ISSN 2250-3153

A Security approach for Data Migration in Cloud Computing

Virendra Singh Kushwah*, Aradhana Saxena**

*Assistant Professor, Department of Computer Science, HIMCS, Mathura
**Assistant Professor, Department of Computer Science, RJIT, Gwalior

***Abstract*--** Cloud computing is a new paradigm that combines several computing concepts and technologies of the Internet creating a platform for more agile and cost-effective business applications and IT infrastructure. The adoption of Cloud computing has been increasing for some time and the maturity of the market is steadily growing. Security is the question most consistently raised as consumers look to move their data and applications to the cloud. I justify the importance and motivation of security in the migration of legacy systems and I carry out an approach related to security in migration processes to cloud with the aim of finding the needs, concerns, requirements, aspects, opportunities and benefits of security in the migration process of legacy systems.

***Index Terms*--** Security; Cloud Computing; Data Migration; Encryption

I. BACKGROUND

1. Overview of Cloud Computing

Cloud computing services such as Amazon EC2 and Windows Azure are becoming more and more popular but it seems many people are still unclear as to what exactly the buzzword "Cloud computing" actually means. In its simplest form, the principle of Cloud computing is the provision of computing resources via a network.

Cloud Computing has become one of the most talked about technologies in recent times and has got lots of attention from media as well as analysts because of the opportunities it is offering. The market research and analysis firm IDC suggests that the market for Cloud Computing services was ` 68000 crore in 2008 and will rise to `178500 crore/year by 2012 [1]. It has been estimated that the cost advantages of Cloud Computing to be three to five times for business applications and more than five times for consumer applications. According to a Gartner press release from June 2008, Cloud Computing will be "no less influential than e-business" [2].

Enterprises have been striving to reduce computing costs and for that reason most of them start consolidating their IT operations and later using virtualization technologies. For the good of the enterprises there is a new technology to help them in this i.e. Cloud Computing. Cloud Computing claims to take enterprises search to a new level and allows them to further reduce costs through improved utilization, reduced administration and infrastructure cost and faster deployment cycles [3].

Cloud Computing is a term used to describe both a platform and type of application. As a platform it supplies, configures and reconfigures servers, while the servers can be physical machines or virtual machines. On the other hand, Cloud Computing describes applications that are extended to be accessible through the internet and for this purpose large data centers and powerful servers are used to host the web applications and web services [3,p2].

The cloud is a metaphor for the Internet and is an abstraction for the complex infrastructure it conceals. There are some important points in the definition to be discussed regarding Cloud Computing. Cloud Computing differs from traditional computing paradigms as it is scalable, can be encapsulated as an abstract entity which provides different level of services to the clients, driven by economies of scale and the services are dynamically configurable [6, p1].

There are many benefits stated of Cloud Computed by different researchers which make it more preferable to be adopted by enterprises. Cloud Computing infrastructure allows enterprises to achieve more efficient use of their IT hardware and software investments.

This is achieved by breaking down the physical barrier inherent in isolated systems, automating the management of the group of the systems as a single entity. Cloud Computing can also be described as ultimately virtualized system and a natural evolution for data centers which offer automated systems management [3, p4].

Enterprises need to consider the benefits, drawbacks and the effects of Cloud Computing on their organizations and usage practices, to make decision about the adoption and use. In the enterprise, the "adoption of Cloud Computing is as much dependent on the maturity of organizational and cultural (including legislative) processes as the technology, per se" [7].
Many companies have invested in Cloud Computing technology by building their public clouds, which include Amazon, Google and Microsoft. These companies are often releasing new features and updates of their services. For instance Amazon Web Services (AWS) released a Security2 and Economics3 center on their website to have academic and community advice regarding these issues [12]. This shows that there are still lots of doubts about the costs and security for enterprises to adopt Cloud Computing. Hence, the issues of economics and security in Cloud Computing for enterprises must be researched. As large organizations are inherently complex hence, it is very important for Cloud Computing to deliver the real value rather than just be a platform for simple tasks such as application testing or running product demos. For this reason, issues around migrating application systems to the cloud and satisfying the needs of key stakeholders should be explored. The stakeholders include technical, project, operations and financial managers as well as the engineers who are going to be developing and supporting the individual systems. For enterprises economics or cost factor is important but at the same time customer relationships, public image, flexibility, business continuity and compliance are of same importance.

2. Types of Cloud Providers

Cloud services are usually divided in the three main types, Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS).

a. Software as a Service (SaaS)

SalesForce. The applications are typically offered to the clients via the Internet and are managed completely by the Cloud provider. That means that the administration of these services such as updating and patching are in the provider‟s responsibility. One big benefit of SaaS is that all clients are running the same software version and new functionality can be easily integrated by the provider and is therefore available to all clients.

b. Platform as a Service (PaaS)

PaaS Cloud providers offer an application platform as a service, for example Google App Engine. This enables clients to deploy custom software using the tools and programming languages offered by the provider. Clients have control over the deployed applications and environment-related settings. As with SaaS, the management of the underlying infrastructure lies within the responsibility of the provider.

c. Infrastructure as a Service (IaaS)

IaaS delivers hardware resources such as CPU, disk space or network components as a service. These resources are usually delivered as a virtualization platform by the Cloud provider and can be accessed across the Internet by the client. The client has full control of the virtualized platform and is not responsible for managing the underlying infrastructure.

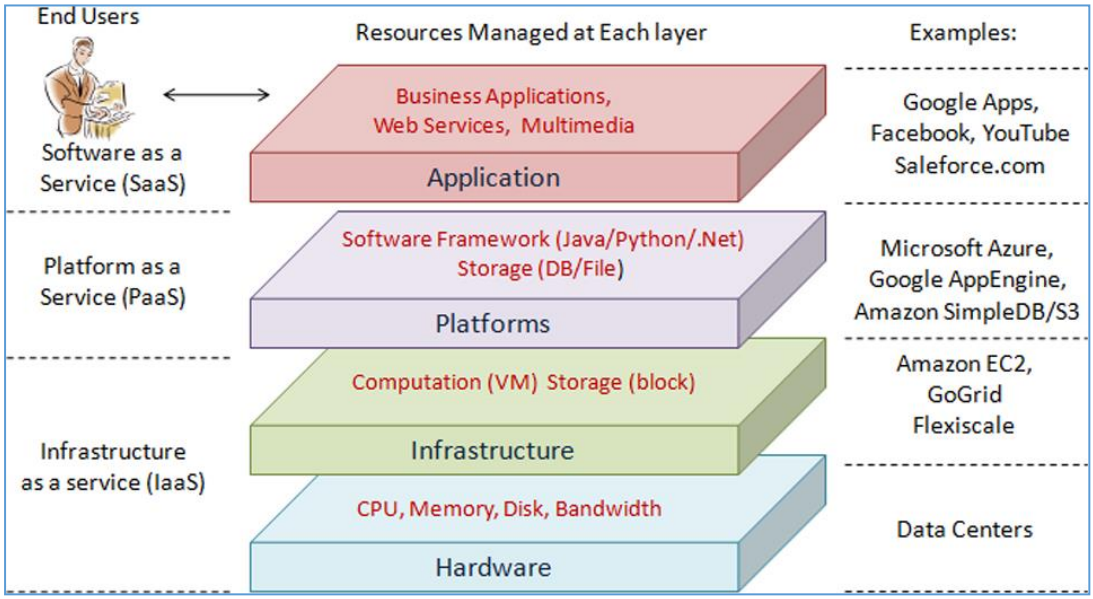

Figure 1: A layered model of Cloud Computing
(Source: Qi Zhang et al., Cloud computing: state-of-the-art and research challenges, J Internet Serv Appl (2010) pp. 7–18)

3. Security issues and challenges

Heightened security threats must be overcome in order to benefit fully from this new computing paradigm. Some security concerns are listed and discussed below:

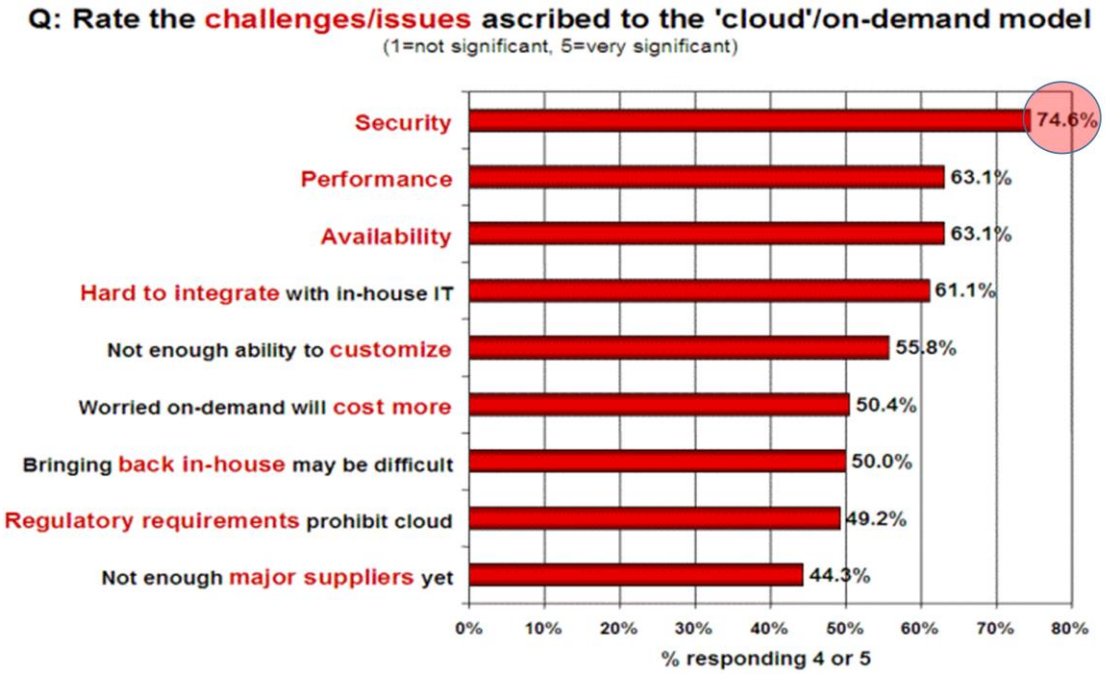

Figure 2: Security is the Major Issue
(Source: http://www.csrc.nist.gov/groups/SNS/cloud-computing/cloud-computing-v26.ppt at slide 17)

a. *Security concern #1:* With the cloud model control physical security is lost because of sharing computing resources with other companies. No knowledge or control of where the resources run.

b. *Security concern #2:* Company has violated the law (risk of data seizure by (foreign) government).

c. *Security concern #3:* Storage services provided by one cloud vendor may be incompatible with another vendor's services if user decides to move from one to the other (e.g. Microsoft cloud is incompatible with Google cloud). (Pearson et al. 2003)

d. *Security concern #4:* Who controls the encryption/decryption keys? Logically it should be the customer.

e. *Security concern #5:* Ensuring the integrity of the data (transfer, storage, and retrieval) really means that it changes only in response to authorized transactions. A common standard to ensure data integrity does not yet exist.

f. *Security concern #6:* In case of Payment Card Industry Data Security Standard (PCI DSS) data logs must be provided to security managers and regulators. [9][10][11]

g. *Security concern #7:* Users must keep up to date with application improvements to be sure they are protected.

h. *Security concern #8:* Some government regulations have strict limits on what data about its citizens can be stored and for how long, and some banking regulators require that customer's financial data remain in their home country.

i. *Security concern #9:* The dynamic and fluid nature of virtual machines will make it difficult to maintain the consistency of security and ensure the audit ability of records.

j. *Security concern #10:* Customers may be able to sue cloud service providers if their privacy rights are violated, and in any case the cloud service providers may face damage to their reputation. Concerns arise when it is not clear to individuals why their personal information is requested or how it will be used or passed on to other parties.

II. SECURITY CHALLENGE IN DATA MIGRATION

1. A perception on the Data Migration

Data migration to a cloud computing environment is in many ways an exercise in risk management. Both qualitative and quantitative factors apply in an analysis. The risks must be carefully balanced against the available safeguards and expected benefits, with the understanding that accountability for security remains with the organization. Too many controls can be inefficient and ineffective, if the benefits outweigh the costs and associated risks. An appropriate balance between the strength of controls and the relative risk associated with particular programs and operations must be ensured.

Data security is another important research topic in cloud computing. Since service providers typically do not have access to the physical security system of data centers, they must rely on the infrastructure provider to achieve full data security. Even for a virtual private cloud, the service provider can only specify the security setting remotely, without knowing whether it is fully implemented. The infrastructure provider, in this context, must achieve the following objectives: (1) *confidentiality*, for secure data access and transfer, and (2) *auditability*, for attesting whether security setting of applications has been tampered or not. Confidentiality is usually achieved using cryptographic protocols, whereas auditability can be achieved using remote attestation techniques. Remote attestation typically requires a trusted platform module (TPM) to generate non-forgeable system summary (i.e. system state encrypted using TPM's private key) as the proof of system security. However, in a virtualized environment like the clouds, VMs can dynamically migrate from one location to another; hence directly using remote attestation is not sufficient. In this case, it is critical to build trust mechanisms at every architectural layer of the cloud. Firstly, the hardware layer must be trusted using hardware TPM. Secondly, the virtualization platform must be trusted using secure virtual machine monitors [14, 15]. VM migration should only be allowed if both source and destination servers are trusted. Recent work has been devoted to designing efficient protocols for trust establishment and management.

2. Need for securing data migration process

Cloud Migration is one of much conversed point where cloud managers face extreme problems at the time of data migration from a company's server to a server that forms cloud elsewhere. Why they face troubles let's find out. As I know, cloud behaves as an interface through which organizations can access data in a virtual environment. Thus, smooth functioning of it depends primarily on how well groomed and knowledgeable cloud providers are in this area.

Moreover, if data migration is not done systematically and properly, it can give rise to problems concerning data and cloud security of company's assets that primarily comprise of data. Thus, hiring cloud providers having sound experience about the field with ample knowledge and skill sets becomes vital for managing cloud more effectively and efficiently.

Example: Suppose an XYZ company wants to shift its data to cloud storage for increased uptime and scalability, it goes to cloud service provider for performing such functions. Now, the cloud provider starts initializing steps for data transfer to cloud, but in between face problems like data crash or unauthorized access by third parties. This is where the problem lies. The proprietor of data that hired cloud manager would not only face reputation losses but also monetary losses. Similar case was experienced when Amazon cloud failure happened and several businesses suffered immense losses due to it.

Thus, securing data remains an utmost priority of cloud managers to prevent global cloud security threats that also include cross-border security concerns.

3. Characteristic of Data Migration

a. Commercial relation exists between clouds
b. Transmission of mass data
c. Many workers which execute transmission process concurrently

III. PROPOSED SOLUTION OF SECURING DATA MIGRATION PROCESS

We have talked about security in cloud computing many times before, explaining why it is just as safe as conventional networking security, even citing its benefits over the conventional. However, there are many who still find cloud computing security lacking.

Individuals which still worry about cloud security are those that fall under the financial institution category like banks, brokers, lenders and the like. They do not trust third party cloud computing providers and vendors, at least not with their most sensitive information and data. They might use cloud computing for some things like websites and applications that they think they can risk security with, but they would never consider parting with direct access of their financial and other similar data.

The biggest reason behind this is simpler than most would imagine as it has something to do with numbers and probability, thought they probably would not admit it is something as basic as that and would rather cite some technical issue like migration and data integrity. Those are valid points, but they are not truly even problems. With ease and security of data migration through cloning and inter-server data transfers with services like Cloud Velocity, migration is truly a no pain no worry process. The real reason as I have said is the probability of a successful attack. Government systems and financial data systems are under attack multiple times a day, and a sizeable majority of these fail at the first lines of defense. The probability of a successful attack is always real, and this probability of success increases as the number of attempts increases.

The process of transitioning all or part of a company's data, applications and services from on-site premises behind the firewall to the cloud, where the information can be provided over the Internet on an on-demand basis. While a cloud migration can present numerous challenges and raise security concerns, cloud computing can also enable a company to potentially reduce

capital expenditures and operating costs while also benefiting from the dynamic scaling, high availability, multi-tenancy and effective resource allocation advantages cloud-based computing offers.

1. Understanding for Distributed file system over clouds

Google File System (GFS) [17] is a proprietary distributed file system developed by Google and specially designed to provide efficient, reliable access to data using large clusters of commodity servers. Files are divided into chunks of 64 megabytes, and are usually appended to or read and only extremely rarely overwritten or shrunk. Compared with traditional file systems, GFS is designed and optimized to run on data centers to provide extremely high data throughputs, low latency and survive individual server failures.

Inspired by GFS, the open source Hadoop Distributed File System (HDFS) [18] stores large files across multiple machines. It achieves reliability by replicating the data across multiple servers. Similarly to GFS, data is stored on multiple geo-diverse nodes. The file system is built from a cluster of data nodes, each of which serves blocks of data over the network using a block protocol specific to HDFS. Data is also provided over HTTP, allowing access to all content from a web browser or other types of clients. Data nodes can talk to each other to rebalance data distribution, to move copies around, and to keep the replication of data high.

2. Prediction based Encryption (PBE)

Predicate Based Encryption (PBE), represents a family of asymmetric encryption schemes that allows for selective fine-grained access control as part of the underlying cryptographic operation. The origins of PBE are in Identity Based Encryption (IBE). In IBE schemes an entity's encryption key is derived from a simple string that represents the entity's own public identity e.g. an email address. For example, given an entity “Virendra” his corresponding encryption key will be *Enc (Virendra) == kushwah.virendra248@gmail.com*. During encryption, the resulting cipher-text will effectively be labelled with the string representing the encryption key, the entity's public identity. An entity's decryption key will be derived from the same string used for the encryption key e.g. Virendra's decryption key will be derived from his e-mail address. On recipt of a ciphertext message the recipient will be able to decrypt the cipher-text if and only if the two identities, contained within the decryption key and cipher-text, are `equal'. PBE schemes offer a richer scheme in which an entity's `identity' can be constructed from a set of attributes and decryption is associated with access policies that offers a more expressive means with which to describe the relation between the attributes.

A solution might be Prediction Based Encryption (PBE) for multicasting. PBE is a combination of both IBE (Identity Based Encryption)[19][20] and ABE (Attribute Based Encryption) [22][24]. In this work, the attributes are used to design user's decryption keys and to encrypt simple text messages. Decryption occurs when a match occurs between the attributes held by the entity (in their Decryption key) and the attributes used to construct a simple text. This matching occurs through the use of predicates, which describe:

- The required attributes needed to decrypt
- The relationship between the attributes.

PBE scheme supports four operations allowing for encryption, decryption and key generation. The precise value for encryption and decryption keys is dependent upon both the construction of the scheme and placement of predicates. A general PBE scheme consists of the four operations[18]:

- **Setup:** initializes the crypto-scheme and generates a master secret key MSK, used to generate decryption keys, and a set of public parameters MPK.

(MSK, MPK):=Setup ()

- **KeyGen:** generates a decryption key Dec (entity) based upon the master secret key and some entity supplied input.

Dec (entity):=KeyGen (MSK, input)

- **Encrypt:** encrypts a plain-text message M using the public parameters and supplied encryption key for an entity.

CT: = Encrypt (M, MPK, Enc (entity))

- **Decrypt:** decrypts a cipher-text if and only if the attributes held by the entity can satisfy the access policy.

M: = Decrypt (CT, MPK, Dec (entity))

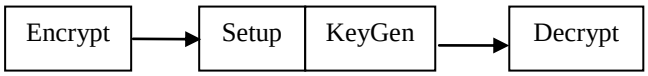

Figure 3. Functioning of the proposed system

3. Working Structure

The working structure of proposed solution can be recognized by the following and important figure. It illustrates entire details toward the security needs.

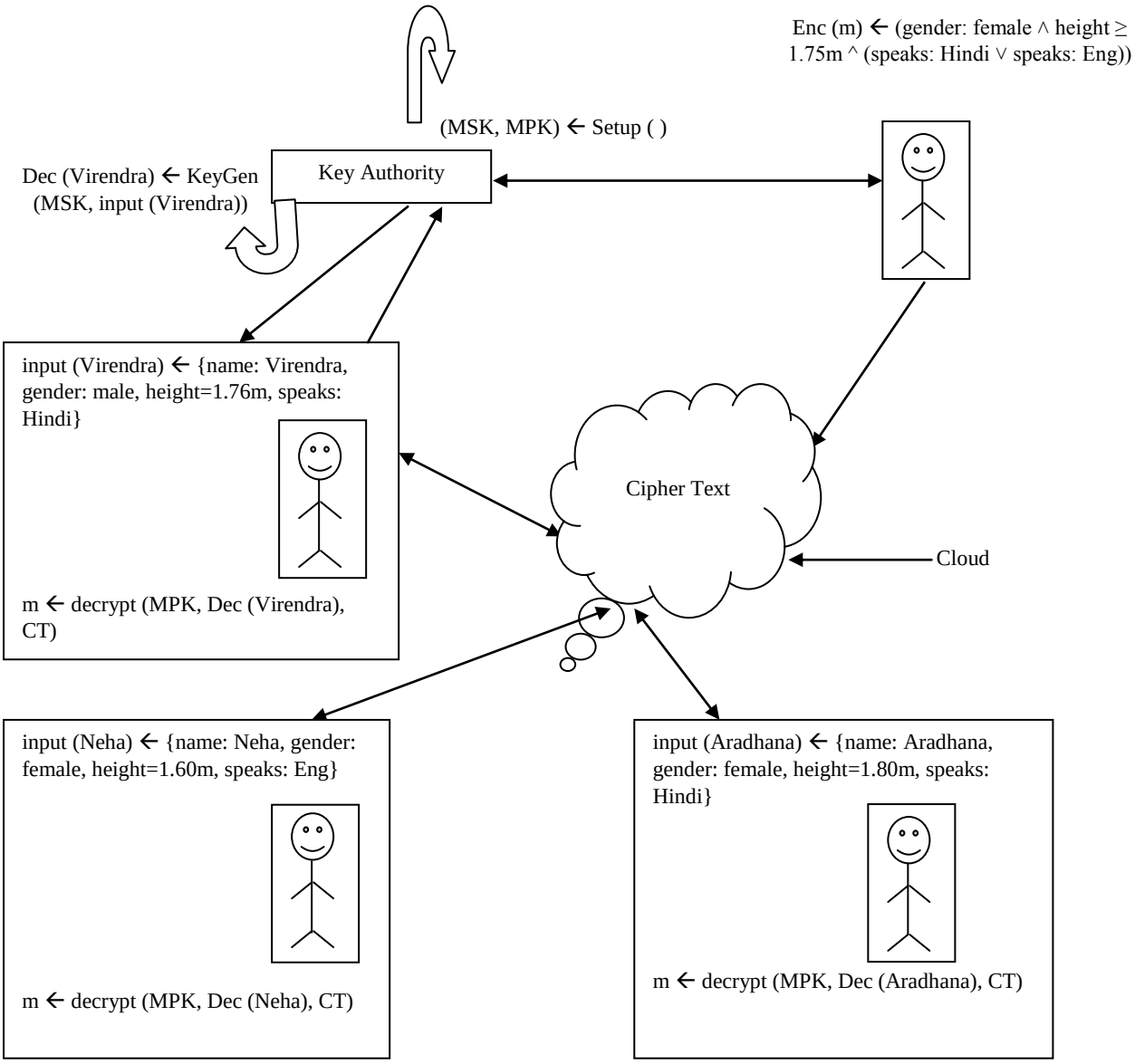

Figure 4: Overall working of the system

The overall working of the proposed solution can be understood by the below diagram. As shown in the diagram, when Virendra, Neha and Aradhana want to communicate for sending and receiving cloud's data. Here, only Aradhana can access the Cloud's data. She is only authorized person who can access the Cloud's data based on Encryption of the message with specific parameters. The encrypted data can be decrypted by its Master Public Key (MPK) as mentioned above.

The working can be under covered by a File System, which is identified by HDFS (Hadoop Distributed File System). This file system creates a layer between the encrypted data and shared link or channel.

IV. CONCLUSION & FUTURE WORK

Cloud is growing because cloud solutions provide users with access to high computational power at a fraction of the cost of buying such a solution outright and which can be acquired on demand; the network becomes an important element in the cloud where users can buy what they need when they need it. Although industry leaders and customers have wide-ranging expectations for cloud computing, privacy and security concerns remain a major impediment to widespread adoption.

The benefits of Cloud computing are the first weapon when organizations or companies are considering moving their applications and services to Cloud, analyzing the advantages that it entails and the improvements that they can get. If the customers decide to incorporate their businesses or part of them to the Cloud, they need to take into account a number of risks and threats that arise, the possible solutions that can be carried out to protect their applications, services and data from those

risks, and some best practices or recommendations which may be helpful when the customers want to integrate their applications in the Cloud.

The future work can be carried out the optimization of security work as an idea to ensure about the work reliability. With the help of LP (Linear Programming), we will optimize the secured data.

ACKNOWLEDGMENT

We would like give thanks to Dr. Jaidhar C.D. (Assistant Professor, Department of CSE, DIAT, Pune) and without his support, this work cannot be completed. Their motivational supports and valuable guidance always encouraged time to time.

REFERENCES

[1]. Gleeson, E. (2009). Computing industry set for a shocking change. Retrieved May 10, 2010 from http://www.moneyweek.com/investment-advice/computing-industry-set-for-ashocking-change-43226.aspx

[2]. Gartner (2008). Gartner Says Cloud Computing Will Be As Influential As E-business. Gartner press release, 26 June 2008. http://www.gartner.com/it/page.jsp?id=707508. Retrieved 3rd May 2010

[3]. Boss, G., Malladi, P., Quan, D., Legregni, L., Hall, H. (2007), Cloud Computing.www.ibm.com/developerworks/websphere/zones/hipods/. Retrieved on 20th May, 2010.

[4]. Foster I, Kesselman C (1998) Computational Grids. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.36.4939

[5]. Foster I, Kesselman, C, Tuecke S (2001) The Anatomy of the Grid: Enabling Scalable Virtual Organization. International Journal of High Performance Computing Applications 15(3):200-222

[6]. Foster I, Zhao Y, Raicu I, Lu S (2008) Cloud Computing and Grid Computing 360-Degree Compared. In: Grid Computing Environments Workshop (GCE'08). oi:10.1109/GCE.2008.4738445

[7]. Fellowes, W. (2008). Partly Cloudy, Blue-Sky Thinking About Cloud Computing. Whitepaper. 451 Group.

[8]. M. Casassa-Mont, S. Pearson and P. Bramhall, "Towards Accountable Management of Identity and Privacy: Sticky Policies and Enforceable Tracing Services", Proc. DEXA 2003, IEEE Computer Society, 2003, pp. 377-382

[9]. https://www.pcisecuritystandards.org/index.shtml

[10]. http://en.wikipedia.org/wiki/Payment_Card_Industry_Data_Security_Standard, 24 January 2010

[11]. J. Salmon, "Clouded in uncertainty – the legal pitfalls of cloud computing", Computing, 24 Sept 2008, http://www.computing.co.uk/computing/features/2226701/clouded-uncertainty-4229153

[12]. Khajeh-Hosseini, A., Greenwood, D., Sommerville, I., (2010). Cloud Migration: A Case Study of Migrating an Enterprise IT System to IaaS. Submitted to IEEE CLOUD 2010

[13]. S. Overby, How to Negotiate a Better Cloud Computing Contract, CIO, April 21, 2010, http://www.cio.com/article/591629/How_to_Negotiate_a_Better_Cloud_Computing_Contract

[14]. Krautheim FJ (2009) Private virtual infrastructure for cloud computing. In: Proc of HotCloud

[15]. Santos N, Gummadi K, Rodrigues R (2009) Towards trusted cloud computing. In: Proc of HotCloud

[16]. Armbrust M et al (2009) Above the clouds: a Berkeley view of cloud computing. UC Berkeley Technical Report

[17]. Ghemawat S, Gobioff H, Leung S-T (2003) The Google file system. In: Proc of SOSP, October 2003 Hadoop Distributed File System, hadoop.apache.org/hdfs

[18]. An article on "Predictions about the future of Cloud Computing" available at http://mediastar91.blogspot.in/2012/04/predictions-about-future-of-cloud.html

[19]. C. Schridde, T. Dornemann, E. Juhnke, B. Freisleben, M. Smith, "An Identity-Based Security Infrastructure for Cloud Environments," 2010 IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS), pp. 644 – 649, 2010.

[20]. J. Y. Sun, C. Z, Y. C. Zhang, and Y. G. Fang, "An Identity-Based Security System for User Privacy in Vehicular Ad Hoc Networks, "IEEE Transactions on Parallel and Distributed Systems, vol. 21, no.9, pp. 1227-1239, 2010.

[21]. A Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters. Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption. In Advances in Cryptology - EUROCRYPT 2010. Springer, 2010.

[22]. M. Pirretti, P. Traynor, P. McDaniel, and B. Waters, "Secure attribute-based systems," Journal of Computer Security, vol. 18, no. 5, pp. 799–837, 2010.

[23]. S. Jahid, P. Mittal, and N. Borisov, "Easier: Encryption-based access control in social networks with efficient revocation," in ASIACCS, Hong Kong, March 2011.

[24]. S. Yu, C. Wang, K. Ren, and W. Lou, "Attribute based data sharing with attribute revocation," in ASIACCS'10, 2010.

[25]. S. Narayan, M. Gagn´e, and R. Safavi-Naini, "Privacy preserving ehr system using attribute-based infrastructure," ser. CCSW '10, 2010, pp. 47–52.

AUTHORS

First Author -- Virendra Singh Kushwah, M.Tech (IS) from ABV-IIITM, Gwalior, HIMCS, Mathura, kushwah.virendra248@gmail.com

Second Author – Aradhana Saxena, M.Tech (IS), from ABV-IIITM, Gwalior RJIT, Gwalior, aradhana298@gmail.com

Correspondence Author – Virendra Singh Kushwah, M.Tech (IS), from ABV-IIITM, Gwalior, HIMCS, Mathura kushwah.virendra248@gmail.com, +91-75000-66166