ISSN: 2278 – 7798
International Journal of Science, Engineering and Technology Research (IJSETR)
Volume 2, Issue 8, August 2013

# Data Security Issues and Strategy on Cloud Computing

Sonam Singh
Department of Computer Science and Engineering, I.F.T.M University, Moradabad Province 244001, India

## Abstract

***"Cloud Computing" is a term, which involves virtualization, distributed computing, networking and web-services. It is a way of offering services to users by allowing them to tap into a massive pool of shared computing resources such as servers, storage and network. User can use services by simply plug into the cloud and pay only for what he uses. All these features made a cloud computing very advantageous and demanding. But the data privacy is a key security problem in cloud computing which comprises of data integrity, data confidentiality and user privacy specific concerns. Most of the persons do not prefer cloud to store their data as they are having a fear of losing the privacy of their confidential data. This paper introduces some cloud computing data security problem and its strategy to solve them which also satisfies the user regarding their data security.***

***Keywords- cloud computing; cloud data security; strategy***

## 1. Introduction

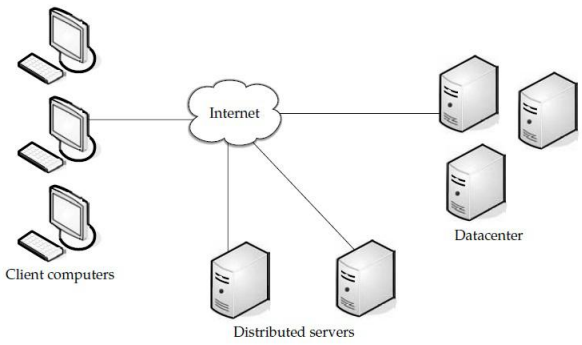
The cloud computing is a term which comes from grid computing, distributed computing, parallel computing, virtualization technology, utility computing and other computer technologies[1]. Cloud computing allows users to connect to a massive, shared pool of computing resources that are provided as a service to users, allowing them to "plug into the cloud" very similar to a utility grid. Such a computing model allows consolidation on the server side and reduces costs by taking advantage of economies of scale. The promise of the cloud is to free users from the tedious and often complex task of managing and provisioning computing resources to run applications. The cloud also brings several additional benefits including (1) a pay-as-you-go cost model, which means that users only pay for what they use, (2) much easier and faster deployment of applications, (3) dynamic and elastic provisioning of resources, which means that unlike a traditional data center setting, there is no need to provision for peak load since applications can grow and shrink processing capacity depending on the load, i.e. the amount of resources provided in the cloud system for the users is increased when they need more and decrease when they need less and 4) a more robust and secure infrastructure that is highly available and elastically scalable. All these features of the cloud come to users at no additional cost, while at the same time providing them with a lot of flexibility in using and provisioning computing resources. The resource can be the computing, storage and other specification service. The users can access their applications and data from anywhere through any device just two things they needed, internet connection and a cloud computing systems interface software on their device which can be as simple as a web browser. Although cloud computing having various advantageous features but the security concerns of cloud remain one of the greatest inhibitors for adoption of Cloud computing. One of the primary concern of cloud computing is the security of user's data.

The security of data is totally relying on cloud provider rather than user. Users have to depend on cloud provider for the security of his own data which may or may not be 100 % guaranteed. So the cloud computing must ensure the security of data stored in the cloud system. Many companies provide the cloud computing platform such as Google, IBM, Microsoft, Amazon, VMware and EMC [2-8]. As the cloud computing systems has various amounts of private and important data of user. Due to the importance of data the hacker pay more attention to get it. By numerous ways the privacy of the data can be breached. And how can a user ensure that the cloud service provider notifies him when a breach occurs. Various companies and organizations are looking forward to cloud computing to expand their on-premises infrastructure, but most cannot afford the risk of compromising the security of their applications and data. Governance and security are crucial to computing on the cloud, whether the cloud system is in firewall or not. The security of cloud computing is the key import problem in the

development of cloud computing. As the cloud computing application has no boundaries and mobility can lead many new security problems. To get over from these issues the cloud provider must apply the strategy which ensures the users that they have security and privacy control over their applications and services and also provide evidences to the users which proves that their data is safe.

## 2. Cloud Computing

Cloud computing is fast becoming a popular option for renting of computing and storage infrastructure services. The cloud computing move the tasks which are implemented in the personal computer and private data center into the larger computing center which are shared with total user and distributed in the internet. It composes applications out of loosely coupled services and one service failure will not disrupt other services. The cloud computing system can be divided into two sections: the front end and the back end. They connect to each other through the internet. The front end is user who use the service provided by the back end which is the cloud section of the system. A Cloud computing system consists of 3 major components such as clients, data center, and distributed servers. Each element has a definite purpose and plays a specific role.




Clients are the End-Users which generally fall into three categories mobile, thin and thick. Thin clients don't do any computation work only display the information. Servers do all the work for them. Thin clients don't have any internal memory whereas thick clients use different browsers like IE or Mozilla Firefox or Google Chrome to connect to the Internet cloud. Data Center is a facility used to house computer systems and associated components, such as telecommunications and storage systems, and Distributed servers are the parts of a cloud which are present throughout the Internet hosting different applications. Cloud Computing uses virtualization technique which is very useful in context of cloud. Virtualisation means" something which isn't real", but gives all the facilities of a real. It is the software implementation of a computer which will execute different programs like a real machine. The virtualization has the ability to run multiple operating systems on a single physical system and share the underlying hardware resources. The cloud computing is developed from many technology such as parallel computing, distributed computing, grid computing and other computer technologies. The grid computing does not rely on virtualization as much as the cloud computing do.

The Cloud Computing model has main three Deployment models, Models are

(1).**Private Cloud:** A cloud platform is dedicated for specific organization. The private cloud is deployed in the company and the security can be made easily. Private clouds are virtualized cloud data centers inside firewall and it is a private space dedicated to system within a cloud data center. Private cloud refers to internal data centers of a business or other organization not made available to the general public. The cloud system infrastructures are owned by an organization which sells cloud services to the general public or to a large industry company.

(2).**Public Cloud:** Available to public users to register and use the available infrastructure. The public cloud is running in the internet and the security is very complex. Public clouds are virtualized data centers outside of firewall and the service provider makes resources available to consumer on demand over the public Internet. The cloud computing is on-demand service and it giving computing capabilities as needed automatically. It can use the service by many machine such as desktop, laptop, PDA and mobile phone.

(3).**Hybrid Cloud:** A Private cloud that can extend to use resources in Public clouds. Hybrid cloud is the composition of two or more clouds and bounded by standard or proprietary technology. Hybrid clouds combine character of both public and private clouds.

The Cloud Service delivery models include,

-- **Infrastructure-as-a-service (*iaas*):** In Infrastructure as a service the consumer gets access to the infrastructure to deploy their applications. It share managed pool of configurable and scalable resources such as network, middleware, database and storage servers. Iaas provides virtual machines and other abstracted hardware and operating systems which may be controlled through a service API. The examples of

Iaas is Amazon Elastic Compute Cloud (EC2), S3, Terre mark Enterprise Cloud, Rack space Cloud.

-- **Platform-as-a-service (*paas*):** The Platforms offered development tools, configuration management and deployment platforms. It gets the deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers and it make raw hardware made available to the user through the Internet but generally includes a specific operating system that is pre-installed and supported by the Cloud vendor. The examples of PaaS are Google App Engine, Amazon Web services, Microsoft Azure, Force

-- **Software-as-a-service (*saas*):** is software offered by a third party provider, available on demand, usually via the Internet configurable remotely. It can reduce expenses, easy to use and access from everywhere. It share instance of a software application as a service accessible via internet browser or client based role access and sharing rules. The service provider hosts the software so the users don't need to install or manage or buy hardware for it. All they have to do is connect and use it. The examples of SaaS are Flickr, Google Docs, Amazon and Cloud Drive, online word processing and spreadsheet tools.

Cloud computing also provides storage services to the users. Cloud storage is a service model in which data is maintained, managed and backed up remotely and made available to users over a network typically the internet [9]. There are many cloud storage providers in the market such as Google, Microsoft, IBM and Amazon. These providers provide storage of two types, 1) Permanent storage in which data will be saved permanently on the cloud. 2) Virtual Storage i.e. the storage on lease. Storage services are available to store all forms of digital data. Storage lies in the category of **Instance storage** which comes with virtual machine images, **Object storage** which provides the storage of binary objects in the form of web-services. It is a generic term that describes an approach to addressing and manipulating discrete units of storage called objects, **Block storage** provides the virtual block devices that can be attached to VM instances and used like local disks, **Semi-structure data storage** used for storing semi or unstructured data with high availability, scalability and high performance, **Relational database storage** uses relational database servers on VM instances in clouds, **Distributed file system** provides distributed storage through file system interfaces. DFS makes easy for users to access and manage files that are physically distributed across a network. **Online drive/Folder service** provides the storage space in the form of a virtual drive or folder on internet [9].

| **Storage Category** | AMAZON | MICROSOFT | GOOGLE |
|---|---|---|---|
| *Instance Storage* | EC2 | Azure VM | – |
| *Object Storage* | S3 | Azure Blob | Google Storage for developers |
| *Block Storage* | EBS | Azure drive | – |
| *Semi-structured DataStorage* | Simple DB | Azure table | Big Table |
| *Relational DB Storage* | RDS | SQL Azure | – |
| *DFS Storage* | – | – | Google File System |
| *Online folder/drive* | – | Sky Drive/Mesh | – |

**Cloud Providers**

On the basis of user's demand of storage cloud provider give their services. Google use GFS for distributed file system storage which is currently most demanding due to the excessive use of social networking sites like Facebook, yahoo etc. GFS consists of multiple nodes which are divided into two types one is Master node and second is large number of chunk servers. Each file is divided into fixed size chunks and chunk servers stored them. Each chunk is assigned a unique 64-bit label by Master node at the time of creation, and logical mapping of files to a constituent chunks are maintained [10]. GFS use Map Reduce which is a programming model for processing large data sets with a parallel, distributed algorithm on a cluster. It comprises of two words, Map which filters and sort the data and Reduce which performs a summary operation on that data. By using Map Reduce library, implementation of HADOOP is done. HADOOP is a free, Java based programming framework which supports the processing of large data sets in a distributed computing environment. It is developed by Apache. HADOOP uses the programming languages PIG and HIVE provides tools HDFS i.e. HADOOP distributed file system and HADOOP Map Reduce. The file in HDFS is split into many files, each of those is replicated and stored on 3 servers for fault tolerance constraints, and the Map

reduce make the decomposition of tasks and integration of results. Amazon provide the EC2 (Elastic Compute Cloud) and S3 (Simple Storage Service). The EC2 can provide many services which ruing in the virtual machine. Microsoft provides storage through VM, drive, table and blob. Blob is a compressed binary file format which stores any kind of file such as multimedia, Text file.

## 3. Cloud Data Security Problem

As vendors provide cloud storage services to the user through network, typically on internet. This gives rise to various security and privacy [11] concerns to user regarding their data. These concerns comprises of,

**1). Access:**

In cloud, the threat of accessing sensitive information of the user is very high. There is also having more chances of data theft from the machine in cloud environment. Hackers and Malicious intruders may hack the cloud accounts and can steal the sensitive data stored in cloud system. Also a person from the company can open the confidential data and modify them unlawfully. Due to these, the privacy of the data is endangered. The cloud system can deploy the cloud center in different area and the data can be stored in different cloud node. The different area has different law so the security management can meet the law risk.

**2).Lack Of User Control:**

As cloud system hide the details of service implementation, technology and the management. They don't provide any information regarding the location of data, which servers are processing and what networks are transmitting the data because of flexibility and scalability of cloud system. Due to this, user can't control the progress of deal with the data which causes the user unsure about their data security and data privacy operated by the cloud in a confidential way. On traditional PC or servers owned by a company or individual, there is control over how the data is stored, restrictions put on who can access it. For cloud computing the data is stored on the server and the third-party company is responsible for deciding the details of data storage. Service provider is responsible to control user data but it is a legal requirement of user to control his data when it is processed or stored.

**3).Lack of Trust:**

Cloud Provider does not provide any audit mechanism to monitor their internal transactions which satisfy the users regarding their data security. How a user ensures that his data is safe without any proof. Cloud is fail in making a trust between user and provider. Lack of trust is the main cause that stops the adoption of cloud by various organizations and peoples. Cloud has mobile and vibrant nature due to which there is no clear aspect that who is legally responsible to ensure privacy of sensitive data put by user on cloud. This also causes lack of trust in cloud. Cloud computing service must be improved in legal protection also.

## 4. Strategy

The data can be encrypted before stored in the cloud system but this is not a total solution to secure the data. Cloud computing storage security is primarily related to data storage isolation, storage place, data recovery and data long term survivability. Once the data is stored in the cloud, the control of the data is transferred to the hands of cloud computing providers. Some unscrupulous businesses can get the customer privacy information by unfair means. The cloud computing services provided for customers are difficult to achieve full transparency. Customers do not understand internal processes of cloud computing and data storage location information. Customers should have the right of the supervision and audit of cloud computing services in order to fully ensure the security of customer data.

This section includes solution to overcome the issues discussed in previous sections. When the data is stored to any location used by cloud provider, on that location or from any other location if someone tries to access the user data by hacking the user account or by opening the encrypted data on machines where data is located. On any of these activities a mail or a message must be sent to the users cell phone or mail-Id which aware the user regarding the activity. The message ask for the permission of giving access to other person by containing a specific code and two links one of which directs to report the problem about the activity to cloud provider so that provider search the cause and satisfies the user by informing and taking action regarding the activity and by clicking on another link user have to enter that specific code and submit it by doing this user give permission to other person to access his data. Without user permission nobody can access his data. This assures the user that his data is secure by making aware of any unauthorised activity done on their confidential data and also provide control of accessing data on user hand.

## 5. Conclusion

Cloud computing has a very fast paced of development it provides easy access to high performance computing resources and storage infrastructure through web services. It shows good prospects and great potential. But the Security concerns – especially data control,

privacy issue and access, can prove to be the hurdles for adoption of clouds. This paper illustrates cloud concepts, its models and services which demonstrate the cloud capabilities such as scalability, elasticity, and platform independent, low-cost and reliability. Address the data security problems of cloud computing and discuss the strategy to solve them in a manner which increases the trust of user on cloud and provides visibility regarding their data security.

## References

[1] Wentao Liu, Research on Cloud Computing Security Problem and Strategy, 978-1-4577-1415-3/12/ ©2012 IEEE.
[2] Amazon Elastic Compute Cloud, http://www.amazon.com/ec2/
[3] Google App Engine, http://appengine.google.com/
[4] Sale force, http://www.salesforce.com/platform/
[5] Microsoft, http://www.microsoft.com/,
[6] VMware, http://www.vmware.com/
[7] IBM Blue Cloud project, http://www03.ibm.com/press/us/en/pressrelease/22613.wss
[8] Open Nebula Project, http://www.opennebula.org/
[9] Xiaoming, Xiaogang, Adarsh, Abhijeet and Pranav. A Survey on Cloud Storage Systems. http://salsahpc.indiana.edu/b649projects/sites/default/files/u10/final%20presentation_v2.pdf
[10]Google File System http://en.wikipedia.org/wiki/Google_File_System
[11]Cloud Security Alliance: http://www.cloudsecurityalliance.org/