Comparing various realizations of the cloud computing paradigm

Bo Pang
Helsinki University of Technology
bpang@cc.hut.fi

Abstract

Cloud Computing is becoming attractive these days. Many big companies, such as Google, Amazon and IBM, have promoted their own cloud computing platforms for customers to deploy their web applications on these platforms. Besides some existing security risks for a traditional web server hosting company, some new security vulnerabilities appear with this new technology, for instance Virtual Machine monitoring and storage isolation. The cloud computing platform needs a better design to mitigate these risks. This paper is a survey, which introduces the concept of cloud computing and some industrial implementation architecture and the threat model for cloud computing and a case study of analyzing three implementations' risks situations. At the end some general guides will be stated.

1 Introduction

Cloud Computing is emerging as the most promising technology of the first decade of 21st century. This new technology brings new concepts to the IT industry. Computing resource, data storage resource, security and management can be packed into goods for sale. Cloud Computing brings the following advantages for customers:

1. No upfront investment for IT infrastructure.
2. Less cost for system maintenance and management.
3. Stronger central security control and data backup and a reliable disasters recovery system.
4. meets scalable and flexible business requirements.
5. Shorter software development lifetime.

Cloud computing is expected to be a "$160 billion addressable market opportunity, including $95 billion in business and productivity applications, and another $65 billion in online advertising" [6]. However, Cloud Computing is a still new concept to even experienced engineers:

1. The related technologies and legislations are not perfect.
2. An other problem is that customers still cannot trust the third party cloud computing service providers to manage their data and application code, and also to guarantee confidentiality, integrity and availability all the time.
3. Many companies' management have no idea about how to choose a suitable Cloud Computing service provider to effectively and efficiently protect their companies' electronic asset.

Hence a general security issues survey about cloud computing industrial is urgently needed.

This paper is a survey of Cloud Computing technology and various implementations' security issue, it analyzes different cloud computing implementations' structure and technologies and security issues. The first section of the paper is about Cloud Computing definitions and advantages. The following section will introduce three different Cloud Computing products' architectures. Then the section three will list the risk that current cloud computing providers meet, the next section is a case study about comparison of diverse Cloud Computing products' security measured by risk model stated at section three, the last part is a conclusion and some security guides.

2 Background

Many big IT companies have been involved in Cloud Computing field. Google first released its Platform as a service(PaaS) platform, Google App Engine, in April 2008. Google App Engine enables you to build and host web applications on the same systems that power Google applications. App Engine offers fast development and deployment; simple administration, with no need to worry about hardware, patches or backups; and effortless scalability [12]. The E-commerce leader company Amazon has its own IaaS platform- Amazon Elastic Compute Cloud (EC2), which is a central part of Amazon.com's cloud computing platform, Amazon Web Services (AWS). EC2 allows users to rent virtual computers on which to run their own computer applications. EC2 enables scalable deployment of applications by providing a web service through which a user can boot an Amazon Machine Image to create a virtual machine, which Amazon calls an "instance", containing any user desired software[9]. An other example is Microsoft's Windows Azure Platform, which is a cloud platform that provides a wide range of Internet services that can be utilized from both on-premises environments and the Internet". It is Microsoft's first step into cloud computing following the launch of the Microsoft Online Services offering. In short, it's Microsoft's PaaS platform [10]. Last but not the least is Eucalyptus. Eucalyptus is an open-source software platform that implements IaaS-style cloud computing using the existing Linux-based infrastructure found in the modern data center. And it offers a interface compatible with Amazon's AWS making it possible to move image file between AWS and the data center. Eucalyptus works with most of the currently available Linux distributions including Ubuntu, Red Hat Enterprise Linux (RHEL), CentOS, SUSE Linux Enterprise Server (SLES), openSUSE, Debian and Fedora. Similarly, Eucalyptus can use a variety of virtualization technologies including VMware, Xen, and KVM to host the cloud service it supports [11].

3 What is Cloud Computing

Cloud Computing is a critical technology for web service development, the following sections will introduce its features and edges.

3.1 Why do people need cloud computing?

Competitive business environments and globalization push middle and small scale enterprises (MSE) to implement new business idea as fast as they can. Thus these enterprises always need high volume IT infrastructure reservation for their upcoming new businesses ideas. The cost of purchasing hardware and software and periphery tools and also the maintenance staff's salary is a huge burden for a MSE. Hence a MSE needs more flexible and elastic solution to manage its IT infrastructure to lower the cost. On the other hand, big companies also need cloud computing. Because a global company has many branches all over the world, their data need to be integrated and analyzed, which demand their company build several data centers to provide 24*7 service at low latency. To solve these problems, cloud computing comes to Chief Technology Officer(CTO)'s sight. The Cloud Computing providers, such as Google, have already owned advanced data centers. Renting Cloud Computing providers' infrastructure may better leverage company's IT budget and solve the geographic predicament. Hence Cloud Computing is on the top of IT outsourcing trend.

3.2 How to use cloud computing?

Cloud Computing includes public clouds, private clouds and hybrid clouds. Public cloud or external cloud represents cloud computing in the mainstream sense, whereby resources are dynamically provisioned on a fine-grained, self-service basis over the Internet, via web applications/web services, from an off-site third-party provider who bills on a fine-grained utility computing basis[6]. On the other hand, the private cloud is built inside the organization and it provides private service to all departments of the organization. The hybrid cloud combines both private cloud and public cloud platform to serve the organization from intranet and Internet. Cloud computing can be divided by other approaches. The common architecture structure of cloud computing is like Figure 1. The cloud computing provider can select which layer of resource to rent. There are three kinds of services: if the providers rent their applications, it is called software as a service (SaaS); if the providers rent their platforms, it is called platform as a service (PaaS); if the providers rent their infrastructures, it is called software as a service (IaaS). The SaaS providers develop web application and serve users to complete certain tasks and charge money for utility. Google Docs and some human resource management software, such as Salesforce, belong in this category. The PaaS instances, such as Google App Engine and Facebook, provide platforms and software stacks for application developers to deploy applications on it and get commission charge from application provider. The last sort, IaaS, supplies a virtual machine (VM) for other service providers to deploy their virtual machine image and gets rent from tenants, and the rent is based on computing, storage and broadband capability, and Amazon web service is the example of this sort provider. Different business requirements need different cloud computing services among these three kinds of cloud computing service providers.

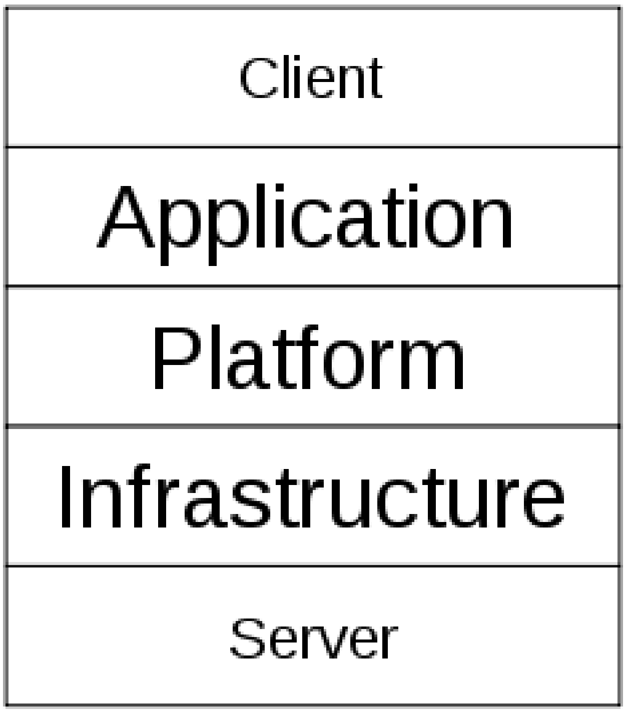

Figure 1: Cloud computing structure [7]

3.3 Advantages of Cloud Computing

The Cloud Computing has some advantages comparing with traditional websites host companies. First there is no upfront investment for IT infrastructure. The Cloud Computing providers always use the pay-as-you-go business model, which means the user's payment will be divided and the future payment can be invested to some other fields to get interests rather than depreciated with the hardware price decreasing. The second is less cost for system maintenance and management. MSEs can hardly hire maintenance staff and operator to guarantee the system run correctly 7*24 hours, so if there is some technology problems happened in in-house system when IT staff is on vacation, it would take a long time to recover from service outage, which may cause huge loss for company. Thus it is cheaper and safer for

MSEs to transfer the maintenance risk to the Cloud Computing providers. The third benefit is stronger central security control and data backup and disasters recovery system. A reputable Cloud Computing provider has more professional security technology team and better and geographic distributed backup system than most enterprises', the data loss and distortion and exposing risk is much low than inhouse system. It is more secure to hold data at Cloud, but the Cloud Computing provider's reputation and SLA are critical to guarantee this feature. Fourthly Cloud Computing meets scalable and flexible business requirements. The Internet market demand is variable and workload of application is unstable, for example every semester's beginning the university's website will always break-down because of heavy request from students, but in holidays the website server is almost idle. The workload tide and ebb cause huge computing and networking resource waste, since universities need to provide extra resource to track the peak of request. If companies underestimated the utilization peak, they would deny users' requests, which is worst situation of business, because business reputation and users' trust is hard to build. Thereby fast resource scale-up and scale-down attributes of Cloud Computing solves this problem. For instance, Google App Engine can scale the computing resource synchronizing with the workload situation, so the Cloud Computing users don't need to worry about their IT resource to be exhausted or wasted. Last but not the least is shorter software development lifetime. When company releases new applications, they don't need to configure and test their product environments, since the Cloud Computing provider takes care of the environment issue. Company may focus on developments more and cut the idea to market period.

4 Realizations of the cloud computing paradigm

Cloud computing technology has been realized by several organizations, the rest sections describes these implementations.

4.1 Amazon web services

As the largest online retailer merchant, Amazon handles millions requests from every corner of the world, which motivates it to implement its own cloud computing infrastructure. And Amazon rents computing resource for others company's business purpose to earn profit and take advantages of cloud computing assets. Amazon web services provide several services including computing capability, persistent data storage and other value added services, such as VPN connection, performance monitor, map reduce calculation and message queue service. Customers can use these services with little friction and only need to pay for the resource they used.

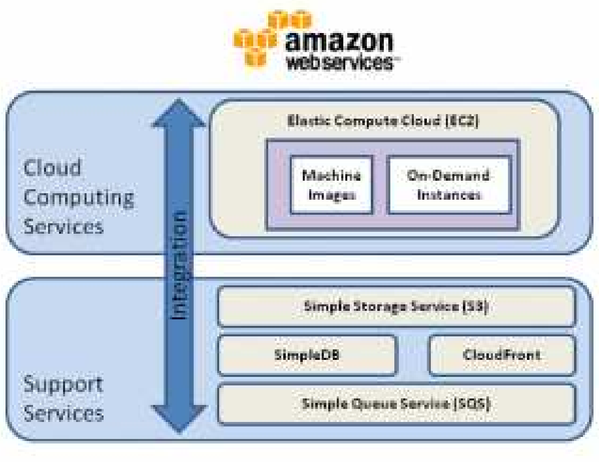

Figure 2: AWS architecture [4]

The architecture of Amazon web service is composed by two components, namely elastic compute cloud (EC2), Simple storage service (S3). Integrating these two components, basic web service can be implemented. Other value-added service components that perform like plugins enhancing other attributes and functions. EC2 builds on Zen virtual machine system and holds and runs customers' virtual machine image. Customer may upload their customized operation system images or use the default image supplied by Amazon. In addition, the customers use client side application to interact or manage the instances in cloud through https or SOAP alternatively. The S3 is the complementary service for EC2, because EC2 has no persistent storage, and considering the data transmission latency, S3 guarantee response speed of data query. The Amazon web service provides more flexibility for costumers to manage operation system and software stack, but customers have to be concern about operation system security and share some part of risk of operation system crash.

4.2 Google App Engine (GAE)

The purpose of Google releasing PaaS platform is not only optimizing utilization efficiency of its infrastructure but also expanding its web service kingdom. Since Google owns massive widget applications and these widgets can be easily integrated to web services running in its cloud, web services hosted by Google App Engine would prefer to mashup Google's widget into their service, Thereby the GAE increases Google's market share rate and influence. Google's platform's core is software stacks. So far GAE support two programming languages, Python and Java. Customers develop and test web services programs on their own computers and then upload and deploy on GAE's cloud environment. The uploaded web service code will be interpreted by GAE's sandbox style interpreter. In GAE, all applications run in a secure environment that provides limited access to the underlying operating system[12]. The limitation for applications includes 1) applications cannot access the Internet by open socket, they can only use URL fetch to communicate with other web service by http; 2) applications cannot start threads or execute any code after sent response to request; 3) applications are not allowed to write to or create file on underlying operation system and they can only read files uploaded by service owners.

Furthermore it is needed to keep web service stateless for

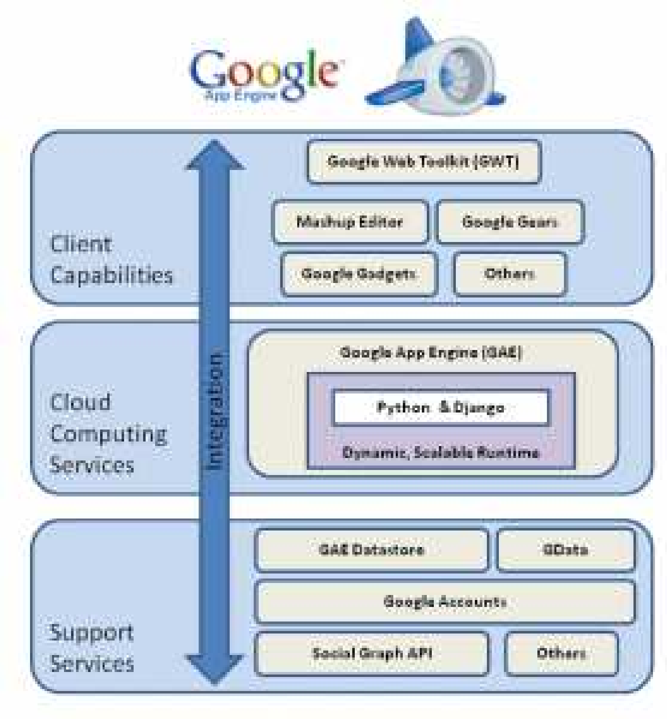

Figure 3: GAE architecture[4]

endowing cloud service with scalability. Therefore GAE requires cloud service designers to store all state and users' data to the App Engine Datastore, memcache or other outside storage services. The Datastore, GAE's persistent storage, is based on bigtable database, which also support Google's searching service. Bigtable database is not a relational database, whereas it is actually a really huge distributed table, contain data in rows. The table 1 is a comparison of Datastore and traditional relational database. From the comparison, the Datastore's performance is not as good as traditional relational database, but its simple structure brings it strong scalability which is the foundation of distributed web service data storage.

GAE, as a PaaS, provides strong platform support for cloud services running on it. The cloud services would combine Google's user authentication and email service to enforce their services' functions and reuse some other service components provided by Google.

4.3 EUCALYPTUS Cloud Platform

The EUCALYPTUS (Elastic Utility Computing Architecture for Linking Your Programs To Useful Systems) project is an open source cloud computing software project, which is aiming to help researchers build their own private cloud. The EUCALYPTUS has three features: First EUCALYPTUS has a modular framework with standard software interfaces, and this enables users choose their preferred components to replace the original modules, which is important for cloud computing researchers to build their experiment environments. Secondly EUCALYPTUS has implemented interface that is totally compatible with Amazon's EC2's. This interface allows customers to transfer their service platforms for mitigating lock-in risk. Thirdly EUCALYPTUS adopts a unique network management system, which offers a virtual network overlay that can isolate different instances' network traffic and connect one instance's subsystem over several clusters within the same Local Area Network.

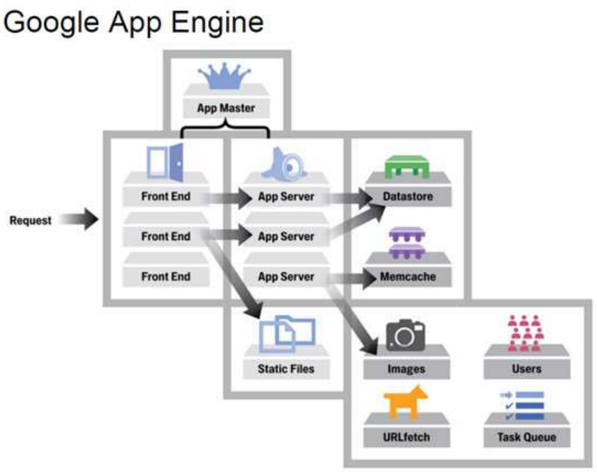

Figure 4: GAE structure [8]

The architecture of EUCALYPTUS is flexible and hierarchical, the figure 5 depicts its hierarchy. The hierarchy is composed of four levels, namely node controller, cluster controller, storage controller and cloud controller, and fourthly client API. The node controller is based on Xen or KVM virtual machine. It controls and monitors virtual machine working while reporting to cluster controller and executing commands received from cluster controller. Cluster controller is in charge of several nodes under it. Besides controls the nodes, cluster controller also manages instance image copies, network overlays and user authentications. Every EUCALYPTUS has only one cloud controller which manages all resource allocation and deallocation and provides interface to tackle authentication and protocol translation. And the storage controller, Walrus, usually combined with cloud control, provides S3 compatible interface to handle input/output data from users, and it also manages VM image storage. The last is client API which is compatible with EC2 client API and communicates with clients' instances in cloud.

5 Security threats to cloud computing

Cloud computing has met several security threats as other new technologies. Some threats only threaten particular service models, and the following paragraph analyzes different types of threats[13].

Abuse and Nefarious use of cloud computing. Affect fields: IaaS and PaaS. For business considering, most cloud computing providers provide frictionless register process, which means anyone who has a valid credit card could register and buy computing resource from providers. The smoothness help normal clients and crackers as well. Cracker may used a stolen credit card to buy computing resource for some computing task which needs intense computing,

Features	Datastore	traditional relational database
SQL support	Just support some basic queries	Fully support
Structure	Hierarchical	Relational
Index	Automatically creating	Manually creating
Average execute speed	Low than 200ms	Low than 100ms
Scalability	Very good	Very hard and needs modification to tables

Table 1: A comparison of Datastore and traditional relational database

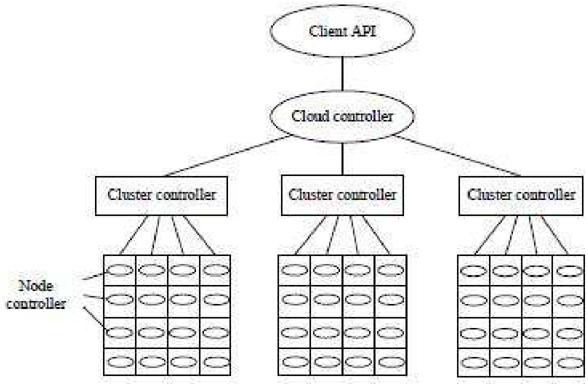

Figure 5: EUCALYPTUS architecture[14]

such as password cracking, DDOS and botnet command center. Moreover a cracker may also harm cloud providers themselves by spamming other organizations, as a defense measure, victim organizations will block all network IP addresses of cloud computing providers, which will impact their business and reputation.

Insecure Interfaces and APIs. Affect fields: all. Cloud computing providers expose a set of API for customers to manage and interact with their instances. However these API may become targets of attackers, if there is no strong authentication method to authenticate customers' identification, or if attacker obtains user's token, password or session, he is able to impersonate and manipulate API to cause damage to user's instance.

Malicious Insiders. Affect fields: all. This is a common problem for all enterprises, but it is especially serious for cloud computing providers, because cloud computing providers host plenty of cloud services. If there is no protective measure and alarm system, the insider of cloud computing provider may access customers' confidential data or manipulate customers' services and even infiltrate customers' organization and asset. Even worse the customers may never know the occurrence happened to their services. Lack of transparency of cloud computing providers' internal situation causes cloud service owners lose control of their service, thereby the customers will prefer the cloud computing providers which reveal more of their security detail and hiring standard and practices.

Shared Technology Issues. Affect fields: IaaS. This threat is based on virtualization hypervisor's vulnerability, which bleach the principle of data isolation. Normally program and data of instance is segregate from other instances'. but, by the means of infiltrating the underlying operation system, attackers may exploit virtual machine's vulnerability and access other tenants' data or network traffic.

Data Loss or Leakage. Affect fields: all. Data lose can be due to several reasons; Authentication, authorization and audit mechanism failed may attracts attackers to tamper the data stored in the cloud. Moreover, disasters happen on data center that does not deploy data recovery mechanism is another reason. And persistence and remanence challenges require cloud computing providers thoroughly wipe the data on storing media before the storing resource is released back the resource pool, otherwise attackers may use special technology to read information remain on them.

Account or Service Hijacking. Affect fields: all. This threat is an old problem. If the authentication encryption method is insufficient, attacker will hijack the session and impersonate users. The cloud computing amplifies the damage, because attackers may use the cloud service under their control as a basement to launch subsequence attack.

Unknown Risk Profile. Affect fields: all. Cloud computing providers always brag their products' feature and functionality, however security issue are seldom mentioned to customers. Customers can not get knowledge about the security structure and situation inside the cloud, which is a latent threat to all the customers whose service and business are totally rely on cloud computing providers' performance.

These threats will be used as metrics to measure the security risks of above three cloud computing implementations. The analysis is depicted by Table 2.

6 Cloud Computing platform implementations analysis

- Amazon web Services(AWS): Although AWS has a strong monitor mechanism, there is no behavior analysis process mentioned in its security white book [5]. Thus the abuse risk exists. The AWS signature version 1 had vulnerability for 7.5 months until next version release and Amazon doesn't provide a visible mechanism for researchers/customers to report security vulnerabilities[3]. Hence their API and interface may contain security vulnerabilities. AWS has established formal policies and procedures to delineate the minimum standards for logical access to AWS platform and infrastructure hosts. AWS requires that staff with potential access to customer data undergo an extensive background check (as permitted by law) commensurate with their position and level of access to data. The policies also identify functional responsibilities for the administration of logical access and security[5]. Therefore Ma-

Threats/platform	AWS	GAE	EUCALYPTUS
Abuse and Nefarious use	exists	Not exists	exists
Insecure APIs	exists	exists	exists
Malicious Insiders	Not exists	exists	N/A
Shared Technology	Not exists	Not exists	Depends on VM
Data Loss or Leakage	Not exists	Not exists	exists
Service Hijacking	Not exists	Not exists	Not exists
Unknown Risk Profile	exists	exists	Not exists

Table 2: A comparison of different implementations' risks

licious Insiders risk is slight. AWS uses S3 database to store all the persistent data and has secure authentication method and firewall that protects transcations, the Data Loss or Leakage risk is not obvious. AWS VPC (Amazon Virtual Private Cloud) provides users IPsec VPN to protect communication and eliminate account leakage or session hijacking. However AWS is a closed source product and the security and design details are obscure, thus Unknown Risk Profile exists.

- Google App Engine: The GAE provides monitoring mechanism and blocks malicious cloud service, hence the abuse risk is trivial[2]. As a massive API owner, GAE is inevitable suffering insecure API, but Google's bug reporting and tracking is good and in time. GAE does not exposure its employment standard and regulations, hence malicious Insiders risk exists. GAE's sandbox policy, forbidding most operations with underlying system, prevents most cross instances interaction from occurring. Google account is a powerful security tool to eliminate the Account or Service Hijacking risk. GAE's security design details are still obscure to public, thus Unknown Risk Profile exists.
- EUCALYPTUS Cloud Platform: at first the abuse risks exist but not serious, because the EUCALYPTUS is normally acting as private cloud, could owners may use register procedure and rigorous rules to mitigate this risk, but there is no instance behavior monitoring functions. If the instance is infiltrated by crackers, administrator can hardly discover the malicious behaviors. In addition, EUCALYPTUS shares the same API with Amazon, hence there is insecure API risk, but EUCALYPTUS's bug reporting system works well. The shared technology issue is caused by vulnerability of virtual machine chose by the clients. Thus clients need to choose secure virtual machine providers and install and configure complying to security best practices. However the network traffic is isolated by EUCALYPTUS's network overlay, so the network traffic is out of this risk. The Data Loss or Leakage risk exists because VM normally doesn't provide data wipe before storage space relocation, thus users have to wipe the sensitive data by themselves before returning storage space. Adopting https connection and encrypted network channel would eliminate Account or Service Hijacking risk. Because EUCALYPTUS is open source private software, risk profile is transparency to its users, there is not Unknown Risk Profile risk.

7 General guilds for Cloud Computing security

From the above analysis, the three deficiencies, namely lack of cloud service behavior monitoring, insecure API and opaqueness of security situation, threaten the security of cloud computing most. To mitigate these risks, some guides and best practices are needed. In Amritw's post [1], some areas are listed below, which should attract focus from cloud computing providers. In my opinion, enhancing these six areas may defense against most threats analyzed by above paragraph.

- Infrastructural security controls
- Transport mechanism and associated controls
- Authentication and authorization access controls
- Monitoring and auditing capabilities
- SLA and methods for deploying security updates throughout the infrastructure
- Transparency across these controls and visibility into how they function on a regular basis

8 Summary

Cloud computing is an indubitably promising technology, which offers IT industry and other industries that needs IT as an auxiliary tool a economic and scalable solution to host web services. The current cloud computing providers have already achieved splendid cloud software products. But there are still a few areas of security risks, thus more efforts are needed to improve cloud computing products in the markets. This paper has a survey of several cloud computing implementations. And it propose a security threat model to measure these implementations' risk situations. From this paper, these implementations' weakness and strength are obvious. And it introduces some basic cloud computing knowledge as well for better understanding the cloud computing concept.

References

[1] "Amazon AWS, Google App Engine, Microsoft Azure, and More C Part 1: Can We Secure The Cloud". Technical report, Jul 2008. `http://techbuddha.wordpress.com/2008/12/`

03/amazon-aws-google-app-engine-microsoft-azure-and-more-part-1-can-we-secure-the-cloud/.

[2] "Cloudsecurity.org Interviews Guido van Rossum: Google App Engine, Python and Security ". Technical report, Jul 2008.

[3] "What is New in the Amazon Cloud?: Security Vulnerability in Amazon EC2 and SimpleDB Fixed". Technical report, Dec 2008. http://cloudsecurity.org/blog/2008/12/18/whats-new-in-the-amazon-cloud-security-vulnerability-in-amazon-ec2-and-simpledb-fixed-75-months-after-notification.html.

[4] "Exploring Cloud Computing Development". Technical report, Jul 2009. http://rdn-consulting.com/blog/2009/02/07/exploring-cloud-computing-development/.

[5] "AWS Security Whitepaper". Technical report, Sep 2010. http://awsmedia.s3.amazonaws.com/pdf/AWS_Security_Whitepaper.

[6] "Cloud computing seen as the future of computer technology". Technical report, Sep 2010. http://www.executivebrief.com/news/cloud-computing-seen-as-the-future-of-computer-technology/.

[7] "cloud computing stack". Technical report, Sep 2010. http://en.wikipedia.org/wiki/File:Cloud_Computing_Stack.svg.

[8] "GAE high level architecture". Technical report, Sep 2010. http://www.dbanotes.net/arch/google_app_engine-arch_intro.html.

[9] "introduction for Amazon Elastic Compute Cloud". Technical report, Sep 2010. http://en.wikipedia.org/wiki/Amazon_Elastic_Compute_Cloud.

[10] "introduction for Azure cloud". Technical report, Sep 2010. http://en.wikipedia.org/wiki/Azure_Services_Platform.

[11] "introduction for eucalyptus". Technical report, Sep 2010. http://www.eucalyptus.com/products/overview.

[12] "introduction for GAE". Technical report, Sep 2010. http://code.google.com/appengine/docs/whatisgoogleappengine.html.

[13] e. a. Jerry Archer. "Top threats to cloud computing v 1.0". Technical report, Cloud Security Alliance, November 1998. http://www.cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf.

[14] Nurmi, D., et al. The Eucalyptus Open-Source Cloud-Computing System. In *Cluster Computing and the Grid, 2009. CCGRID '09. 9th IEEE/ACM International Symposium on*, pages 124 – 131, Jun 2009.