Cloud Computing Security and Privacy

Muhammad Adeel Javaid

Member Vendor Advisory Council CompTIA
*Corresponding Author: ajaviad@gmail.com

Copyright © 2014 Horizon Research Publishing. All rights reserved.

Abstract The cloud computing paradigm is still evolving, but has recently gained tremendous momentum. However, security and privacy issues pose as the key roadblock to its fast adoption. In this paper we present security and privacy challenges that are exacerbated by the unique aspects of clouds and show how they're related to various delivery and deployment models. We discuss various approaches to address these challenges, existing solutions, and work needed to provide a trustworthy cloud computing environment.

Keywords Cloud Security, Cloud Threats, Cloud Privacy

1. Introduction

As with any change in IT infrastructure, it creates new risk and new opportunities, Cloud Computing is not any different from that. Shared, on-demand nature of cloud computing put it to some unique risk that was not experienced before. (Chonka A, et al., 2010) mentioned, as security experts they can see the same mistakes that occurred during the development of Internet is happening here too. They pointed out functionality and performance got higher priority than security. (Subashini S, 2010) has done a very good survey on the security issues in different service delivery models but their main focus was not on the top security threats or concerns which new customers need to know before knowing cloud architecture or service delivery models.

Cloud computing inherits the characteristics of Grid computing but also provides a dynamic and flexible environment, that is scalable, robust and resistant to rapid changes of conditions. This is achieved through its inherent characteristics, such as automatic recovery, self-monitoring, self-management, automatic reconfiguration, the possibility of setting SLAs, and high scalability.

The use of the Cloud suggests that the data of the user are stored and maintained in a remote site. There is therefore a major challenge to create cloud-based services that reduce the risk of security and privacy. It is necessary to include privacy mechanisms in early designs and to try and find patching solutions afterwards.

It is a fact that both private businesses and public sector organizations (Bezos 2008; Amazon EC2 Feature) have awareness for the necessity of privacy. In addition, there is always the possibility to design cloud computing infrastructure with a guarantees on privacy and security similar to developed technologies such as Web Services (TC3 2011) and Grid Computing. When evaluating privacy in the usage of Cloud Computing, there are two main factors that should be taken into consideration. The first factor is to ensure privacy during communication. This is achieved through the usage of encryption in communication with the Cloud. Data Encryption is used to ensure the confidentiality of personal information. Encryption methods such as symmetric key cryptography or public key cryptography with the usage of certificates make sensitive personal data accessible only to those who are duly authorized, thus ensuring privacy in digital communications.

The other factor that ensures data privacy is the usage of equivalent mechanisms for storing data in the Cloud, so that it is only readable by authorized users.

To conduct our research we will undertake a study of two well known Cloud Service architectures and their design about security and privacy of data. Since security is usually achieved through the usage of cryptography in communication, we will also examine the methods including the proposed generic secure architectures that could be implemented for ensuring privacy in the storage systems of the Cloud. We will also try to focus on top threats and to propose their workable solutions.

2. Amazon Web Services

Amazon Web Services (AWS) “delivers a highly scalable cloud computing platform with high availability and dependability, and the flexibility to enable customers to build a wide range of applications” (Amazon AWS). To enable security and privacy from end to end, AWS are provided in accordance with best security practices and appropriate security features.

Security within Amazon EC2 is provided on various layers: the operating system (OS) of the host system, the virtual instance operating system or guest OS, a firewall, and signed API calls. Each of these items builds on the

capabilities of the others. The goal is to protect data contained within Amazon EC2 from being intercepted by unauthorized systems or users and to provide Amazon EC2 without sacrificing the flexibility in configuration that customers demand.

In order to secure data at rest, physical security and data encryption are necessary. Amazon is trying to protect customer's data at rest by using multiple layers of physical security measures. For example, only specific people of Amazon personnel have physical access to Amazon datacenters. However, the data in S3 is not automatically encrypted, while every user must first encrypt its sensitive data and then upload it to Amazon S3. Amazon encourages users to do all these. When an object is deleted from Amazon S3, the removal starts immediately and lasts several seconds. After that, there is no remote access to the deleted object.

Users maintain full control over who has access to their data. They may grant access to their Amazon S3 data to other AWS accounts by AWS Account ID or email, or DevPay Product ID. Users can also grant access to their Amazon S3 data to all AWS Accounts or to everyone (enabling anonymous access). Figure-1 below shows architecture of Amazon Web Services.

Figure 1. AWS Architecture

3. Eucalyptus

The Eucalyptus is open source software which was created to provide computing services using clusters of computers. The current interface is compatible with the interface of Amazon EC2. The Eucalyptus is implemented using common tools and basic technologies of Linux and Web services, in order to be easily installed and maintained (Canonical 2011). It started originally as a university project on the "Computer Science Department of the University of California, Santa Barbara", but now is supported by the company "Eucalyptus Systems" - a company founded by the original creators of Eucalyptus.

The architecture of Eucalyptus is simple, flexible by design and hierarchical. Essentially, the system allows a user to start, monitor, access and terminate virtual machines (Chase 2003 ; Keahey 2005 ; McNett 2007) using a simulation of the Amazon EC2 interface. This means that users of Eucalyptus interact with the system using the exact

same tools and interface used to interact with Amazon EC2. It currently supports virtual machines running on Xen (Nurmi 2009), and KVM (Hoff 2009) hypervisor, but future version are expected to add support for VMware and other virtualization software (Lilly 2009).

The Eucalyptus implements each module of the system in the form of a separate web service. There are subsystems, each one with its own Web-service interface that constitutes the Eucalyptus system (See Figure-2 below).

Figure 2. The hierarchical structure of Eucalyptus

3.1. Node Controller

The Node Controller (NC) runs on each node that is designed to run virtual machine instances. The NC controls uses the software in the node (eg, operating system and hypervisor) according to requests sent from the Cluster Controller.

NC collects information for local node resources, such as the number of cores, the size of memory, the available disk space, and also monitors and the state of local virtual machine instances. The information is forwarded to the Cluster Controller upon request.

The Cluster Controller monitors virtual machine instances on a node using the commands run instance and terminate instance at the NC node. Upon verification of the sender of the request (for example, only the administrator and the user who started the instance is authorized to terminate an instance) and after checking the available resources, the NC executes the command using the hypervisor. To start a new VM, the NC makes a copy of the required files of the VM (the kernel, the file system and memory) from a remote repository or a local cache, creates an access point in the virtual layer network and instructs the hypervisor to start the new instance. To stop a running instance, the NC instructs

the hypervisor to terminate the VM, destroying the access point to the virtual network and deletes the files associated with this instance. It should be noted that in Eucalyptus, various NC in the same cluster form a separately managed entity called Availability Zone.

3.2. Cluster Controller

The Cluster Controller (CC) usually runs in the front-end computer of the cluster or any computer with network connection to both the nodes running NC and the computer running the Cloud Controller (CIC). Many of the features of CC are similar to the functions of NC but have to do with groups of VM. The CC has three main operations: scheduling incoming requests for new instances, monitoring the virtual network layer (VDE) of instances and collecting / reporting information for a group of NC.

When the CC receives a number of requests for new VM instances, communicates with each NC using the command describe Resources and sends the command run instances to the first NC that has enough free resources. When a CC receives the describe Resources command, it also receives information about a number of features (processor cores, memory and disk space) that describe the resources required by the new instance. By using this information, the CC calculates the number of parallel instances of this type that can reside in the NC and reports this number to the CIC.

3.3. Storage Controller (WALRUS)

The Eucalyptus system includes the Walrus, a data storage service that uses Web services (Achis2, Mule) and has a compatible interface with Amazon Simple Storage Service (S3). The Walrus is implemented through an HTTP interface called REST and a SOAP interface which is compatible with S3. The Walrus offers two types of functions:

Users who have access to Eucalyptus can use the Walrus to transfer data to and from the cloud, as well as instances that have been launched in nodes.

It also functions as storage for VM images. The images of the file system, the core and the memory used to create VM instance in nodes can be transferred to the Walrus storage and then be used by nodes.

Users may use the tools of the S3 (created by Amazon or others) to transfer data to and from the Walrus. The system uses the certificates stored in the database of the Cloud Controller.

Like in the S3, the Walrus supports parallel and serial data transfers. To strengthen multithreading, the Walrus does not undertake the task of data validity checking. As in S3, users must certify that the data transferred to the Walrus is valid and that there are no duplicate records in the same objects. For that reason, the Walrus returns an MD5 checksum of the object saved. Once the user has been identified as a valid user of Eucalyptus, all requests for reading and writing are sent by HTTP.

The Walrus also serves as a storage service. The images of

the file system, the kernel and memory of VM are incorporated into packets and are transported using the EC2 tools offered by Amazon. These tools compress and then encrypt the image with the use of certificates and then split the image into several pieces as described in the "image description file" or otherwise manifest of the VM. The Walrus has a duty to verify the validity and decrypt the image transferred from users. When a node controller requests an image from the Walrus, before the instantiation of the VM, it sends a transfer request that is validated with the use of appropriate certificates. Then the image is checked for validity, it is decrypted and then is sent to the node. In order to improve efficiency and because usually VM images are quite large, the Walrus has a cache with already decrypted images. The cache is deleted after some time or when the image manifest is overwritten.

3.4. Cloud Controller

The components of the Eucalyptus that were described above are managed by the Cloud Controller (CIC). The CIC is a collection of web services that may be divided in three categories:

3.5. Resource Services

These are services that manage the distribution of system resources to different functions, allowing users to manage the properties of virtual machines, and record both system components and virtual resources.

3.6. Data Services

These services check the user and system data and provide a configurable environment for the user to make requests for resource allocation.

3.7. Interface Services

Manage protocols transformations, user authentication and provide tools for managing the system.

3.8. Virtual Control Overlay

In Eucalyptus, all virtual machines are interconnected and at least one has internet access so that the owner can connect and interact with them. In a cloud, where the VM may be used by many different users, only virtual machines belonging to the same group are allowed to communicate with each other.

4. Security Issues and Challenges

External network attacks in the cloud are increasing at a notable rate. Malicious user outside the Cloud often performs DoS or DDoS attacks to affect the availability of

Cloud services and resources. Port scanning, IP spoofing, DNS poisoning, phishing are also executed to gain access of Cloud resources. A malicious user can capture and analyze the data in the packets sent over this network by packet sniffing. IP spoofing occurs when a malicious user impersonates a legitimate users IP address where they could access information that they would not have been able to access otherwise. Availability is very important. Not having access to services when needed can be a disaster for anyone especially in the case of being denied service. This can occur when exhaustion of the host servers causes requests from legitimate consumers to be denied. This can cost a company large amounts of money and time if the services they depend on to operate are not available.

Internal attacker (authorized user) can easily get access to other user's resources without being detected. An insider has higher privileges and knowledge (related to network, security mechanism and resources to attack) than the external attacker. Therefore, it is easy for an insider to penetrate an attack than external attackers.

4.1. Vulnerabilities

In Cloud, existing vulnerabilities, threats, and associated attacks raise several security concerns. Vulnerabilities in Cloud can be defined as the loopholes in the security architecture of Cloud, which can be exploited by an adversary via sophisticated techniques to gain access to the network and other infrastructure resources. Figure-3 below gives an overview of factors contributing to risk according to the Open Group's risk taxonomy.

The Open Group's taxonomy uses the same two top-level risk factors as ISO 27005: the likelihood of a harmful event (here, *loss event frequency*) and its consequence (here, *probable loss magnitude*). The probable loss magnitude's subfactors (in Figure 3 above) influence a harmful event's ultimate cost. The loss event frequency subfactors

(on the left) are a bit more complicated. A loss event occurs when a threat agent (such as a hacker) successfully exploits a vulnerability. The frequency with which this happens depends on two factors:

- The frequency with which threat agents try to exploit a vulnerability. This frequency is determined by both the agents' motivation (What can they gain with an attack? How much effort does it take? What is the risk for the attackers?) and how much access ("contact") the agents have to the attack targets.
- The difference between the threat agents' attack capabilities and the system's strength to resist the attack.

This second factor brings us towards a useful definition of vulnerability.

4.2. Defining Vulnerability

According to Open Group's risk taxonomy "Vulnerability is the probability that an asset will be unable to resist the actions of a threat agent. Vulnerability exists when there is a difference between the force being applied by the threat agent, and an object's ability to resist that force."

We will now discuss major Cloud specific vulnerabilities, which pose serious threats to Cloud computing.

4.2.1. Abuse and Suspicious Use of Cloud

Users receive various types of services from cloud providers, such as bandwidth and storage capacity. Some service providers provide their users with a free trial period that gives hackers and malicious users an opportunity to access the cloud with nefarious intentions. Spammers and other criminals take advantage of the simple procedures, relatively anonymous access and convenient registration to access the cloud and launch malicious attacks (Hicks, 2009a).

Figure 3. Open Group's risk taxonomy

Some of the impacts of their attacks include cracking passwords, decoding and building rainbow tables. Hackers and cybercriminals also take advantage of the fact that cloud service providers have limited fraud detection capabilities to launch dynamic attack points, conduct malicious data hosting and execute botnet commands. For example, some hackers use flash files and other content rich applications that allow them to install malware by utilizing users' browsers (Spitzner, 2011)

To lessen this threat, service providers need to establish stricter initial registration and validation processes. Credit card fraud monitoring and coordination efforts should be enhanced, as well as stepping up introspection of customer network traffic. Another area requiring rigorous monitoring is the public blacklists for the service provider's network blocks. This will assist in protecting the integrity of clients' data, some of which could be sensitive (Backhouse & Dhillon, 2000).

4.2.2. Insecure Interfaces and Application Programming Interfaces

Cloud customers are able to interact and manage cloud services through interfaces and APIs. Service providers need to secure their service models because they play an important role in the orchestration, provisioning, and management and monitoring of the processes of running the cloud. Customers on the other hand, must be aware of the security risks associated with the use, implementation and management as well as monitoring of the cloud services (Schreiber, 2004).

The security of these service models affects the security and availability of cloud services, and therefore, they should include features of authentication, access controls, encryption and monitoring of activities. A weak set of APIs and interfaces results to a variety of security issues, including malicious or unidentified access, API dependencies, inflexible access controls, reusable tokens/passwords, limited monitoring/logging capabilities, anonymous access, clear text transmission and/or authentication of content and improper authorizations. These security issues affect the confidentiality, integrity, availability and accountability of a data in the cloud. The security model of cloud providers interfaces should be analyzed to ensure they are effective. Service providers should implement strong authentication processes and access controls and in concert with encryption transmission. They should also be aware and understand the chain associated with the API (Flavio & Roberto, 2010).

4.2.3. Malicious Insiders

Advanced persistent threats are some of the threats facing cloud computing in this era. Insider attacks are becoming a common occurrence in this technology and are orchestrated by employees at the provider or user's site (Anderson & Rainie, 2010). A malicious insider can access cryptographic keys and files as well as passwords that can allow them to participate in information theft, damage and fraud.

Insiders can also by-pass or bend of security control and engage in other acts of security breaches, and hence,

compromise the security controls established to safeguard information systems from attacks against integrity, confidentiality and availability of IT systems, networks and the data in the cloud. The problem of insider attacks has been on the rise especially due to lack of transparency in the processes and procedures of providing cloud services (Underwood, 2012). This means that a provider may not reveal how access is granted to employees and how their access is monitored and how reporting and policy compliance is analyzed.

Additionally, customers may not understand the hiring practices of their service providers that could make room for an adversary, hackers or other cyber-criminals to steal confidential information or even take control of the cloud. The level of access granted could be an avenue for accessing confidential data or taking control of the cloud services with little or no risk of detection. Malicious insider attacks pose a great threat to the brand reputation and productivity as well as the financial well-being of an organization (Underwood, 2012). The overall premise behind insider attack is the use of much unsophisticated methods of tricking or coercing users into doing something they would normally not do. For example, an insider may send a phishing email to a user who ends up downloading a malicious code onto their computer or whichever device they are using. According to Spitzner (2011), one hundred percent of all successful compromises are a result of an insider assisting the attacker.

It is imperative that organizations implement policies to mitigate the threat of an insider taking part in an advanced persistent threat. Separation of duties is an important concept in the internal controls, though sometimes difficult and costly to implement. This objective is achieved by allocating tasks and associate privileges to multiple people for a specific security process. Separation of duties in the IT organization is fundamental and it's the mandate of firms to apply it for regulatory purposes (Backhouse & Dhillon, 2000). As a result, IT organizations should lay more emphasis on separation of duties in all their functions, especially security.

Separation of duties achieves two objectives in relation to security. First, it prevents conflict of interest, wrongful acts, errors, fraud and abuse that occur in case of conflict of interest (Backhouse & Dhillon, 2000). The other very important objective is the detection of control failures including information theft, security breaches and by-pass or bending of security controls.

4.2.4. Virtualized Technology

Due to the cloud virtualization, service providers reside their users' applications on virtual machines based on shared infrastructure. The virtual machines are not designed to accommodate a multi—tenant architecture and are based on the physical hardware of the cloud provider (Karthick et. al., 2011). Overlooked flaws in technology have enabled guest operating systems to obtain unauthorized levels of control and influence on the platform. In order to maintain the security of the users, service providers are isolating their

VMs to prevent one malicious VM from affecting the others under the same provider.

In order to provide virtual memory as well as schedule CPU caches policies to VMs, VMs are managed by hypervisor. In addition, since the hypervisor is the main way of managing a virtualized cloud platform, hackers are targeting it in order to gain access to VMs and physical hardware because it resides in between the two (Siani, 2009). This means that an attack on hypervisor can damage both the VMs and hardware. Service providers should employ strong isolation to ensure that VMs are not able to access or affect the operations of other users running under the same cloud service provider. Moreover, service providers should implement the best practices for configuration and monitor the environment for unauthorized activity. The authentication and access operations should be strengthened in addition to enforcing procedures for patching and vulnerability remediation. Vulnerability scanning and configuration audits should also be promoted.

4.2.5. Data Loss or Leakage

Data loss refers to compromised data that may include deletion or alteration of records without first making a backup of the original content which can be intentional or unintentional (Farell, 2008). It can also occur in the course of unlinking part of a record from the larger context, unauthorized access of confidential data and loss of an encoding key. Data loss and leakage can result from various issues including: operational failures; inconsistent use of encryption keys; authorization and audit controls; political issues; disposal challenges; insufficient authentication; inconsistent software keys; persistence and remanence challenges; risk of association; disaster recovery and unreliable data storage (Tim et.al., 2009). Unreliable data storage occurs when data is stored on unreliable media that will be impossible to recover if data is lost. Inconsistent use of encryption keys will lead to both loss of the data and unauthorized access of the data by illegal users, resulting to destruction of sensitive and confidential information.

An example of data loss is the case of twitter hacks. Twitter online hackers accessed accounts and numerous sensitive corporate documents housed in Google's online web office service Google Docs were stolen. The security of twitter data was not efficient enough, since the entire company data was only a single password crack away from discovery. This shows the effect of a data leakage on a brand, its reputation and the trust of partners, employees and users. Loss of core intellectual property has noncompliance and legal consequences beside the competitive and financial implications involved (Farell, 2008).

Cloud providers need to specify backup strategies and generate strong key generation, management, storage and destruction practices. The integrity of data in transit should be protected and such data should be encrypted if possible (Flavio & Roberto, 2010). API access should be strongly controlled, while providers should be expected to wipe persistent media before releasing it into the pool.

4.2.6. Account or Service Hijacking

Account or service hijacking occurs when hackers gain unauthorized access and control users' accounts usually by using stolen credentials. Such attacks include fraud, phishing and exploitation of software vulnerabilities. Attackers can use stolen credentials and spy on users' sensitive activities, return falsified information, redirect information to illegitimate sites, manipulate data and hence compromise the integrity, confidentiality and availability of the cloud computing services (George, 2011). Attackers are using users' accounts as a new base to leverage cloud service providers' reputation by launching constant attacks. Monitoring should be proactive to detect unauthorized activity and prohibit users and services from sharing credentials. Clients should also understand the security policies of cloud providers.

4.2.7. Unknown Risk Profile

With the innovation of cloud computing, organizations are less involved with acquiring and/or maintain hardware and software. However, users need to understand software versions, code updates, intrusion attempts and other security practices (Hicks, 2009a). While these features and their functionality may be well advertised when adopting cloud-computing services, the details of compliance to the internal security procedures, patching, auditing, logging and configuration hardening may not be clear. Users need clarification about how and where their data and related logs are stored since an unknown risk profile may include serious threats. Infrastructural details should be partially or fully disclosed, as well as should the logs and data.

4.2.8. Session Riding and Hijacking

Session hijacking refers to the use of a valid session to obtain unauthorized access to the information in a computer system or theft of a user authentication cookie used in accessing a remote server. The cookie is also used in web application technologies weaknesses in the web application structure and is easily accessible to hackers, giving them an opportunity to carry out a wide variety of malicious activities (Brohi & Bamiah, 2011). Session riding refers to hackers sending commands to web applications on behalf of users by tricking the targeted user. Hackers send an email to the user or coerce him to visit a specially crafted website.

Session riding executes online transactions, deletes user data, and sends spam to an intranet through the internet. Other outcomes of session riding are changes to the system and network configurations as well as opening the firewall (Tim et. al., 2009). Additionally, web technologies and refinement evolve new ways of compromising data, provide access to otherwise secure networks and pose threats to the smooth running of online business.

4.2.9. Virtual Machine Escape

Cloud computing servers make use of the same web applications, OS and enterprise as localized VMs and physical servers. The probability of a malware or hacker to

remotely exploit vulnerabilities in these systems and applications is a great threat to virtualized cloud computing environments (Hartig, 2009). Locating multiple VMs jointly increases the attack surface and the risk of compromise from VM – to VM. Security systems should be capable of detecting intrusion and malicious activity at VM level, regardless of where the VM is located within the virtualized cloud environment.

VM escape is a significant vulnerability that enables guest-level VM to attack the host VM. An attacker runs a code on a VM allowing an OS to run within it to break out and interact with the hypervisor (Schreiber, 2004). This enables an attacker to access the host OS and all other VMs running on that particular host. The complexity of Hypervisors and VMs may increase the threat to attack surface that weakens security, such as check pointing, migration of VMs and paging.

4.2.10. Reliability and Availability of Service

Another significant cloud computing vulnerability is reliability and availability. In the event of glitches in infrastructure such as failure to cloud storage, data may be inaccessible or even lost (Flavio & Roberto, 2010). With more services being developed on top of cloud, computing infrastructures, a failure or outage can create a domino effect by interrupting many internet based applications and services. This raises questions such as what forms of settlement exist for stakeholders in cases of failures, what is the responsibility of cloud providers and what procedures should be put in place to overcome these issues.

4.2.11. Insecure Cryptography

Organizations have come up with algorithms and cryptographic mechanisms as means to secure their data. However, attackers have discovered means to decode these security mechanisms and hack them. This is so because it is common for crucial flaws to exist in the implementation of cryptographic algorithms, which can change strong encryption into weak encryption that is vulnerable to attack (Spitzner, 2011).

Although VMs provide more flexible and efficient set-up than traditional servers in cloud providers' data centers, they still lack enough access to generate random numbers required to properly encrypt a data, and this becomes a big problem. How do computers generate truly random numbers that cannot be replicated or guessed? In PCs, mouse movements and key strokes are monitored by an OS to gather bits of a data that is collected in an entropy pool (Underwood, 2012). In servers with neither a mouse nor a keyboard,

random numbers are pulled from the movements of the computer's hard drive. According to Schreiber (2004), the random numbers gathered from the movements of VMs' internal clock is not enough to generate strong encryption keys.

4.2.12. Data Protection and Portability

Services are offered on contractual bases between a client and the provider. However, it is not clear what happens when the contract is terminated and the client does not want to continue. Is the client's sensitive data going to be deleted or is it going to be misused by the provider. The other issue is what would happen to the services and the client's data if the provider were to go out of business for any reason. Data portability therefore becomes a main weakness in cloud computing (Karthick et. al., 2011)

4.2.13. Vendor Lock In

New business models and immature service providers has raised the risk of business failure. Lock-in makes a client unable to deal with another provider without a good amount of switching costs due to dependence on one provider for products and services. Before the process of selecting a provider, clients need to be sure of their potential provider (Hartig, 2009). Clients may also remain locked in to one provider due to lack of standards. They do not have the freedom to migrate easily from one provider to another because of the heterogeneous policies and standards set by each provider.

5. Cloud Computing Security Pattern

A cloud computing security pattern is shown in Figure-4 below. Services provided by the Cloud Computing environment are not under direct control and therefore a few control families become more significant. Controls in the CA series increase in importance to ensure oversight and assurance given that the operations are being "outsourced" to another provider. SA-1/4/5 are crucial to ensure that acquisition of services are managed correctly. CP-1 helps ensure a clear understanding of how to respond in the event of interruptions to service delivery. The RA controls are very important to understand the risks associated with the service in a business context, but may be challenging to implement, depending on the supplier and the degree of visibility into their operations.

Figure 4. Cloud Computing Security Pattern

Figure 5. Security-aware cloud architecture

6. A Generic Secure Cloud Architecture

A security-aware cloud architecture is shown in Fig. 5 below. The Internet cloud is envisioned as a massive cluster of servers. These servers are provisioned on demand to perform collective web services or distributed applications using datacenter resources. Cloud platform is formed dynamically by provisioning or de-provisioning, of servers, software, and database resources. Servers in the cloud can be physical machines or virtual machines. User interfaces are applied to request services. The provisioning tool carves out the systems from the cloud to deliver on the requested service.

In addition to building the server cluster, cloud platform demand distributed storage and accompanying services. The cloud computing resources are built in datacenters, which are typically owned and operated by a third-party provider. Consumers do not need to know the underlying technologies. In a cloud, software becomes a service. The cloud demands a high-degree of trust of massive data retrieved from large datacenters. We need to build a framework to process large scale data stored in the storage system. This demands a distributed file system over the database system. Other cloud resources are added into a cloud platform including the storage area networks, database systems, firewalls and security devices. Web service providers offer special APIs that enable developers to exploit Internet clouds. Monitoring and metering units are used to track the usage and performance of resources provisioned.

The software infrastructure of a cloud platform must handle all resource management and do most of the maintenance, automatically. Software must detect the status of each node, server joining and leaving and do the tasks accordingly. Cloud computing providers, like Google and Microsoft, have built a large number of datacenters all over the world. Each datacenter may have thousands of servers. The location of the datacenter is chosen to reduce power and cooling costs. Thus, the datacenters are often built around hydroelectricity power stop. The cloud physical platform builder concerns more about the performance/price ratio and reliability issues than the sheer speed performance.

7. Layered Cloud Architectural Development

The architecture of a cloud is developed at three layers: infrastructure, platform, and application as demonstrated in Fig. 6 below. These three development layers are implemented with virtualization and standardization of hardware and software resources provisioned in the cloud.

The services to public, private, and hybrid clouds are conveyed to users through the networking support over the Internet and intranets involved. It is clear that the infrastructure layer is deployed first to support IaaS type of services. This infrastructure layer serves as the foundation to build the platform layer of the cloud for supporting PaaS services. In turn, the platform layer is a foundation to implement the application layer for SaaS applications. Different types of cloud services demand to apply the resources at the, separately.

The infrastructure layer is built with virtualized compute, storage and network resources. The abstraction of these hardware resources is meant to provide the flexibility demanded by users. Internally, the virtualization realizes the automated provisioning of resources and optimizes the infrastructure management process. The platform layer is for general-purpose and repeated usage of the collection of software resources. This layer provides the users with an environment to develop their applications, to text the operation flows, and to monitor the execution results and performance. The platform should be able to assure the users with scalability, dependability, and security protection. In a way, the virtualized cloud platform serves as a “system middleware” between the infrastructure and application layers of the cloud.

The application layer is formed with a collection of all needed software modules for SaaS applications. Service applications in this layer include daily office management work, such as information retrieval, document, processing, and calendar and authentication services, etc. The application layer is also heavily used by enterprises in business marketing and sales, consumer relationship management (CRM), financial transactions, supply chain management, etc. It should be noted that not all cloud services are restricted to a single layer. Many applications may apply resources at mixed layers. After all, the three layers are built from bottom up with a dependence relationship.

From the provider’s perspective, the services at various layers demand different amounts of function support and resource management by the providers. In general, the SaaS demands the most work from the provider, the PaaS in the middle, and IaaS the least. For an example, Amazon EC2 provides not only virtualized CPU resources to users but also the management of these provisioned resources. Services at the application layer demands more work from the providers. The best example is the Salesforce CRM service in which the provider supplies not only the hardware at the bottom layer and the software at the top layer but also provides the platform and software tools for user application development.

Figure 6. Layered Architectural Development of Cloud Platform

Table 1. Physical and Cyber Security Protection at Cloud/Datacenters

Protection Schemes	Brief Description and Deployment Suggestions
Secure datacenters and computer buildings	Chose hazard-free location, enforce building safety. Avoid windows, keep buffer zone around the site, bomb detection, camera surveillance, earthquake-proof, etc.
Use redundant utilities at multiple sites	Multiple power and supplies, alternate network connections, multiple databases at separated sites, data consistency, data watermarking, user authentication, etc.
Trust delegation and Negotiation	Cross certificates must be used to delegate trust across PKI domains for various data- centers. Trust negotiation among that CAs need to resolve policy conflicts
Worm containment and DDoS Defense	Internet worm containment and distributed defense against DDoS attacks are necessary to secure all datacenters and cloud platforms.
Reputation system for Datacenters	Reputation system could be built with P2P technology. One can build a hierarchy of reputation systems from datacenters to distributed file systems.
Fine-grain file access control	This refers to fine-grain access control at the file or object level. This adds up the security protection beyond firewalls and intrusion detection systems.
Copyright Protection and Piracy prevention	Piracy prevention achieved with peer collusion prevention, filtering of poisoned contents, non-destructive read, alteration detection, etc.
User and data privacy Protection	Use double authentication, biometric identification, intrusion detection and disaster recovery, privacy enforcement by data watermarking, data classification, etc

8. Cloud Defense Methods

Virtualization enhances cloud security. But virtual machines (VMs) add an additional layer of software that could become a single-point of failure. With virtualization, a single physical machine can be divided or partitioned into multiple VMs (e.g. server consolidation). This provides each VM with better security isolation and each partition is protected from DoS attacks by other partitions. Security attacks in one VM are isolated and contained from affecting the other VMs. VM failures do not propagate to other VMs. Hypervisor provides the visibility of the guest OS, with complete guest isolation. Fault containment and failure isolation of VMs provide a more secure and robust

environment.

In Table-1 below, eight protection schemes are listed to secure public clouds and datacenters. Malicious intrusions may destroy valuable hosts, network, and storage resources. Internet anomalies found in routers, gateways, and distributed hosts may stop cloud services. Trust negotiation is often done at the SLA level. Public Key Infrastructure (PKI) service could be augmented with datacenter reputation systems. Worm and DDoS attacks must be contained. It is harder to establish security in the cloud due to the fact all data and software are shared by default. However, it is possible to add social tools like reputation systems to prove any trust model chosen to deploy.

In virtual machine safe cloning, we anticipate an attack by

knowing when an attack has occurred. The snapshot control is based on the defined RPO. We need new security mechanisms to protect cloud. For example, one can apply secured information logging, migrate over secured VLAN, and apply ECC based encryption for secured migration. The sandbox provides a safe execution platform for running the programs. Further, Sandbox can provide a tightly controlled set of resources for the guest operating systems, which allows a security testbed to run the untested code and programs from the untrusted third party vendors. Figure 7 below shows cloud service models and corresponding security measures.

9. Reputation-Guided Protection of Datacenters

Figure-8 gives an overview of design options of reputation systems. The public opinion on the character or standing (such as honest behavior or reliability) of an entity could be the reputation of a person, an agent, a product, or a service. It represents a collective evaluation by a group of people/agents and resource owners. Many reputation systems have been proposed in the past mainly for P2P, multi-agent, or e-commerce systems.

To address the reputation systems for cloud services, a systematic approach is based on the design criteria and administration of the reputation systems. Figure-8 shows a

two-tier classification of existing reputation systems that have been proposed in recent years. Most of them were designed for peer-to-peer or social networks. These reputation systems can be converted for protecting cloud computing applications. In general, the reputation systems are classified as centralized or distributed depending on how they are implemented. In a centralized system, a single central authority is responsible for managing the reputation system, while the distributed model involves multiple control centers working collectively. Reputation-based trust management and techniques for securing P2P and social networks could be merged to defend datacenters and cloud platforms against attacks from the open network.

A centralized reputation system is easier to implement, but demand more powerful and reliable server resources; while a distributed reputation system is much more complex to build. Distributed systems are more scalable and reliable to handle failures. At the second tier, reputation systems are further classified by the scope of reputation evaluation. The user-oriented reputation systems focus on individual users or agents. Most P2P reputation system belongs to this category. In datacenters, the reputation is modeled for the resource site as a whole. This reputation applies to products or services offered by the cloud. Commercial reputation systems have been built by e-Bay, Google, and Amazon in connection with the services they provided. These are centralized reputation systems.

Figure 7. Cloud Service Models and Security Measure

Figure 8. Design options of reputation systems for social networks and cloud platforms.

The distributed reputation systems are most developed by the academic research communities.

Aberer and Despotovic have proposed a model to manage trust in P2P systems. The EigenTrust reputation system was developed at Stanford University using trust matrix approach. The PeerTrust system was developed at Georgia Institute of Technology for supporting e-commerce applications. The PowerTrust system was developed at University of Southern California based on Power law characteristic of Internet traffic for P2P applications. Vu, et al proposed a QoS-based ranking system for P2P transactions.

To redesign the above reputation systems for protecting datacenters offers new opportunities for their expanded applications beyond the P2P networks. Data consistency is checked across multiple databases. Copyright protection secures wide-area content distributions. To separate user data from specific SaaS programs, the providers take the most responsibility in maintaining data integrity and consistency. Users can switch among different services using their own data. Only the users have the keys to access the requested data. The data objects must be uniquely named to ensure global consistency. To ensure data consistency, unauthorized updates of data objects by other cloud users are prohibited.

The reputation system can be implemented with a trust overlay network. A hierarchy of P2P

reputation system is suggested to protect cloud resources at the site level and data objects at the file level. This demands both coarse-grain and fine-grained access control of shared resources. These reputation systems keep track of security breaches at all levels. The reputation system must be designed to benefit both cloud users and the datacenters. Data objects used in cloud computing reside in multiple datacenters over a storage-area network (SAN).

In the past, most reputation systems were designed for peer-to-peer social networking or for on-line shopping services. These reputation systems can be converted to protect cloud platform resources or user applications on the cloud. A centralized reputation system is easier to implement,

but demand more powerful and reliable server resources. Distributed reputation systems are more scalable and reliable to handle failures. The five security mechanisms presented earlier can be greatly assisted by using the reputation system specifically designed for datacenters.

10. Conclusions

In cloud computing, information storage, transmission and processing are purchased as a commodity from a service provider. Although security issues in storage and transmission can be addressed to a reasonable extent using standard tools, protecting data being processed is another story. Although fully homomorphic encryption would enable processing of encrypted data, the theoretical breakthrough is far from practical. Therefore, efforts need to be put into securing data processing in limited cases, such as searching in encrypted data.

Another fundamental issue in cloud computing is the abolishment of physical constraints that helped securing the data in the past. Like in a transition from paper voting in polling stations to Internet voting, this brings additional security challenges, and it is not even clear beforehand that all of these can be solved.

Meanwhile, the increased outsourcing and specialization not only affects companies, but also individuals. In this case, cloud computing not only means a different way for individuals to store their data, but also a different way to manage the external representations of their identity. In this “outsourced identity”, people *in-tend* to get themselves socially sorted and social sorting cannot be seen as problematic in and of itself. Instead, rules should be put in place to distinguish good from bad sorting, notably by requiring informed consent.

In order to provide ethical foundations for the complex interaction of governments, industries and the social environment in the age of cloud computing, re-research is needed in the relation between information technology and

social sustainability. The precautionary principle may serve as a tool to translate successful approaches from environmental ethics to information ethics in the cloud.

REFERENCES

- [1] K. Aberer and Z. Despotovic, "Managing Trust in a Peer-to-Peer Information System", ACM CIKM International Conference on Information and Knowledge Management, 2001.
- [2] M. Al-Fares, A. Loukissas, and A. Vahdat, "A Scalable, Commodity Datacenter Network Architecture," Proc. of the ACM SIGCOMM 2008 Conference on Data Communication, Seattle, WA, August 17–22, 2008.
- [3] Amazon EC2 and S3, "Elastic Compute Cloud (EC2) and Simple Scalable Storage (S3)",
- [4] http://en.wikipedia.org/wiki/Amazon_Elastic_Compute_Cloud, http://spatten_presentations.s3.amazonaws.com/s3-on-rails.pdf Chapter 7, Cloud Architecture and Datacenter Design (57 pages) in Distributed Computing: Clusters, Grids and Clouds, All rights reserved by Kai Hwang, Geoffrey Fox, and Jack Dongarra, May 2, 2010 7 - 51
- [5] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, R. Karp, A. Rabkin, I. Stoica, and M. Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing", Technical Report No. UCB/EECS-2009-28, University of California at Berkeley, USA, Feb. 10, 2009.
- [6] I. Arutyun, et al, "Open Circus: A Global Cloud Computing Testbed", IEEE Computer Magazine, April 2010, pp35-43.
- [7] L. Barroso, J. Dean, and U. Holzle, "Web search for a planet: the architecture of the Google cluster," IEEE Micro, April 2003. doi:10.1109/MM.2003.1196112
- [8] L. Barroso and U. Holzle, The Datacenter as A Computer: An Introduction to The Design of Warehouse-Scale Machines, Morgan Claypool Publisher, 2009
- [9] G. Boss, P. Milladi, et al, "Cloud Computing- The BlueCloud Project", [www.ibm.com/ developer works/ websphere/zones/hipods/](http://www.ibm.com/developer/works/websphere/zones/hipods/), Oct. 2007.
- [10] M. Cai, K. Hwang, J. Pan and C. Papadupolous, "WormShield: Fast Worm Signature Generation with Distributed Fingerprint Aggregation", IEEE Trans. of Dependable and Secure Computing (TDSC), Vol.4, No. 2, April/June 2007, pp.88-104.
- [11] Chase, et al, "Dynamic Virtual Clusters in a Grid Site Manager", IEEE 12-th Symposium on High-Performance Distributed Computing (HPDC), 2003.
- [12] Y. Chen, K. Hwang, and W. S. Ku, "Collaborative Detection of DDoS Attacks over Multiple Network Domains", IEEE Trans. on Parallel and Distributed Systems, Vol. 18, No.12, Dec. 2007, pp.1649-1662.
- [13] T. Clou, Introduction To Computing Computing: Business and Technology, Active Book Press, 2010.
- [14] C. Clark, K. Fraser, J. Hansen, E. Jul, I. Pratt, and A. Warfield, "Live Migration of Virtual Machines", Proc. of Symp. on Networked Systems Design and Implementation. Boston, May 2, 2005. 273 – 286.
- [15] Cloud Computing Tutorial, <http://www.thecloudtutorial.com>. Jan. 2010.
- [16] Cloud Security Alliance, "Security guidance for Critical Areas of Focus in Cloud Computing", April 2009
- [17] C. Collberg and C. Thomborson, "Watermarking, Temper-Proofing, and Obfuscation-Tools for Software Protection", IEEE Trans. Software Engineering, Vol.28, 2002, pp.735-746
- [18] A. Costanzo, M. Assuncao, and R. Buyya, "Harnessing Cloud Technologies for a Virtualized Distributed Computing Infrastructure", IEEE Internet Computing, Sept. 2009.
- [19] D. Dyer, "Current Trends/Challenges in Datacenter Thermal Management—a Facilities Perspective," presentation at IThERM, San Diego, CA, June 1, 2006.
- [20] Q. Feng, K. Hwang, and Y. Dai, "Rainbow Product ranking for Upgrading e-Commerce", IEEE Internet Computing, Sept. 2009.
- [21] I. Foster, Y. Zhao, J. Raicu, and S. Lu, "Cloud Computing and Grid Computing 360-Degree Compared," Grid Computing Environments Workshop, 12-16 Nov. 2008.
- [22] Google Inc., "Efficient Data Center Summit, April 2009". Available at <http://www.google.com/corporate/green/datacenters/summit.html>.
- [23] Green Grid, "Quantitative analysis of power distribution configurations for datacenters".