Cloud Computing Technology and Security 22

22.1 Introduction

Cloud computing as a technology is difficult to define because it is evolving without a clear start point and no clear prediction of its future course. Even though this is the case, one can say that it is a continuous evolution of a computer network technology going beyond the client-server technology. It is a technology extending the realms of a computer network creating an environment that offers scalability, better utilization of hardware, on-demand applications and storage, and lower costs over the long run through the creation of virtual servers cloned from existing instances each offering near instantaneous increase in performance, allowing companies to react quickly and dynamically to emerging demands. The "cloud" or "cloud solution," as the technology is commonly referred to, can either be hosted on-site by the company or off-site such as Microsoft's SkyDrive and Samsung's S-Cloud.

The cloud technology seems to be in flax; hence it may be one of the foundations of the next generation of computing. Keep watching! It may be in that in the next few years, a grid of a few cloud infrastructure may provide computing for millions of users. This is a broader view of cloud computing. Cloud computing technology consists of and rests on a number of sound, fundamental, and proven technologies including virtualization, service-oriented architectures, distributed computing, grid computing, broadband networks, Software as a Service, browser as a platform, free and open-source software, autonomic systems, Web application frameworks, and service-level agreements [1]. We will discuss many of these technologies in depth in this chapter.

First, let us start by trying to give a broad but specific view of the technology, what it is composed of, and how it works. We will start by a more specific definition given by the National Institute of Standards and Technology (NIST). According to NIST [1], cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources like networks, servers, storage, applications, and services that can be rapidly provisioned and released with minimal management effort or service provider

J.M. Kizza, *Guide to Computer Network Security*, Computer Communications and Networks, DOI 10.1007/978-3-319-55606-2_22

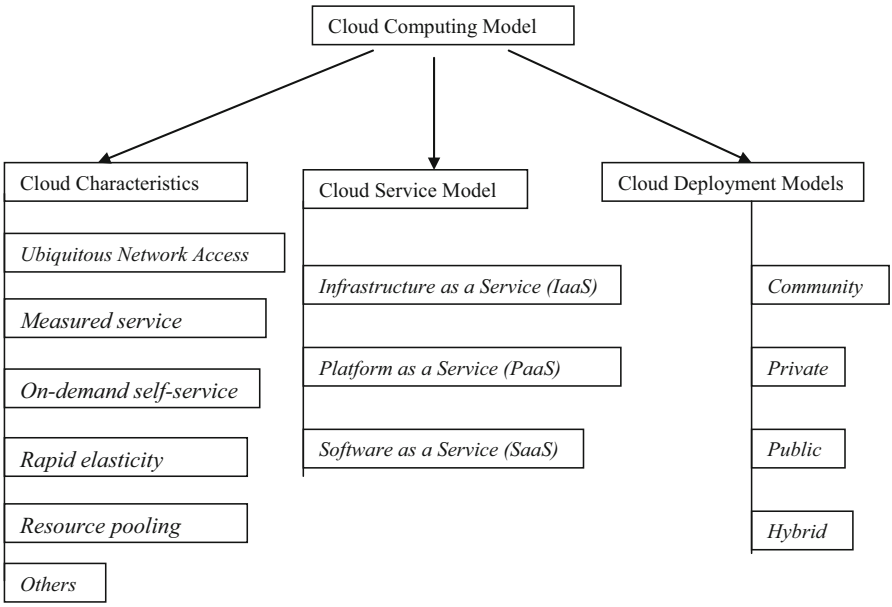

Fig. 22.1 Broad view of a cloud computing model

interaction. So for the remainder of this chapter, we are going to focus on this model of computing and discuss its benefits and security concerns. This computing model as shown in Fig. 22.1 is composed of a number of essential characteristics, three service models, and four deployment models.

22.2 Cloud Computing Infrastructure Characteristics

Traditionally data center computing models were mainly based on a client-server model architecture and design relying firmly on a three-tier architecture design that included access, distribution, and core switches connecting relatively few clients and meeting limited client needs compared to today's cloud service models. In most cases, each server was dedicated to either a single or limited applications and had IP addresses and media access control addresses. This static nature of the application environment worked well and lent itself to manual processes for server deployment or redeployment. According to both Jim Metzler and Steve Taylor of Network World [2], they primarily used a spanning tree protocol to avoid loops. But because of the dramatic advances in the previous years in virtualization technology, distributed computing, rapid improvements, and access to high-speed Internet have all dramatically changed the staid nature of the data center. Today's data center, providing cloud services, is but staid; it is bursting with activities and services with distinctly new characteristics that differentiate it from traditional cousin. For example, its services are now on demand, by the minute or the hour;

it is elastic, because users can have as much or as little of a service as they want at any given time; and the service is fully managed by the provider, that is, the consumer needs nothing but a personal computer and Internet access. Let us now briefly look at each one of these characteristics:

Ubiquitous Network Access

Like in the previous section on on-demand self-service, the advances and use of virtualization technology and the availability and access to high-speed Internet have also helped to change the nature of access to the computing services sought by customers and the increase in the number of services a customer can select from. With more choice also came the high specialization and quality of services that a customer can expert.

Measured Service

Because cloud services are flexible and on demand and are elastic, it is important, therefore, for these services to be metered. The concept of metered services allows customers to get what they want in the required amounts at the time they want the service. One of the most popular characteristics of cloud computing technology is measured or metered service for most, if not all, of the cloud services including storage, processing, bandwidth, and active user accounts. This pick-what-you-can-afford-to-pay-for principle based on metering results in an automatic control and optimization of cloud technology resource use based on the type of service, and these statistics can be reported as needed, thus providing transparency for both the provider and consumer.

On-Demand Self-Service

With the rapid and unprecedented use of virtualization technology and the availability and access to high-speed Internet, the traditional and all other models of acquisition of computing services that demanded perpetual ownership of software or computing hardware and long contracts with employees that helped to use the service and the need for redundancy and outsourcing of services all diminished and turned into a more flexible model where consumers of computing services were no longer restricted to having one of the rigid traditional models of either ownership, outsources, or boxed services. Now, a consumer is able to not only automatically provision any computing services and capabilities as needed but also to determine the time and how long to use the provisioned services.

Rapid Elasticity

Computing service elasticity means the ability to resize and dynamically scale the virtualized resources at hand such as servers, processors, operating systems, and others to meet the customer's on-demand needs. The provider must make sure that

there are resources at hand that must meet elastic capabilities to ensure that end users' requests are continually and promptly met. Amazon's EC2 is a good example of a Web service interface that allows the customer to obtain and configure capacity with minimal effort.

Resource Pooling

Increased flexibility, access, and ease of use usually lead to high and varied demands of services from customers. To meet these new demands, providers usually respond by offering a variety of and polling of system resources and services. As noted in the NIST report, the provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.

Others

There are other characteristics common to cloud computing beyond the five we have discussed above. Among these are:

- Massive scale —that the cloud offers the resources at a massive scale on demand.
- Virtualization—in fact this is the linchpin of the cloud technology. The cloud is possible because of virtualization of the fundamental functionalities of the physical machine.
- Free software—or near free software as needed from the cloud.
- Autonomic computing—in a sense that you scale computing resources at a time you want them on the fly.
- Multi-tenancy—because of cloud's massive-scale and easy access of those resources, cloud computing can accommodate a large number of users at a time.

22.3 Cloud Computing Service Models

Infrastructure as a Service (IaaS) The process of providing the customer with the ability and capability to manage and control, via a Web-based virtual server instance API, system resources such as starting, stopping, accessing, and configuring the virtual servers, operating systems, applications, storage, processing, and other fundamental computing resources is referred to as Infrastructure as a Service (IaaS). In doing all these, however, the consumer does not have access nor control of the underlying physical cloud infrastructure.

Platform as a Service (PaaS) This is a set of software and product development tools hosted on the provider's infrastructure and accessible to the customer via a Web-based virtual server instance API. Through this instance, the customer can

create applications on the provider's platform over the Internet. Accessing the platform via the Web-based virtual instance API protects the resources because the customer cannot manage or control the underlying physical cloud infrastructure including network, servers, operating systems, or storage.

Software as a Service (SaaS) Ever since the beginning of computing software, over the years, the key issue that has driven software development has been the issue of the cost of software. Trying to control the cost of software has resulted into software going through several models. The first model was the home-developed software where software users developed their own software based on their needs and they owned everything and were responsible for updates and management of it. The second model, the traditional software model, was based on packaged software where the customer acquired a more general-purpose software from the provider with a license held by the provider and the provider being responsible for the updates while the customer being responsible for its management. However, sometimes, software producers provide additional support services, the so-called premium support, usually for additional fees. Model three was the open-source model led by a free software movement starting around the late 1980s. By the late 1980s, free software turned into open source with the creation of the Open Source Initiative (OSI). Under the name "open-source" philosophy, some for-profit "free software" started to change the model from a purely free software to some form of payment for support of the updates of the software. The open-source software model transformed the cost of software remarkably. Model four consisted of software outsourcing.

The outsourcing model was in response to the escalating cost of software associated with software management. The component of software management in the overall cost of software was slowly surpassing all the costs of other components of software including licensing and updates. In model four, however, software is still licensed from the software company on a perpetual basis; support fees are still paid; however, the software producer takes on the responsibility of the management of that software.

Software model five is Software as a Service (SaaS). Under this model, there is a different way of purchasing. Under SaaS, there is the elimination of the upfront license fee. All software applications are retained by the provider, and the customer has access to all applications of choice from the provider via various client devices through either a thin client interface, such as a Web browser, a Web portal, or a virtual server instance API. The cloud user's responsibilities and actual activities in the use of and operations of the requested cloud services are limited to user-specific application configuration settings, leaving the management and control of the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities to the cloud provider.

Three Features of SaaS Applications

In particular, Software as a Service has the following features:

- Scalability—in that it can handle growing amounts of work in a graceful manner.
- Multi-tenancy—in that one application instance may be serving hundreds of companies. This is different from the client-server model from which the cloud computing model grows out of in which each customer is provisioned their own server running one instance.
- Metadata-driven configurability—customers can configure their application through metadata.

22.4 Cloud Computing Deployment Models

There are three cloud deployment models which are actually cloud types. These are the public, the private, and the hybrid models.

Public Clouds The public clouds provide access to computing resources for the general public over the Internet allowing customers to self-provision resources typically via a Web service interface on a pay-as-you-go basis. One of the benefits of public clouds is to offer large pools of scalable resources on a temporary basis without the need for capital investment in infrastructure by the user.

Private Cloud Unlike public clouds, private clouds give users immediate access to computing resources hosted within an organization's infrastructure and premises. Users, who are usually in some form of a relationship with the cloud owner, choose and scale collections of resources drawn from the private cloud, typically via Web service interface, just as with a public cloud. Also the private cloud is deployed within and uses the organization's existing resources and is always behind the organization's firewall subject to the organization's physical, electronic, and procedural security measures. In this case, therefore, private clouds offer a higher degree of security.

Hybrid Cloud A hybrid cloud combines the computing resources of both the public and private clouds.

22.5 Virtualization and Cloud Computing

In this chapter, we are going to discuss virtualization in depth so here we are only defining the concept and showing its role in cloud computing. Virtualization is a process of creating something in effect and performance but not in reality—hence virtual. In computing, virtualization can be used on both software and hardware. In software, virtualization has been used in operating systems where the underlying operating systems create a number of virtual operating systems, not only clones of itself but even others, to run on the underlying machine and perform tasks at a higher performance level. In hardware, virtualization is being used to create new resources like servers, storage devices, and others. The potential power of

virtualization in substantially increasing the performance of computing systems such as hardware and software through division of the underlying physical computing resources into many equally powerful virtual machines, thus scaling up the performance and creating elasticity of many computing system. With virtualization, computation and applications can be moved easily from physical to virtual machines. This transfer of computation and applications from one machine to another both physical and virtual is not truly a new idea. The client-server computing model is an example of this. But while in the client-server model, the applications running on the servers are managed by the service providers, in cloud computing model, computations and application may scale out to many other servers owned by the cloud provider. This is possible through the power of virtualization. Virtualization is a fundamental feature in cloud computing. Virtualization allows applications from different customers to run on different virtual machines, hence providing separation and protection.

22.6 Benefits of Cloud Computing

Cloud computing as a model of computing is very exciting and has tremendous benefits for those who dare to use it. It is not only exciting when you come to learn it, but it also has an array of benefits including but not limited to leveraging on a massive scale, homogeneity, virtualization, low-cost software, service orientation, and advanced security technologies.

Reduced Cost For all cloud computing benefits to a company, perhaps the biggest benefit is cost savings. Whether it is a small-, medium-, or large-scale manufacturing business, there are essential cost benefits in using a cloud model for most of the company's computing needs. The biggest issue here is the fact that cloud computing is operated remotely off company premises except a few devices needed for accessing the cloud resources via a Web portal. This means that company personnel can do the same amount of work on fewer computers by having higher utilization, save on not housing data centers on premises, save on personnel for running the data center, and save on expenses that would normally be essential for running a data center on the premises. There is documentary evidence to support these views from industry experts. In the words of Greg Papadopoulos, CTO, Sun Microsystems [3], hosting providers bring "brutal efficiency" for utilization, power, security, service levels, and idea-to-deploy time. George Reese, founder of Valtira and enStratus, states that using cloud infrastructures saves 18–29% before considering that you no longer need to buy for peak capacity [3]. And according to Dan Farber, editor in chief, CNET News, we are at the beginning of the age of planetary computing where billions of people will be wirelessly interconnected, and the only way to achieve that kind of massive-scale usage is by massive-scale, *brutally efficient* cloud-based infrastructure [3]. And finally there are savings on power consumption since there are few computers on premises. Currently, servers are

used at only 15% of their capacity in many companies, and 80% of enterprise software expenditure is on installation and maintenance of software. Use of cloud applications can reduce these costs from 50 to 90% [3].

Automatic Updates Our economy is now an online economy because most of, if not all, businesses are now online and depend on software applications for day-to-day services. Software is continuously changing, and as business functionalities change, software need to be changed or updated. The cost of software updates and management has always been on the rise, usually surpassing the cost of new software. For companies to stay competitive and in many cases afloat, they must be consistently updating and changing software. The business of software updates and software management and licensing is a big drain on company resources. So having automatic updates and management from the cloud provider can be a great relief to any company. But updates are not limited to only software. Also not worrying about hardware updates is cost effective for companies.

Green Benefits of Cloud Computing There has been a vigorous debate about cloud computing energy consumption, and this debate is continuing, pitting those claiming that cloud computing is gobbling up resources as large cloud and social networking sites need daily megawatts of power to feed insatiable computing needs and those who claim that the computing model is indeed saving power from millions of servers left idling daily and consuming more power. We will discuss this more in the coming sections. For now, we think that there are indeed savings in power consumption by cloud computing.

Remote Access With a Web portal access to the cloud, company employees may be able to work while they are on the road, at home, or in the office. This is of great benefit to the company so that there is no downtime because somebody is not in the office.

Disaster Relief Many companies live in constant fear of disasters occurring when they have company vital data stored on premises. No one likes to be a victim of large-scale catastrophes such as devastating hurricanes, earthquakes, fires, and of course terrorist attacks. Such misfortunes can create havoc to the companies' vital data and disrupt operations even if there were limited physical damage. Additionally, there are smaller disasters like computer crashes and power outages that can also wreak havoc on a company vital data. While this is possible, there are many companies, especially small ones, that may not even have any disaster recovery plan, and some who have it may not be able to execute it effectively. This fear can be overcome with investments in cloud technology. Company's vital backup data can be safely stored on secure data centers on the cloud instead of the company's server room.

Self-Service Provisioning Cloud computing allows users to deploy their own virtual sets of computing resources like servers, network, storage, and others, as

needed without the delays, competency, and complications typically involved in physical resource acquisition, installation, and management. The cloud owners, irrespective of their physical location, not only can provide all the computing resources your organization needs but also have the necessary capacity needed to monitor, manage, and respond to the organization's daily and hour-by-hour infrastructure, software, and platform requirements.

Scalability Because of the minute-by-minute monitoring capability of cloud computing of an organization's computing needs and the ability to increase or reduce the required resources as the demand increases or decreases, cloud computing offers the best infrastructure, platform, and software scalability that cannot be matched in any owned computing facility.

Reliability and Fault Tolerance Because the cloud provider, with qualified professionals and experience, monitors the computing requirements of a client company and can easily scale to demand, cloud computing offers a high degree of reliability and fault tolerance.

Ease of Use To attract more customers, cloud providers must make the user interface as friendly as possible so that customers can scale into the cloud with least efforts.

Skills and Proficiency Some of the most sought-after assets from a cloud provider are profaneness, professionalism, and a vast skills set provided to the customers. Companies, especially small ones, would pay a high price to get an employee with the skills set, efficiency, proficiency, and experience found with cloud center staff.

Response Time Depending on the bandwidth at the company Web portal, cloud computing services normally have speed because the computing resources provided are modern and powerful to be able to accommodate large number of users.

Mobility Because of Web portal interface to the cloud, cloud computing essentially is a mobile computing platform, allowing the users to access their applications.

Increased Storage Storage is cloud computing's main function. Because of this, it is cheap and readily scalable to need.

Others Benefits Other benefits beyond those we discussed above include providing a high quality of service (QoS); providing a high-quality, well-defined, and stable industry standard API; and on-demand availability of computing resources based on "at hand" financial constraints.

Security We are going to discuss this more in the coming section, but cloud computing, because of its individual virtual machines created per use, has already a built-in security provision. In addition to these built-in provisions due to

virtualization, the cloud model also offers a strong authentication regime at the browser interface gateway, a security mechanism that is individually and quickly set up and torn down as needed, and a strong validation and verification scheme that is expensive to deploy at an individual client-server model.

22.7 Cloud Computing, Power Consumption and Environmental Issues

As we briefly discussed in the last section, there is a heated debate ongoing pitting those claiming that cloud computing is gobbling up resources as large cloud and social networking sites need daily megawatts of power to feed insatiable computing needs and those who claim that the computing model is indeed saving power from millions of servers left idling daily and consuming more power. Let us not interject ourselves into the debate.

In the paper "Make IT Green: Cloud Computing and its Contribution to Climate Change" [4], Greenpeace called attention to the growing, power-hungry data center footprint, citing estimates that cloud computer sites could consume up to 622.6 billion kWh (kilowatts per hour) of power and also that the quintessential cloud computing devices like the smartphones including Apple iPad and the Androids offering online access to the cloud and social networks can contribute to a much larger carbon footprint of the information technology sector than previously estimated. Greenpeace has supporters in their camp, and these supporters are professionals adding weight to the debate. For example, in his paper "The Environmental Cost of Cloud Computing: Assessing Power Use and Impacts," Jonathan Koomey [5], professor of civil and environmental engineering at Stanford University, claims that by 2009, the cloud was responsible for 1–2% of the world's electricity use [6]. All major cloud providers including the Google, Facebook, Amazon, Yahoo, Microsoft, Apple, and others are involved in one way or the other in dealing with power producers to provide them with enormous amounts of power to run the data centers. For example, in 2010, Facebook signed an agreement with PacifiCorp, a utility provider in the northwest, to provide it with power for its Oregon data center. This did not go well with Greenpeace because PacifiCorp gets most of its energy from coal-fired power stations, the known largest source of greenhouse gas emissions in the United States.

In their climate projection for year 2020 report titled "SMART 2020: Enabling the Low Carbon Economy in the Information Age" [5], the Climate Group and the Global e-Sustainability Initiative (GeSI) noted the growth of 9.5% a year in ICT electricity consumption and GHG emissions by 2020. Projections like these are adding fuel to the debate about cloud computing technology and the environment [7].

The other side of the debate is no less vigorous. They are holding the view that because cloud computing technology makes companies share pooled resources and facilities, this is indeed helping to contain the expected degradation of the environment resulting from the projected, relentless, and viral growth of information

technology device use as projected in the Climate Group and the Global e-Sustainability Initiative (GeSI) 2007 report titled "SMART 2020: Enabling the Low Carbon Economy in the Information Age" [5]. In that report it is projected that:

- PC ownership will quadruple between 2007 and 2020 to 4 billion devices, and emissions will double over the same period, with laptops overtaking desktops as the main source of global ICT emissions expected to reach 22% mark.
- Mobile phone ownership will almost double to nearly 5 billion accounts by 2020, but emissions will only grow by 4%.
- Broadband uptake will treble to almost 900 million accounts over the same period, with emissions doubling over the entire telecom infrastructure.

In the same report, the group makes a compelling case for ICT's significant potential to deliver climate and energy solutions, estimating that ICTs could cut 7.8 GtCO2 of global greenhouse gas emissions by 2020, a 15% reduction. We concur with this report in saying that new innovations in cloud computing technologies together with increased awareness and use of renewable energy can make cloud technology a greener technology and make it use less power.

22.8 Cloud Computing Security, Reliability, Availability, and Compliance Issues

The cloud computing model as we know it today did not start overnight. The process has taken years moving through seven software models beginning with in-house software, licensed software normally referred to as the traditional model, open source, outsourcing, hybrid, Software as a Service, and finally the Internet model, the last two being part of the cloud computing model. When one carefully examines the cloud servicing model, one does not fail to notice the backward compatibilities or the carryovers of many of the attributes that characterized software through all the models. While this brings the benefits of each one of those software models, also many, if not all, of the software complexity and security issues in those models were carried over into the cloud computing model. Because of this, our first thought was to discuss the security issues in the cloud computing model through the prism of these models. It is tempting, but we are going to follow a different path while keeping the reader rooted into the different software models. Security is and continues to be a top issue in the cloud computing model. The other three related issues are performance, compliance, and availability. We will discuss all four in this section, but since security is the number one issue, we will address it first.

We want to start the discussion of cloud computing security by paraphrasing Greg Papadopoulos, CTO of Sun Microsystems, who said that cloud users normally "trust" cloud service providers with their data like they trust banks with their money. This means that they expect the three issues of security, availability, and

performance to be of little concern to them as they are with their banks [3]. To give a fair discussion of the security of anything, one has to focus on two things that are the actors and their roles in the process you are interested in securing and the application or data in play. The application or data is thought of in relation to the state it is in at anyone's time. For example, the states for both data and application can be are either in motion between the remote hosts and the service provider's hypervisors and servers or in the static state when it is stored at remote hosts, usually on the customer's premises or in the service provider's servers. The kind of security needed in either one of these two states is different.

22.8.1 Cloud Computing Actors, Their Roles, and Responsibilities

In the cloud computing model, the main players are the cloud provider, the customer who is the data owner and who seeks cloud services from the cloud provider, and the user who may or may not be the owner of the data stored in the cloud. The first two players have delegated responsibilities to all who work on their behalf. To fully understand these delegated responsibilities assigned to each one of these, we need to look at first the marginal security concerns resulting from the peripheral system access control that always result in the easiest breach of security for any system, usually through compromising user accounts via weak passwords. This problem is broad affecting both local and outsourced cloud solutions. Addressing this and all other administrative and user security concern requires companies offering and using cloud solutions to design an access control regime that covers and requires every user, local or remote, to abide by these access policies including the peripheral ones like the generation and storage of user passwords. Additionally, employees need to be informed of the danger of picking easy passwords and to understand the danger of writing a password down or divulging a password to anyone. Access control administration is so important that cloud providers spend large amounts of resources to design a strong assess control regimes. For example, let us look at the access control of the three major cloud providers, namely, Amazon Web Services (AWS), Microsoft Windows Azure, and Rackspace.

Amazon Web Services

With Amazon Web Services (AWS) EC2 and S3, one can remotely self-provision what they want seamlessly. This kind of easiness, while great, created a set of security problems unless there is a strong access control regime in place. For Amazon, the solution is through use of Amazon Identity and Access Management (IAM). This allows the account owner to create multiple accounts for other authorized users on a single amazon account. Then as usual, each user is assigned permissions on the main account, accessible via user ID and passwords based on the user's role and responsibility in the customer's company. Based on the traditional access control, fine-grained security can be attained for all service users.

Microsoft Windows Azure

Microsoft Azure on the other hand has several security modules including [8]:

- Azure Security Center—to prevent, detect, and respond to threats with increased visibility and control over the security of your Azure resources. This is done by:
  - Understand the cloud security state—Letting the user to get a central view of the security state of all Azure resources. This quickly lets the user make sure that all security controls are in place, and the user can quickly identify any resources needing attention.
  - Take control of cloud security—Using a security policy-driven monitoring of all security configurations to guide resource owners through the process of implementing their required controls.
  - Easily deploy integrated cloud security solutions—Rapidly enable a range of security solutions from Microsoft and its partners, including industry-leading firewalls and antimalware. Use streamlined provisioning to easily deploy security solutions.
  - Detect threats and respond fast—Staying ahead of current and emerging cloud threats requires an integrated, analytics-driven approach. By combining Microsoft global threat intelligence and expertise with insights into cloud security-related events across a customer Azure deployments, Azure Security Center helps the customer detect actual threats early, and it reduces false positives. Cloud security alerts offer insights into the attack campaign, including related events and impacted resources, and suggest ways to remediate issues and recover quickly.

Rackspace

In Rackspace, client authentication is done by the Cloud Authentication Service, also known as *Auth*. Auth allows each client needing authentication to obtain an authentication *token* and a list of regional service endpoints to the various services available in the cloud. Users must authenticate their credentials, but once authenticated they can create/delete containers and objects within that account. Since the Cloud Files system is designed to be used by many different customers, each user account is the user's portion of the Cloud Files system. Each client authentication is provided via a ReST (see below) interface which requires two headers, *X-Auth-User* and *X-Auth-Key* or *X-Auth-Token* with values for the *username* and *API access key*, respectively. Clients obtain this *token*, along with the Cloud Servers API URL, by first using the Rackspace Cloud Authentication Service [17].

Request: ReST
To authenticate, the client provides the following in x-headers:

- Username as *X-Auth-User x-header*
- API access key (from the Rackspace Cloud Control Panel in client Account API Access section) as *X-Auth-Key*

Response
Upon successful authentication, an HTTP status 204 is returned with the X-Storage-Url, X-CDN-Management-Url, and X-Auth-Token headers. Any 2xx response is a good response. For example, a 202 response means the request has been accepted. Also, additional x-headers may be returned. These additional headers are related to other Rackspace services and can be ignored. An HTTP status of 401 (unauthorized) is returned upon authentication failure. All subsequent container/object operations against Cloud Files should be made against the URI specified in X-Storage-Url or X-CDN-Management-Url and must include the X-Auth-Token header [17].

After these exchanges, the client is ready to use the cloud.

22.8.2 Security of Data and Applications in the Cloud

Let us next look at the security of data and applications in the cloud. To do this we need to focus first on the security and role of the hypervisor and then the servers on which user services are based. A hypervisor also called *virtual machine manager (VMM)* is one of the many hardware virtualization techniques allowing multiple operating systems, termed *guests*, to run concurrently on a host computer. The hypervisor is piggybacked on a kernel program, itself running on the core physical machine running as the physical server. The hypervisor presents to the guest operating systems a virtual operating platform and manages the execution of the guest operating systems. Multiple instances of a variety of operating systems may share the virtualized hardware resources. Hypervisors are very commonly installed on server hardware, with the function of running guest operating systems, that themselves act as servers. The security of the hypervisor therefore involves the security of the underlying kernel program and the underlying physical machine, the physical server, and the individual virtual operating systems and their anchoring virtual machines.

Hypervisor Security
The key feature of the cloud computing model is the concept of virtualization. We covered virtualization in the previous chapter. It is this virtualization that gives the cloud the near instant scalability and versatility that makes cloud computing so desirable a computing solution by companies and individuals. The core of virtualization in cloud computing is the easy process of minting of virtual machines on demand by the hypervisor. The hypervisor allocates resources to each virtual machine it creates, and it also handles the deletion of virtual machines. Since each virtual machine is initiated by an instance, the hypervisor is a bidirectional conduit into and out of every virtual machine. The compromise of either, therefore, creates a danger to the other. However, most hypervisors are constructed in such a way that

there is a separation between the environments of the sandboxes (the virtual machines) and the hypervisor. There is just one hypervisor, which services all virtual sandboxes, each running a guest operating system. The hypervisor runs as part of the native monolithic operating system, side by side with the device drivers, file system, and network stack, completely in kernel space. So, one of the biggest security concerns with a hypervisor is the establishment of covert channels by an intruder—"hyperjacking." If an intruder succeeds in establishing a covert channel, either by modifying file contents or through timing, it is possible for information to leak from one virtual machine instance to another [9].

Also since the hypervisor is the controller of all virtual machines, it, therefore, becomes the single point of failure in any cloud computing architecture. That is, if an intruder compromises a hypervisor, then the intruder has control of all the virtual machines the hypervisor has allocated. This means that the intruder can even create or destroy virtual machines at will. For example, the intruder can perform a denial-of-service attack, by bringing down the hypervisor which then brings down all virtual machines running on top of the hypervisor.

The processes of securing virtual hosts differ greatly from processes used to secure their physical counterparts. Securing virtual entities like a hypervisor, virtual operating systems, and corresponding virtual machines is more complex. To understand hypervisor security, let us first discuss the environment in which the hypervisor works. Recall that a hypervisor is part of a virtual computer system (VCS). In his 1973 thesis in the Division of Engineering and Applied Physics, Harvard University, Robert P. Goldberg defines a virtual computer system as a hardware-software duplicate of a real existing computer system in which a statistically dominant subset of the virtual processor's instructions execute directly on the host processor in native mode. He also gives two parts to this definition, the environment and implementation [10].

Environment The virtual computer system must simulate a real existing computer system. Programs and operating systems which run on the real system must run on the virtual system with identical effect. Since the simulated machine may run at a different speed from the real one, timing-dependent processor and I/O code may not perform exactly as intended.

Implementation Most instructions being executed must be processed directly by the host CPU without recourse to instruction by instruction interpretation. This guarantees that the virtual machine will run on the host with relative efficiency. It also compels the virtual machine to be similar or identical to the host and forbids tampering with the control store to add an entirely new order code.

In the environment of virtual machines, a hypervisor is needed to control all the sandboxes (virtual machines). Generally in practice, the underlying architecture of the hypervisor determines if there is a desired true separation between the sandboxes or not. Robert P. Goldberg classifies two types of hypervisor [11]:

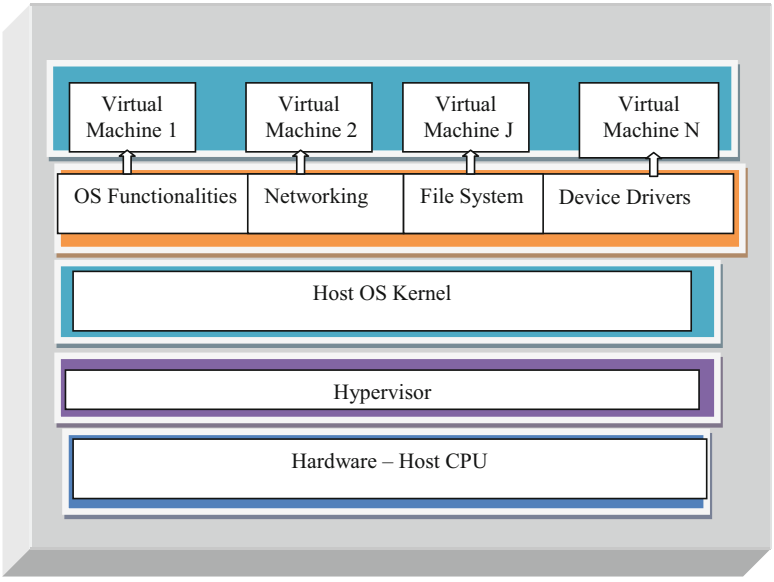

Fig. 22.2 Type-1 hypervisor

Type-1 (Or Native, Bare Metal) Hypervisors run directly on the host's hardware to control the hardware and to manage guest operating systems. See Fig. 22.2. All guest operating systems then run on a level above the hypervisor. This model represents the classic implementation of virtual machine architectures. Modern hypervisors based on this model include Citrix XenServer, VMware ESX/ESXi, and Microsoft Hyper-V. The most common commercial hypervisors are based on a monolithic architecture below.

The underlying hypervisor services all virtual sandboxes, each running a guest operating system. The hypervisor runs as part of the native monolithic operating system, side by side with the device drivers, file system, and network stack, completely in kernel space.

Type-2 (Or Hosted) Hypervisors run just above a host operating system kernel such as Linux, Windows, and others as in Fig. 22.3. With the hypervisor layer as a distinct second software level, guest operating systems run at the third level above the hardware. The host operating system has direct access to the server's hardware like host CPU, memory, and I/O devices and is responsible for managing basic OS services. The hypervisor creates virtual machine environments and coordinates calls to CPU, memory, disk, network, and other resources through the host OS. Modern hypervisors based on this model include KVM and VirtualBox.

The discussion so far highlights the central role of the hypervisor in the operations of virtual machine systems, and it points to its central role in securing all virtual machine systems. Before we look at what can be done to secure it, let us ask ourselves what security breaches can happen to the hypervisor. There are

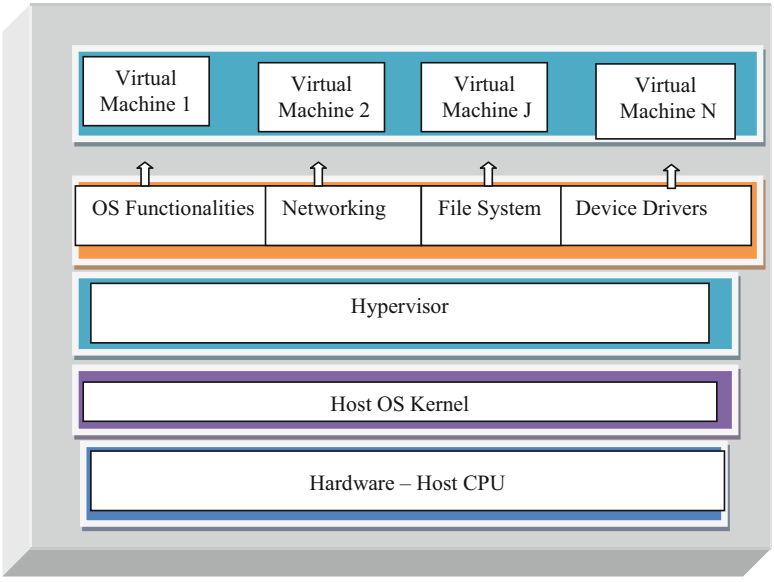

Fig. 22.3 Type-2 hypervisor

several of these security breaches most severe involving self-installation of both malware and rootkits masquerading as they are the hypervisor.

Hacking the Hypervisor

In his blog "Yes, Hypervisors Are Vulnerable," Neil MacDonald, vice president and a Gartner Fellow [12], observes the following about hypervisor and the vulnerabilities associated with it:

- The virtualization platform (hypervisor/VMM) is a software written by human beings and will contain vulnerabilities. Microsoft, VMware, Citrix, and others, all of them will and have had vulnerabilities.
- Some of these vulnerabilities will result in a breakdown in isolation that the virtualization platform was supposed to enforce.
- Bad guys will target this layer with attacks. The benefits of a compromise of this layer are simply too great.
- While there have been a few disclosed attacks, it is just a matter of time before a widespread publicly disclosed enterprise breach is tied back to a hypervisor vulnerability.

As we observed in Chap. 21, there has been a growing increase in the virtualization vulnerabilities. Published papers have so far shown that the security of hypervisors can be undermined. As far back as 2006, Samuel T. King, Peter M. Chen, Yi-Min Wang, Chad Verbowski, Helen J. Wang, and Jacob R. Lorch demonstrated in their paper, "SubVirt: Implementing malware with virtual machines," the use of type of

malware, which they called a virtual machine-based rootkit (VMBR), installing a virtual machine monitor underneath an existing operating system and hoisting the original operating system into a virtual machine [14].

In fact in this study, the authors demonstrated a malware program that started to act as its own hypervisor under Windows. According to the IBM X-Force *2010 Mid-Year Trend and Risk Report*, which disclosed a 10-year virtualization vulnerability trend from 1999 through 2009, there were 373 reported vulnerabilities affecting virtualization solutions during the period with a steady growth trend starting around 2002 and peaking in 2008 to 100 and falling off by 12 percent in 2009. What do we learn from all these? We learn that the hypervisor layer of virtualization, playing the core role in the virtualization process, is very vulnerable to hacking because this is the weakest link in the data center. Therefore, attacks on hypervisor are on the rise. Data from the IBM X-Force *2010 Mid-Year Trend and Risk Report* show that every year since 2005, vulnerabilities in virtualization server products, the hypervisors, have overshadowed those in workstation products, an indication of the hackers' interest in the hypervisors. The report further shows that 35% of the server virtualization vulnerabilities are vulnerabilities that allow an attacker to "escape" from a guest virtual machine to affect other virtual machines or the hypervisor itself. Because, as Fig. 22.2 shows, the hypervisor in *type-1* environment is granted CPU privilege to access all system I/O resources and memory, this makes it a security threat to the whole cloud infrastructure. Just one vulnerability in the hypervisor itself could result in a hacker gaining access to the entire system, including all the guest operating systems. Because malware runs below the entire operating system, there is a growing threat of hackers using malware and rootkits to install themselves as a hypervisor below the operating system thus making them more difficult to detect. Once installed and operating below the main operating system, there is evidence that malware can intercept any operations of the operating system without the antivirus software detecting them, although there is a debate with some scholars disputing this. Since 2009, there have been efforts in producing kernel-mode anti-rootkit products, such as HooSafe from Microsoft and North Carolina State University and Rackspace's NoVirusThanks Anti-Rootkit [14].

In *type-2* hypervisor configuration (Fig. 22.3), the microkernel architecture is designed specifically to guarantee a robust separation of application partitions. This architecture puts the complex virtualization program in user space; thus every guest operating system uses its own instantiation of the virtualization program. In this case, therefore, there is a complete separation between the sandboxes (virtual boxes), thus reducing the risks exhibited in type-1 hypervisors.

An attack, therefore, on *type-2* hypervisors can bring down one virtual box, not more, and cannot bring down the cloud infrastructure as is the case in *type-1* hypervisors.

According to Samuel T. King et al., overall, virtual machine-based rootkits are hard to detect and remove because their state cannot be accessed by software running in the target system. Further, VMBRs support general-purpose malicious services by allowing such services to run in a separate operating system that is protected from the target system [13].

Securing Load Balancers

For every hypervisor, there is a load balancer, used to route traffic to different virtual machines to help spread traffic evenly across available machines. Load balancers in a hypervisor play a vital role of ensuring a fair distribution of available load to all virtual machines especially during high traffic and ensuring the full utilization of the cloud infrastructure. Elastic load balancers play a central in the cloud infrastructure along the following lines [14]:

- It listens to all traffic destined for the internal network and distributes incoming traffic across the cloud infrastructure.
- Automatically scales its request-handling capacity in response to incoming application traffic.
- It creates and manages security groups associated with each instance and provides additional networking and security options if and when needed.
- It can detect the health of the virtual machines, and if it detects unhealthy load-balanced virtual machine, it stops routing traffic to it and spreads the load across the remaining healthy virtual machines.
- It supports the ability to stick user sessions to specific virtual machines.
- It supports SSL termination at the load balancer, including offloading SSL decryption from application virtual machines, centralized management of SSL certificates, and encryption to backend virtual machines with optional public key authentication.
- It supports use of both the Internet Protocol version 4 and 6 (IPv4 and IPv6).

Due to the load balancer's ability to listen and process all traffic that is destined to the internal network of the cloud, it is a prime target for attackers. If a load balancer was compromised, an attacker could listen to traffic and may compromise secure traffic destined to outside the network. Additionally, if the load balancer is compromised along with a virtual machine, traffic could be directed to an unsecure internal server where further attacks are launched [15]. Because the load balancer is a single point in the cloud infrastructure, it is very vulnerable to denial-of-service attacks, if it is compromised. This can lead to cloud activity disruption.

What is the best way to secure the load balancer from attacks then? A load balancer is normally secured through proper configuration and monitoring of the balancer's logs. This is achieved through restriction of access to administration of the balancer itself by configuring the load balancer to only accept administrative access over a specific administrative network. This administrative network should be connected to the administrative only network. Limiting access over the administrator network greatly limits the number of users with access to the load balancer [16].

Virtual Operating System Security

Besides the hypervisor, the virtualization system also hosts virtual servers each running either a guest operating system or another hypervisor. And on the peripheral of the virtual machine system are the consoles and hosts. Through each one of these resources, the virtual machine system can be susceptible to security vulnerabilities. Let us briefly look at these since they were covered in more detail in the previous chapter.

Host Security

Through hosts like workstations, user gains access to the virtual machine system, hence to the cloud. Two problems are encountered here:

- Escape-to-hypervisor vulnerabilities—that allow intruders to penetrate the virtual machine from the host
- Escape-to-host vulnerabilities—that allow vulnerabilities in the virtual machine to move to the hosts

Guest Machines

Guest machines running guest operating system can also pose a security problem to the cloud. However, as we saw in the previous chapter, vulnerabilities in the guest virtual machines are confined to that machine, and they rarely affect other machines in the system.

22.8.3 Security of Data in Transition: Cloud Security Best Practices

With the vulnerabilities in the cloud we have discussed above, what is the best way to protect the user of the cloud? For a cloud customer, the key areas of concerns are virtualization technology security vulnerabilities that may be encountered during the use of the cloud that may affect the customer and unauthorized access to customer data and other resources stored or implemented in the cloud, whether the cloud provider uses strong enough encryption to safeguard customer data, secure access, and use of cloud applications and secure cloud management. Let us next discuss the best practices that try to address some of these concerns.

22.8.4 Service-Level Agreements (SLAs)

A service-level agreement (SLA) is a service contract between the provider of a service and the client defining the level of expected service in terms of security, availability, and performance. The cloud service-level agreements (SLAs) are a series of service contracts between cloud providers and clients to define the level(s) of service based on the types of services sought by the client because the effectiveness of these contracts depends on how well maximized and tailored these services are to the particular needs of each client. For example, the security of services sought by a client may depend on the tier of cloud offering the client is using. To see how involved and intricate these documents can be, take an example of security concerns. For IaaS, the security responsibilities are shared with the provider responsible for physical, environmental, and virtualization security, while the client takes care of the security in applications, operating system, and others. Now if we change the service model to SaaS, the provider is responsible for almost every aspect of security.

22.8.5 Data Encryption

The moment data leaves your endpoint Web-cloud access point in your location, it travels via a public network and stored in shared environment—the cloud. In a public or in a shared environment, data can be intercepted and infiltrated by intruders from within and outside the cloud and during transmission from man-in-the-middle cryptanalysis. To prevent these kinds of breaches, strong encryptions and authentications regimes are needed. Encryption to safeguard any kinds of data breaches required a strong access control and authentication to all Web-based cloud resource interface, encryption of all administrative access to the cloud hypervisor, and all access to applications and data.

22.8.6 Web Access Point Security

Most cloud access instances are Web based. Most security breaches to stored data originated from Web applications. There is therefore a need for strong security controls in the cloud APIs.

22.8.7 Compliance

Because most clouds are either public, community, or hybrids and clients using these clouds usually are in businesses that deal with personal data, cloud providers must observe a number of compliance regulations including FISMA, HIPAA, SOX, and SAS 70 II for clouds based in the United States and the Data Protection Directive for clouds based in the EU. In addition, providers accepting payments using credit card must comply with PCI DSS. Let us briefly look at these [16]:

The *Federal Information Security Management Act of 2002 (FISMA)* is a United States federal law enacted in 2002 as Title III of the E-Government Act of 2002 (Pub.L. 107–347, 116 Stat. 2899). The act recognized the importance of information security to the economic and national security interests of the United States. It requires federal agencies to develop, document, and implement an agency-wide program to provide information security for the information and information systems that support the operations and assets of the agency, including those provided or managed by another agency, contractor, or other source. So cloud providers intending to contract with or affiliate with any organization that provides services to the U.S. Federal government must adhere to FISMA.

The *Health Insurance Portability and Accountability Act of 1996* (HIPAA; Pub.L. 104–191, 110 Stat. 1936, enacted August 21, 1996) was enacted by the United States Congress 1996. It has two parts: Title I which protects health insurance coverage for workers and their families when they change or lose their jobs and Title II, also known as the Administrative Simplification (AS) provisions, requires the establishment of national standards for electronic health care transactions and national identifiers for providers, health insurance plans, and

employers. The AS also address the security and privacy of health data. Cloud providers must abide by this provision.

Exercises

1. What is cloud computing?
2. Discuss the software models predating cloud computing.
3. Discuss the major models in cloud computing.
4. What are the benefits of cloud computing over Software as a Service (SaaS)?
5. Define and discuss Software as a Service (SaaS), Infrastructure as a Service (IaaS), and storage as a service.
6. Describe the seven business models of software.
7. Discuss the services that make up/describe cloud computing.
8. Discuss the differences between clouding computing and virtualization.
9. Discuss four business applications best suited for cloud computing.
10. To determine what business applications should go on the cloud, you need to estimate the return on investment for that application. What can you consider when computing ROI?
11. List and discuss three characteristics an application must have in order to be considered suited for the cloud.
12. What is MapReduce? Describe the structure and working of MapReduce.
13. What is Hadoop? Describe the three subprojects of Hadoop.
14. The structure and working of MapReduce are as follows:
    Structure
    - Map—for each input (Ki,Vi) produce zero or more output pairs (Km,Vm).
    - Combine—optional intermediate aggregation (less M->R data transfer).
    - Reduce—for input pair (Km, list(V1,V2... Vn)) produce zero or more output pairs (Kr,Vr).

Describe the functions in the dataflow of MapReduce that you need to write to make it usable.

Advanced Exercises

Cloud Computing Semester Projects

Research Guidelines

For these projects you can work either in groups or individually. Study the problems and use MapReduce/Hadoop, and try to find meaningful answers to the chosen problem through research and development.

Write a 20-double-spaced-page paper summarizing your research. The paper should be at a publishable level, meaning that it must have an abstract and must follow either IEEE or ACM scientific presentation guidelines. The paper must also be well referenced.

1. *Text Mining and Sentiment Analysis for Online Forums, Hotspot Detection, and Forecast*

Text sentiment analysis, or emotional polarity computation as it is commonly referred to also, has become a flourishing frontier in the text mining community. The research involves a careful study of online forums for hotspots, to detect and forecast options forming in these hotspots using sentiment analysis and text mining techniques and relevant tools.

Using MapReduce or Hadoop, split the chosen text into subtexts and analyze the emotional polarity of each subtext and to obtain a value for each subtext. Then try to group together the subtext to obtain value for the original text. Do this to as many texts as you can in the corpus or forum.

Then, use these analyzed pieces and a good and unsupervised text mining approach to group the forums into various clusters, with the center of each representing a hotspot forum within the current time span. The data sets used in your research should come from all social networks and the open Web.

Comment on whether your results demonstrate a good way of forecasting and whether it achieves highly consistent results. List ten top hotspot forums you found.

2. *Unlock the Power of the Mobile Cloud Through Development of New Powerful Cloud Mobile Hybrid Application Development*

Rapid advances and marriage between computing and telecommunication technologies have created an ubiquitous landscape of powerful smart mobile computing devices capable of and are progressively being used in a pay-as-you-go computing model, now called mobile cloud computing. This model of cloud computing is increasingly using a growing amount of big data, particularly in data-intensive domains. Define and develop a new class of cloud mobile hybrid (CMH) applications, collective applications that have a cloud-based back end and a mobile device front end that are capable of advancing and unlocking the huge potential of the mobile cloud landscape and its capabilities.

3. *Protect the Mobile Cloud Through Development of Applications*

Rapid advances and marriage between computing and telecommunication technologies have created an ubiquitous landscape of powerful smart mobile computing devices capable of and are progressively being used in a pay-as-you-go computing model, now called mobile cloud computing. This new model of cloud computing and the resulting mobile cloud landscape is currently highly unsecure from rogue application to unscrupulous users; it is a Wild West in there. Define and develop a new class of secure cloud mobile hybrid (CMH) applications.

4. *Opinion and Public Sentiment Setters and Leaders*

Rapid advances and marriage between computing and telecommunication technologies have created an ubiquitous landscape of powerful smart mobile computing devices capable of and are progressively being used as Internet-ready devices capable of and powerful enough to perform many of the Internet functionalities that are currently confined within the spheres of their big brothers, the PC and laptops. This hybrid new landscape is increasingly being used as trend and public opinion setters, thanks to Twitter, Facebook, and the like. Most opinion and trendsetting takes plan in social network groups or clusters. Use modern tools dealing with big data and text mining techniques to develop an analysis that identifies social groups for leaders, followers, and special hot topics and trends within the social networks. You can use your findings to comment on public options based on demographics of society.

5. *Real-Time Cloud Notifier*

Cloud Notify is a Web application that takes advantage of the cloud to provide management and notification services to users. Using Cloud Notify, users can create one or more topics and assign subscribers to those topics. Users can then send notifications either via text message or e-mail to the subscribers. The system should be able to bounce and inform the management console who is off and who has seen the message and the locations of those who have not seen the message. The system should be able to re-notify—two more times in a given period of time—and send a warning message to the management console.

References

1. Mell P, Grance T. The NIST definition of cloud computing, NIST special publication 800–145. http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf
2. Metzler J, Taylor S. The data center network transition: wide area
3. Networking Alert. Network World, August 08, 2011. http://www.networkworld.com/newsletters/frame/2011/080811wan1.html?source=nww_rss
4. Mell P, Grance T. Effectively-and-securely-using-the-cloud-computing-paradigm. http://www.scribd.com/doc/13427395/
5. Greenpeace. Make IT Green: cloud computing and its contribution to climate change, Greenpeace, USA. http://www.greenpeace.org/international/Global/international/planet-2/report/2010/3/make-it-green-cloud-computing.pdf
6. Koomey J. The environmental cost of cloud computing: assessing power use and impacts. http://www.slideshare.net/gigaom/jonathan-koomey-the-environmental-cost-of-cloud-computing
7. Climate Group and the Global e-Sustainability Initiative (GeSI). SMART 2020: enabling the low carbon economy in the information age. http://gesi.org/article/43
8. Azure Security Center. https://azure.microsoft.com/en-us/services/security-center/
9. Larry Dignan Virtualization: what are the security risks? http://www.zdnet.com/article/virtualization-what-are-the-security-risks/
10. Bob Violino. Five cloud security trends experts see for 2011. http://www.csoonline.com/article/647128/five-cloud-security-trends-experts-seefor-2011

11. Goldberg RP (1973) Architectural principles for virtual computer systems. Harvard University, Cambridge, MA, pp 22–26
12. Wikipedia. http://en.wikipedia.org/wiki/Hypervisor
13. MacDonald N (2011) Yes, hypervisors are vulnerable, January 26, 2011. http://blogs.gartner.com/neil_macdonald/2011/01/26/yes-hypervisors-are-vulnerable/
14. King ST, Chen PM, Wang YM, Verbowski C, Wang HJ, Lorch JR (2006) SubVirt: implementing malware with virtual machines. http://web.eecs.umich.edu/~pmchen/papers/king06.pdf
15. Elastic Load Balancing, *AWS*. http://aws.amazon.com/elasticloadbalancing/
16. Global Information Assurance Certification Paper. https://www.giac.org/paper/gsna/119/ids-load-balancer-security-audit-administrators-perspective/103792
17. Wikipedia. http://en.wikipedia.org/wiki/Cloud_computing. Authenticate to the Rackspace Cloud. https://developer.rackspace.com/docs/cloud-files/v1/getting-started/authenticate/#authenticate-to-cloud