Difficulty In Analysis Of Data In Cloud From Forensic Point Of View

Asmita Ballal
TGPCET
Nagpur

Prof. Sulabha Patil
TGPCET
Nagpur

Dr. R. V. Dharaskar
MPGI
Nanded

Abstract

Cloud environment is offering different services to the users and more and more companies are working to tap the benefits being provided by this environment. However, the data that is stored in cloud is scattered and is administered by different countries having varied time zones, laws, regulations etc. The distributed data in cloud poses various difficulties from forensic point of view. A comprehensive review of the difficulties in analysis of data in Cloud environments is discussed in this paper.

1. Introduction

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.[1] Cloud computing is an evolution and combination of decades of technology,resulting in a model of convenient, on-demand, elastic, location-independent computing resources. Migration to cloud computing involves replacing much of the traditional IT hardware found in an organization's data centre (including servers, racks, network switches and air conditioning units) with virtualized, remote, on-demand software services, configured for the particular needs of the organization. These services can be hosted and managed by the user organization (on a reduced hardware base), or by a third-party provider. Consequently, the software and data comprising the organization's application may be physically stored across many different locations, potentially with a wide geographic distribution.

Telecommunication companies

routinely generate and store enormous amounts of high-quality data, have a very large customer base, and operate in a rapidly changing and highly competitive environment.[2]

Telecommunication companies utilize data mining to improve their marketing efforts, identify fraud, and better manage their telecommunication networks. However, these companies also face a number of data mining challenges due to the enormous size of their data sets, the sequential and temporal aspects of their data, and the need to predict very rare events—such as customer fraud and network failures—in real-time.

Telecommunication companies maintain data about the phone calls that traverse their networks in the form of call detail records, which contain descriptive information for each phone call. Despite of all the hype surrounding the cloud, enterprise customers are still reluctant to deploy their business in the cloud.[3] Security is one of the major issues which reduces the growth of cloud computing and complications with data privacy and data protection continue to plague the market. The impact of cloud computing solutions in the telecommunication industry is caused by the decoupling of applications,also known as service provisioning, from networking services, i.e. packet forwarding and routing[5]. This paper describes the various security issues of cloud computing.

2. CLOUD COMPUTING

A cloud has several uses, offering a variety of services and can be deployed in more than one way. Consequently, several definitions of cloud computing have been proposed (Mell & Grance, 2011; Schubert, et al., 2010; Wyld, 2009). [1]Schubert et al. define cloud computing as: a 'cloud' is an elastic execution environment of resources involving multiple stakeholders and providing a metered service at multiple granularities for a specified level of quality (of service). (Schubert, et al., 2010)

There are three main levels of service for users of cloud computing (Mell & Grance, 2011):

- in the Software as a Service (SaaS) model, a client can make use of software applications made available from the cloud provider.

- the Platform as a Service (PaaS) model provides an application programming interface (API) for clients to create and host custom-built applications.PaaS also includes cloud providers offering database management systems such as Amazon SimpleDBiii.

- the Infrastructure as a Service (IaaS) model is the leasing of virtualized computing resources such as processing power, volatile memory and persistent storage space to host virtual machines. IaaS products include Amazon EC2iv, which allows clients to create and launch virtual machines running a variety of operating systems.

The storage of data on these services is not by the user, but instead by the cloud owner. In addition to the different levels of deployment, a cloud can be categorised by its organizational deployment, with consequent impact on the geographical location and storage architecture of data held:

- in a private cloud, the infrastructure is operated solely by the organization who owns the cloud. This cloud will likely be found within the same premises as the owning organization and be within its administrative control, and include only that same organization's data;

- a community cloud is shared between several organizations, either because of a common organizational goal, or in order to pool IT resources. Community clouds may be located within one or more of the community organization's premises, and will be administered by the community;

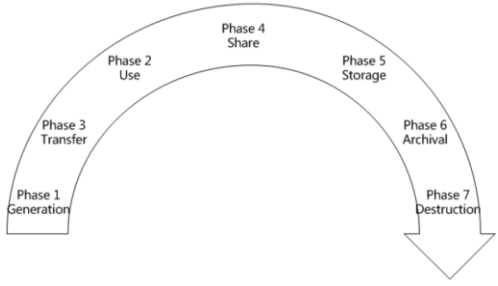
- public clouds will usually be owned by a provider organization, which will maintain the cloud facilities in one or more corporate data centres. The administrative control of the cloud resources will therefore reside with the provider, rather than the user. Consumers will lease virtual storage and compute resources from the provider as required. A public cloud will therefore likely contain data from more than one user; and

- a hybrid cloud is a composition of two or more of the above deployment options. Hybrid clouds can be used to provide load balancing to multiple clouds. For example, an organization may have exhausted the available resources within its private cloud, and so incorporate resources available on lease from a public cloud.

The delivery models form the core of the cloud and they exhibit certain characteristics like on-demand self-service, multi-tenancy, ubiquitous network, measured service and rapid elasticity which are shown in the top layer.[3] These fundamental elements of the cloud require security which depends and varies with respect to the deployment model that is used, the way by which it is delivered and the character it exhibits. Some of the fundamental security challenges are data storage security, data transmission security, application security and security related to third-party resources.

3. Data Security and Privacy Issues

The content of data security and privacy protection in cloud is similar to that of traditional data security and privacy protection. It is also involved in every stage of the data life cycle.[4]Data life cycle refers to the entire process from generation to destruction of the data. The data life cycle is divided into
seven stages.



3.1. Data Generation

Data generation is involved in the data ownership. In the traditional IT environment, usually users or organizations own and manage the data. But if data is to be migrated into cloud, it should be considered that how to maintain the data ownership.For personal private information, data owners are entitled to know what personal information being collected, and in some cases, to stop the collection and use of personal information.

3.2. Data Transfer

For data transmission across enterprise boundaries, both data confidentiality and integrity should be ensured in order to prevent data from being tapped and tampered with by unauthorized users. In other words, only the data encryption is not enough. Data integrity is also needed to be ensured. Therefore it should ensure that transport protocols
provide both confidentiality and integrity.

3.3. Use

Due to the multi-tenant feature of cloud computing models, the data being processed by cloud based applications is stored together with the data of other users. Unencrypted data in the process is a serious threat to data security.

3.4.Share

The data owners can authorize the data access to one party, and in turn the party can further share the data to another party without the consent of the data owners. Therefore, during data sharing, especially when shared with a third party, the data owners need to consider whether the third party continues to maintain the original protection measures and usage restrictions.

3.5.Storage

The data in the cloud may be divided into: (1) The data in IaaS environment, such as Amazon's Simple Storage Service;
(2) The data in PaaS or SaaS environment related to cloud based applications. In order to ensure the effective of encryption, there needs to consider the use of both encryption algorithm and key strength. As the cloud computing environment involving large amounts of data transmission, storage and handling, there also needs to consider processing speed and computational efficiency of encrypting large amounts of data. Ideally,the keys are managed by the data owners. But at present, because the users have not enough expertise to manage the keys, they usually entrust the key management to the cloud providers. As rapid elasticity feature of cloud computing resources, the users don't know where their data is being stored. To migrate out of or into the cloud storage will consume the user's network utilization (bandwidth) and an amount of time. And some cloud providers, such as Amazon, will require users to pay transfer fees. How to directly verify the integrity of data in cloud storage without having to first download the data and then upload the data is a great challenge. As the data is dynamic in cloud storage, the traditional technologies to ensure data integrity may not be effective.

3.6.Archival

Archiving for data focuses on the storage media, whether to provide off-site storage and storage duration. If the data is stored on portable media and then the media is out of control,the data are likely to take the risk of leakage. If the cloud service providers do not provide off-site archiving, the availability of the data will be threatened.

3.7.Destruction

Due to the physical characteristics of storage medium, the data deleted may still exist and can be restored. This may result in inadvertently disclose of sensitive information.

4. Issues in SAAS

In the SaaS model, enterprise data is stored at the SaaS provider's data center, along with the data of other enterprises.[3]

The cloud provider might, additionally, replicate the data at multiple locations across countries for the purposes of maintaining high availability. The following key security elements should be carefully considered as an integral part of the SaaS application development and deployment process:

- Data security- The SaaS vendor must adopt additional security checks to ensure data security and prevent breaches due to security vulnerabilities in the application or through malicious employees. This involves the use of strong encryption techniques for data security and fine-grained author- ization to control access to data.
- Network security- All data flow over the network needs to be secured in order to prevent leakage of sensitive information. This involves the use of strong network traffic encryption techniques such as Secure Socket Layer (SSL) and the Transport Layer Security (TLS) for security.
- Data locality- Due to compliance and data privacy laws in various countries, locality of data is of utmost importance in many enterprise architecture. There's also the question of whose jurisdiction the data falls under, when an investigation occurs.
- Data integrity- SaaS applications are multi-tenant applications hosted by a third party. SaaS applications usually expose their functionality via XML based APIs. Also, in SOA based environments, many on-premise applications expose their functionality via SOAP and REST web services as well. One of the biggest challenges with web services is transaction management. At the protocol level, HTTP (Hyper Text Transfer Protocol) does not support transactions or guaranteed delivery, so the only option is to implement these at the API level.
- Data segregation- As a result of multi-tenancy multiple users can store their data using the applications provided by SaaS. In such a situation, data of various users will reside at the same

location. Intrusion of data of one user by another becomes possible in this environment. A SaaS model should therefore ensure a clear boundary for each user's data. The boundary must be ensured not only at the physical level but also at the application level.

- Data access- The SaaS model must be flexible enough to incorporate the specific policies put forward by the organization. The model must also be able to provide organizational boundary within the cloud because multiple organization will be deploying their business processes within a single cloud environment.[9]
- Authentication and authorization- With SaaS, the software is hosted outside of the corporate firewall. Many a times user credentials are stored in the SaaS providers' databases and not as part of the corporate IT infrastructure.
- Data confidentiality- Cloud computing involves the sharing or storage by users of their own information on remote servers owned or operated by others and accesses through the Internet or other connections. The entire contents of a user's storage device may be stored with a single cloud provider or with many cloud providers. Whenever an individual, a business, a government agency, or any other entity shares information in the cloud, privacy or confidentiality questions arise.
- Web application security- SaaS is software deployed over the internet and/or is deployed to run behind a firewall in local area network or personal computer. [8]The key characteristics include Network-based access to, and management of, commercially available software and managing activities from central locations rather than at each customer's site, enabling customers to access application remotely via the Web.
- Data breaches- Since data from various users and business organizations lie together in a cloud environment, breaching into the cloud environment will potentially attack the data of all the users. Thus the cloud becomes a high value

target[10].

- Virtualization vulnerability- Ensuring that different instances running on the same physical machine are isolated from each other is a major task of virtualization which is not met completely in today's scenario. The other issue is the control of administrator on host and guest operating systems.
- Availability- A multi-tier architecture needs to be adopted, supported by a load-balanced farm of application instances, running on a variable number of servers. Resiliency to hardware/software failures, as well as to denial of service attacks, needs to be built from the ground up within the application.
- Backup- The SaaS vendor needs to ensure that all sensitive enterprise data is regularly backed up to facilitate quick recovery in case of disasters. Also the use of strong encryption schemes to protect the backup data is recommended to prevent accidental leakage of sensitive information.
- Identity management and sign-on process- ID management is a broad administrative area that deals with identifying individuals in a system (such as a country, a network or an organization) and controlling the access to the resources in that system by placing restrictions on the established identities.

5. Challenges specific to telecom sector

In addition to the above challenges,there are a number of additional issues that are specific to the telecommunication industry in regard to: legal and regulatory, business and market, and technical aspects. In particular, the author [Leonardo et.al] evaluate the legal and regulatory challenges arising from data protection legislation, data retention legislation, lawful interception and protection of the national infrastructure. The EU legislation requires the data controller to know who processes which data and for what purpose.In the current model, the telecommunication systems are located where law enforcement has jurisdiction. When it comes to cloud services this is not necessarily true, as services may be produced in different jurisdictions, such as one or more different countries. Firstly, the telecommunication provider must

be able to deliver the data and must therefore contractuallyass ure that the cloud service provider will make the data available upon request. Secondly, a challenge arises when the legal framework in two or more jurisdictions imply different rules concerning the preconditions under which data may beprovided. It is of vital interest to the telecommunication providers that they provide their customers with access to cloud computing services without actually losing the control over the customers' data and identity to the cloud computing provider.

The telecommunication providers challenge is to share only the minimum amount of customers data with the cloud service providers in order to prevent them from taking control over the customers and their data.[6] It is a major challenge to manage and control such an environment where all components need to trust each other. Another reason for concern is the control plane of the mobile networks whose reliable operation depends on certain security assumptions, such as having only trusted base stations.

One of the challenges for integrating cloud computing and telecommunications services is the development of standardized interfaces.[7] Interoperability is crucial for the commoditization of cloud computing services, and the standardization of interfaces is the related challenge. Furthermore, in the case of integration of multiple or cascading cloud computing platforms, or the advent of cloud computing marketplaces, the trust assessment of the network and of its services has the potential to turn into a complex and challenging problem.

6. Current Cloud Data Management

Whereas resource control in Grid environments is enforced by system administrators, the situation is different in the context of Clouds, where usershave the control of the remote virtual resources.[14] This raises some additional security concerns about control policies, as clients have to rely on the security tools of the Cloud service providers.To take the example of Nimbus , GSImechanisms are used to authenticate and authorize client requests, VM image files, resource requests,reservation and usage times for users. Authorization is done based on the role information contained in the issuer's Virtual Organization Membership Service credentials and attributes. More security mechanisms (e.g., intrusion detectors) are needed to protect the virtual host from attacks.Hadoop Distributed File System (HDFS)[15] , the default back end for the Hadoop Map/Reduce framework, implements security as a rudimentary file and directory permission mechanism. Since both

clients and servers need to be authenticated for keeping data secure from unauthorized access, HDFS relies on Kerberos as the underlying authentication system. Moreover,even if a typical user does not have full access to the filesystem, HDFS is vulnerable to various attacks that it cannot detect, such as Denial of Service. In Amazon Simple Storage Service (S3)[16] , the data storage and management infrastructure for Amazon's Elastic Compute Cloud [17], the users can decide how, when and to whom the information stored in Amazon Web Services is accessible. Amazon S3 API provides access control lists (ACLs) for write and delete permissions on both objects and objects containers, denoted buckets. Regarding data transfers, data in transit is protected from being intercepted, as the access is allowed only via SSL encrypted endpoints.Uusers may encrypt them before uploading so as to make sure the data are not tampered with.

7. Conclusion

As described in the paper, though there are extreme advantages in using a cloud-based system, there are yet many practical problems which have to be solved. Cloud computing is a disruptive technology with profound implications not only for Internet services but also for the IT sector as a whole. Still, several outstanding issues exist, particularly related to service-level agreements (SLA), security and privacy, and power efficiency. As described in the paper, currently security has lot of loose ends which scares away a lot of potential users. The authors have listed a set of privacy,security and trust issues that must be taken into account before cloud computing solutions can be fully integrated and deployed by telecommunication providers.There is also the need for legal issues regarding clouds including data retention and privacy laws to be re-examined, following the widespread adoption of cloud technologies.

8. References

[1] George Grispos,Tim Storer, William Glisson."Calm Before the Storm: The Challenges of Cloud Computing", *Digital Forensics*.

[2] Gary M Weiss,"Data Mining in the Telecommunications Industry",*IGI Global* 2009 (486-491)

[3] S. Subashini , V.Kavitha ," A survey on security issues in service delivery models of cloud computing " *Network and Computer Applications*,July 2010.

[4] Deyan Chen ,Hong Zhao,"Data Security and Privacy Protection Issues in Coud Computing",*2012 International Conference on Computer Science and Electronics Engineering*.

[5] Leonardo A. Martucci, Albin Zuccatoy, Ben Smeetsxz, Sheikh M. Habibk, Thomas Johanssonz, Nahid Shahmehr," Privacy, Security and Trust in Cloud Computing-The Perspective of the Telecommunication Industry".

[6] N. Klapisz, Ed.," Valuation drivers in the telecommunication industry" *Ernst & Young, 2011.*

[7] S. M. Habib, S. Ries, and M. Muhlhauser, "Towards a trust management system for cloud computing," *Int. Joint Conf. of IEEE TrustCom/IEEE ICESS/FCST*,IEEE Computer Society, 2011, pp. 933–939.

[8] Zalewski M,"Browser security handbook",2009 /http://code.google.com/p/browsersec/S [accessed on:19February2010].

[9] Blaze M, Feigenbaum J, loannidis J, Keromytis AD,"The role of trust management in distributed systems security, secure Internet programming, issues for mobile and distributed objects" Berlin: *Springer-Verlag*; 1999. p. 185-210.

[10] Bernard Golden,"Defining private clouds",2009 /http://www.cio.com/article/492695/Defining_Private_Clouds_Part_OneS [accessed on:11January2010].

[11] Mell, P, & Grance, T.," The NIST Definition of Cloud Computing", *National Institute of Standards and Technology, Special Publication 800-145* Retrieved February 2, 2012, from http://csrc.nist.gov/publications/nistpubs/800-145/sp800-145.pdf00-145/sp800-145.pdf

[12] Wyld, D. C. "Moving to the Cloud: An Introduction to Cloud Computing in Government",*IBM Center for the Business of Government - E-government series report* Retrieved February 2, 2012, from http://www.businessofgovernment.org/sites/default/files/moving%20to%20the%20cloud.pdfs/default/files/moving%20to%20the%20cloud.pdf

[13] Schubert, L, Jeffery, K., & Neidecker-Lutz, B,"The Future of Cloud Computing: Opportunities For European Cloud Computing Beyond 2010", *European Commission Information and Society Theme - Expert Group Report* Retrieved February 2, 2012, from http://cordis.europa.eu/fp7/ict/ssai/docs/cloud-report-final.pdf

[14] Cristina Basescu, Catalin Leordeanu, Alexandru Costan,Alexandra Carpen-Amarie,Gabriel Antoniu," Managing Data Access on Clouds",*The 25th International Conference on Advanced Information Networking and Applications* (AINA-2011) (2011) 459-466" DOI : 10.1109/AINA.2011.61

[15]"HDFS. the Hadoop distributed file system,"http://hadoop.apache.org/common/docs/r0.20.1/hdfs_design.html.

[16] Amazon Simple Storage Service (S3). http://aws.amazon.com/s3/.

[17] Amazon Elastic Compute Cloud (EC2), http://aws.amazon.com/ec2/.