driven Cloud

[9] Optimization Architecture”, Proceedings of ACM in SAC’10, pp.380 – 385.

[10] Roedig, U., Ackermann, R., Steinmetz, R.: Evaluating and Improving Firewalls for IP Telephony