Πανεπιστήμιο Πειραιώς – Τμήμα Ψηφιακών Συστημάτων

Πρόγραμμα Μεταπτυχιακών Σπουδών

«Ασφάλεια Ψηφιακών Συστημάτων»

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	**Cloud Security Services and Privacy**
Ονοματεπώνυμο Φοιτητή	**Ελευθέριος Σοφικίτης**
Πατρώνυμο	**Σπυρίδων**
Αριθμός Μητρώου	**MTE0928**
Επιβλέπων	**Κ. Λαμπρινουδάκης, Επίκουρος Καθηγητής**

Ημερομηνία Παράδοσης Ιανουάριος 2013

Table of Contents

ABSTRACT

Cloud computing is a double-edged sword from the privacy and security standpoints. Despite its potential to provide a low cost security, organizations may increase risks by storing sensitive data in the cloud. In this thesis, is analyzed how the cloud's characteristics such as newness, nature of the architecture, and attractiveness and vulnerability as a cybercrime target are tightly linked to privacy and security. It is tried to approach the issue from both sides, and the one of the user and the one of Cloud Service Provider. Also the XACML language and an a policy sample are presented providing that way that XACML policies may help to secure a heterogeneous environment as the one of Cloud computing.

Introduction

Ideally, cloud computing and the security that it affords, should align, but they usually do not. It has become a common mantra in the high-technology industry to chant “cloud computing good” while at the same time saying “cloud security bad.” But what does that really mean? Exactly what is wrong with security in cloud computing? The purpose of this thesis is to answer those questions through a systematic investigation of what constitutes cloud computing and what security it offers. As such, this thesis also explores the implications of cloud computing security on privacy, auditing, and compliance for both the cloud service provider (CSP) and the customer. Is security in cloud computing a bad thing? The answer depends on what you use cloud computing for, and your expectations. If you are a large organization with significant resources to devote to a sophisticated information security program, you need to overcome a number of security, privacy, and compliance challenges that we explore later in the book. However, if you are a small to medium-size business (SMB), the security of cloud computing might look attractive, compared to the resources you can afford to spend on information security today.

Chapter 1. The Evolution of Cloud Computing

To understand what cloud computing is and is not, it is important to understand how this model of computing has evolved. As Alvin Toffler notes in his famous book, The Third Wave (Bantam, 1980), civilization has progressed in waves (three of them to date: the first wave was agricultural societies, the second was the industrial age, and the third is the information age). Within each wave, there have been several important subwaves. In this post-industrial information age, we are now at the beginning of what many people feel will be an era of cloud computing.

In his book The Big Switch (W.W. Norton & Co., 2008), Nicholas Carr discusses an information revolution very similar to an important change within the industrial era. Specifically, Carr equates the rise of cloud computing in the information age to electrification in the industrial age. It used to be that organizations had to provide their own power (water wheels, windmills). With electrification, however, organizations no longer provide their own power; they just plug in to the electrical grid. Carr argues that cloud computing is really the beginning of the same change for information technology. Now organizations provide their own computing resources (power). The emerging future, however, is one in which organizations will simply plug in to the cloud (computing grid) for the computing resources they need. As he puts it, “In the end the savings offered by utilities become too compelling to resist, even for the largest enterprises. The grid wins.” In fact, Part 2 of his book is about “living in the cloud” and the benefits it provides. (Carr also discusses at length some of the perceived negative consequences to society of this big switch, specifically some of the darker aspects this change brings to society.)

Carr is not alone in arguing for the benefits of cloud computing, but he has put forth what is arguably the most articulate statement of those benefits thus far. And although he focuses specifically on the economic benefits of cloud computing, he does not discuss information security problems associated with "the big switch.

Figure 1 displays cloud computing and cloud service providers (CSPs) as extensions of the Internet service provider (ISP) model. In the beginning (ISP 1.0), ISPs quickly proliferated to provide access to the Internet for organizations and individuals. These early ISPs merely provided Internet connectivity for users and small businesses, often over dial-up telephone service. As access to the Internet became a commodity, ISPs consolidated and searched for other value-added services, such as providing access to email and to servers at their facilities (ISP 2.0). This version quickly led to specialized facilities for hosting organizations' (customers') servers, along with the infrastructure to support them and the applications running on them. These specialized facilities are known as collocation facilities (ISP 3.0). Those facilities are "a type of data center where multiple customers locate network, server, and storage gear and interconnect to a variety of telecommunications and other network service provider(s) with a minimum of cost and complexity." As collocation facilities proliferated and became commoditized, the next step in the evolution was the formation of application service providers (ASPs), which focused on a higher value-added service of providing specialized applications for organizations, and not just the computing infrastructure (ISP 4.0). ASPs typically owned and operated the software application(s) they provided, as well as the necessary infrastructure.

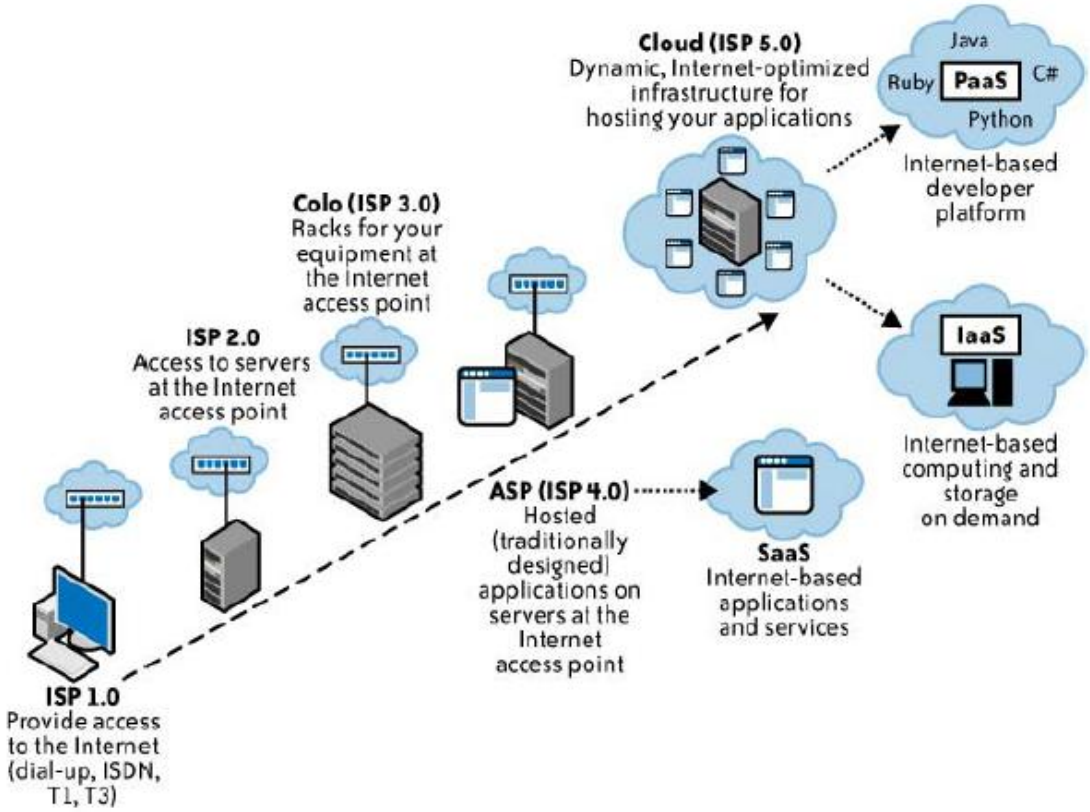


Figure 1

Although ASPs might appear similar to a service delivery model of cloud computing that is referred to as software-as-a-service (SaaS), there is an important difference in how these services are provided, and in the business model. Although ASPs usually provided

services to multiple customers (just as SaaS providers do today), they did so through dedicated infrastructures. That is, each customer had its own dedicated instance of an application, and that instance usually ran on a dedicated host or server. The important difference between SaaS providers and ASPs is that SaaS providers offer access to applications on a shared, not dedicated, infrastructure.

With increasing attention, some would say hype, now being paid to cloud computing, companies are increasingly claiming to be “cloudy.” Suddenly, many companies are claiming to operate “in the cloud.” Serious cloud washing is underway. Similarly, a number of computing groups have announced their efforts to promote some facet of cloud computing.

1.1 What Is Cloud Computing?

1.1.1 Cloud Computing Defined

Definition of cloud computing is based on five attributes: multitenancy (shared resources), massive scalability, elasticity, pay as you go, and self-provisioning of resources.

Multitenancy (shared resources)

Unlike previous computing models, which assumed dedicated resources (i.e., computing facilities dedicated to a single user or owner), cloud computing is based on a business model in which resources are shared (i.e., multiple users use the same resource) at the network level, host level, and application level.

Massive scalability

Although organizations might have hundreds or thousands of systems, cloud computing provides the ability to scale to tens of thousands of systems, as well as the ability to massively scale bandwidth and storage space.

Elasticity

Users can rapidly increase and decrease their computing resources as needed, as well as release resources for other uses when they are no longer required.

Pay as you go

Users pay for only the resources they actually use and for only the time they require them.

Self-provisioning of resources

Users self-provision resources, such as additional systems (processing capability, software, storage) and network resources.

One of the attributes of cloud computing is elasticity of resources. This cloud capability allows users to increase and decrease their computing resources as needed, as next figure illustrates. There is always an awareness of the baseline of computing resources, but predicting future needs is difficult, especially when demands are constantly changing. Cloud computing can offer a means to provide IT resources on demand and address spikes in usage.

Interest in the cloud is growing because cloud solutions provide users with access to supercomputer-like power at a fraction of the cost of buying such a solution outright. More importantly, these solutions can be acquired on demand; the network becomes the supercomputer in the cloud where users can buy what they need when they need it. Cloud computing identifies where scalable IT-enabled capabilities are delivered as a service to customers using Internet technologies.

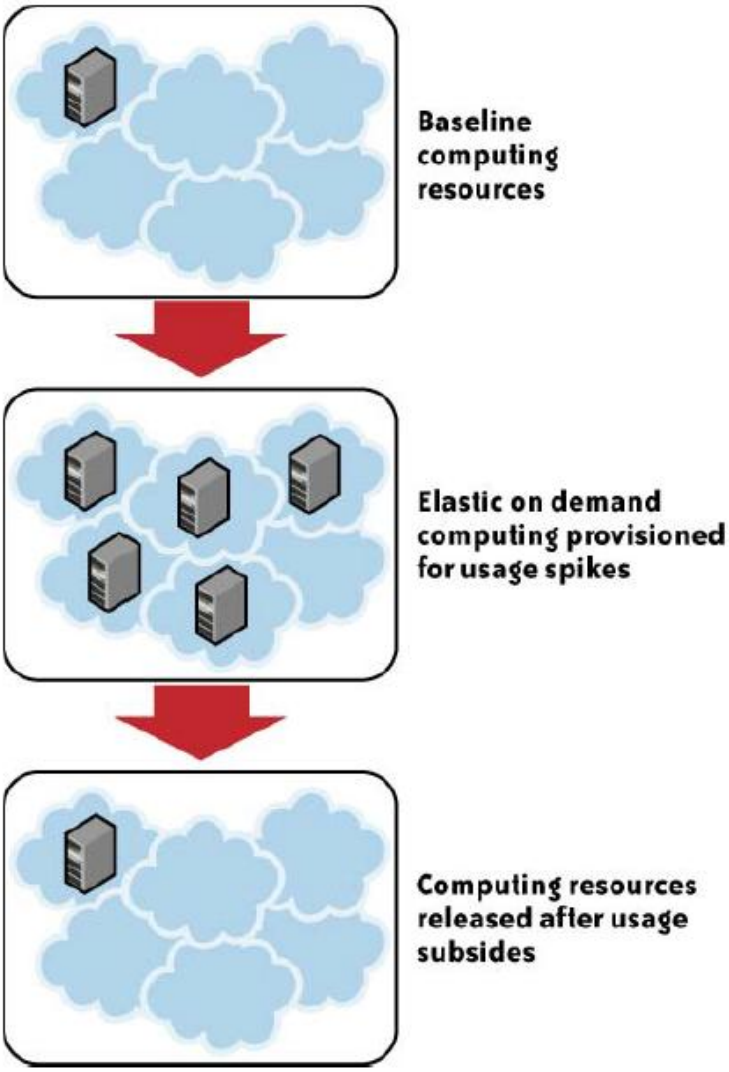

Baseline computing resources
Elastic on demand computing provisioned for usage spikes
Computing resources released after usage subsides

Figure 2

1.2 The SPI Framework for Cloud Computing

A commonly agreed upon framework for describing cloud computing services goes by the acronym "SPI." This acronym stands for the three major services provided through the cloud: software-as-a-service (SaaS), platform-as-a-service (PaaS), and infrastructure-as-a-service (IaaS).

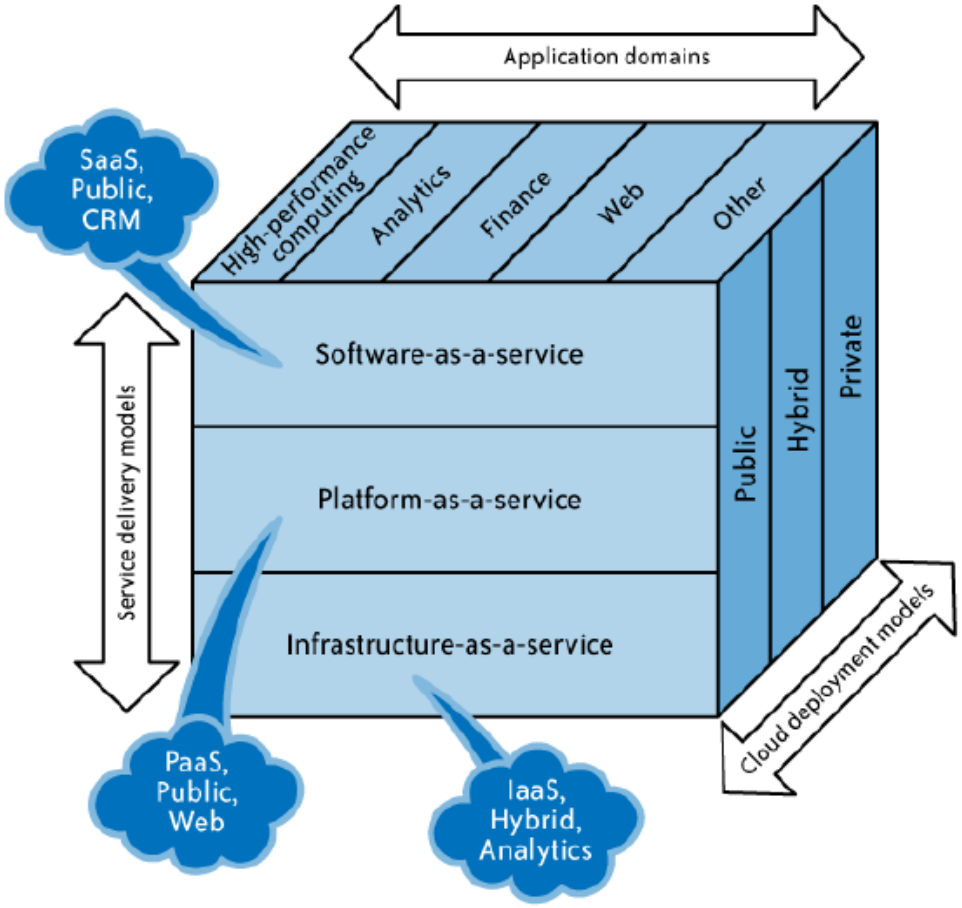


Figure 3

1.2.1 Relevant Technologies in Cloud Computing

Cloud computing isn't so much a technology as it is the combination of many preexisting technologies. These technologies have matured at different rates and in different contexts, and were not designed as a coherent whole; however, they have come together to create a technical ecosystem for cloud computing. New advances in processors, virtualization technology, disk storage, broadband Internet connection, and fast, inexpensive servers have combined to make the cloud a more compelling solution.

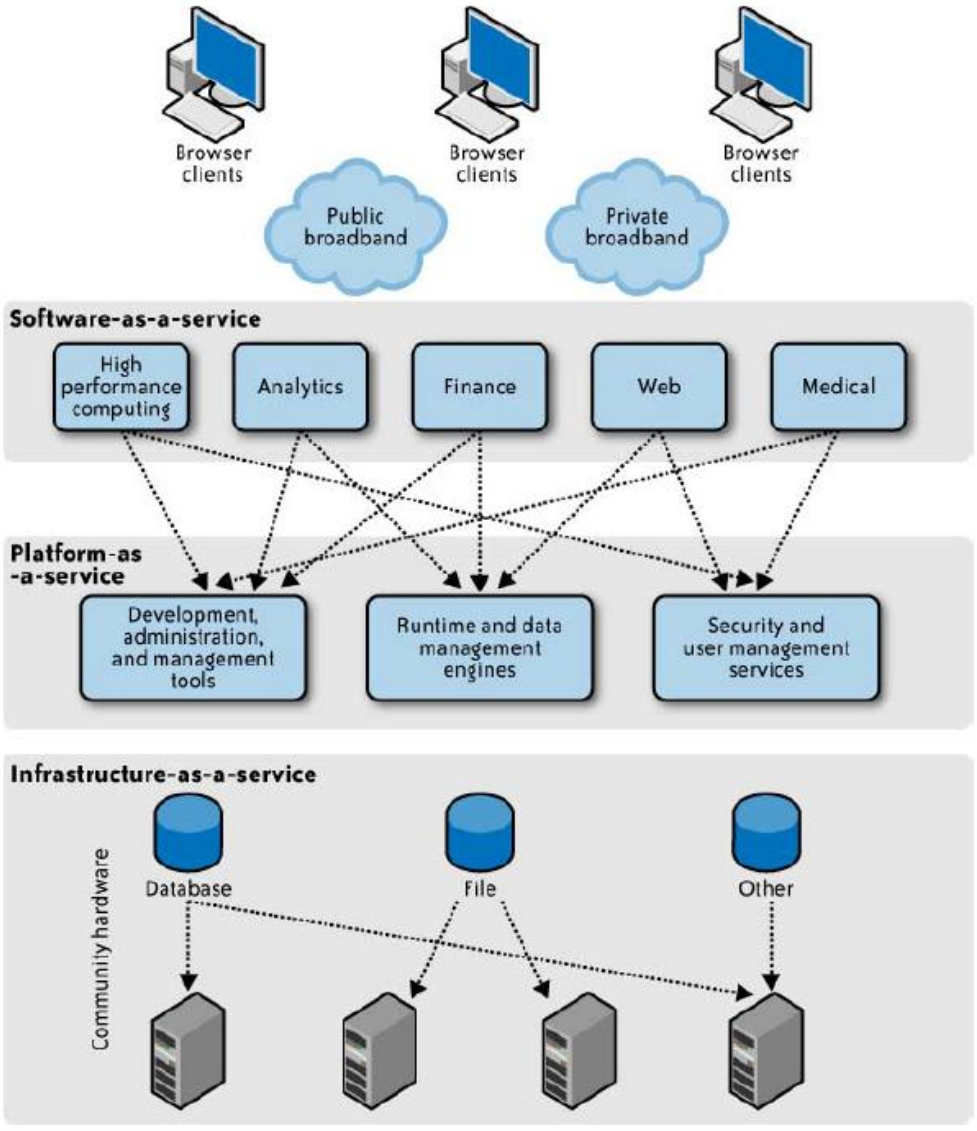


Figure 4. Architecture for relevant technologies

Cloud access devices

The range of access devices for the cloud has expanded in recent years. Home PCs, enterprise PCs, network computers, mobile phone devices, custom handheld devices, and custom static devices (including refrigerators) are all online. Interestingly, the growth of the iPhone and the proliferation of applications available from its App Store illustrate an improvement in terms of access to the cloud. This greater access is resulting in greater use and growth of services within the cloud. For example, you can now use Skype through the iPhone, thus bringing this peer-to-peer network much closer to users, and Salesforce.com has introduced an application that allows users to access its services from the iPhone, as well as many other vendors.

Browsers and thin clients

Users of multiple device types can now access applications and information from wherever they can load a browser. Indeed, browsers are becoming increasingly sophisticated. Enterprise applications, such as SAP and Oracle, can be accessed through a browser interface—a change from when a client (a so-called “fat”) application needed to be loaded onto the desktop. The general population has become more familiar with the browser function and can use a discrete application, where the context is intuitive, without requiring training or user guides.

High-speed broadband access

A critical component of the cloud is the broadband network, which offers the means to connect components and provides one of the substantial differences from the utility computing concept of 30 years ago. Broadband access is now widely available, especially in global metropolitan areas. Nearly pervasive wireless access (e.g., WiFi, cellular, emerging WiMAX) is available, which has established mobile devices as entry points to the IT resources of the enterprise and the cloud.

Data centers and server farms

Cloud-based services require large computing capacity and are hosted in data centers and server farms. These distributed data centers and server farms span multiple locations and can be linked via internetworks providing distributed computing and service delivery capabilities. A number of examples today illustrate the flexibility and scalability of cloud computing power. For instance, Google has linked a very large number of inexpensive servers to provide tremendous flexibility and power. Amazon's Elastic Compute Cloud (EC2) provides virtualization in the data center to create huge numbers of virtual instances for services being requested. Salesforce.com provides SaaS to its large customer base by grouping its customers into clusters to enable scalability and flexibility.

Storage devices

Decreasing storage costs and the flexibility with which storage can be deployed have changed the storage landscape. The fixed direct access storage device (DASD) has been replaced with storage area networks (SANs), which have reduced costs and allowed a great deal more flexibility in enterprise storage. SAN software manages integration of storage devices and can independently allocate storage space on demand across a number of devices.

Virtualization technologies

Virtualization is a foundational technology platform fostering cloud computing, and it is transforming the face of the modern data center. The term virtualization refers to the abstraction of compute resources (CPU, storage, network, memory, application stack, and database) from applications and end users consuming the service. The abstraction of infrastructure yields the notion of resource democratization—whether infrastructure, applications, or information— and provides the capability for pooled resources to be made available and accessible to anyone or anything authorized to utilize them via standardized methods.

Virtualization technologies enable multitenancy cloud business models by providing a scalable, shared resource platform for all tenants. More importantly, they provide a dedicated resource view for the platform's consumers. From an enterprise perspective, virtualization offers data center consolidation and improved IT operational efficiency. Today, enterprises have deployed virtualization technologies within data centers in various forms, including OS virtualization (VMware, Xen), storage virtualization (NAS, SAN), database virtualization, and application or software virtualization (Apache Tomcat, JBoss, Oracle App Server, WebSphere).

From a public cloud perspective, depending on the cloud services delivery model (SPI) and architecture, virtualization appears as a shared resource at various layers of the virtualized service (e.g., OS, storage, database, application).

Figure 6 illustrates OS virtualization and the layers of the virtualization environment as defined by Sun Microsystems. IaaS providers including Amazon (EC2), ServePath (GoGrid), and Sun Cloud employ this type of virtualization, which enables customers to run instances of various operating system flavors in a public cloud. The virtualization platform shown in **Figure 6** is the Sun xVM hypervisor environment that virtualizes shared hardware resources for the guest or virtual server operating systems (Linux, Solaris, and Microsoft Windows) hosted on the hypervisor. The hypervisor is a small application that runs on top of the physical machine hardware layer. It implements and manages the virtual CPU (vCPU), virtual memory (vMemory), event channels, and memory shared by the resident virtual machines (VMs). It also controls I/O and memory access to devices.

In Xen, as well as Sun xVM (which is based on the work of the Xen community), a VM is called a domain, whereas in the VMware virtualization product it is referred to as a guest OS.

In **Figure 6**, the VMs are labeled as dom0 and domU1, domU2, and domU3. Dom0 is used to manage the other user domains (domU1, etc.). VMware employs a similar mechanism, and calls it as "service console." Management through dom0 or the service console consists of creating, destroying, migrating, saving, or restoring user domains. An operating system running in a user domain is configured so that privileged operations are executed via calls to the hypervisor.

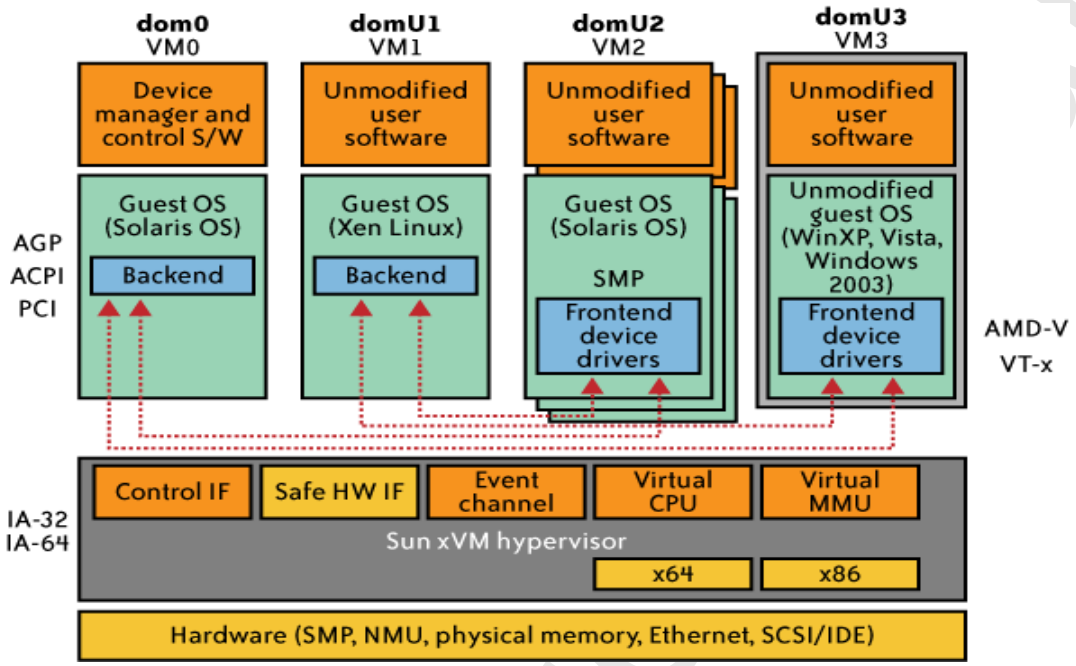


Figure 6. Sun xVM hypervisor enviroment

APIs

A suitable application programming interface (API) is another enabler for the cloud computing services delivery model (**Figure 7**). APIs empower users by enabling features such as selfprovisioning and programmatic control of cloud services and resources. Depending on the type of cloud services delivery model (SPI), an API can manifest in different forms, ranging from simple URL manipulations to advanced SOA-like programming models. APIs also help to exploit the full potential of cloud computing and mask the complexity involved in extending existing IT management processes and practices to cloud services.

APIs offered by IaaS cloud service providers (CSPs) such as Amazon EC2, Sun Cloud, and GoGrid allow users to create and manage cloud resources, including compute, storage, and networking components. In this case, use of the API is via HTTP. The GET, POST, PUT, and DELETE requests are used, although most tasks can be accomplished with GET and POST. In some cases, resource representations are in JavaScript Object Notation (JSON). For example, Sun's cloud specification of the Sun Cloud API includes:

- Common behaviors that apply across all requests and responses
- Resource models, which describe the JSON data structures used in requests and responses
- Requests that may be sent to cloud resources, and the responses expected

All aaS developers need to become familiar with specific APIs to deploy and manage software modules to the aaS platform. SaaS services typically do not offer APIs other than for basic export and import functionality using browsers or scripts that use HTTP(S) and web URI manipulation methods.

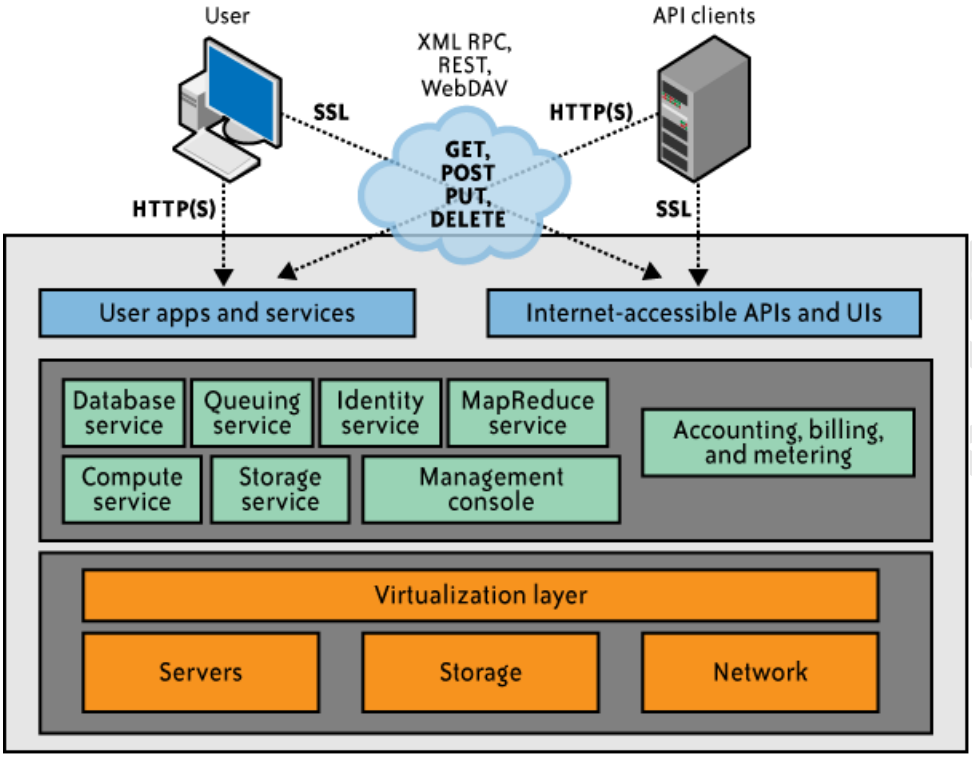


Figure 7. API enabler for cloud computing

Today, one of the key challenges that cloud customers face is the fact that each CSP has a unique API. As a result, cloud applications are not portable across clouds, and it is very difficult to achieve interoperability among applications running across clouds (including your private cloud). Since APIs are unique to a cloud service, architects, developers, and data center staff members must become familiar with platform-specific features. Although there is no cloud API standard, standardization efforts are mushrooming and are driven by vendor as well as user communities. One such effort is Universal Cloud Interface (UCI), an attempt to create an open and standardized cloud interface for the unification of various cloud APIs. The UCI forum claims that the goal is to achieve a singular programmatic point of contact that can encompass the entire infrastructure stack, as well as emerging cloudcentric technologies, all through a unified interface. As of this writing, we are not aware of any concerted effort by CSPs to develop a ubiquitous and consistent API across clouds—and that makes porting an application and sharing data across clouds a monumental task. It is also important to realize that market incentives for CSPs are geared toward locking their customers into their cloud offerings. This may make easy interoperability difficult to achieve.

1.3 The Traditional Software Model

Traditional software applications are based on a model with large, upfront licensing costs and annual support costs. Increasing the number of users can raise the base cost of the package due to the need for additional hardware server deployments and IT support. Licensing costs are often based on metrics that are not directly aligned with usage (server type, number of CPUs, etc., or some physical characteristic) and are not virtual. A typical enterprise software package requires hardware deployment, servers, and backup and network provisioning to accommodate the number of users on- and off-campus. Security architecture is also taxed in

an effort to protect this valuable resource from unauthorized access. Traditional software applications tend to be highly customizable, which comes at a cost—in both dollars and manpower.

1.4 The Cloud Services Delivery Model

As we noted earlier, a cloud services delivery model is commonly referred to as an SPI and falls into three generally accepted services (**Figure 8**).

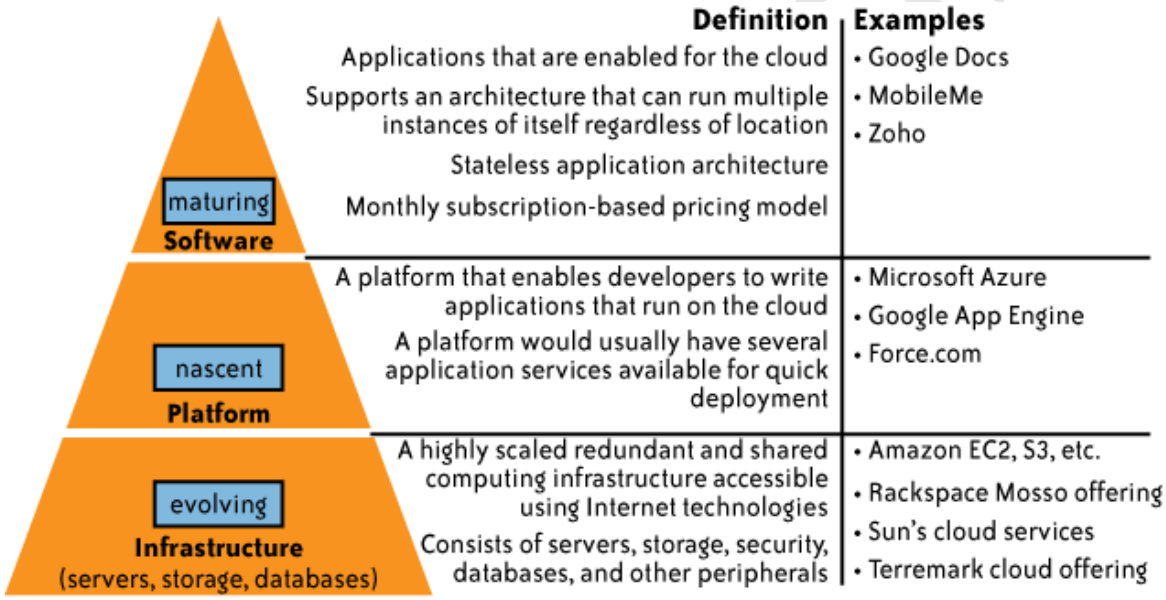


Figure 8. Cloud services delivery model

1.4.1 The Software-As-a-Service Model

Traditional methods of purchasing software involved the customer loading the software onto his own hardware in return for a license fee (a capital expense, known as CapEx). The customer could also purchase a maintenance agreement to receive patches to the software or other support services. The customer was concerned with the compatibility of operational systems, patch installations, and compliance with license agreements.

In a SaaS model, the customer does not purchase software, but rather rents it for use on a subscription or pay-per-use model (an operational expense, known as OpEx). In some cases, the service is free for limited use. Typically, the purchased service is complete from a hardware, software, and support perspective. The user accesses the service through any authorized device. In some cases, preparatory work is required to establish company-specific data for the service to be fully used and potentially integrated with other applications that are not part of the SaaS platform.

Key benefits of a SaaS model include the following:

- SaaS enables the organization to outsource the hosting and management of applications to a third party (software vendor and service provider) as a means of reducing the cost of application software licensing, servers, and other infrastructure and personnel required to host the application internally.
- SaaS enables software vendors to control and limit use, prohibits copying and distribution, and facilitates the control of all derivative versions of their software. SaaS centralized control often allows the vendor or supplier to establish an ongoing revenue stream with multiple businesses and users without preloading software in each device in an organization.
- Applications delivery using the SaaS model typically uses the one-to-many delivery approach, with the Web as the infrastructure. An end user can access a SaaS application via a web browser; some SaaS vendors provide their own interface that is designed to support features that are unique to their applications.
- A typical SaaS deployment does not require any hardware and can run over the existing Internet access infrastructure. Sometimes changes to firewall rules and settings may be required to allow the SaaS application to run smoothly.
- Management of a SaaS application is supported by the vendor from the end user perspective, whereby a SaaS application can be configured using an API, but SaaS applications cannot be completely customized.

A typical SaaS offering is SaaS over a public network, in which a SaaS-based application is delivered via the Internet to the organization's firewall. The single most important architectural difference between the traditional software model and the SaaS model is the number of tenants the application supports. The traditional software model is an isolated, single-tenant model, which means a customer buys a software application and installs it on a server. The server runs only that specific application and only for that single customer's end user group. The SaaS model is a multitenant architecture model, which means the physical backend hardware infrastructure is shared among many different customers, but logically is unique for each customer. Multitenant architecture design maximizes the sharing of resources across tenants, but is still able to securely differentiate data belonging to each tenant. For example, when a user at one company accesses customer information by using a SaaS Customer Relationship Management (CRM) application, the application instance that the user connects to can accommodate users from dozens, or even hundreds, of other companies—all completely unbeknownst to any of the other users. SaaS solutions are very different from application service provider (ASP) solutions. There are two main explanations for this:

- ASP applications are traditional, single-tenant applications, but are hosted by a third party. They are client/server applications with HTML frontends added to allow remote access to the application.
- ASP applications are not written as Net-native applications. As a result, their performance may be poor, and application updates are no better than self-managed premise-based applications.

By comparison, SaaS applications are multitenant applications that are hosted by a vendor with expertise in the applications and that have been designed as Net-native applications and are updated on an ongoing basis.

1.4.2 The Platform-As-a-Service Model

In a platform-as-a-service (PaaS) model, the vendor offers a development environment to application developers, who develop applications and offer those services through the provider's platform. The provider typically develops toolkits and standards for development, and channels for distribution and payment. The provider typically receives a payment for providing the platform and the sales and distribution services. This enables rapid propagation of software applications, given the low cost of entry and the leveraging of established channels for customer acquisition.

PaaS is a variation of SaaS whereby the development environment is offered as a service. The developers use the building blocks (e.g., predefined blocks of code) of the vendor's development environment to create their own applications.

PaaS solutions are development platforms for which the development tool itself is hosted in the cloud and accessed through a browser. With PaaS, developers can often build web applications without installing any tools on their computer, and can then deploy those applications without any specialized system administration skills.

PaaS systems are useful because they enable lone developers and start-up companies to deploy web-based applications without the cost and complexity of buying servers and setting them up. The benefits of PaaS lie in greatly increasing the number of people who can develop, maintain, and deploy web applications. In short, PaaS offers to democratize the development of web applications in much the same way that Microsoft Access democratized the development of the client/server application.

Today, building web applications requires expert developers with three highly specialized skill sets:

I) Backend server development (e.g., Java/J2EE)

II) Frontend client development (e.g., JavaScript/Dojo)

III) Website administration

PaaS offers the potential for general developers to build web applications without needing specialized expertise, which allows an entire generation of Microsoft Access, Lotus Notes, and PowerBuilder developers to build web applications without too steep a learning curve.

The alternative to PaaS is to develop web applications using desktop development tools, such as Eclipse or Microsoft Access, and then manually deploy those applications to a cloud-hosting provider, such as Amazon Web Services (AWS).

At a minimum, a PaaS solution should include the following elements:

- A PaaS development studio solution should be browser-based.
- An end-to-end PaaS solution should provide a high-productivity integrated development environment (IDE) running on the actual target delivery platform so that debugging and test scenarios run in the same environment as production deployment.
- A PaaS solution should provide integration with external web services and databases.
- A PaaS solution must provide comprehensive monitoring of application and user activity, to help developers understand their applications and effect improvements.
- Scalability, reliability, and security should be built into a PaaS solution without requiring additional development, configuration, or other costs. Multitenancy (the ability for an application to automatically partition state and data to service an arbitrary number of users) must be assumed without additional work of any sort.
- A PaaS solution must support both formal and on-demand collaboration throughout the entire software life cycle (development, testing, documentation, and operations), while maintaining the security of source code and associated intellectual property.
- A PaaS solution should support pay-as-you-go metered billing.

PaaS platforms also have functional differences from traditional development platforms, including:

Multitenant development tools

Traditional development tools are intended for a single user; a cloud-based studio must support multiple users, each with multiple active projects.

Multitenant deployment architecture

Scalability is often not a concern of the initial development effort and is left instead for the system administrators to handle when the project deploys. In PaaS, scalability of the application and data tiers must be built-in (e.g., load balancing and failover should be basic elements of the developing platform).

Integrated management

Traditional development solutions (usually) are not associated with runtime monitoring, but in PaaS the monitoring ability should be built into the development platform.

Integrated billing

PaaS offerings require mechanisms for billing based on usage that are unique to the SaaS world.

1.4.3 The Infrastructure-As-a-Service Model

In the traditional hosted application model, the vendor provides the entire infrastructure for a customer to run his applications. Often, this entails housing dedicated hardware that is purchased or leased for that specific application. The IaaS model also provides the infrastructure to run the applications, but the cloud computing approach makes it possible to offer a pay-peruse model and to scale the service depending on demand. From the IaaS provider's perspective, it can build an infrastructure that handles the peaks and troughs of its customers' demands and add new capacity as the overall demand increases. Similarly, in a hosted application model, the IaaS vendor can cover application hosting only, or can extend to other services (such as application support, application development, and enhancements) and can support the more comprehensive outsourcing of IT.

The IaaS model is similar to utility computing, in which the basic idea is to offer computing services in the same way as utilities. That is, you pay for the amount of processing power, disk space, and so on that you actually consume. IaaS is typically a service associated with cloud computing and refers to online services that abstract the user from the details of infrastructure, including physical computing resources, location, data partitioning, scaling, security, backup, and so on. In cloud computing, the provider is in complete control of the infrastructure. Utility computing users, conversely, seek a service that allows them to deploy, manage, and scale online services using the provider's resources and pay for resources the customer consumes. However, the customer wants to be in control of the geographic location of the infrastructure and what runs on each server.

Features available for a typical IaaS system include:

Scalability

The ability to scale infrastructure requirements, such as computing resources, memory, and storage (in near-real-time speeds) based on usage requirements

Pay as you go

The ability to purchase the exact amount of infrastructure required at any specific time

Best-of-breed technology and resources

Access to best-of-breed technology solutions and superior IT talent for a fraction of the cost

1.5 Cloud Deployment Models

The term cloud is a metaphor for the Internet and is a simplified representation of the complex, internetworked devices and connections that form the Internet. Private and public clouds are subsets of the Internet and are defined based on their relationship to the enterprise. Private and public clouds may also be referred to as internal or external clouds; the differentiation is based on the relationship of the cloud to the enterprise.

The public and private cloud concepts are important because they support cloud computing, which enables the provisioning of dynamic, scalable, virtualized resources over Internet connections by a vendor or an enterprise IT organization to customers for a fee. The end users who use the services offered via cloud computing may not have knowledge of, expertise in, or control over the technology infrastructure that supports them.

The majority of cloud computing infrastructure consists of reliable services delivered through data centers and built on servers with different levels of virtualization technologies. The services are accessible anywhere that access to networking infrastructure is available. The cloud appears as a single point of access for all consumer computing needs. Commercial offerings should meet the quality of service requirements of customers and typically offer service-level agreements (SLAs). Open standards are critical to the growth of cloud computing, and open source software has provided the foundation for many cloud computing implementations (e.g., the use of Xen in AWS).

1.5.1 Public Clouds

Public clouds (or external clouds) describe cloud computing in the traditional mainstream sense, whereby resources are dynamically provisioned on a fine-grained, self-service basis over the Internet, via web applications or web services, from an off-site, third-party provider who shares resources and bills on a fine-grained, utility-computing basis.

A public cloud is hosted, operated, and managed by a third-party vendor from one or more data centers. The service is offered to multiple customers (the cloud is offered to multiple tenants) over a common infrastructure (**Figure 9**).

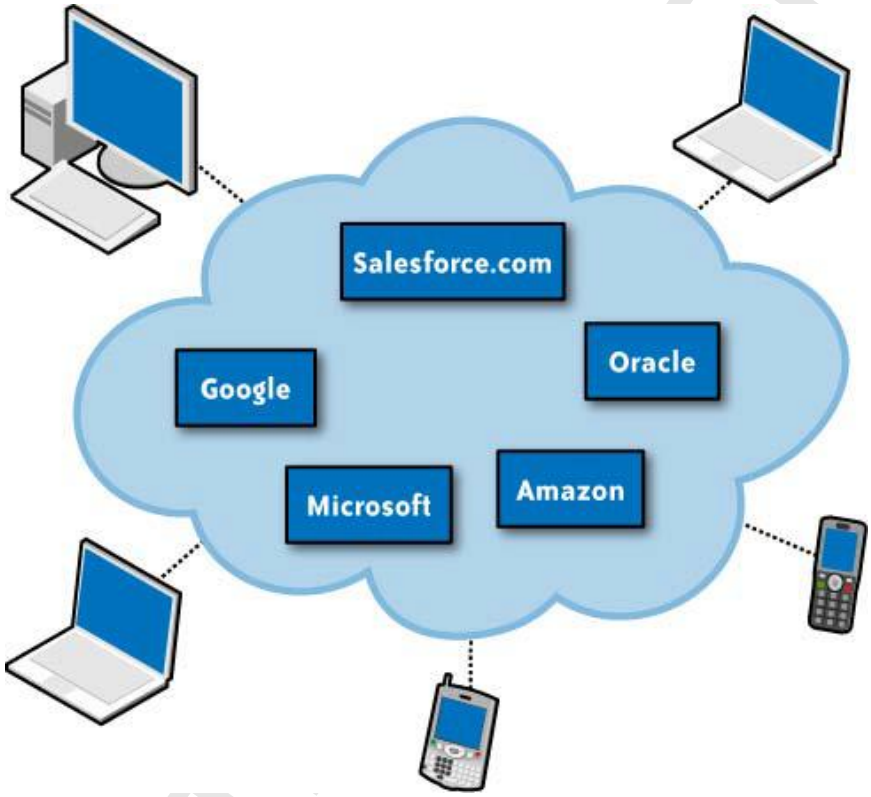


Figure 9. Public cloud

In a public cloud, security management and day-to-day operations are relegated to the third party vendor, who is responsible for the public cloud service offering. Hence, the customer of the public cloud service offering has a low degree of control and oversight of the physical and logical security aspects of a private cloud.

1.5.2 Private Clouds

Private clouds and internal clouds are terms used to describe offerings that emulate cloud computing on private networks. These (typically virtualization automation) products claim to deliver some benefits of cloud computing without the pitfalls, capitalizing on data security, corporate governance, and reliability concerns. Organizations must buy, build, and manage

them and, as such, do not benefit from lower upfront capital costs and less hands-on management. The organizational customer for a private cloud is responsible for the operation of his private cloud.

Private clouds differ from public clouds in that the network, computing, and storage infrastructure associated with private clouds is dedicated to a single organization and is not shared with any other organizations (i.e., the cloud is dedicated to a single organizational tenant). As such, a variety of private cloud patterns have emerged:

Dedicated

Private clouds hosted within a customer-owned data center or at a collocation facility, and operated by internal IT departments

Community

Private clouds located at the premises of a third party; owned, managed, and operated by a vendor who is bound by custom SLAs and contractual clauses with security and compliance requirements

Managed

Private cloud infrastructure owned by a customer and managed by a vendor. In general, in a private cloud operating model, the security management and day-to-day operation of hosts are relegated to internal IT or to a third party with contractual SLAs. By virtue of this direct governance model, a customer of a private cloud should have a high degree of control and oversight of the physical and logical security aspects of the private cloud infrastructure—both the hypervisor and the hosted virtualized OSs. With that high degree of control and transparency, it is easier for a customer to comply with established corporate security standards, policies, and regulatory compliance.

1.5.3 Hybrid Clouds

A hybrid cloud environment consisting of multiple internal and/or external providers is a possible deployment for organizations. With a hybrid cloud, organizations might run non-core applications in a public cloud, while maintaining core applications and sensitive data in-house in a private cloud (Figure 10).

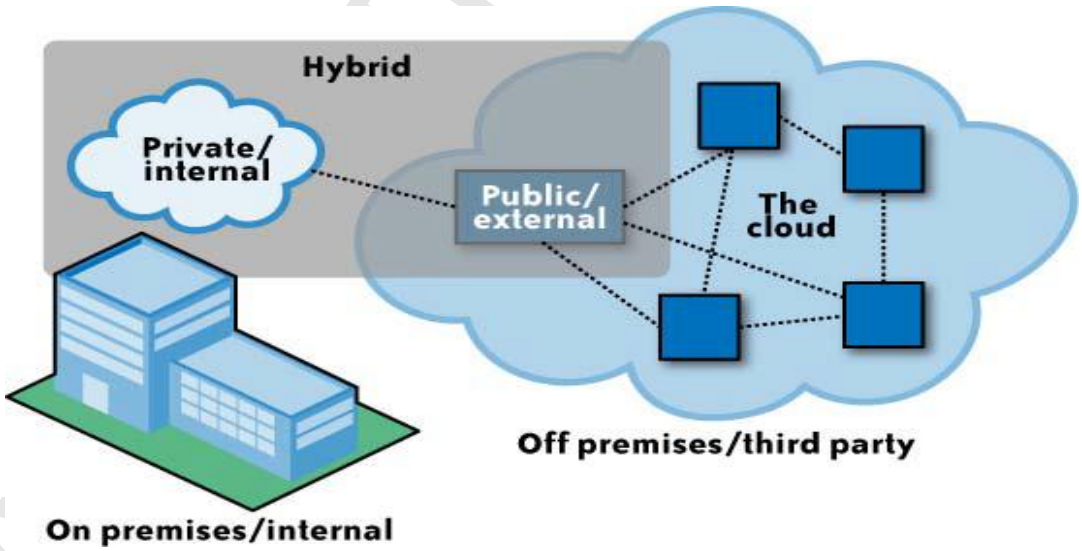


Figure 10. Hybrid cloud

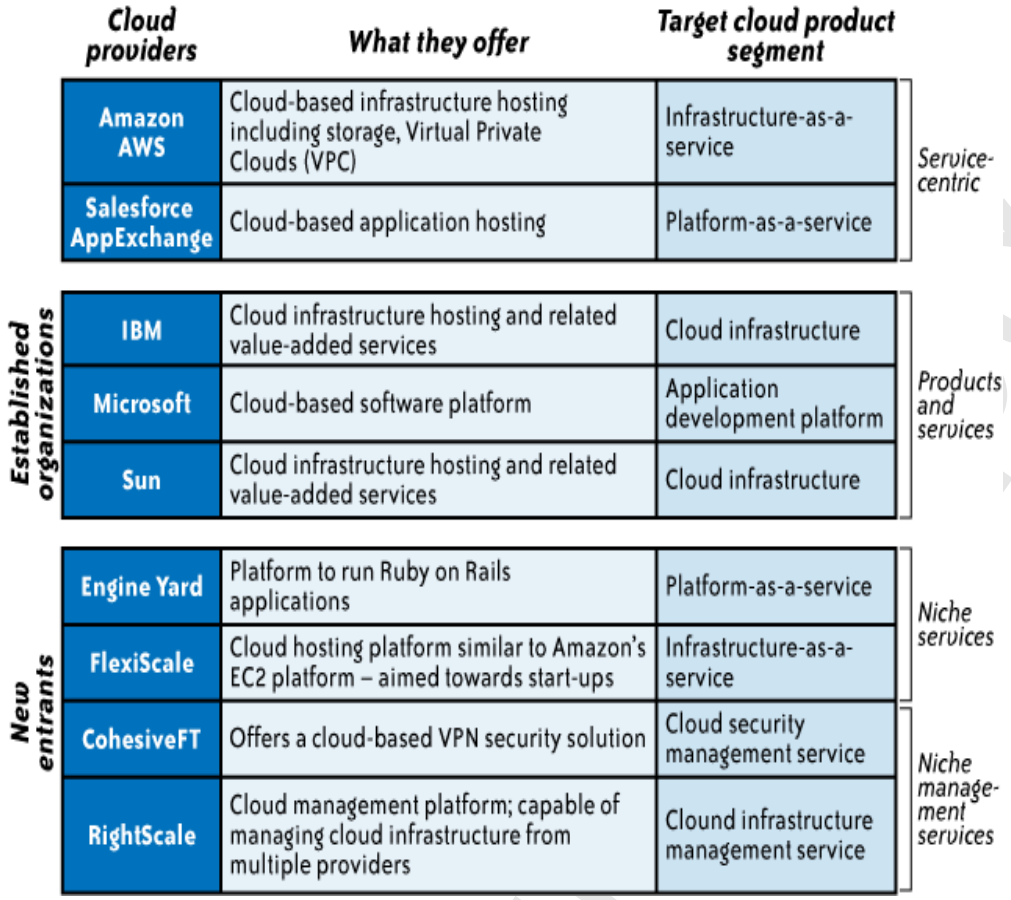


Figure 11. Lists some examples of CSPs and their services.

Services provided through the integration of cloud components are evolving, barriers are being overcome, and enablers are being developed. A major concern is to trust that a company's or an individual's information is both secure and private. Establishing this trust is a major milestone in the adoption of the full range of cloud computing; see the next section for more details.

1.6 Barriers to Cloud Computing Adoption in the Enterprise

Although there are many benefits to adopting cloud computing, there are also some significant barriers to adoption.

Security

Because cloud computing represents a new computing model, there is a great deal of uncertainty about how security at all levels (e.g., network, host, application, and data levels) can be achieved. That uncertainty has consistently led information executives to state that security is their number one concern with cloud computing.

Privacy

The ability of cloud computing to adequately address privacy regulations has been called into question. Organizations today face numerous different requirements attempting to protect the privacy of individuals' information, and it is not clear (i.e., not yet established) whether the cloud computing model provides adequate protection of such information, or whether organizations will be found in violation of regulations because of this new model.

Connectivity and Open Access

The full potential of cloud computing depends on the availability of high-speed access to all. Such connectivity, rather like electricity availability, globally opens the possibility for industry and a new range of consumer products. Connectivity and open access to computing power and information availability through the cloud promotes another era of industrialization and the need for more sophisticated consumer products.

Reliability

Enterprise applications are now so critical that they must be reliable and available to support 24/7 operations. In the event of failure or outages, contingency plans must take effect smoothly, and for disastrous or catastrophic failure, recovery plans must begin with minimum disruption. Each aspect of reliability should be carefully considered when engaging with a CSP, negotiated as part of the SLA, and tested in failover drills. Additional costs may be associated with the required levels of reliability; however, the business can do only so much to mitigate risks and the cost of a failure. Establishing a track record of reliability will be a prerequisite for widespread adoption.

Interoperability

The interoperability and portability of information between private clouds and public clouds are critical enablers for broad adoption of cloud computing by the enterprise. Many companies have made considerable progress toward standardizing their processes, data, and systems through implementation of ERPs. This process has been enabled by scalable infrastructures to create single instances, or highly integrated connections between instances, to manage the consistency of master and transaction data and produce reliable consolidated information. Even with these improved platforms, the speed at which businesses change may still outpace the ability of IT organizations to respond to these changes. SaaS applications delivered through the cloud provide a low-capital, fast-deployment option. Depending on the application, it is critical to integrate with traditional applications that may be resident in a separate cloud or on traditional technology. The standard for interoperability is either an enabler or a barrier to interoperability, and permits maintenance of the integrity and consistency of a company's information and processes.

Independence from CSPs

Examples exist of IT outsourcing contracts that have effectively locked a customer into a service that does not meet current or evolving needs at a speed and cost that are acceptable to meet business goals. This could be caused by a number of factors, and is a concern if limited options exist for quickly engaging an alternative provider supplier to meet the needs without large transition or penalty costs. A CSP may hold valuable data and business rules that cannot be easily migrated to a new provider. Standards to enable migration and plug and play of cloud components can help. For example, companies today depend less on the browser provider, but may depend on a proprietary data-based structure. Separating storage IaaS providers from processing providers can help with provider flexibility. There are downsides to going to a componentized approach, because the customer may become the integrator of these services. However, these may be the skills that enterprises should develop to balance the scalability of cloud computing with acceptable price performance and risk.

Economic Value

The growth of cloud computing is predicated on the return on investment that accrues. It seems intuitive that by sharing resources to smooth out peaks, paying only for what is used, and cutting upfront capital investment in deploying IT solutions, the economic value will be there. There will be a need to carefully balance all costs and benefits associated with cloud computing—in both the short and long terms. Hidden costs could include support, disaster recovery, application modification, and data loss insurance. There will be threshold values whereby consolidating investments or combining cloud services makes sense; for example, it might not be efficient or cost-effective to utilize multiple autonomous SaaS applications. Each may contract for disaster recovery program services. There is a point where economies of scale mean these functions should be combined in a similar service. Application usage may begin with a low volume of transactions that can be supported with semi-automated master data management. As usage expands and interoperability requirements for the business process become more onerous, a new approach is needed. This evolution may be the most

cost effective approach; however, there is a risk that the business transition costs from one solution to another may change the cost and benefit equation, and hence the solution that should be employed.

IT Governance

Economic value is an aspect of IT governance. Effective governance processes that align IT and the business are critical to set the appropriate context for making investment decisions and to balance short-term and long-term needs.

Changes in the IT Organization

The IT organization will be affected by cloud computing, as has been the case with other technology shifts. There are two dimensions to shifts in technology. The first is acquiring the new skill sets to deploy the technology in the context of solving a business problem, and the second is how the technology changes the IT role. During the COBOL era, users rarely programmed, the expectations of the user interface varied, and the adaptability of the solution was low. Training was delivered in separate manuals and the user used the computer to solve problems only down predefined paths. With the advent of fourth-generation languages, roles within IT, such as system analyst and programmer, became merged into analyst/programmer, users started to write their own reports, and new applications, including operational data stores, data entry, and query programs, could be rapidly deployed in weeks. IT's role will change once again: the speed of change will impact the adoption of cloud technologies and the ability to decompose mature solutions from hype to deliver real value from cloud technology; and the need to maintain the controls to manage IT risk in the business will increase.

Political Issues Due to Global Boundaries

In the cloud computing world, there is variability in terms of where the physical data resides, where processing takes place, and from where the data is accessed. Given this variability, different privacy rules and regulations may apply. Because of these varying rules and regulations, by definition politics becomes an element in the adoption of cloud computing, which is effectively multijurisdictional.

For cloud computing to continually evolve into a borderless and global tool, it needs to be separated from politics. Currently, some major global technological and political powers are making laws that can have a negative impact on the development of the global cloud. For example, as a result of the USA Patriot Act, Canada has recently asked that its government not use computers in the global network that are operating within U.S. borders, fearing for the confidentiality and privacy of the Canadian data stored on those computers. Cloud computing depends largely on global politics to survive. Imagine if the telecommunications companies in the United States get their way and do away with the current Internet standard of network neutrality completely. Having data throttled and information filtered goes against the basic concept of cloud computing and global knowledge. You can't have a working cloud of information and services to draw from and build on if someone or something is constantly manipulating the data held within it, or worse, if something is blocking it from your view to achieve a hidden agenda. Politics are affecting the scalability of the Internet, the availability of Internet access, the free flow of information, and the cloud-based global economy on a daily basis. We already know the concept works; it was instrumental in crunching the massive amounts of data needed to complete the Human Genome Project. That project has netted answers to the question of where hundreds of diseases and traits come from, and would not have been possible in such a short time without the computer sharing allowed by cloud computing and available via the Internet.

Chapter 2. Infrastructure Security

In this Chapter you can see the threats, challenges and guidance assosiated with securing an organization's core IT infrastructure at the network, host, and application levels. Information security practitioners commonly use this approach; therefore, it is readily familiar to them. We discuss this infrastructure security in the context of SPI service delivery models (SaaS, PaaS, and IaaS). Non-information security professionals are cautioned not to simply equate infrastructure security to infrastructure-as-a-service (IaaS) security. Although infrastructure security is more highly relevant to customers of IaaS, similar consideration should be given to providers' platform-as-a-service (PaaS) and software-as-a-service (SaaS) environments, since they have ramifications to your customer threat, risk, and compliance management. Another dimension is the cloud business model (public, private, and hybrid clouds), which is orthogonal to the SPI service delivery model; what we highlight is the relevance of discussion points as they apply to public and private clouds. When discussing public clouds the scope of infrastructure security is limited to the layers of infrastructure that move beyond the organization's control and into the hands of service providers (i.e., when responsibility to a secure infrastructure is transferred to the cloud service provider or CSP, based on the SPI delivery model). Information in this chapter is critical for customers in gaining an understanding of what security a CSP provides and what security you, the customer, are responsible for providing.

2.1 Infrastructure Security: The Network Level

When looking at the network level of infrastructure security, it is important to distinguish between public clouds and private clouds. With private clouds, there are no new attacks, vulnerabilities, or changes in risk *specific to this topology* that information security personnel need to consider. Although your organization's IT architecture may change with the implementation of a private cloud, your current network topology will probably not change significantly. If you have a private extranet in place (e.g., for premium customers or strategic partners), for practical purposes you probably have the network topology for a private cloud in place already. The security considerations you have today apply to a private cloud infrastructure, too. And the security tools you have in place (or should have in place) are also necessary for a private cloud and operate in the same way. **Figure 12** shows the topological similarities between a secure extranet and a private cloud. However, if you choose to use public cloud services, changing security requirements will require changes to your network topology. You must address how your existing network topology interacts with your cloud provider's network topology. There are four significant risk factors in this use case:

• Ensuring the confidentiality and integrity of your organization's data-in-transit to and from your public cloud provider

• Ensuring proper access control (authentication, authorization, and auditing) to whatever resources you are using at your public cloud provider

• Ensuring the availability of the Internet-facing resources in a public cloud that are being used by your organization, or have been assigned to your organization by your public cloud providers

• Replacing the established model of network zones and tiers with domains We will discuss each of these risk factors in the sections that follow.

2.1.1 Ensuring Data Confidentiality and Integrity

Some resources and data previously confined to a private network are now exposed to the Internet, and to a shared public network belonging to a third-party cloud provider.

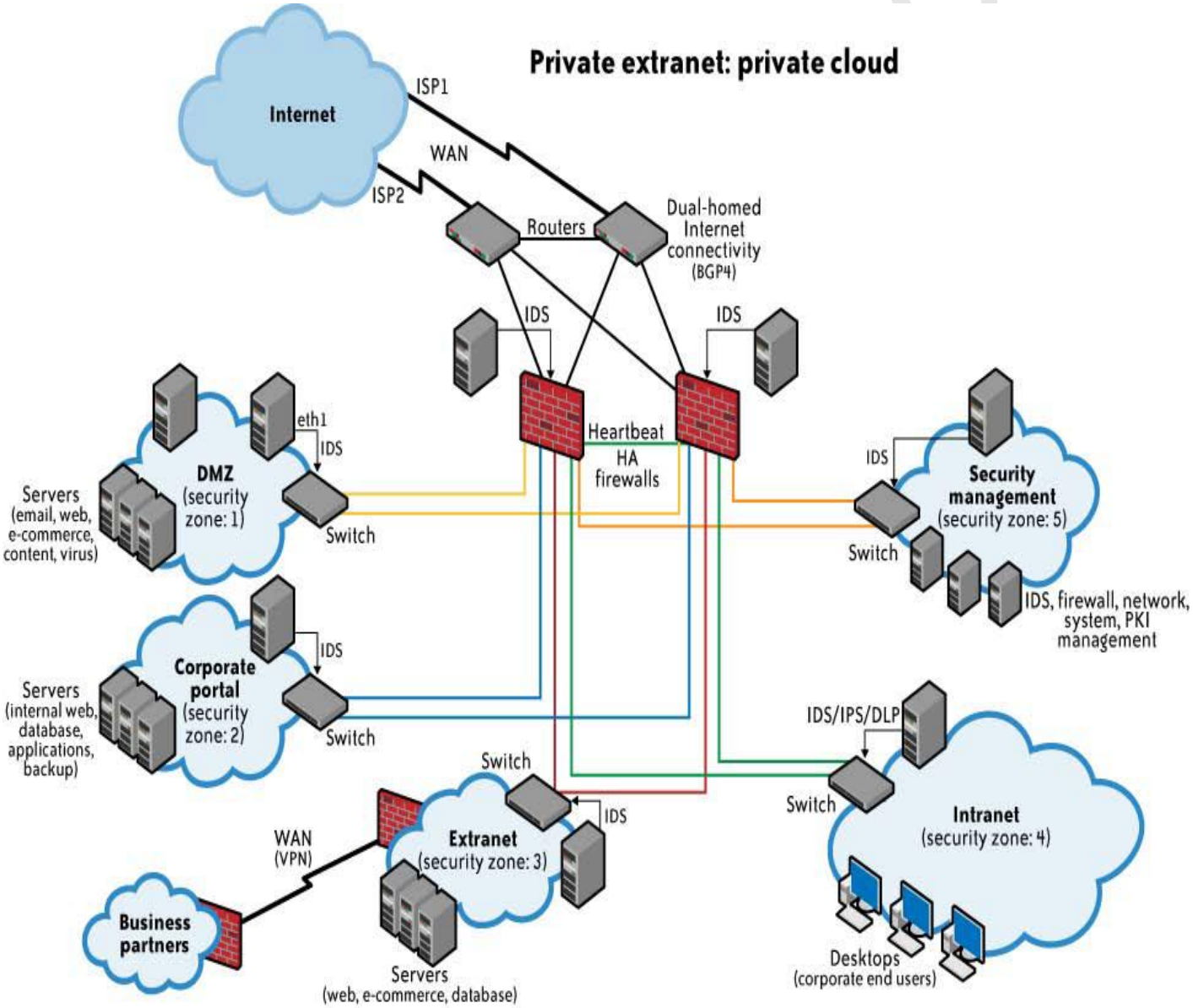


Figure 12. Generic network topology for private cloud computing

An example of problems associated with this first risk factor is an Amazon Web Services (AWS) security vulnerability reported in December 2008.[1] In a blog post, the author detailed a flaw in the digital signature algorithm used when "... making Query (aka REST) requests to Amazon SimpleDB, to Amazon Elastic Compute Cloud (EC2), or to Amazon Simple Queue Service (SQS) over HTTP." Although use of HTTPS (instead of HTTP) would have mitigated the integrity risk, users not using HTTPS (but using HTTP) did face an increased risk that their data could have been altered in transit without their knowledge.

2.1.2 Ensuring Proper Access Control

Since some subset of these resources (or maybe even all of them) is now exposed to the Internet, an organization using a public cloud faces a significant increase in risk to its data. The ability to audit the operations of your cloud provider's network (let alone to conduct any realtime monitoring, such as on your own network), even after the fact, is probably non-existent. You will have decreased access to relevant network-level logs and data, and a limited ability to thoroughly conduct investigations and gather forensic data.

An example of the problems associated with this second risk factor is the issue of reused (reassigned) IP addresses. Generally speaking, cloud providers do not sufficiently "age" IP addresses when they are no longer needed for one customer. Addresses are usually reassigned and reused by other customers as they become available. From a cloud provider's perspective this makes sense. IP addresses are a finite quantity and a billable asset. However, from a customer's security perspective, the persistence of IP addresses that are no longer in use can present a problem. A customer can't assume that network access to its resources is terminated upon release of its IP address. There is necessarily a lag time between the change of an IP address in DNS and the clearing of that address in DNS caches. There is a similar lag time between when physical (i.e., MAC) addresses are changed in ARP tables and when old ARP addresses are cleared from cache; an old address persists in ARP caches until they are cleared. This means that even though addresses might have been changed, the (now) old addresses are still available in cache, and therefore they still allow users to reach these supposedly non-existent resources. Recently, there were many reports of problems with "non-aged" IP addresses at one of the largest cloud providers; this was likely an impetus for an AWS announcement of the Amazon Elastic IP capabilities in March 2008[2]. (With Elastic IP addresses, customers are given a block of five routable IP addresses over which they control assignment.) Additionally, according to Simson Garfinkel:

A separate ongoing problem with the load balancers causes them to terminate any TCP/IP connection that contains more than 231 bytes. This means that objects larger than 2GB must be stored to S3 in several individual transactions, with each of those transactions referring to different byte ranges of the same object3.

However, the issue of "non-aged" IP addresses and unauthorized network access to resources does not apply only to routable IP addresses (i.e., resources intended to be reachable directly from the Internet). The issue also applies to cloud providers' internal networks for customer use and the assignment of non-routable IP addresses.[4] Although your resources may not be directly reachable from the Internet, for management purposes your resources must be accessible within the cloud provider's network via private addressing.

(Every public/Internetfacing resource also has a private address.) Other customers of your cloud provider may not be well intentioned and might be able to reach your resources internally via the cloud provider's networks.[5] As reported in *The Washington Post*, AWS has had problems with abuses of its resources affecting the public and other customers.[6] Some products emerging onto the market[7] will help alleviate the problem of IP address reuse, but unless cloud providers offer these products as managed services, customers are paying for yet another third-party product to solve a problem that their cloud provider's practices created for them.

2.1.3 Ensuring the Availability of Internet-Facing Resources

Reliance on network security has increased because an increased amount of data or an increased number of organizational personnel now depend on externally hosted devices to ensure the availability of cloud-provided resources. Consequently, the three risk factors enumerated in the preceding section must be acceptable to your organization. BGP[8] prefix hijacking (i.e., the falsification of Network Layer Reachability Information) provides a good example of this third risk factor. Prefix hijacking involves announcing an autonomous system[9] address space that belongs to someone else without her permission. Such announcements often occur because of a configuration mistake, but that misconfiguration may still affect the availability of your cloud-based resources. According to a study presented to the North American Network Operators Group (NANOG) in February 2006, several hundred such misconfigurations occur per month[10]. Probably the best known example of such a misconfiguration mistake occurred in February 2008 when Pakistan Telecom made an error by announcing a dummy route for YouTube to its own telecommunications partner, PCCW, based in Hong Kong. The intent was to block YouTube within Pakistan because of some supposedly blasphemous videos hosted on the site. The result was that YouTube was *globally* unavailable for two hours[11].

In addition to misconfigurations, there are deliberate attacks as well. Although prefix hijacking due to deliberate attacks is far less common than misconfigurations, it still occurs and can block access to data. According to the same study presented to NANOG, attacks occur fewer than 100 times per month. Although prefix hijackings are not new, that attack figure will certainly rise, and probably significantly, along with a rise in cloud computing. As the use of cloud computing increases, the availability of cloud-based resources increases in value to customers. That increased value to customers translates to an increased risk of malicious activity to threaten that availability

DNS[12] attacks are another example of problems associated with this third risk factor. In fact, there are several forms of DNS attacks to worry about with regard to cloud computing. Although DNS attacks are not new and are not directly related to the use of cloud computing, the issue with DNS and cloud computing is an increase in an organization's risk at the network level because of increased external DNS querying (reducing the effectiveness of "split horizon" DNS configurations[13]) along with some increased number of organizational personnel being more dependent on network security to ensure the availability of cloud-provided resources being used.

Although the "Kaminsky Bug"[14] (CVE-2008-1447, "DNS Insufficient Socket Entropy Vulnerability") garnered most of the network security attention in 2008, other DNS problems impact cloud computing as well. Not only are there vulnerabilities in the DNS protocol and in implementations of DNS,[15] but also there are fairly widespread DNS cache poisoning attacks whereby a DNS server is tricked into accepting incorrect information. Although many people thought DNS cache poisoning attacks had been quashed several years ago, that is not true,

and these attacks are still very much a problem—especially in the context of cloud computing. Variants of this basic cache poisoning attack include redirecting the target domain's name server (NS), redirecting the NS record to another target domain, and responding before the real NS (called *DNS forgery*).

A final example of problems associated with this third risk factor is denial of service (DoS) and

distributed denial of service (DDoS) attacks. Again, although DoS/DDoS attacks are not new and are not directly related to the use of cloud computing, the issue with these attacks and cloud computing is an increase in an organization's risk at the network level because of some increased use of resources external to your organization's network. For example, there continue to be rumors of continued DDoS attacks on AWS, making the services unavailable for hours at a time to AWS users[16]. (Amazon has not acknowledged that service interruptions are in fact due to DDoS attacks.)

However, when using IaaS, the risk of a DDoS attack is not only external (i.e., Internet-facing).

There is also the risk of an internal DDoS attack through the portion of the IaaS provider's network used by customers (separate from the IaaS provider's corporate network). That internal (non-routable) network is a shared resource, used by customers for access to their non-public instances (e.g., Amazon Machine Images or AMIs) as well as by the provider for management of its network and resources (such as physical servers). If I were a rogue customer, there would be nothing to prevent me from using my customer access to this internal network to find and attack other customers, or the IaaS provider's infrastructure—and the provider would probably not have any detective controls in place to even notify it of such an attack. The only preventive controls other customers would have would be how hardened their instances (e.g., AMIs) are, and whether they are taking advantage of a provider's capabilities to firewall off groups of instances (e.g., AWS).

2.1.4 Replacing the Established Model of Network Zones and Tiers with Domains

The established isolation model of network zones and tiers no longer exists in the public IaaS and PaaS clouds. For years, network security has relied on zones, such as intranet versus extranet and development versus production, to segregate network traffic for improved security. This model was based on exclusion—only individuals and systems in specific roles have access to specific zones. Similarly, systems within a specific tier often have only specific access within or across a specific tier. For example, systems within a presentation tier are not allowed to communicate directly with systems in the database tier, but can communicate only with an authorized system within the application zone. SaaS clouds built on public IaaS or PaaS clouds have similar characteristics. However, a public SaaS built on a private IaaS (e.g., Salesforce.com) may follow the traditional isolation model, but that topology information is not typically shared with customers.

The traditional model of network zones and tiers has been replaced in public cloud computing with "security groups," "security domains," or "virtual data centers" that have logical separation between tiers but are less precise and afford less protection than the formerly established model. For example, the security groups feature in AWS allows your virtual machines (VMs) to access each other using a virtual firewall that has the ability to filter traffic based on IP address (a specific address or a subnet), packet types (TCP, UDP, or ICMP), and ports (or a range of ports). Domain names are used in various networking contexts and

application-specific naming and addressing purposes, based on DNS. For example, Google's App Engine provides a logical grouping of applications based on domain names such as *mytestapp.test.mydomain.com* and *myprodapp.prod.mydomain.com*.

In the established model of network zones and tiers, not only were development systems logically separated from production systems at the network level, but these two groups of systems were also physically separated at the host level (i.e., they ran on physically separated servers in logically separated network zones). With cloud computing, however, this separation no longer exists. The cloud computing model of separation by domains provides logical separation for addressing purposes only. There is no longer any "required" physical separation, as a test domain and a production domain may very well be on the same physical server. Furthermore, the former logical network separation no longer exists; logical separation now is at the host level with both domains running on the same physical server and being separated only logically by VM monitors (hypervisors).

2.1.5 Network-Level Mitigation

Given the factors discussed in the preceding sections, what can you do to mitigate these increased risk factors? First, note that network-level risks exist regardless of what aspects of "cloud computing" services are being used (e.g., software-as-a-service, platform-as-a-service, or infrastructure-as-a-service). The primary determination of risk level is therefore not which aaS is being used, but rather whether your organization intends to use or is using a public, private, or hybrid cloud. Although some IaaS clouds offer virtual network zoning, they may not match an internal private cloud environment that performs stateful inspection and other network security measures.

If your organization is large enough to afford the resources of a private cloud, your risks will decrease—assuming you have a true private cloud that is internal to your network. In some cases, a private cloud located at a cloud provider's facility can help meet your security requirements but will depend on the provider capabilities and maturity. You can reduce your confidentiality risks by using encryption; specifically by using validated implementations of cryptography for data-in-transit. Secure digital signatures make it much more difficult, if not impossible, for someone to tamper with your data, and this ensures data integrity.

Availability problems at the network level are far more difficult to mitigate with cloud computing—unless your organization is using a private cloud that is internal to your network topology. Even if your private cloud is a private (i.e., non-shared) external network at a cloud provider's facility, you will face increased risk at the network level. A public cloud faces even greater risk. But let's keep some perspective here—greater than what? Even large enterprises with significant resources face considerable challenges at the network level of infrastructure security. Are the risks associated with cloud computing actually higher than the risks enterprises are facing today? Consider existing private and public extranets, and take into account partner connections when making such a comparison. For large enterprises without significant resources, or for small to medium-size businesses (SMBs), is the risk of using public clouds (assuming that such enterprises lack the resources necessary for private clouds) really higher than the risks inherent in their current infrastructures? In many cases, the answer is probably no - there is *not* a higher level of risk. **Figure 12** lists security controls at the network level.

Threat outlook	Low (with the exception of DoS attacks)
Preventive controls	Network access control supplied by provider (e.g., firewall), encryption of data in transit (e.g., SSL, IPSec)
Detective controls	Provider-managed aggregation of security event logs (security incident and event management, or SIEM), network-based intrusion detection system/intrusion prevention system (IDS/IPS)

Figure 12. Security controls at the network level

2.2 Infrastructure Security: The Host Level

When reviewing host security and assessing risks, you should consider the context of cloud services delivery models (SaaS, PaaS, and IaaS) and deployment models (public, private, and hybrid). Although there are no known new threats to hosts that are specific to cloud computing, some virtualization security threats—such as VM escape, system configuration drift, and insider threats by way of weak access control to the hypervisor—carry into the public cloud computing environment. The dynamic nature (elasticity) of cloud computing can bring new operational challenges from a security management perspective. The operational model motivates rapid provisioning and fleeting instances of VMs. Managing vulnerabilities and patches is therefore much harder than just running a scan, as the rate of change is much higher than in a traditional data center.

In addition, the fact that the clouds harness the power of thousands of compute nodes, combined with the homogeneity of the operating system employed by hosts, means the threats can be amplified quickly and easily—call it the “velocity of attack” factor in the cloud. More importantly, you should understand the trust boundary and the responsibilities that fall on your shoulders to secure the host infrastructure that you manage. And you should compare the same with providers’ responsibilities in securing the part of the host infrastructure the CSP manages.

2.2.1 SaaS and PaaS Host Security

In general, CSPs do not publicly share information related to their host platforms, host operating systems, and the processes that are in place to secure the hosts, since hackers can exploit that information when they are trying to intrude into the cloud service. Hence, in the context of SaaS (e.g., Salesforce.com, Workday.com) or PaaS (e.g., Google App Engine, Salesforce.com, Force.com) cloud services, host security is opaque to customers and the responsibility of securing the hosts is relegated to the CSP. To get assurance from the CSP on the security hygiene of its hosts, you should ask the vendor to share information under a nondisclosure agreement (NDA) or simply demand that the CSP share the information via a controls assessment framework such as SysTrust or ISO 27002. From a controls assurance perspective, the CSP has to ensure that appropriate preventive and detective controls are in place and will have to ensure the same via a third-party assessment or ISO 27002 type assessment framework.

Since virtualization is a key enabling technology that improves host hardware utilization, among other benefits, it is common for CSPs to employ virtualization platforms,

including Xen and VMware hypervisors, in their host computing platform architecture. You should understand how the provider is using virtualization technology and the provider's process for securing the virtualization layer.

Both the PaaS and SaaS platforms abstract and hide the host operating system from end users with a host abstraction layer. One key difference between PaaS and SaaS is the accessibility of the abstraction layer that hides the operating system services the applications consume. In the case of SaaS, the abstraction layer is not visible to users and is available only to the developers and the CSP's operations staff, where PaaS users are given indirect access to the host abstraction layer in the form of a PaaS application programming interface (API) that in turn interacts with the host abstraction layer. In short, if you are a SaaS or a PaaS customer, you are relying on the CSP to provide a secure host platform on which the SaaS or PaaS application is developed and deployed by the CSP and you, respectively.

In summary, host security responsibilities in SaaS and PaaS services are transferred to the CSP. The fact that you do not have to worry about protecting hosts from host-based security threats is a major benefit from a security management and cost standpoint. However, as a customer, you still own the risk of managing information hosted in the cloud services. It's your responsibility to get the appropriate level of assurance regarding how the CSP manages host security hygiene.

2.2.2 IaaS Host Security

Unlike PaaS and SaaS, IaaS customers are primarily responsible for securing the hosts provisioned in the cloud. Given that almost all IaaS services available today employ virtualization at the host layer, host security in IaaS should be categorized as follows:

Virtualization software security.

The software layer that sits on top of bare metal and provides customers the ability to create and destroy virtual instances. Virtualization at the host level can be accomplished using any of the virtualization models, including OS-level virtualization (Solaris containers, BSD jails, Linux-VServer), paravirtualization (a combination of the hardware version and versions of Xen and VMware), or hardware-based virtualization (Xen, VMware, Microsoft Hyper-V). It is important to secure this layer of software that sits between the hardware and the virtual servers. In a public IaaS service, customers do not have access to this software layer; it is managed by the CSP only.

Customer guest OS or virtual server security

The virtual instance of an operating system that is provisioned on top of the virtualization layer and is visible to customers from the Internet; e.g., various flavors of Linux, Microsoft, and Solaris. Customers have full access to virtual servers.

2.2.3 Virtualization Software Security

Since the CSP manages the virtualization software that sits on top of the hardware, customers will have neither visibility nor access to this software. Hardware or OS virtualization enables the sharing of hardware resources across multiple guest VMs without interfering with each other so that you can safely run several operating systems and applications at the same time on a single computer. For the purpose of simplicity, we made an assumption that IaaS services are using “bare metal hypervisor” technologies (also known as type 1 hypervisors), such asVMware ESX, Xen, Oracle VM, and Microsoft’s Hyper-V.

These hypervisors support a varietyof guest OSs, including Microsoft Windows, various Linux “flavors,” and Sun’s OpenSolaris.Given that hypervisor virtualization is the essential ingredient that guarantees compartmentalization and isolation of customer VMs from each other in a multitenant environment, it is very important to protect the hypervisors from unauthorized users. A new arms race between hacker and defender (CSP) in the realm of virtualization security is already underway. Since virtualization is very critical to the IaaS cloud architecture, any attack that could compromise the integrity of the compartments will be catastrophic to the entire customer base on that cloud. A recent incident at a tiny UK-based company called Vaserv.com exemplifies the threat to hypervisor security. By exploiting a zero-day vulnerability in HyperVM, a virtualization application made by a company called Lxlabs, hackers destroyed 100,000 websites hosted by Vaserv.com. The zero-day vulnerability gave the attackers the ability to execute sensitive Unix commands on the system, including rm -rf, which forces a recursive delete of all files.

Evidently, just days before the intrusion, an anonymous user posted on a hacker website called milw0rm a long list of yet-unpatched vulnerabilities in Kloxo, a hosting control panel that integrates into HyperVM. The situation was worse for approximately 50% of Vaserv’s customers who signed up for unmanaged service, which doesn’t include data backup. It remains unclear whether those website owners will ever be able to retrieve their lost data. CSPs should institute the necessary security controls, including restricting physical and logical access to hypervisor and other forms of employed virtualization layers. IaaS customers should understand the technology and security process controls instituted by the CSP to protect the hypervisor. This will help you to understand the compliance and gaps with reference to your host security standard, policies, and regulatory compliances. However, in general, CSPs lack transparency in this area and you may have no option but to take a leap of faith and trust CSPs to provide an “isolated and secured virtualized guest OS.”

Threats to the hypervisor

The integrity and availability of the hypervisor are of utmost importance and are key to guaranteeing the integrity and availability of a public cloud built on a virtualized environment. A vulnerable hypervisor could expose all user domains to malicious insiders. Furthermore, hypervisors are potentially susceptible to subversion attacks. To illustrate the vulnerability of the virtualization layer, some members of the security research community demonstrated a “Blue Pill” attack on a hypervisor. During Black Hat 2008 and Black Hat DC 2009[17] Joanna Rutkowska, Alexander Tereshkin, and Rafal Wojtczuk from Invisible Things Lab demonstrated a number of ways to compromise Xen’s virtualization.[18] Although Rutkowska and her team have identified problems with Xen implementations, generally they seem quite positive about the Xen approach. But their demonstration does illustrate the complexity of securing virtualized systems and the need for new approaches to protect hypervisors from such attacks. Since virtualization layers within public clouds for the most part are proprietary and closed source (although some may employ a derivative of open source virtualization software such as Xen), the source code of software used by CSPs is not available for scrutiny by the security research community.

2.2.4 Virtual Server Security

Customers of IaaS have full access to the virtualized guest VMs that are hosted and isolated from each other by hypervisor technology. Hence customers are responsible for securing and ongoing security management of the guest VM. A public IaaS, such as Amazon's Elastic Compute Cloud (EC2), offers a web services API to perform management functions such as provisioning, decommissioning, and replication of virtual servers on the IaaS platform. These system management functions, when orchestrated appropriately, can provide elasticity for resources to grow or shrink in line with workload demand.

The dynamic life cycle of virtual servers can result in complexity if the process to manage the virtual servers is not automated with proper procedures. From an attack surface perspective, the virtual server (Windows, Solaris, or Linux) may be accessible to anyone on the Internet, so sufficient network access mitigation steps should be taken to restrict access to virtual instances. Typically, the CSP blocks all port access to virtual servers and recommends that customers use port 22 (Secure Shell or SSH) to administer virtual server instances. The cloud management API adds another layer of attack surface and must be included in the scope of securing virtual servers in the public cloud. Some of the new host security threats in the public IaaS include:

• Stealing keys used to access and manage hosts (e.g., SSH private keys)

• Attacking unpatched, vulnerable services listening on standard ports (e.g., FTP, NetBIOS, SSH)

• Hijacking accounts that are not properly secured (i.e., weak or no passwords for standard accounts)

• Attacking systems that are not properly secured by host firewalls

• Deploying Trojans embedded in the software component in the VM or within the VM image (the OS) itself

Securing virtual servers

The simplicity of self-provisioning new virtual servers on an IaaS platform creates a risk that insecure virtual servers will be created. Secure-by-default configuration needs to be ensured by following or exceeding available industry baselines. Securing the virtual server in the cloud requires strong operational security procedures coupled with automation of procedures. Here are some recommendations:

• Use a secure-by-default configuration. Harden your image and use a standard hardened image for instantiating VMs (the guest OS) in a public cloud. A best practice for cloudbased applications is to build custom VM images that have only the capabilities and services necessary to support the application stack. Limiting the capabilities of the underlying application stack not only limits the host's overall attack surface, but also greatly reduces the number of patches needed to keep that application stack secure.

• Track the inventory of VM images and OS versions that are prepared for cloud hosting. The IaaS provider provides some of these VM images. When a virtual image from the IaaS provider is used it should undergo the same level of security verification and hardening for hosts within the enterprise. The best alternative is to provide your own image that conforms to the same security standards as internal trusted hosts.

• Protect the integrity of the hardened image from unauthorized access.

• Safeguard the private keys required to access hosts in the public cloud.

• In general, isolate the decryption keys from the cloud where the data is hosted—unless they are necessary for decryption, and then only for the duration of an actual decryption activity. If

your application requires a key to encrypt and decrypt for continuous data processing, it may not be possible to protect the key since it will be collocated with the application.

• Include no authentication credentials in your virtualized images except for a key to decrypt the filesystem key.

• Do not allow password-based authentication for shell access.

• Require passwords for sudo[19] or role-based access (e.g., Solaris, SELinux).

• Run a host firewall and open only the minimum ports necessary to support the services on an instance.

• Run only the required services and turn off the unused services (e.g., turn off FTP, print services, network file services, and database services if they are not required).

• Install a host-based IDS such as OSSEC or Samhain.

• Enable system auditing and event logging, and log the security events to a dedicated log server. Isolate the log server with higher security protection, including accessing controls.

• If you suspect a compromise, shut down the instance, snapshot your block volumes, and back up the root filesystem. You can perform forensics on an uncompromised system later.

• Institute a process for patching the images in the cloud—both offline and instantiated images.

• Periodically review logs for suspicious activities. **Figure 13** lists security controls at the host level.

Threat outlook	**High**
Preventive controls	Host firewall, access control, patching, hardening of system, strong authentication
Detective controls	Security event logs, host-based IDS/IPS

Figure 13. Security controls at the host level

2.3 Infrastructure Security: The Application Level

Application or software security should be a critical element of your security program. Most enterprises with information security programs have yet to institute an application security program to address this realm. Designing and implementing applications targeted for deployment on a cloud platform will require that existing application security programs reevaluate current practices and standards. The application security spectrum ranges from standalone single-user applications to sophisticated multiuser e-commerce applications used by millions of users. Web applications such as content management systems (CMSs), wikis, portals, bulletin boards, and discussion forums are used by small and large organizations.

A large number of organizations also develop and maintain custom-built web applications for their businesses using various web frameworks (PHP[20], .NET[21], J2EE[22] Ruby on Rails, Python, etc.). According to SANS, until 2007 few criminals attacked vulnerable

websites because other attack vectors were more likely to lead to an advantage in unauthorized economic or information access. Increasingly, however, advances in cross-site scripting (XSS) and other attacks have demonstrated that criminals looking for financial gain can exploit vulnerabilities resulting from web programming errors as new ways to penetrate important organizations. In this section, we will limit our discussion to web application security: web applications in the cloud accessed by users with standard Internet browsers, such as Firefox, Internet Explorer, or Safari, from any computer connected to the Internet.

Since the browser has emerged as the end user client for accessing in-cloud applications, it is

important for application security programs to include browser security into the scope of application security. Together they determine the strength of end-to-end cloud security that helps protect the confidentiality, integrity, and availability of the information processed by cloud services.

2.3.1 Application-Level Security Threats

According to SANS, web application vulnerabilities in open source as well as custom-built applications accounted for almost half the total number of vulnerabilities discovered between November 2006 and October 2007. The existing threats exploit well-known application vulnerabilities (e.g., the OWASP Top 10;), including cross-site scripting (XSS), SQL injection, malicious file execution, and other vulnerabilities resulting from programming errors and design flaws. Armed with knowledge and tools, hackers are constantly scanning web applications (accessible from the Internet) for application vulnerabilities. They are then exploiting the vulnerabilities they discover for various illegal activities including financial fraud, intellectual property theft, converting trusted websites into malicious servers serving client-side exploits, and phishing scams. All web frameworks and all types of web applications are at risk of web application security defects, ranging from insufficient validation to application logic errors.

It has been a common practice to use a combination of perimeter security controls and network- and host-based access controls to protect web applications deployed in a tightly controlled environment, including corporate intranets and private clouds, from external hackers. Web applications built and deployed in a public cloud platform will be subjected to a high threat level, attacked, and potentially exploited by hackers to support fraudulent and illegal activities. In that threat model, web applications deployed in a public cloud (the SPI model) must be designed for an Internet threat model, and security must be embedded into the Software Development Life Cycle (SDLC); **(Figure 14).**

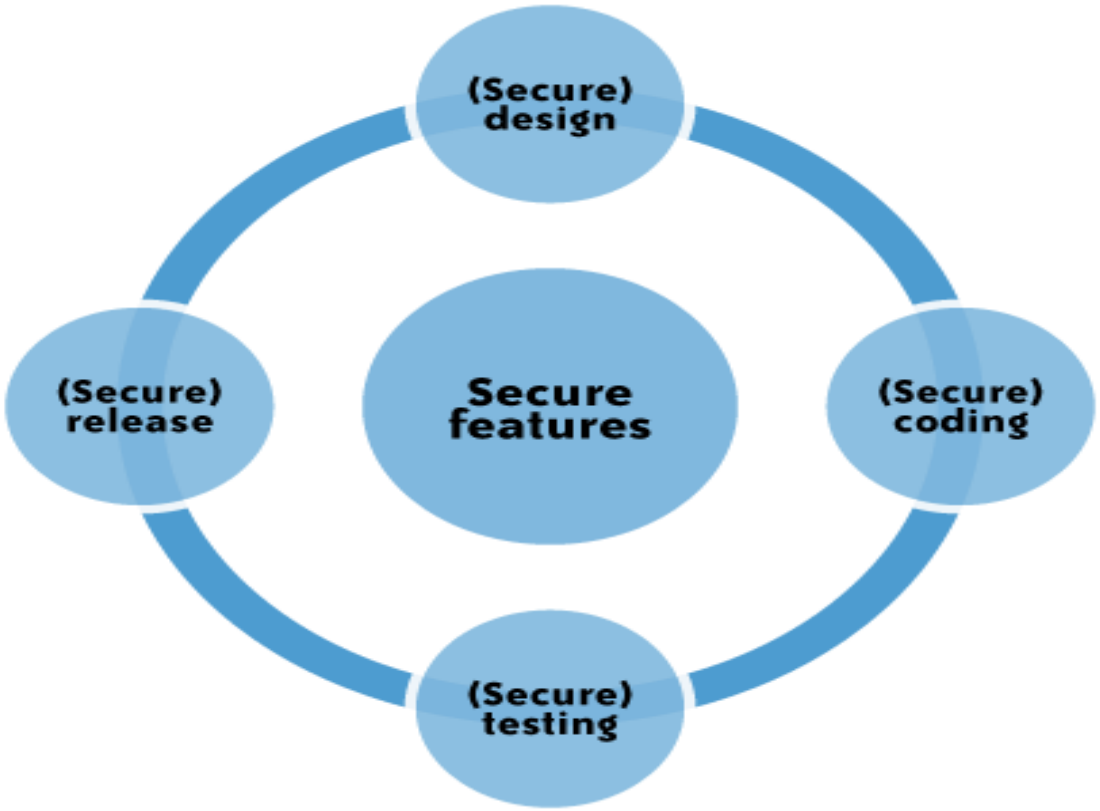


Figure 14. The SDLC

2.3.2 DoS and EDoS

Additionally, you should be cognizant of application-level DoS and DDoS attacks that can potentially disrupt cloud services for an extended time. These attacks typically originate from compromised computer systems attached to the Internet (routinely, hackers hijack and control computers infected by way of viruses/worms/malware and, in some cases, powerful unprotected servers). Application-level DoS attacks could manifest themselves as high-volume web page reloads, XML[23] web services requests (over HTTP or HTTPS), or protocol-specific requests supported by a cloud service. Since these malicious requests blend with the legitimate traffic, it is extremely difficult to selectively filter the malicious traffic without impacting the service as a whole. For example, a DDoS attack on Twitter on August 6, 2009, brought the service down for several hours **(Figure 15).**

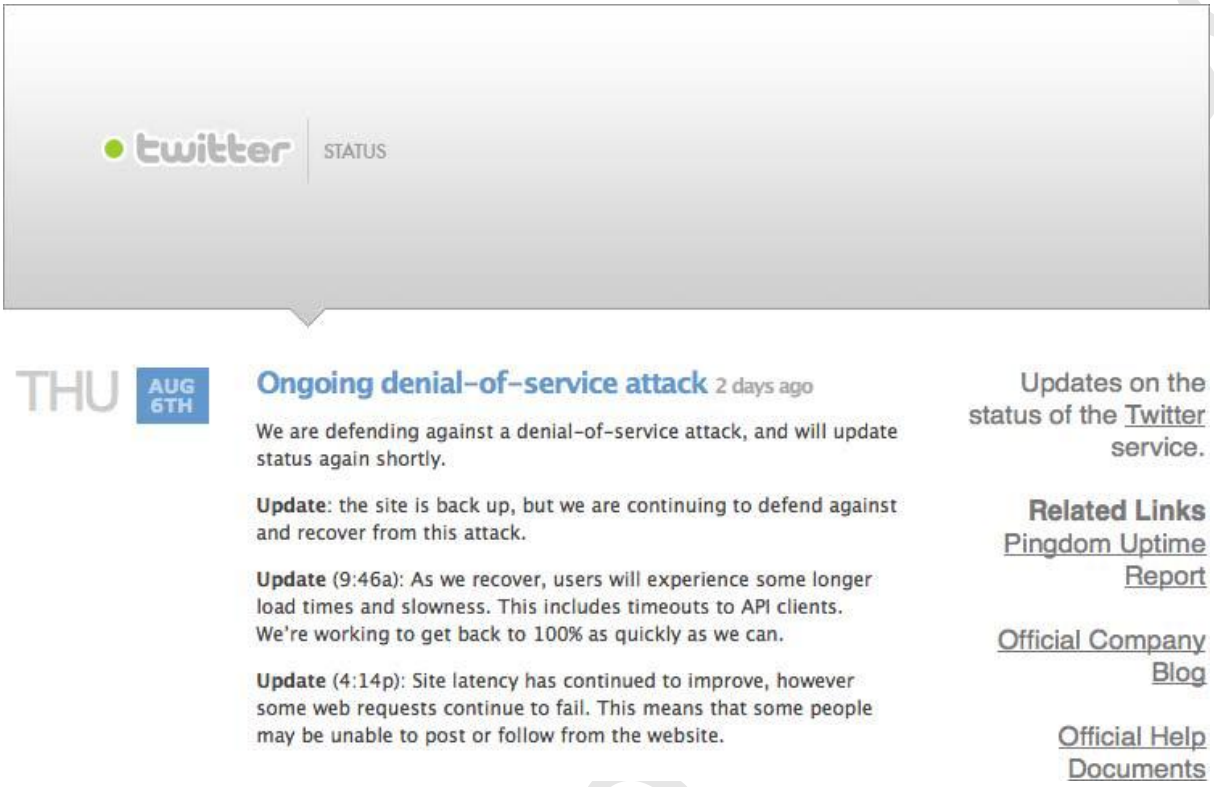


Figure 15. DDoS attack on Twitter

Apart from disrupting cloud services, resulting in poor user experience and service-level impacts, DoS attacks can quickly drain your company's cloud services budget. DoS attacks on pay-as-you-go cloud applications will result in a dramatic increase in your cloud utility bill: you'll see increased use of network bandwidth, CPU, and storage consumption. This type of attack is also being characterized as *economic denial of sustainability* (EDoS).[24]

The low barriers for small and medium-size enterprises to adopt cloud computing for legitimate use are also leveling the field for hackers. Using hijacked or exploited cloud accounts, hackers will be able to link together computing resources to achieve massive amounts of computing without any of the capital infrastructure costs. In the not-so-distant future, you might witness DoS attacks launched from IaaS or PaaS clouds against other cloud services (such hostile and offensive cloud models are being characterized as *dark clouds*).

2.3.3 End User Security

You, as a customer of a cloud service, are responsible for end user security tasks—security procedures to protect your Internet-connected PC—and for practicing "safe surfing." Protection measures include use of security software, such as anti-malware, antivirus, personal firewalls, security patches, and IPS-type software on your Internet-connected computer. The new mantra of "the browser is your operating system" appropriately conveys the message that browsers have become the ubiquitous "operating systems" for consuming cloud services. All Internet browsers routinely suffer from software vulnerabilities that make

them vulnerable to end user security attacks. Hence, our recommendation is that cloud customers take appropriate steps to protect browsers from attacks. To achieve end-to-end security in a cloud, it is essential for customers to maintain good browser hygiene. The means keeping the browser (e.g., Internet Explorer, Firefox, Safari) patched and updated to mitigate threats related to browser vulnerabilities. Currently, although browser security add-ons are not commercially available, users are encouraged to frequently check their browser vendor's website for security updates, use the auto-update feature, and install patches on a timely basis to maintain end user security[25].

2.3.4 Who Is Responsible for Web Application Security in the Cloud?

Depending on the cloud services delivery model (SPI) and service-level agreement (SLA), the scope of security responsibilities will fall on the shoulders of both the customer and the cloud provider. The key is to understand what your security responsibilities are versus those of the CSP. In that context, recent security surveys have highlighted the fact that lack of transparency in security controls and practices employed by CSPs is a barrier to cloud adoption.

To start with, cloud customers do not have the transparency required in the area of software vulnerabilities in cloud services. This prevents customers from managing the operational risk that might come with the vulnerabilities. Furthermore, by treating their software as proprietary, CSPs are impeding security researchers from analyzing the software for security flaws and bugs. (The exception is cloud providers that are operating on open source software.)

Due to this lack of transparency, customers are left with no choice but to trust their CSPs to disclose any new vulnerability that may affect the confidentiality, integrity, or availability of their application. For example, as of March 2009, no prominent IaaS, PaaS, or SaaS vendors are participating in the Common Vulnerability and Exposures (CVE) project. Case in point: AWS took 7.5 months to fix a vulnerability that Colin Percival reported in May 2007[26]. This vulnerability was a cryptographic weakness in Amazon's request signing code that affected its database API (SimpleDB) and EC2 API services, and it was not made public until after it was fixed in December 2008. (Colin does acknowledge that Amazon took this issue seriously at all times, and the lengthy timeline was simply due to the large amount of work involved in rolling out a patch to the affected services.) Enterprise customers should understand the vulnerability disclosure policy of cloud services and factor that into the CSP risk assessment. The following sections discuss the web application security in the context of the SPI cloud service delivery model.

2.3.5 SaaS Application Security

The SaaS model dictates that the provider manages the entire suite of applications delivered to users. Therefore, SaaS providers are largely responsible for securing the applications and components they offer to customers. Customers are usually responsible for operational security functions, including user and access management as supported by the provider. It is a common practice for prospective customers, usually under an NDA, to request information

related to the provider's security practices. This information should encompass design, architecture, development, black- and white-box application security testing, and release management. Some customers go to the extent of hiring independent security vendors to perform penetration testing (black-box security testing) of SaaS applications (with consent from the provider) to gain assurance independently.

However, penetration testing can be costly and notall providers agree to this type of verification. Extra attention needs to be p aid to the authentication and access control features offered by SaaS CSPs. Usually that is the only security control available to manage risk to information. Most services, including those from Salesforce.com and Google, offer a web-based administration user interface tool to manage authentication and access control of the application. Some SaaS applications, such as Google Apps, have built-in features that end users can invoke to assign read and write privileges to other users. However, the privilege management features may not be advanced, fine-grained access and could have weaknesses that may not conform to your organization's access control standard.

One example that captures this issue is the mechanism that Google Docs employs in handling images embedded in documents, as well as access privileges to older versions of a document. Evidently, embedded images stored in Google Docs are not protected in the same way that a document is protected with sharing controls. That means if you have shared a document containing embedded images, the other person will always be able to view those images even after you've stopped sharing the document. A blogger[27] discovered this access control quirk and brought it to Google's attention. Although Google has acknowledged the issue, its response conveys that it believes[28] those concerns do not pose a significant security risk to its users.

Another incident related to Google Docs was a privacy glitch[29] that inappropriately shared access to a small fraction (Google claims 0.05% of the documents were affected) of word processing and presentation documents stored on its Google Apps cloud service. Though the documents were shared only with people whom the Google Docs users had already shared documents, rather than with the world at large, the problem illustrates the need to evaluate and understand cloud-specific access control mechanisms.

Cloud customers should try to understand cloud-specific access control mechanisms— including support for strong authentication and privilege management based on user roles and functions—and take the steps necessary to protect information hosted in the cloud. Additional controls should be implemented to manage privileged access to the SaaS administration tool, and enforce segregation of duties to protect the application from insider threats. In line with security standard practices, customers should implement a strong password policy—one that forces users to choose strong passwords when authenticating to an application[30].

It is a common practice for SaaS providers to commingle their customer data (structured and unstructured) in a single virtual data store and rely on data tagging to enforce isolation between customer data. In that multitenant data store model, where encryption may not be feasible due to key management and other design barriers, data is tagged and stored with a unique customer identifier. This unique data tag makes it possible for the business logic embedded in the application layer to enforce isolation between customers when the data is processed. It is conceivable that the application layer enforcing this isolation could become vulnerable during software upgrades by the CSP. Hence, customers should understand the virtual data store architecture and the preventive mechanisms the SaaS providers use to guarantee the compartmentalization and isolation required in a virtual multitenant environment.

Established SaaS providers, such as Salesforce.com, Microsoft, and Google, are known to invest in software security and practice security assurance as part of their SDLC. However, given that there is no industry standard to assess software security, it is almost

impossible to benchmark providers against a baseline[31]. **Figure 16** lists security controls at the application level.

Threat outlook	Medium
Preventive controls	Identity management, access control assessment, browser hardened with latest patches, multifactor authentication via delegated authentication, endpoint security measures including antivirus and IPS
Detective controls	Login history and available reports from SaaS vendors

Figure 16. Security controls at the application level

2.3.6 PaaS Application Security

PaaS vendors broadly fall into the following two major categories:

• Software vendors (e.g., Bungee, Etelos, GigaSpaces, Eucalyptus)

• CSPs (e.g., Google App Engine, Salesforce.com's Force.com, Microsoft Azure, Intuit QuickBase)

Organizations evaluating a private cloud may utilize PaaS software to build a solution for internal consumption. Currently, no major public clouds are known to be using commercial off-the-shelf or open source PaaS software such as Eucalyptus (Eucalyptus does offer a limited experimental pilot cloud for developers at Eucalyptus.com[32], however). Therefore, given the nascent stage of PaaS deployment, we will not discuss software security of standalone PaaS software in this chapter. Nonetheless, it is recommended that organizations evaluating PaaS software perform a risk assessment and apply the software security standard similar to acquiring any enterprise software.

By definition, a PaaS cloud (public or private) offers an integrated environment to design, develop, test, deploy, and support custom applications developed in the language the platform supports. PaaS application security encompasses two software layers:

• Security of the PaaS platform itself (i.e., runtime engine)

• Security of customer applications deployed on a PaaS platform

Generally speaking, PaaS CSPs (e.g., Google, Microsoft, and Force.com) are responsible for securing the platform software stack that includes the runtime engine that runs the customer applications. Since PaaS applications may use third-party applications, components, or web services, the third-party application provider may be responsible for securing their services. Hence, customers should understand the dependency of their application on all services and assess risks pertaining to third-party service providers. Until now, CSPs have been reluctant to share information pertaining to platform security using the argument that such security information could provide an advantage for hackers. However, enterprise customers should demand transparency from CSPs and seek information necessary to perform risk assessment and ongoing security management.

PaaS application container

In the multitenant PaaS service delivery model, the core security tenets are containment and isolation of multitenant applications from each other. In that model, access to your data should be restricted to your enterprise users and to applications that you own and manage. The security model of the PaaS platform runtime engine is the CSP's intellectual property, and it is essential to delivering the "sandbox" architecture in a multitenant computing model. Hence, the sandbox characteristic of the platform runtime engine is central in maintaining the confidentiality and integrity of your application deployed in the PaaS. CSPs are responsible for monitoring new bugs and vulnerabilities that may be used to exploit the PaaS platform and break out of the sandbox architecture. This type of situation is the worst case scenario for a PaaS service; the privacy implications for customer-sensitive information are undesirable and could be very damaging to your business. Hence, enterprise customers should seek information from the CSP on the containment and isolation architecture of the PaaS service.

Network and host security monitoring outside the PaaS platform is also the responsibility of the PaaS cloud provider (i.e., monitoring of a shared network and system infrastructure hosting customer applications). PaaS customers should understand how PaaS CSPs are managing their platform, including updating of the runtime engine and change, release, and patch management.

2.3.7 Customer-Deployed Application Security

PaaS developers need to get familiar with specific APIs to deploy and manage software modules that enforce security controls. Furthermore, given that the API is unique to a PaaS cloud service, developers are required to become familiar with platform-specific security features— available to them in the form of security objects and web services for configuring authentication and authorization controls within the application. When it comes to PaaS API design, currently no standard is available, nor is there any concerted effort by CSPs to develop a ubiquitous and consistent API across clouds—and that makes porting of an application across PaaS clouds a monumental task. Currently, the Google App Engine supports only Python and Java, and Salesforce.com's Force.com supports only a proprietary language called Apex. (Apex differs from languages such as C++, Java, and .NET. Unlike those languages, Apex is much more limited in scope and is specific to building business applications on the Force.com platform.) In this regard, cloud services have the potential to retain customers more forcefully than traditional software licensing. The lack of an API standard has ramifications for both security management and portability of applications across the cloud.

Developers should expect CSPs to offer a set of security features, including user authentication, single sign-on (SSO) using federation, authorization (privilege management), and SSL or TLS support, made available via the API. Currently, there is no PaaS security management standard: CSPs have unique security models, and security features will vary from provider to provider. In the case of the Google App Engine, a developer using Python or Java objects can configure the user profile and select HTTPS as a transport protocol. Similarly, Force.com offers an Apex API to configure security parameters, manipulate various runtime configurations, and assign certain TCP ports for application-to-application connection-type interactions using Apex objects[33]. Based on our assessment of major PaaS CSPs, the security features available to PaaS applications are limited to basic security configuration—SSL configuration, basic privilege management, and user authentication using the provider's identity store. In only a few cases, user federation is supported using the Security Assertion Markup Language (SAML). **Figure 17** lists security controls applicable to PaaS applications.

Threat outlook	Medium
Preventive controls	User authentication, account management, browser hardened with latest patches, endpoint security measures including antivirus and IPS
Detective controls	Application vulnerability scanning

Figure 17. Security controls applicable to PaaS applications

2.3.8 IaaS Application Security

IaaS cloud providers (e.g., Amazon EC2, GoGrid, and Joyent) treat the applications on customer virtual instances as a black box, and therefore are completely agnostic to the operations and management of the customer's applications. The entire stack—customer applications, runtime application platform (Java, .NET, PHP, Ruby on Rails, etc.), and so on—runs on the customer's virtual servers and is deployed and managed by customers. To that end, customers have full responsibility for securing their applications deployed in the IaaS cloud. Hence, customers should not expect any application security assistance from CSPs other than basic guidance and features related to firewall policy that may affect the application's communications with other applications, users, or services within or outside the cloud.

Web applications deployed in a public cloud must be designed for an Internet threat model, embedded with standard security countermeasures against common web vulnerabilities (e.g., the OWASP Top 10). In adherence with common security development practices, they should also be periodically tested for vulnerabilities, and most importantly, security should be embedded into the SDLC. Customers are solely responsible for keeping their applications and runtime platform patched to protect the system from malware and hackers scanning for vulnerabilities to gain unauthorized access to their data in the cloud. It is highly recommended that you design and implement applications with a "least-privileged" runtime model (e.g., configure the application to run using a lower privileged account).

Developers writing applications for IaaS clouds must implement their own features to handle authentication and authorization. In line with enterprise identity management practices, cloud applications should be designed to leverage delegated authentication service features supported by an enterprise Identity Provider (e.g., OpenSSO, Oracle IAM, IBM, CA) or third-party identity service provider (e.g., Ping Identity, Symplified, TriCipher). Any custom implementations of Authentication, Authorization, and Accounting (AAA) features can become a weak link if they are not properly implemented, and you should avoid them when possible.

In summary, the architecture for IaaS hosted applications closely resembles enterprise web applications with an *n*-tier distributed architecture. In an enterprise, distributed applications run with many controls in place to secure the host and the network connecting the distributed hosts. Comparable controls do not exist by default in an IaaS platform and must be added through a network, user access, or as application-level controls. Customers of IaaS clouds are responsible for all aspects of their application security and should take the steps necessary to protect their application to address application-level threats in a multitenant and hostile Internet environment.

Figure 18 lists security controls applicable to IaaS applications.

Threat outlook	High
Preventive controls	Application developed using security-embedded SDLC process, least-privileged configuration, timely patching of application, user authentication, access control, account management, browser hardened with latest patches, endpoint security measures including antivirus, IPS, host-based IDS, host firewall, and virtual private network (VPN) for administration
Detective controls	Logging, event correlation, application vulnerability scanning and monitoring

Figure 18. Security controls applicable to IaaS applications

2.3.9 Public Cloud Security Limitations

Customers evaluating the public cloud should keep in mind that there are limitations to the public cloud when it comes to support for custom security features. Security requirements such as an application firewall, SSL accelerator, cryptography, or rights management using a device that supports PKCS 12 are not supported in a public SaaS, PaaS, or IaaS cloud. In the future, IaaS and PaaS providers may offer some of these more sophisticated security features, depending on customer demand. In general, any mitigation controls that require deployment of an appliance or locally attached peripheral devices in the public IaaS/PaaS cloud are not feasible at this time.

References

1. This issue was reported on the blog of Colin Percival, “Daemonic Dispatches,” on December 18, 2008. “AWS signature version 1 is insecure”. There was no public acknowledgment of this issue on the AWS website, nor any public response to Percival’s blog posting.
2. “Announcing Elastic IP Addresses and Availability Zones for Amazon EC2”. Though announced in March 2009, the Elastic IP service became available October 22, 2008.
3. Section 3.3, “An Evaluation of Amazon’s Grid Computing Services: EC2, S3 and SQS,” by Simson L. Garfinkel; TR-08-07, Computer Science Group, Harvard University, Cambridge, Massachusetts.
4. RFC 1918, “Address Allocation for Private Internets,” for further information.
5. For example, “Instance Addressing and Network Security” in the Amazon Elastic Compute Cloud Developer Guide (API Version 2008-12-01).
6. “Amazon: Hey Spammers, Get Off My Cloud!” reported in The Washington Post, July 1, 2008.
7. An example is CohesiveFT’s VPN-Cubed, but this product is not available as a cloud provider service from most cloud providers—which would mean yet another third-party solution to integrate into your cloud environment. However, cloud provider AWS does offer this product as a service.
8. Border Gateway Protocol is an interdomain routing protocol used in the core of the Internet. You can find more information about BGP in RFC 4271, “A Border Gateway Protocol 4 (BGP-4).”
9. According to RFC 1930, “Guidelines for Creation, Selection, and Registration of an Autonomous System (AS),” an autonomous system is a connected group of one or more IP prefixes run by one or more network operators that has a single and clearly defined routing policy.
10. “Short-Lived Prefix Hijacking on the Internet” by Peter Boothe, James Hiebert, and Randy Bush, presented at NANOG 36 in February 2006.
11. For example, “Pakistan Cuts Access to YouTube Worldwide” in The New York Times, February 26, 2008.
12. DNS stands for Domain Name System. RFCs 1034, “Domain Names—Concepts and Facilities,” and 1035, “Domain Names—Implementation and Specification.”
13. That is not to say that internal DNS systems are entirely free of attacks—just that they are safer than external DNS systems and queries using them. For example, the paper “Corrupted DNS Resolution Paths: The Rise of a Malicious Resolution Authority,” written by members of the faculty of the Georgia Institute of Technology.
14. The Kaminsky Bug was named after the security researcher who discovered the problem, Dan Kaminsky of IOActive. A good non-technical explanation of the bug and of attempts to mitigate it through efforts with the vendor community is available in the article “Fresh Phish,” published in the October 2008 issue of IEEE’s Spectrum magazine.
15. For example, US-CERT Vulnerability Note VU#800113, “Multiple DNS implementations vulnerable to cache poisoning.” As of December 31, 2008, the National Vulnerability Database lists 312 vulnerabilities for the DNS protocol and implementations of DNS. The National Vulnerability Database is sponsored by the U.S. Department of Homeland Security’s US-CERT, and NIST.
16. For example, “Rumor: Amazon Hit With Denial-of-Service Attack, Again,” posted June 6, 2008 at http://www.appscout.com/2008/06/rumor_amazon_hit_with_denialof.php.
17. Black Hat DC 2009.
18. http://theinvisiblethings.blogspot.com/2008/08/our-xen-0wning-trilogy-highlights.html.
19. http://en.wikipedia.org/wiki/Sudo.
20. http://en.wikipedia.org/wiki/PHP.
21. http://msdn.microsoft.com/netframework/.
22. http://en.wikipedia.org/wiki/J2EE.

23. XML stands for eXtensible Markup Language; http://en.wikipedia.org/wiki/XML

24. http://rationalsecurity.typepad.com/blog/2009/01/a-couple-of-followups-on-my-edos-economic-denial-of-sustainability-concept.html.

25. A good reference for browser security is Google's Browser Security Handbook.

26. http://www.daemonology.net/blog/2008-12-18-AWS-signature-version-1-is-insecure.html.

27. Google Docs access control issue: http://peekay.org/2009/03/26/security-issues-with-google-docs/.

28. Google Docs access control response to a weakness issue: http://googledocs.blogspot.com/2009/03/just-to-clarify.html.

29. Google Docs privacy glitch: http://www.techcrunch.com/2009/03/07/huge-google-privacy-blunder-shares-your-docs-without-permission/.

30. http://en.wikipedia.org/wiki/cloud.

31. The Payment Application Data Security Standard (PA-DSS) is applicable only to organizations that store, process, or transmit cardholder data—with guidance for software developers and manufacturers of applications and devices used in those transactions.

32. http://open.eucalyptus.com/wiki/EucalyptusPublicCloud.

33. http://www.salesforce.com/us/developer/docs/api/Content/sforce_api_concepts_security.htm.

Chapter 3. Privacy

"You can have security and not have privacy, but you cannot have

privacy without security."

—Tim Mather

Particularly in less regulated industries (those other than health care and financial services) responsibility and accountability for privacy is often (erroneously) assigned to IT instead of the business unit that owns the data. In many cases, it is treated as a checkbox to verify among several other burdensome requirements. Infrastructure and data security in public cloud computing is, for many organizations (e.g., large enterprises), likely to be less robust than their own current capabilities. With this likely less-secure, greater-risk security posture, it follows that the risk of a privacy breach is also increased. It should, however, be noted that many small and medium-size businesses (SMBs) have limited IT and dedicated information security resources, and as a result they place limited focus on this area. For these organizations, the security afforded by a public cloud service provider (CSP) can be greater. Even a seemingly small data breach can have a considerable financial impact (e.g., cost of incident response and possible forensic investigation, restitution to victims of identity theft, punitive damages), as well as long-term consequences such as negative publicity and loss of customer confidence. Despite the all-too-familiar headlines, privacy considerations are often not proportional to the level of inherent risk.

3.1 What Is Privacy?

The concept of privacy varies widely among (and sometimes within) countries, cultures, and jurisdictions. It is shaped by public expectations and legal interpretations; as such, a concise definition is elusive if not impossible. Privacy rights or obligations are related to the collection, use, disclosure, storage, and destruction of personal data (or personally identifiable information—PII). At the end of the day, privacy is about the accountability of organizations to data subjects, as well as the transparency to an organization's practice around personal information. Likewise, there is no universal consensus about what constitutes personal data. For the purposes of this discussion, we will use the definition adopted by the Organization for Economic Cooperation and Development (OECD): any information relating to an identified or identifiable individual (data subject)[1].

Another definition gaining popularity is the one provided by the American Institute of Certified Public Accountants (AICPA) and the Canadian Institute of Chartered Accountants (CICA) in the Generally Accepted Privacy Principles (GAPP) standard: "The rights and obligations of individuals and organizations with respect to the collection, use, retention, and disclosure of personal information."

3.2 What Is the Data Life Cycle?

Personal information should be managed as part of the data used by the organization. It should be managed from the time the information is conceived through to its final disposition. Protection of personal information should consider the impact of the cloud on each of the following phases as detailed in the next figure.

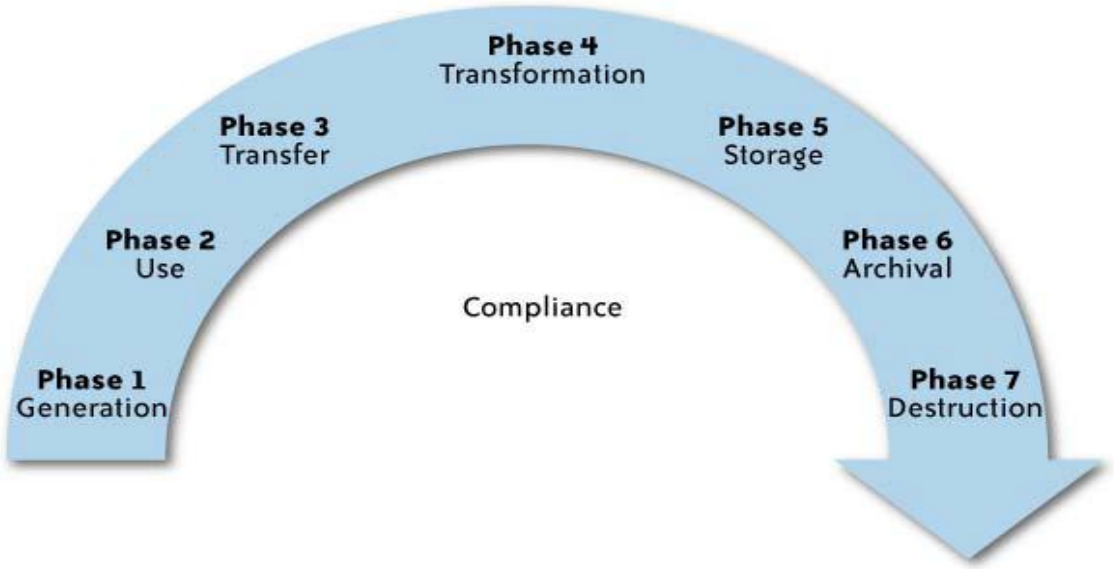


Figure 19

The components within each of these phases are:

Generation of the information:

- Ownership: Who in the organization owns PII, and how is the ownership maintained if the organization uses cloud computing?
- Classification: How and when is PII classified? Are there limitations on the use of cloud computing for specific data classes?
- Governance: Is there a governance structure to ensure that PII is managed and protected through its life cycle, even when it is stored or processed in a cloud computing environment?

Use

- Internal versus external: Is PII used only within the collecting organization, or is it used outside the organization (e.g., in a public cloud)?
- Third party: Is the information shared with third parties (e.g., subcontractors or CSPs)?
- Appropriateness: Is the use of the information consistent with the purpose for which it was collected? Is the use within the cloud appropriate based on the commitments the organization made to the data subjects?
- Discovery/subpoena: Is the information managed in the cloud in a way that will enable the organization to comply with legal requirements in case of legal proceedings?

Transfer

- Public versus private networks: When information is transferred to a cloud is the organization using public networks, and is it protected appropriately? (PII should always be protected to address the risk level and legal requirements.)
- Encryption requirements: Is the PII encrypted? Some laws require that PII will be encrypted when transmitted via a public network (and this will be the case when the organization is using a public cloud).
- Access control: Are there appropriate access controls over PII when it is in the cloud?

Transformation

- Derivation: Are the original protection and use limitations maintained when data is transformed or further processed in the cloud?
- Aggregation: Is data in the cloud aggregated so that it is no longer related to an identifiable individual (and hence is no longer considered PII)?
- Integrity: Is the integrity of PII maintained when it is in the cloud?

Storage

- Access control: Are there appropriate controls over access to PII when stored in the cloud so that only individuals with a need to know will be able to access it?
- Structured versus unstructured: How is the data stored to enable the organization to access and manage the data in the future?
- Integrity/availability/confidentiality: How are data integrity, availability, and confidentiality maintained in the cloud?
- Encryption: Several laws and regulations require that certain types of PII should be stored only when encrypted. Is this requirement supported by the CSP?

Archival

- Legal and compliance: PII may have specific requirements that dictate how long it should be stored and archived. Are these requirements supported by the CSP?
- Off-site considerations: Does the CSP provide the ability for long-term off-site storage that supports archival requirements?
- Media concerns: Is the information stored on media that will be accessible in the future? Is the information stored on portable media that may be more susceptible to loss? Who controls the media and what is the organization's ability to recover such media from the CSP if needed?
- Retention: For how long will the data be retained by the CSP? Is the retention period consistent with the organization's retention period?

Destruction

- Secure: Does the CSP destroy PII obtained by customers in a secure manner to avoid potential breach of the information?
- Complete: Is the information completely destroyed? Does the destruction completely erase the data, or can it be recovered?

The impact differs based on the specific cloud model used by the organization, the phase of personal information in the cloud, and the nature of the organization. The following analysis provides some of these considerations; however, every organization should consider performing a Privacy Impact Assessment (PIA) before embarking on a cloud computing initiative that involves personal information.

3.3 What Are the Key Privacy Concerns in the Cloud?

Privacy advocates have raised many concerns about cloud computing. These concerns typically mix security and privacy. Here are some additional considerations to be aware of:

Access

Data subjects have a right to know what personal information is held and, in some cases, can make a request to stop processing it. This is especially important with regard to marketing activities; in some jurisdictions, marketing activities are subject to additional regulations and are almost always addressed in the end user privacy policy for applicable organizations. In the cloud, the main concern is the organization's ability to provide the individual with access to all personal information, and to comply with stated requests. If a data subject exercises this right to ask the organization to delete his data, will it be possible to ensure that all of his information has been deleted in the cloud?

Compliance

What are the privacy compliance requirements in the cloud? What are the applicable laws, regulations, standards, and contractual commitments that govern this information, and who is responsible for maintaining the compliance? How are existing privacy compliance requirements impacted by the move to the cloud? Clouds can cross multiple jurisdictions; for example, data may be stored in multiple countries, or in multiple states within the United

States. What is the relevant jurisdiction that governs an entity's data in the cloud and how is it determined?

Storage

Where is the data in the cloud stored? Was it transferred to another data center in another country? Is it commingled with information from other organizations that use the same CSP? Privacy laws in various countries place limitations on the ability of organizations to transfer some types of personal information to other countries. When the data is stored in the cloud, such a transfer may occur without the knowledge of the organization, resulting in a potential violation of the local law.

Retention

How long is personal information (that is transferred to the cloud) retained? Which retention policy governs the data? Does the organization own the data, or the CSP? Who enforces the retention policy in the cloud, and how are exceptions to this policy (such as litigation holds) managed?

Destruction

How does the cloud provider destroy PII at the end of the retention period? How do organizations ensure that their PII is destroyed by the CSP at the right point and is not available to other cloud users? How do they know that the CSP didn't retain additional copies? Cloud storage providers usually replicate the data across multiple systems and sites—increased availability is one of the benefits they provide. This benefit turns into a challenge when the organization tries to destroy the data—can you truly destroy information once it is in the cloud? Did the CSP really destroy the data, or just make it inaccessible to the organization? Is the CSP keeping the information longer than necessary so that it can mine the data for its own use?

Audit and monitoring

How can organizations monitor their CSP and provide assurance to relevant stakeholders that privacy requirements are met when their PII is in the cloud?

Privacy breaches

How do you know that a breach has occurred, how do you ensure that the CSP notifies you when a breach occurs, and who is responsible for managing the breach notification process (and costs associated with the process)? If contracts include liability for breaches resulting from negligence of the CSP, how is the contract enforced and how is it determined who is at fault?

3.4 Who Is Responsible for Protecting Privacy?

There are conflicting opinions regarding who is responsible for security and privacy. Some publications assign it to providers; (2) but although it may be possible to transfer liability via contractual agreements, it is never possible to transfer accountability. Ultimately, in the eyes of the public and the law, the onus for data security and privacy falls on the organization that collected the information in the first place—the user organization. This is true even if the user organization has no technical capability to ensure that the contractual requirements with the CSP are met.

History and experience have proven that data breaches have a cascading effect. When an organization loses control of users' personal information, the users are responsible (directly or indirectly) for subsequent damages resulting from the loss. Identity theft is only one of the possible effects; others may include invasion of privacy or unwelcome solicitation. When an affected individual is dealing with the fallout, he will likely blame the one who made the decision to use the service, as opposed to the provider of the service. Full reliance on a

third party to protect personal data is irresponsible and will inevitably lead to negative consequences.

Responsible data stewardship requires an in-depth understanding of the technology underlying cloud computing and the legal requirements and implications. As such, a cross functional team is critical to adequately maintain security and privacy.

The accountability model (discussed earlier in this chapter) is similar to discussions around privacy in outsourcing or subcontracting relationships, and the conclusion is similar:

- Organizations can transfer liability, but not accountability.
- Risk assessment and mitigation throughout the data life cycle is critical.
- Knowledge about legal obligations and contractual agreements or commitments is imperative.

There are, however, many new risks and unknowns; thus, the overall complexity of privacy protection in the cloud represents a bigger challenge.

3.5 Changes to Privacy Risk Management and Compliance in Relation to Cloud Computing

3.5.1 Collection Limitation Principle

This principle specifies that collection of personal data should be limited to the minimum amount of data required for the purpose for which it is collected. Any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.

In the privacy arena, lack of specifics on data collection with providers creates misunderstandings down the road. For instance, one global outsourcer said, “Clients come in expecting the right things in security, but the wrong things in privacy. They are expecting best practices, but they don’t know what they are.” There are comprehensive security frameworks and standards (such as the ISO 27000 series, NIST guidelines, etc.), and organizations know how to implement them. There is no universally adopted privacy standard—instead, there are conflicting laws, regulations, and views on what privacy is and what it requires from organizations to protect it. Many organizations want to do what they perceive to be “the right thing”; however, their perception may be different from the law. As a result, there may be different expectations regarding what privacy means between the organization and the CSP, and no agreed best practices

It is essential that service-level agreements (SLAs) are initially defined before any information is provided or shared, because it is very hard to negotiate them later. If you start the request for proposal (RFP) process with an SLA target, you will be able to disqualify providers who cannot meet your stated needs. Well-defined security and privacy SLAs should be part of the statement of work (SOW). Ensure that your SLAs have teeth with specific penalty clauses. Do not cede command of service-level negotiation to the provider.

Moreover, organizations face the risk that, as different data elements about individuals are collected and later merged, the combined information is more than needed and the original purpose as well as the organization may be in potential violation of local laws.

3.5.2 Use Limitation Principle

This principle specifies that personal data should not be disclosed, made available, or otherwise used for purposes other than those with the consent of the data subject, or by the authority of law.

Cloud computing places a diverse collection of user and business information in a single location. As data flows through the cloud, strong data governance is needed to ensure that the original purpose of collection and limitation on use is attached to the data. This is critical when organizations create a centralized database, because future applications can easily combine the data via expanded views that are utilized for new purposes never approved by data subjects.

The ability to combine data from multiple sources increases the risk of unexpected uses by governments. Governments in different countries could ask CSPs to report on particular types of behaviors or to monitor activities of particular types or categories of users. The possibility that a CSP could be obliged to inform a government or a third party about user activities might be troubling to the provider as well as to its users.

3.5.3 Security Principle

Security is one of the key requirements to enable privacy. This principle specifies that personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification, or disclosure of data.

3.5.4 Retention and Destruction Principle

This principle specifies that personal data should not be retained for longer than needed to perform the task for which it was collected, or as required by laws or regulations. Data should be destroyed in a secure way at the end of the retention period.

How long data should be retained and when it should be destroyed is still a challenge for most companies. Data growth has led to definitions of policies and procedures for data retention and destruction. Most policies have been driven or imposed by legislation and regulations, such as the Health Insurance Portability and Accountability Act of 1996 (HIPAA), the Sarbanes-Oxley Act (SOX), and other federal and state compliance requirements.

The actual deletion process is sometimes loosely defined. But when data copies, data backups, or archives are deleted, are they really gone? Deleting a file only marks the space (or blocks) it occupies as usable. Until the blocks are actually overwritten, the data is still there and can be retrieved. In fact, the disk space occupied by deleted files must be overwritten with other data several times before the entirety of the files is deemed irretrievable (a minimum of seven times per the U.S. federal government's guidelines).

In many cases, disk or tape media is reused to store more data; therefore, data deletion typically does not constitute much of an issue. However, when leased IT assets, such as servers or disk arrays, must be returned, when obsolete systems are replaced, or when storage media has reached end-of-life, special care must be taken to ensure that any data once stored is irretrievable.

Encryption can play a key role in the destruction process. Encrypted data can be destroyed even when organizations lose track of their data by destroying the encryption key—data can no longer be decrypted and hence is rendered inaccessible. This is especially beneficial when the data is kept by CSPs—encrypted data can be destroyed without the involvement of the CSPs.

The problem begins when there is a lack of clearly defined policies around data destruction in cloud computing. Virtual storage devices can be reallocated to new users without deleting the data, and then allocated to new users. Personal information stored in this device may now be available to the new user, potentially violating individual rights, laws, and regulations. Servers or disks can be decommissioned without much thought as to whether data is still accessible.

There are several approved methods of data destruction, including media destruction, disk degaussing, multiple data overwrites with random byte patterns, and destruction of keying material for encrypted data.

3.5.5 Transfer Principle

This principle specifies that data should not be transferred to countries that don't provide the same level of privacy protection as the organization that collected the information. In a cloud computing environment, infrastructure is shared between organizations; therefore, there are threats associated with the fact that the data is stored and processed remotely, and there is increased sharing of platforms between users, which increases the need to protect privacy of data stored in the cloud. Another feature of cloud computing is that it is a dynamic environment; for example, service interactions can be created in a more dynamic way than in traditional e-commerce. Services can potentially be aggregated and changed dynamically by customers, and service providers can change the provisioning of services. In such scenarios, personal and sensitive data can move around within a single CSP infrastructure and across CSP organizational boundaries.

The goal of integrated services provided by multiple CSPs is to enhance the possibility of data transfer to third parties. This transfer should be disclosed to the data subject prior to collection. In many cases there is a need for unambiguous consent by the individual to the data transfer. Typically the organization is required to agree to the provider's standard terms of service without any scope for negotiation. The terms are likely to be biased in the provider's favor, and the organization may not know all the entities that are involved in the process, and hence is rendered unable to provide an accurate notice to the data subjects.

The transfer challenge is further complicated because data can be anywhere in the world. Usually, a company computing in the cloud does not know in what country its data resides at any given time. Instead of its data being stored on the company's servers, data is stored on the service provider's servers, which could be in Europe, China, or anywhere else. This tenet of cloud computing conflicts with various legal requirements, such as the European laws that require that a company know where the personal data in its possession is at all times, and there may be a need to report to data protection authorities on the data transfer. In some cases there may be a need to pre approve the transfer by data subjects.

The U.S. Safe Harbor Program—perhaps the most common means of compliance with EU requirements imposed when transferring the personal data of EU citizens to the United States— may not satisfy a multinational's EU legal obligations, because in cloud computing data could be stored on servers outside of both Europe and the United States, making the Safe Harbor Program ineffective. Furthermore, the Safe Harbor option may not be available for certain organizations not regulated by the Federal Trade Commission, such as those in the financial services industry. This may be the case even if the CSP is registered under the Safe Harbor Program.One cloud computing application service provider (ASP) offers its customers the option to store their data only on European servers (for a higher fee, naturally). However, it is an impractical solution because it limits the very flexibility and efficiency that cloud computing is designed to provide. Given the enormous potential and

benefits of computing in the cloud, it seems that, once again, the law needs to catch up with technology.

3.5.6 Accountability Principle

This principle states that an organization is responsible for personal information under its control and should designate an individual or individuals who are accountable for the organization's compliance with the remaining principles. Accountability within cloud computing can be achieved by attaching policies to data and mechanisms to ensure that these policies are adhered to by the parties that use, store, or share that data, irrespective of the jurisdiction in which the information is processed.

The way to move onward is for organizations to value accountability and build mechanisms for accountable, responsible decision making while handling data. Specifically, accountable organizations ensure that obligations to protect data are observed by all processors of the data, irrespective of where that processing occurs.

3.6 Legal and Regulatory Implications

Across the globe, the legal and regulatory requirements for data privacy range from strictly enforced to non-existent, which can prove to be a daunting challenge for multinational companies or those serving customers from multiple jurisdictions. Some programs such as the OECD Guidelines (3) and the European Union Data Protection Directive (4) are principle-based, where personal data processing is not permitted, except as directed in the statutes, whereas in countries such as the United States, certain types of processing are restricted, but activities are generally considered lawful unless specifically prohibited by applicable state and federal regulations. The jurisdiction of these laws is determined differently in different countries and states. Some of the laws are based on the location of the organization, some on the physical location of the data center, and some on the location of the data subjects. The only universal consistency is that the law has not caught up with the technology.

To further compound the challenge of processing personal data in a global environment, some requirements are conflicting. For example, compliance with the U.S. Federal Rules of Civil Procedure (FRCP) can breach the EU Directive. Differing attitudes on privacy have been the force behind countless cross-jurisdictional legal battles, international trade barriers, and longstanding political disputes.

References

1 http://www.oecd.org/document/18/0,3343,en_2649_34255_1815186_1_1_1_1,00.html.
2 http://www.privacyrights.org/ar/cloud-computing.htm.
3 OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data
4 EU Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and the free movement of such data

Chapter 4. Examples of Cloud Service Providers

4.1 Amazon Web Services (IaaS)

Amazon Web Services (AWS) provides infrastructure-as-a-service (IaaS) offerings in the cloud for organizations requiring computing power, storage, and other services. According to Amazon, AWS allows you to "take advantage of Amazon.com's global computing infrastructure," which is the heart of Amazon.com's retail business and transactional enterprise. AWS offers a number of infrastructure-related services, including the following:

Elastic Compute Cloud (EC2)

EC2 is a web service that provides resizable compute capacity in the cloud. EC2 allows scalable deployment of applications by providing a web services interface through which customers can create virtual machines (VMs)—that is, server instances—on which the customer can load any software of her choice. A customer can create, launch, and terminate server instances as needed, paying by the hour for active servers.

Simple Storage Service (S3)

S3 provides a web services interface that can be used to store and retrieve unlimited amounts of data, at any time, from anywhere on the Web.

Simple Queue Service (SQS)

SQS is a distributed queue messaging service that supports the programmatic sending of messages via web services applications as a way to communicate over the Internet. The intent of SQS is to provide a scalable hosted message queue that resolves issues arising from the common producer-consumer problem or connectivity between producers and consumers.

CloudFront

CloudFront is a content delivery network that delivers your content using a global network of edge locations. Requests for objects are automatically routed to the nearest edge location, so content is delivered with the best possible performance. CloudFront works with S3 which durably stores the original, definitive versions of files.

SimpleDB

SimpleDB is a web service providing the core database functions of data indexing and querying. This service works in close conjunction with S3 and EC2, collectively providing the ability to store, process, and query data sets in the cloud, making web-scale computing easier and more cost-effective for developers. There are a number of potential use cases to consider when discussing AWS, as outlined in **Figure 20**.

Use case	Use case description	Service(s)
Web/application hosting	Web/application vendors can leverage the AWS infrastructure for computing power and storage as an alternative to internally hosting their applications. This can result in cost savings and efficiencies associated with managing infrastructure and time to market.	EC2, S3, SimpleDB, SQS
Backup and storage	Organizations can leverage AWS as an option for managing internal backup and storage as an alternative to an on-site storage infrastructure. Though storage hardware costs are generally decreasing, the size of productivity and media files is growing, which is resulting in exponential increases in storage needs	S3
Content delivery	Organizations involved in content delivery, such as streaming media, can leverage the AWS worldwide network of edge servers to minimize degradation of delivery and service	CloudFront, S3
Highperformance computing	Organizations that have high-performance computing requirements can leverage AWS computing power on demand to process large amounts of data without having to create an internal infrastructure. This can result in cost savings and efficiencies associated with usage and time to market.	EC2, S3
Media hosting	Organizations that are involved in the distribution and storage of media files can leverage AWS to offset the unpredictable requirements related to storage and processing	EC2, S3, SQS, CloudFront
MapReduce	This is a web service that enables businesses,	EC2, S3

	researchers, data analysts, and developers to process a vast amount of data utilizing a Hadoop framework	
Cloud “bursting”	This is the ability to deal with rapid spikes in processing demands.	EC2

Figure 20

With regard to pricing, AWS is based on consumption as defined by the type of service provided, and the rates are posted online. Additionally, AWS services are based on platform flexibility, allowing the customer to choose the appropriate operating system, programming model, and so forth to meet her needs.

4.2 Google (SaaS, PaaS)

Google App Engine is Google’s platform-as-a-service (PaaS) offering for building and hosting web applications on the Google infrastructure. Currently, the supported programming languages are Python and Java. App Engine is free up to a certain level of used resources, after which fees are charged for additional storage, bandwidth, or CPU cycles required by the application.

Google Apps is Google’s software-as-a-service (SaaS) offering for business email and ollaboration. It features several applications with similar functionality to traditional office suites, including Gmail, Google Calendar, Talk, Docs, and Sites. Additionally, Google Apps has a number of security and compliance products to provide email security and compliance for existing email infrastructures. The Standard Edition is free and offers the same amount of storage as regular Gmail accounts; the Premier version is based on a per-user license model and associated storage level. There are a number of potential use cases to consider when discussing Google services, as noted

in **Figure 21**.

Use case	**Use case description**	**Service(s)**
Messaging	Organizations can leverage Google Apps for internal email and calendar services without the investment and maintenance of a messaging architecture.	Gmail, Google Calendar
Securing existing email systems	Organizations can leverage Google Apps for securing existing email systems by filtering out messaging threats including spam and	Google Email Security

	viruses without the investment and maintenance of hardware and software	
Email retention and legal discovery for existing email systems	Organizations can leverage Google Apps for managing email retention with a searchable archive so that they can locate email quickly in the event of legal discovery without the investment and maintenance of hardware and software.	Google Email Archiving and Discovery
Collaboration	Organizations can leverage Google Apps for office productivity and collaboration without the need to install software on local machines and/or servers.	Google Docs, Google Sites
Application development	Organizations can leverage the Google App Engine platform to develop custom applications based on Java and Python, and the associated services, without investing in internal infrastructure.	App Engine

Figure 21

4.3 Microsoft Azure Services Platform (PaaS)

Azure Services Platform is Microsoft's PaaS offering that is part of the company's strategy of lessening its emphasis on the desktop and shifting more resources to web-based products. It provides an operating system called Windows Azure that serves as a runtime for the applications and provides a set of services that allows development, management, and hosting of managed applications at Microsoft data centers. The platform includes the following services:

.NET Services

A set of developer-oriented services that provide basic pieces required by many cloudbased applications (access control, service bus, workflow, etc.).

SQL Services

A set of services that extend the capabilities of Microsoft SQL Server into the cloud as a web-based, distributed relational database. It provides web services that enable relational queries, search, and data synchronization with mobile users, remote offices, and business partners.

Live Services

A set of services that provide developers the ability to connect their applications to Windows Live users. In addition, Live Services let users log in using Live ID, access and share contacts, feed content into Windows Live, and so on. In regard to pricing, the Azure Services Platform is based on a consumption model including compute time, storage, API calls, and so forth. There are a number of potential use cases to consider when discussing the Azure ServicesPlatform, as noted in **Figure 22**.

Use case	Use case description	Services
Application vendor to offer SaaS version	Organizations can leverage the Azure Services Platform to enhance the functionality of existing applications without investing in internal infrastructure. For example, instead of continuing to leverage the in-house deployment model, the vendor can leverage the Azure Services Platform to develop a SaaS version of the product.	Windows Azure, .NET Services, SQL Services
Application development	Organizations can leverage the Azure Services Platform to develop custom applications based on Windows Azure and associated services without investing in internal infrastructure.	Windows Azure, .NET Services, SQL Services

Figure 22. Azure Services Platform use cases

4.4 Proofpoint (SaaS, IaaS)

Proofpoint provides SaaS and IaaS services in the cloud related to securing the enterprise email infrastructure, with solutions for email security, archiving, encryption, and data loss prevention. Proofpoint's solutions are priced on a per-user, per-year basis, depending on the specific product features deployed. Proofpoint offers a number of SaaS and IaaS services, including the following:

Enterprise

The Enterprise email security and data loss prevention solution provides security for both inbound and outbound email, without the need for on-premises hardware or software. This customizable solution can be deployed with a variety of options including a "Protection" bundle (with antispam, antivirus, email firewall, and email policy enforcement features), a "Privacy" bundle (with data protection features including detection of private identity, health care and financial information detection, preconfigured data protection policies, and incident management), and an "Encryption" bundle that adds policy-based email encryption features.

Shield

The Shield SaaS connection management and frontline spam protection service defends against malicious and spam email connections, reducing inbound spam volumes and preventing denial of service (DoS) and directory harvest attacks.

Archive

Archive is an on-demand email archiving solution that addresses email storage management, legal discovery, and regulatory compliance. Patented encryption technology is used to ensure that messages are secure while being transmitted to Proofpoint's data centers and also while stored in the archive. At the same time, archived messages remain fully searchable by authorized users. The solution makes it possible for enterprises to create and enforce legal holds during e-discovery (i.e., to find all relevant email related to a legal case and to ensure the retention of that data during a lawsuit), and it gives end users easy, self-service access to their historical email.

There are a number of potential use cases to consider when discussing Proofpoint services, as noted in **Figure 23.**

Use case	Use case description	Service
Inbound email security	Organizations can leverage Proofpoint email security solutions to block spam, viruses and other malware, phishing attacks, and inappropriate content in incoming email messages, and enforce basic corporate email policies for outgoing email.	Enterprise
Data loss prevention	Organizations can protect confidential information from inappropriate distribution via email. Outgoing email and attachments are scanned for confidential information and blocked from transmission if they are found to contain such content. Blocked messages can be quarantined for review by security or compliance personnel.	Enterprise
Compliance with data protection regulations	Organizations can ensure that private information such as customer financial data and personal health care information is protected against inappropriate exposure. Outgoing email and attachments are scanned for the presence of protected financial, health care, or identity data and then automatically encrypted or blocked as appropriate.	Enterprise
Email archiving	Organizations can enforce corporate policies for email retention, ease email storage burdens on local servers, and enable rapid searching of historical email for e-discovery (including the ability to enforce legal holds) and give end users essentially an "unlimited" inbox.	Archive

Figure 23. Proofpoint use cases

4.5 RightScale (IaaS)

RightScale provides IaaS-related services in the cloud to assist organizations in managing cloud deployments offered by other CSPs, including vendors such as AWS, FlexiScale, and GoGrid. The RightScale Cloud Management Platform allows organizations to manage and maintain their cloud deployments through one web-based management platform, while at the same time taking advantage of offerings by more than one CSP. RightScale's pricing is based on a number of editions from Developer through Enterprise level, and associated features and server times.

N O T E

Server usage and other charges from cloud infrastructure providers, such as AWS, are billed separately by the cloud provider and are not included in monthly RightScale usage fees.

The RightScale Cloud Management Platform includes the following:

Cloud Management Environment

The Cloud Management Environment provides control, administration, and life cycle support for cloud deployments via a dashboard for real-time management of deployments across one or more clouds, including public and private clouds. The dashboard provides transparent access to and control over all aspects of cloud deployment, including the ServerTemplates, underlying scripts, input parameters, real-time monitoring, and automatic or manual response.

Cloud Ready ServerTemplates

ServerTemplates and the Best Practice Deployment Library help to simplify deployment management. ServerTemplates, developed by RightScale, incorporate standard cloud configurations for common application deployment components such as scalable web and application servers, database master/slave pairs, and grids for batch processing. Partner ServerTemplates, developed by RightScale partners, help incorporate RightScale's partners' applications, tools, and components into deployments. Customer ServerTemplates can be cloned, customized for specific needs, and then saved in a custom library. Over time, an organization will build a repository of ServerTemplates representing valuable corporate knowledge for the organization.

Adaptable Automation Engine

The Adaptable Automation Engine executes and manages deployments that adapt to situations as required by system demand, system failure, or other specified events. As demand changes, servers can be added or decommissioned. As components fail, existing servers can adopt their roles or the system can deploy new servers. As queues fill or empty, grids can expand or contract automatically. Active monitoring, alerts, and escalations ensure real-time adaptation based on the rules and automatic responses defined by the organization.

Multi-cloud Engine

The Multi-cloud Engine interacts with cloud infrastructure application programming interfaces (APIs) and manages the unique aspects of each cloud. As a result, organizations are not locked into any one cloud; instead, they are free to choose among several cloud providers, deploy across multiple clouds, or move an application from one cloud to another.

There are a number of potential use cases to consider when discussing RightScale services, as

noted in **Figure 24.**

Use case	Use case description	Service
Complexity in managing the cloud infrastructure	Organizations can leverage RightScale as an option for managing the complexities involved in deploying and managing a CSP's infrastructure services. This can result in efficiencies associated with managing services and time to market, as the organization can focus on core strengths rather than learning how to deploy within the CSP's environment.	Cloud Management Platform
Single management platform	Organizations can leverage RightScale as an option for managing and maintaining cloud deployments through one management platform. This can result in efficiencies and costs savings related to personnel costs as the organization can more effectively and efficiently address head count related to managing cloud deployments.	Cloud Management Platform
Portability	Organizations can leverage RightScale to manage cloud infrastructure APIs and the unique aspects of each cloud so that they can freely choose among a variety of CSP offerings based on their unique needs, manage and migrate deployments across these clouds (public or private), and avoid vendor lock-in.	Cloud Management Platform

Figure 24. RightScale use cases

4.6 Salesforce.com (SaaS, PaaS)

Salesforce.com is a provider of SaaS-based CRM products, as well as having a PaaS offering, Force.com. Salesforce.com's CRM solution is divided into several applications including Sales, Marketing, Service, and Partners. Pricing is on a per-user basis, and the rates and different support packages are posted online.

Salesforce.com has more recently begun to provide PaaS-based services through the Force.com platform. Force.com allows external developers to create add-on applications that integrate into the main Salesforce.com applications, and are hosted on Salesforce.com's infrastructure. Applications are built using Apex, a proprietary programming language for the Force.com platform. Pricing is on a per-developer basis, and different support packages allow for varied levels of storage, API calls, and so forth. AppExchange is a directory of applications built for Salesforce.com by third-party developers which users can purchase and add to their Salesforce environments. As of May 2009, approximately 800 applications are available from more than 450 independent software vendors (ISVs) via AppExchange.

There are a number of potential use cases to consider when discussing Salesforce.com services, as noted in **Figure 25**.

Use case	Use case description	Service(s)
On-demand CRM	Organizations can leverage Salesforce.com CRM applications to centralize, manage, and efficiently share prospective client information without investing in internal infrastructure.	CRM
Extend functionality of Salesforce.com CRM	Organizations can leverage Force.com to develop add-on applications that extend the functionality of the Salesforce.com CRM or leverage the existing directory of applications within AppExchange without investing in internal infrastructure.	Force.com, AppExchange
Application development	Organizations can leverage the Force.com platform to develop custom applications based on the Force.com platform without investing in internal infrastructure.	Force.com

Figure 25. Salesforce.com use cases

4.7 Sun Open Cloud Platform

As the company that coined the phrase “The Network is the Computer,” Sun Microsystems envisions a world of many clouds, both public and private, that are open and compatible. Sun takes an inclusive view that there are many different types of clouds, and many different applications that can be built using them. To that end, according to Sun, it plans to offer an extensive portfolio of products (hardware and software) and services under the umbrella of an “Open Cloud Platform” to foster open communities and partner ecosystems.

According to Sun, the Open Cloud Platform is an open architecture (APIs, open format) and infrastructure encompassing technologies such as Java, MySQL, OpenSolaris, and Open Storage software. Sun believes that its Open Cloud Platform offering will foster an ecosystem of partners, developers, and others, because cloud computing can be successful only if you can leverage maximum reuse of others’ technologies and components. According to Sun, the Open Cloud Platform will offer the necessary cloud service ingredients (hardware, software, and management capabilities) to help customers and partners wishing to become CSPs for any of the cloud delivery models—SaaS, PaaS, or IaaS (SPI). Sun is working with service providers and enterprises to build their own clouds to service their respective customers and users.

One of the things driving cloud computing is the wide availability of open source software and components; developers can rapidly assemble applications out of open source components and run them in the cloud. According to Sun, it has developed foundational technologies (software and hardware) to enable the three emerging cloud business models: public clouds, private clouds, and hybrid clouds.

Sun’s foundation technologies include OpenSolaris, MySQL, the open source GlassFish application server, Crossbow (a network virtualization technology and an OpenSolaris component), the Sun xVM hypervisor (based on the open source Xen), the Solaris Zetta File System (ZFS), the Sun xVM VirtualBox, and NetBeans (an IDE for developers). Sun’s hardware portfolio encompasses an array of servers based on X86, SPARC, and energy-efficient chipmultithreaded (CMT) UltraSPARC processors and Open Storage with a range of densities and I/O capacities.

Sun's implementation of a public cloud, initially targeting the developer community, is an IaaS, with a public compute and storage infrastructure service (future delivery). Developers will access the Sun public cloud services from a web browser to provision resources on their platform of choice—Linux, Windows, or OpenSolaris operating systems. For its initial offering, Sun plans to support a RESTful API for creating and managing cloud resources, including compute, storage, and networking components. Sun will also provide client libraries for Java, Ruby, and Python development. Sun's X86 virtual box supports the Open Virtualization Format (OVF), which makes VMs portable across clouds that support the OVF open standard.

Sun's cloud offerings also include Project Kenai (beta):

Project Kenai host[s] projects and code to be deployed on [the] Sun Cloud, [and] facilitates collaboration with like-minded developers to access or initiate projects directly from the NetBeans. Project Kenai also has [a] repository of APIs for the Sun Cloud service. These APIs are posted for review (at http://www.kenai.com) and comment using the Creative Commons license.

There are a number of potential use cases to consider when discussing the Sun cloud platform,

as noted in **Figure 26**.

Use case	Use case description	Service(s)
High-performance computing and elasticity	Organizations that have high-performance computing requirements can leverage the Sun Open Cloud Platform to process large amounts of data.	Sun Open Cloud Platform, private cloud services
Development and testing	Organizations, start-ups, social network developers, and enterprises trying to experiment with disruptive ideas or wanting to experiment with hosting applications in the public cloud can leverage the Sun public cloud services and open source components to develop applications in Java, Ruby, Python, and MySQL.	Sun compute and storage cloud, Project Kenai, NetBeans, VirtualBox
Surge computing	Organizations can offload an overburdened IT infrastructure (temporarily or permanently) to accommodate peak loads, batch processing jobs, or anticipated spikes in demand for services.	Sun Open Cloud Platform, compute and storage cloud, Project Kenai, NetBeans, VirtualBox

Figure 26. Sun use cases

4.8 Workday (SaaS)

Workday is a provider of SaaS-based human resources and financial management products. Workday pricing is on a per-user basis and functionality. Workday's solutions are divided into several modules, including the following:

Human Capital Management

Workday's HR and Human Capital Management software is designed to help companies organize, staff, pay, and develop the global workforce.

Payroll

Workday Payroll allows companies to group employees, manage payroll calculation rules, and pay employees according to organizational, policy, and reporting needs.

Worker Spend Management

Workday Worker Spend Management combines Workday Expense, Procurement, and Business Resource Management capabilities into one solution that extends Workday Human Capital Management and helps companies understand and manage total workforce cost—spend on, by, and for workers.

Financial Management

Workday Financial Management offers a financial services solution to address internal and external requirements by combining support for business and HR accounting transactions, a framework for internal control and audit, and robust financial reporting and businessperformance management.

Benefits Network

Workday Benefits Network provides HR organizations with a catalog of prebuilt ntegrations that connect to benefits providers, giving HR management organizations the ability to evaluate, select, and offer the most appropriate plans for their workforce.

Table 9-8 notes a use case to consider when discussing Workday services.

Use case	Use case description	Services
On-demand HR	Organizations can manage various aspects of HR and financial management processes without investing in internal infrastructure.	HR and Financial Management modules

Figure 27. Workday use case

4.9 Summary

As we have discussed, the cloud computing market is becoming increasingly crowded each day. Amazon, Google, Microsoft, Salesforce.com, and Sun are considered some of the key players in the cloud computing market, but they represent only a handful of the providers in this space. **Figure 28** summarizes their respective service offerings and focus areas

CSP	Offering	Focus area
Amazon	Core offerings include the AWS infrastructure related to servers, storage and bandwidth, databases, and messaging for interfaces. Differentiators vis-à-vis competitors: • Supports varied operating systems/programming languages • Content Delivery Network	SMB focus
Google	For creating and running web applications. Supports only Python and Java (does not support Microsoft and others). Also provides SaaS-related productivity applications	SMB focus
Microsoft	Operating system with a set of developer services. Allows the building of new cloud applications and the enhancement of existing applications for the cloud. Platform for Microsoft application development	Enterprise/SMB Focus
Proofpoint	Core offerings include on-demand services related to email security and archiving.	Enterprise/SMB Focus
RightScale	Core offering is a Cloud Management Platform for managing the cloud infrastructure from multiple vendors. Differentiators vis-à-	SMB focus

	vis competitors: • Transparent access and control over multiple cloud offerings to best meet organizational needs • Portability	
Salesforce.com	Allows the building and integration of business and CRM applications within the Salesforce.com infrastructure. Supports only the Apex proprietary programming language	Enterprise focus
Sun	Core offerings include infrastructure related to servers, storage, and databases. Differentiators vis-à-vis competitors: • Supports varied operating systems/programming languages • Open cloud concept to support other CSPs (public, private, and hybrid clouds) • Virtual data center capabilities	Enterprise/SMB focus
Workday	Core offerings include on-demand services related to HR.	Enterprise/SMB Focus

Figure 28

The question we haven't asked yet is: what is the current focus group for CSPs? Are their customers at the enterprise level or at the small and medium-size business (SMB) level? It would appear that the cloud is already a viable and sensible solution for many SMBs. However, as we have discussed in this book, there are still a number of questions around security, availability, and so on. Will these issues need to be overcome before cloud computing takes off or enterprise-level organizations? We will address these issues and others in the concluding chapter. However, security-as-a-service is security offered as a service and delivered in the cloud, as opposed to the security of CSP offerings discussed in the rest of the book. Additionally, in Chapter 11 we will look at the impact if cloud computing on the role of traditional corporate IT. How is corporate IT being affected by cloud computing, and what is the relationship between the two.

Chapter 5. Security as a cloud

So far, we have adressed the security provided by Cloud Service Providers (CSPS) as well as the security provided by customers using cloud services. In this chapter, the focus is on security provided as cloud services; that is, security delivered through the cloud, also known as security-as-a-service.

Just like software-as-a-service (SaaS), the business model with security-as-a-service is subscription-based. In addition, security-as-a-service is also sometimes referred to as "SaaS," which is how we will address it specifically in this chapter.

With SaaS, there are two emerging provider types. The first type comprises established information security vendors who are changing their delivery methods to include services delivered through the cloud. The second type comprises start-up information security companies that are also emerging in this field as pure, play CSPs—that is, these companies provide security only as a cloud service, and do not provide traditional client/server security products for networks, hosts, and/or applications.

Among established information security companies that are changing their business models to also include SaaS, the most prominent are traditional anti-malware vendors. However, other established information security companies are also involved in the delivery of SaaS, especially with regard to email filtering.

Origins

Three points of impetus help to explain how security-as-a-[cloud] service began. The earliest impetus is a decade old now: spam, or unsolicited email. As early as 1999, companies (such as Postini[1] were offering email services as follows:

Postini was founded with the idea that email should be better. While email is the most popular Internet resource, service providers and software developers aren't making email better, and worst of all, aggressive marketers are targeting any email user as a potential customer. Postini services are designed to extend the capability of your service providers' email offering. Junk mail services are only the first step. Over the coming months, you will see more Postini services that make email even better[2].

A number of other companies now provide email filtering services, both standalone security companies, as well as many Internet service providers (ISPs) which are often reselling the services of standalone security companies with their own brand.

A second impetus for SaaS is managed security services (MSSs). Managed security service providers (MSSPs) have been providing outsourced services to customers for several years, whereby the MSSPs manage an organization's network security devices, such as firewalls and intrusion detection systems (IDSs). The impetus for using MSS was, and is, the same as cloud computing: lower costs compared to in-house solutions through shared resources. The difference between MSSPs and CSPs, however, is that the shared resources for MSSPs are personnel, and not infrastructure. Additionally, because many organizations are not staffed to handle round-the-clock support for such services and do not have the expertise to fully staff such positions, the shared services (i.e., personnel) model of MSSPs can be financially attractive. The MSSP model became an impetus for CSPs because it broke the strong but informal barrier to outsourcing parts of an organization's information security program. And in this case, that outsourcing also meant off-premises management of information security devices. (Although outsourcing information security is often an option, initially it tended to be outsourcing on-premises—that is, within the customer's own facilities—as opposed to off premises. Of course, now outsourcing can be on-premises, off-premises, onshore, offshore, and other variations in delivery.)

Although this network security work is outsourced in this model, the responsibility for a customer's security remains with the customer. It is the customer who is responsible for managing and monitoring the MSSP, and the customer dictates what security policies are to be enforced. The MSSP monitors and manages devices (e.g., firewalls, IDSs) and data flows (e.g., Web, content, or email filtering). But these devices (including the devices that manage and monitor data flows) belong to the customer. As a result, cost savings and efficiency improvements go only so far. Although this is a subscription-type service (an operational

expense, or OpEx), there is still the associated capital expense (CapEx) of the customer's onpremises hardware. With cloud computing, CapEx is further reduced because most of the devices and the monitoring and management are the responsibilities of the SaaS provider.

A third point of impetus for SaaS is the declining organizational efficiency of trying to provide security on the endpoints directly. Not only is there is a huge proliferation of endpoints, but they have so many configuration variables that organizational IT departments simply cannot manage them effectively. Additionally, because many of these endpoints are mobile, trying to troubleshoot configuration problems and keep security software up-to-date is a huge task. Add to those problems the fact that many mobile devices lack sufficient resources (e.g., processing power, memory, and storage capacity) to adequately handle today's endpoint protection suites and the endpoint protection situation is not looking positive.

Because of these issues, and the explosive growth in malware, protecting endpoints on the endpoints is an increasing problem. For example, "In 2008, Symantec detected 1,656,227 malicious code threats.... This represents over 60 percent of the approximately 2.6 million malicious code threats that Symantec has detected over time[3]." This has led to a change in thinking about how to protect those endpoints. Instead of protecting endpoints on the endpoints, why not protect them through the cloud? That is, why not clean the traffic to and from the endpoints as it transits the cloud? Instead of dealing with all the complications of trying to monitor and manage the endpoints themselves, move the monitoring and management of traffic to and from the endpoint (not the monitoring and management of devices themselves) to the cloud.

This concept of moving anti-malware protection to the cloud, instead of being endpointresident, gained considerable traction with the presentation of a paper at the July 2008 USENIX Conference in San Jose, California. That paper, titled "CloudAV: N-Version Antivirus in the Network Cloud" and available at[4], showed that cloud-based antivirus (i.e., anti-malware) provides 35% better detection against recent threats than endpoint-based single engines, and an overall detection rate of 98%. That overall detection rate is significantly better than the results of a single engine running on an endpoint. (Endpoints are generally limited to running only a single anti-malware engine at a time because of constraints on endpoint resources, as well as incompatibilities of running multiple engines.)

Today's Offerings

Today's offerings in the SaaS segment involve several services to improve information security: email filtering (including backup, archival, and e-discovery[5]); web content filtering; vulnerability management; and identity-as-a-service (spelled in this chapter as IdaaS).

Email Filtering

SaaS for email primarily involves cleansing spam, phishing emails, and malware included in email from an organization's incoming email stream, and then delivering that clean email securely to the organization so that it is effectively not repolluted. The touted benefits of this approach are not only more comprehensive security for clients due to the use of multiple engines, but also better performance of those client devices (because the anti-malware runs in the cloud and not on the endpoint directly), as well as far better anti-malware management.

The anti-malware management is superior to endpoint solutions because that anti-malware is OS- and processor-agnostic, so it can be managed centrally through the cloud rather than working with multiple management systems, probably from multiple anti-malware vendors.

This cleansing-in-the-cloud service has corollary benefits: reduced bandwidth used by email, reduced loads on organizational email servers, and improved effectiveness of a (recipient) organization's own anti-malware efforts.

Although most attention on SaaS involving email tends to focus on inbound email, it is also often used with outgoing email. Many organizations want to ensure that they are not inadvertently sending malware-infected emails, and cleansing outbound email through SaaS is a good method for preventing such problems and embarrassments. Additionally, outside SaaS email can be used to enforce organizational policies around the encryption of email (e.g., between specified [email] domains, such as those belonging to business partners or customers).

This email encryption is generally performed at the (email) server-to-server level so that individual user actions and key management are not required. This is accomplished by

using either Secure Sockets Layer (SSL) or Transport Layer Security (TLS) on network communications at the transport layer.

A further benefit of SaaS anti-malware is the collective intelligence that is gained from the visibility of all malware threats to all endpoints across an enterprise, irrespective of type (e.g., server, desktop, laptop, or mobile device), location, OS, or processor architecture. Having this greater view in a timely manner is a significant help to organizational information security teams.

SaaS for email also includes email backup and archiving. This service usually involves storing and indexing an organization's email messages and attachments in a centralized repository. That centralized repository allows an organization to index and search by a number of parameters, including date range, recipient, sender, subject, and content. These capabilities are particularly useful for e-discovery purposes, which can be extremely expensive without such capabilities.

Web Content Filtering

As endpoints belonging to an organization—whether they are within an organization's facilities, at home, or on the road—try to retrieve web traffic, that traffic is diverted to a SaaS provider that scans for malware threats and ensures that only clean traffic is delivered to end users. Organizations can also enforce their web content policies by allowing, blocking, or throttling traffic (use of bandwidth for that traffic reduced). Because of the number of websites accessible today, earlier URL filtering solutions deployed on organizations' premises are increasingly inefficient. SaaS providers supplement that URL filtering with the examination of Hypertext Transfer Protocol (HTTP) header information, page content, and embedded links to better understand site content. Additionally, these services use a collective reputation scoring system to bolster the accuracy of this filtering.

SaaS for web content also involves scanning outbound web traffic for sensitive information (e.g., ID numbers, credit card information, intellectual property) that users could send externally without appropriate authorization (data leakage protection). Web traffic is also scanned for content analysis, file type, and pattern matching to prevent data exfiltration.

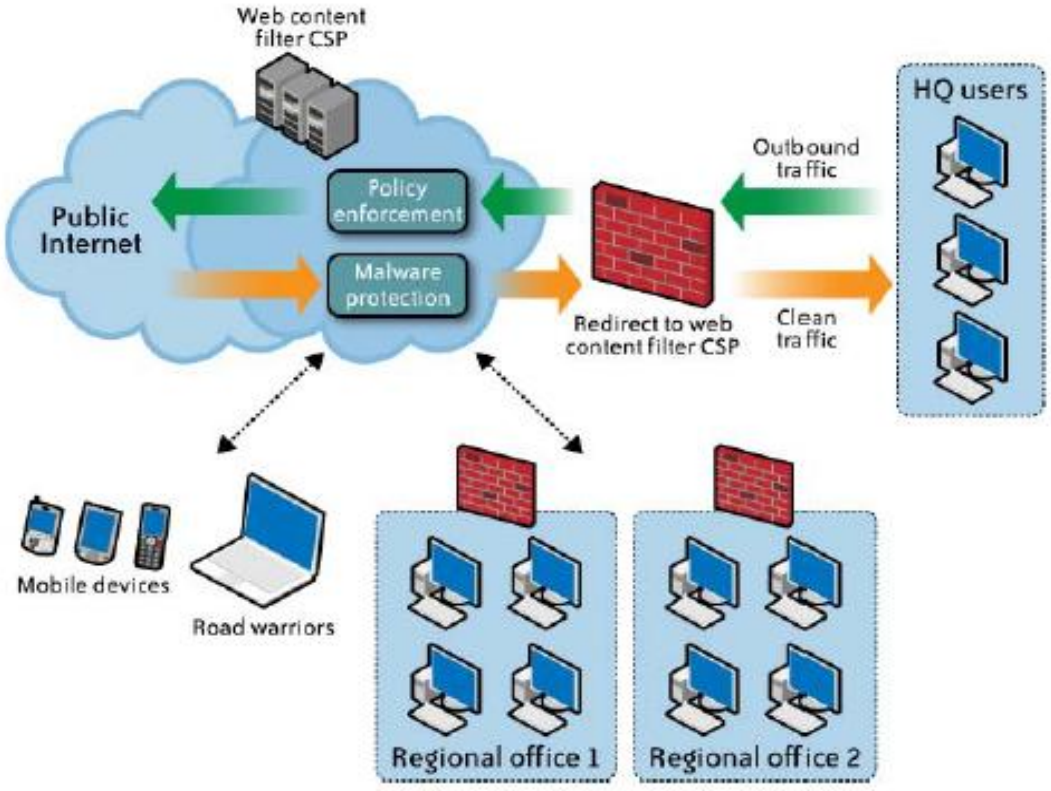


Figure 29

Vulnerability Management

As the Internet-facing presence of organizations has grown in size and complexity, as well as in importance to their operations, ensuring the secure configuration and operation of the systems involved has become more difficult and more important. There are SaaS providers that discover, prioritize, and assess systems for vulnerabilities, and then report and remediate those vulnerabilities and verify the systems' secure operation. Such information is also used to monitor for and report on compliance with some regulatory requirements (e.g., the Payment Card Industry's Data Security Standard).

Identity Management-As-a-Service

Identity management-as-a-service (IDaaS) only recently emerged as an example of SaaS, in comparison to email filtering, web content filtering, and vulnerability management, which are more established as SaaS offerings. There are some significant deficiencies in today's identity and access management (IAM) capabilities with regard to uses in cloud computing (e.g., scalability). IDaaS attempts to provide some IAM services in the cloud. Today's relatively early IDaaS offering tends to focus on authentication, because this is the most critical problem for customers; However, the most significant problem for CSPs concerns IDaaS providers, and developing some form of collaborative meta system. (Just as meta directories did not scale within organizations, virtual directories will not scale to a cloud level.) IDaaS providers will also need to provide other IAM services for cloud customers, including authorization (groups and roles at a minimum), provisioning, and auditing.

References

1 Postini was acquired by Google in September 2007. http://googleblog.blogspot.com/2007/09/weve-officially-acquired-postini.html.

2 According to the Internet Archive's Wayback Machine for Postini.com, May 10, 2000. MessageLabs, another email filtering company, founded in 1999 and acquired by Symantec in May 2008, even used to refer to itself as an application service provider—an earlier evolution of cloud computing

3 Symantec's Global Internet Security Threat Report: "Trends for 2008," Volume XIV, published April 2009, p. 10.

4 http://www.eecs.umich.edu/fjgroup/pubs/cloudav-usenix08.pdf

5 E-discovery refers to discovery in civil litigation of information in an electronic format. Because so much of an organization's information is transmitted via email, email becomes an obvious best place to start an e-discovery program, though e-discovery is not (or should not be) limited to email.

CHAPTER 6. The Impact of Cloud Computing on the Role of Corporate IT

Cloud Computing has the potential to be the next disruptive technology with consequence of significant change. Depending on the perspective and situation of the organization or the individual, this represents both opportunity and crisis. Such change may be resisted, even if it is a good idea and it works. The role of the corporate IT department will be impacted significantly by a company's adoption of cloud computing. The degree of change will relate to the current approach to IT governance and management, and to the level and speed of adoption. What are the driving forces and resisting forces that will drive adoption, and which is stronger? In this simplistic analysis, the driving forces need to outweigh the resistance to make individuals and organizations adopt the cloud as the enabling technology platform of the next decade or so. Understanding the driving forces to maximize benefit and the way resistance is managed will impact the speed of adoption and the ultimate role that corporate IT professionals perform. In this chapter we discuss:

• Driving forces for adopting the cloud

• Resisting forces to maintain the status quo

• How cloud computing will affect the role of IT

6.1 Why Cloud Computing Will Be Popular with Business Units

There are a number of reasons why business units will see cloud computing as an alternative way of using information technology. These reasons may well change the role of IT in the near future, and some of IT's traditional service delivery models and organizational structures will need to be changed to accommodate the power of computing that can be easily deployed through cloud computing. Some of the reasons include the following:

• Cloud computing is a low-cost solution.

• Cloud computing offers responsiveness and flexibility.

• The IT expense matches the transaction volumes.

• Business users are in direct control of technology decisions.

• The line between home computing applications and enterprise applications will blur. We explore each of these in turn.

6.1.1 Low-Cost Solution

First and foremost, cloud technologies have to be cost-effective in terms of total cost, and they must improve the ratio between maintenance cost and discretionary spending on value-added projects. Most of the annual budgets in the vast majority of IT departments today are consumed by maintenance and depreciation—providing no new value add. This balance between maintenance and depreciation versus new value add is critical; there is no benefit in reducing infrastructure costs and paying more for application development through increased cost of integration. This leads to the importance of taking a holistic view of the true costs of IT: including integration cost, reporting cost, disaster recovery planning, IT staff costs, and the cost of swapping out a poorly performing cloud service provider (CSP). Compelling cost benefit equations will drive the adoption; disastrous contractual relationships will slow the adoption for individuals and companies that are impacted. The potential that cloud computing has for economies of scale and innovation will provide a strong cost-effectiveness driver. The low cost of technology offered by CSPs will encourage business units to go directly to CSPs rather than using traditional IT departments (because doing the latter would cost more and most likely it would take longer to adopt such changes). This may have a profound impact on the role of IT. IT may no longer be seen as an implementation provider, but rather as a risk advisor and guidance provider. This would require different IT skill sets and a new IT structure to support business units. Instead of silos of groups or liaison roles, IT needs to be more closely embedded within business units and be seen as part of the business unit instead of aligned to IT. Core IT would provide training and mentorship for these IT resources and edducate them on the compliance requirements as they adopt the new technologies.

6.1.2 Responsiveness / Flexibility

Provided that IT meets availability and reliability goals within acceptable costs, perhaps a more critical driver than cost is responsiveness and flexibility. A new company is acquired, a new product is launched, a layoff happens, and the sales force is reorganized. All of these events can occur within short notice, command resources, and can consume an entire IT budget. Technology support for process improvement becomes a lower priority than keeping the lights on and responding to a “must-do” project. In the acquisition scenario, it may well be that your payroll provider is more experienced at combining two companies’ payrolls than an in-house IT department would be. It may be that the email provider is also well versed in taking on such a consolidation project and can scale the infrastructure with ease. The CSP’s experience may lead to a cost-effective solution, and be more responsive than that of traditional IT departments. The argument for CSP responsiveness can be seen by what is in many ways a virtual IT organization. This “delegation” to a CSP gives IT management the bandwidth to deal with the hard stuff, such as people, combining processes, product lines, and getting value out of the acquisition. This argument also applies to the IT business-as-usual operation. If most of IT’s funds are spent on routine or commodity functions, performance of critical business processes will degenerate. If users do not have the ability to adapt or take advantage of evolving technology, satisfaction with IT will also decline. Responsiveness may also be enhanced by a broad customer base for an application contributing new solutions to run on the platform and extend software-as-aservice (SaaS) functionality. This allows a business group to adapt to new regulations, respond to new requirements, and find a better way of doing things. Do all companies then end up with the

same systems and processes? Maybe some of that happens, but the way they combine and apply components leads to tremendous variety in operational approach. The bigger risk is getting left behind by not taking advantage of the responsiveness, flexibility, and adaptability that a combination of CSPs can offer.

6.1.3 IT Expense Matches Transaction Volume

A company may have a critical and urgent requirement to mobilize a new sales force, in a new region. The requirements include having a high degree of visibility into the sales pipeline to manage sales execution and the quality of the demand signal. Other investments will be made based on this data. On top of the responsiveness argument is the impact on cash flow, one of the most important business metrics. A company can buy as many Salesforce.com applications as it needs to support a hopefully growing but possibly dwindling sales force. In another example, a company performs Sarbanes-Oxley (SOX) Section 404 IT control testing once a year. Renting an IT Governance, Risk, and Compliance (GRC) application for that period may reduce costs, and the company may take advantage of increasing levels of automation and adaptation to new regulations through an evolving SaaS provider Matching investment to revenue using a SaaS model is an attractive proposition. Some enterprise resource planning (ERP) vendors are using this model to price on-premises software, validating the strength of this approach but also offering an alternative scenario to achieve this benefit. That's a good thing from a cloud adoption perspective, as it increases competition in responding to this business driver.

6.1.4 Business Users Are in Direct Control of Technology Decisions

In the future, business users will be able to purchase services from a service catalog, and they will be in control of the services they use. In this scenario, there may be little to zero touch from the IT department in transacting services, and costs may be directly billed to the business user for transactions and services consumed. In that scenario, business users would have an incentive to discontinue obsolete functions, so accountability and alignment of IT costs will be improved. Will business users really make those kinds of decisions? Should they be allowed to? What are the implications of this? Answers will vary across the various industries and the relative maturity of the CSP customer.

6.1.5 The Line Between Home Computing Applications and Enterprise Applications Will Blur

In many scenarios, knowledge worker tools delivered by the Internet are used more efficiently to run the home than to run the workplace. We tend to collaborate more effectively with our friends at home than we do with our colleagues at work. It's a hassle to key in my telephone numbers when I change phones, so I store them in the cloud. My smart phone gives me stock quotes, and allows me to make trades from wherever I am. In my private life, I assess the value of the application and invest time in assimilating it into my way of operating. If it does not work or I don't like it, I don't use it or I find an alternative. In this way, adopting tools in your personal life on a self-selecting basis from CSPs can educate and raise expectations. The personal productivity of the knowledge worker in business life and in private life feeds off each other and adds another push to adoption. The boundary between personal and work life merges.

6.2 Potential Threats of Using CSPs

A number of threats from CSPs may promote the existing role of the IT function and dissuade businesses from using CSPs.

Vested Interest of Cloud Providers

CSPs have made considerable investments in data centers and infrastructure. The cost of capturing the customer has been expended and needs to be recouped. The price for initial service may have been low. The business model relies on a continual and expanding revenue stream from each customer. The big CSPs become bigger as cloud services grow. They "partner" with their clients, but no one customer holds very much sway. In some cases, the CSP, often out of necessity, uses proprietary technology. This can be a significant risk that may exist if a CSP goes bankrupt, or starts to raise prices to compensate for loss of revenue, or is unresponsive to business needs. In some cases, the customer may be locked in, and exiting from a CSP may prove costly.

Loss of Control Over the Use of Technologies

Loss of control may be a reality if competitive forces are not maintained during the entire life of a cloud service, and it is costly to switch CSPs. Assume that cost-effectiveness and responsiveness can be maintained. Should a company outsource its critical IT function to a third party or many third parties? It may not be clear whether one or many are optimal. However, as customers rely more on the CSP, they may have less control over their use of technology. The IT function will most likely resist this change, as it has a direct impact on its function.

Perceived High Risk of Using Cloud Computing

Through cloud computing's association with the Internet, and the fact that it is a new service, there is a perception that cloud computing has significant risks and challenges. A central question that drives uncertainty toward the adoption of cloud computing concerns where the data is being processed or stored at any given time. Any replacement of in-house services with CSP-based services may add measurably to this risk.

Portability and Lock-in to Proprietary Systems for CSPs

The deployment of cloud services offers the possibility of spending less money on routine IT operations, and more in adding value to the business. A prerequisite for this is realized cost improvement in data center operations. This requires migration costs from a data center to a CSP to be sufficiently low. It also relies on lower upfront costs to establish a new customer on the platform. This will enable contract time frames to be shorter and be cost-effective for both parties. Lock-in concerns will need to be removed by better standards to migrate data and allow multiple service providers to cooperate in meeting customer needs. If this flexibility does not exist as hardware and storage becomes cheaper, this will not be passed on to the consumer; new customers will get better rates, fueling dissatisfaction. Supplier dominance in the market would inhibit widespread growth. This concern is explored in the Open Cloud Manifesto[1].

Lack of Integration and Componentization

Prior to ERP, the scope of package solutions comprised individual applications, such as finance, payroll, or manufacturing. The packages sat alongside custom applications and most of the data exchange was through custom programs. The level of automation was low and the level of integration was low. ERPs sought to improve integration, and they became a central component of the majority of IT strategies. One driving force for ERP vendors was to extend their software footprint, and effectively lock in a customer to that vendor. Coexistence of ERP implementation strategies came later. This was partly driven by customer demand for new functionality and the reaction time of ERP vendors, and by mixed ERP environments that developed through user companies' acquisitions including a different ERP platform. ERP vendors are now responding with more open, non-proprietary architectures. That also makes it possible for ERP vendors who acquire niche software providers to integrate that code as is, without rewriting their application and leaving existing customers high and dry. So, tools and standards that enable integration, componentization of applications, and service-oriented architecture have led to marked shifts in ERP positioning. These position changes by ERP vendors are examples of the way customer demand and expediency drive changes. There are parallels with the adoption of cloud computing.

ERP Vendors Offer SaaS

The major ERP vendors now have SaaS offerings. Such expanded offerings in the past have been conceived as both an offensive and a defensive measure. Emphasis on the middle market in the past five years was driven partly to keep Microsoft from establishing an enterprise ERP presence as well as to gain additional revenue. Customers can assume that both dynamics are operating in the SaaS scenario, and as base ERPs become more componentized and sophisticated this will help with interoperability issues.

6.3 A Case Study Illustrating Potential Changes in the IT Profession Caused by Cloud Computing

To illustrate a company's adoption of cloud computing and its impact on IT, we have painted a picture of a fictional cloud-enabled company called Nimbus Systems, a small to medium-size business (SMB). Here are some factors pertaining to the company:

• Core ERP is hosted in a private cloud in a data center in Colorado. Functions are limited to finance, reporting, master data maintenance, budgeting, and planning.

• There is no single order management solution. Customer interaction is highly customized and demand comes through in a variety of modes, but uses a standard format.

• Direct procurement uses strategic sourcing agreements, online auctions, and links with the supply chain. This is a hosted B2B application with a combination of public and private clouds. Vendor records and product catalogs are stored in this application.

• Manufacturing is outsourced to 23 companies worldwide.

• Distribution is outsourced to two global carriers.

• A hosted supply chain solution orchestrates the “virtual supply chain,” and is managed by the planning department in the Bahamas.

• Three alternative sales force automation systems are available on a SaaS model. Customer data is stored in the ERP.

• Indirect purchasing and travel management are cloud applications.

• Netbook computers are the default workstations, with all “desktop” applications and storage in the cloud.

• Collaboration tools are available on a pay-per-use basis as a service paid directly by the business consumer.

• R&D has retained its own research machines, but product development is in a cloud application that brings all parties the design, source, and price, and plans product launch information.

• Each product shipped is tailored to a specific customer need. A custom variant configuration application has been developed and the rules are maintained in-house.

• All documents, emails, and voice mails are archived and managed by a specialist CSP.

• HR, payroll, and benefits management are outsourced.

• The help desk is maintained in-house, but all second- and third-level support is routed tothe appropriate third-party provider.

What is interesting about this scenario is that the individual components already exist in production at different companies. The Nimbus Systems IT department does not operate or house the infrastructure; building of applications is limited to the specialized customer-facing application, and support and maintenance functions are provided by the CSP. The Nimbus Systems IT organization now spends fewer resources on building and maintaining commoditized functions and more on the differentiated functions. The skills have become more oriented around architecture, procurement, accreditation, a common vocabulary, and inspection and monitoring. In many cases, these are new skills in the IT department that need to be acquired. For IT services to support the business goals of Nimbus Systems there is a more sophisticated need for IT governance and management.

In the Nimbus Systems model, someone from the company needs to take responsibility for critical functions such as:

• Developing the IT strategy, and in particular, any shared investments to support business goals • Defining the architecture and standards

• Adding new suppliers and services

• Negotiating with individual suppliers

• Maintaining the service catalog

• Integrating services to form an end-to-end process

• Monitoring data integrity, security, and privacy

• Conducting disaster recovery planning

• Monitoring IT costs and alignment to delivered value

• Conducting IT supplier management and contingency planning if a supplier fails or is unresponsive

A critical component will be an approach to the governance of IT to create an environment for services to be delivered effectively. ISO/IEC 38500:2008[2], corporate governance of information technology standard, provides a framework for effective governance of IT to assist those at the highest level in an organization to understand and fulfill their legal, regulatory, and ethical obligations in respect of the organization's use of IT. This standard provides guiding principles for directors of organizations on the effective, efficient, and acceptable use of IT within their organizations. It is organized into three prime sections: Scope, Framework, and Guidance[3].

The framework comprises definitions, principles, and a model. It sets out six principles for good corporate governance of IT:

• Responsibility

• Strategy

• Acquisition

• Performance

• Conformance

• Human behavior

It also provides guidance to those advising, informing, or assisting directors.

An established governance mechanism will be essential to prevent cloud computing from becoming the next generation of "shadow IT" (i.e., IT functions performed outside of IT, and not under the control of the established IT department).

IT governance creates an environment for effective IT management processes to be established and operated. Implementing service management frameworks, such as the Information

Technology Infrastructure Library (ITIL), provides a mechanism to manage the portfolio of services and ensure that there is comprehensive coverage of IT processes such as disaster recovery planning, change control, and capacity management. It will be useful to refer to some of the ITIL definitions and concepts to make the explicit connection with the ITIL service management framework and the requirements that cloud computing will place on IT management.

ITIL v3 defines a set of IT management processes and functions as part of a service life cycle **(Figure 30)** IT organizations are covering these functions in some way, perhaps not formalized and not with a service management orientation.

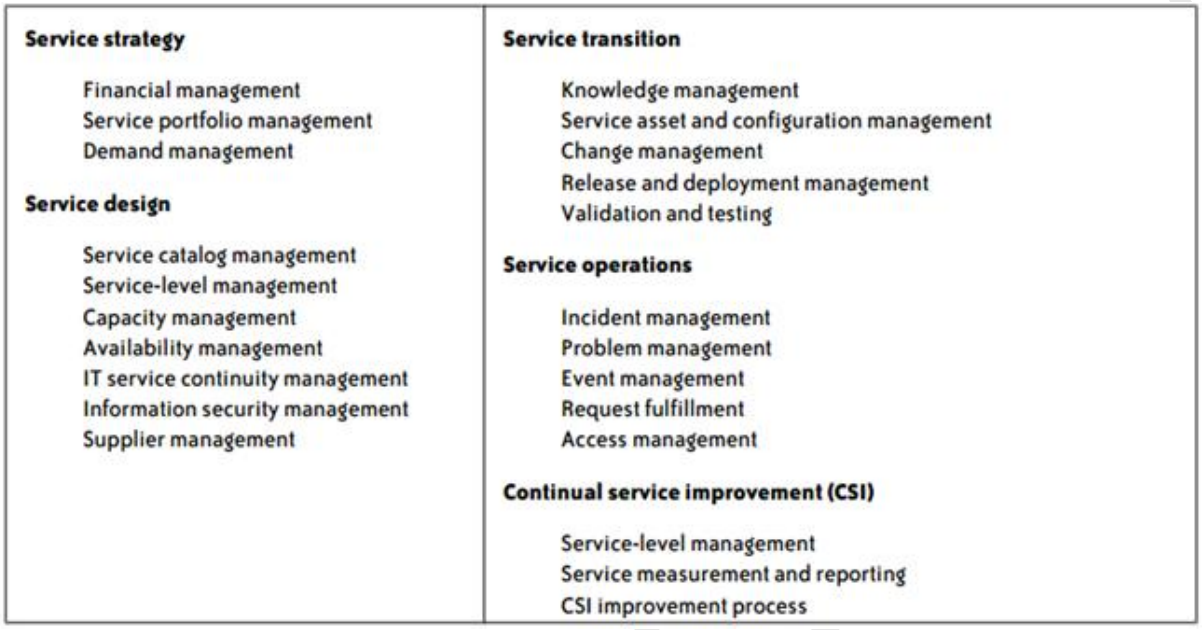


Figure 30. ITIL management processes and functions

We have listed these processes to make the following key points:

Responsibility for specific aspects of IT may be delegated to managers within the organization. However, accountability for the effective and acceptable use and delivery of IT by an organization remains with the directors and cannot be delegated, according to ISO/IEC 38500:2008.

Responsibility for managing IT will be delegated throughout the corporation and, in some cases, to third parties. This is true now and will continue to be true as the driving forces of cloud adoption place more control in the customer's hands. The company itself is accountable for IT service effectiveness, not the CSP.

ITIL is process-oriented and the processes are designed to support a services-oriented approach from internal and external providers. This means ownership for processes such as information security and availability is managed across services, and responsibility can be made clear for those complicating aspects of the cloud architecture.

A critical concept concerns the service portfolio containing all services (live, in-process, and retired), the service catalog (live services), and service-level agreements (SLAs) that define the agreement of the service to be delivered to meet a specific customer need. Within the structure of a service catalog is both a business service component and a technical service catalog. This supports the concept of business managers who are responsible for an organization's business objectives and performance and for engaging IT to support those goals and appropriate IT services. These concepts support the increased level of ownership for defining, procuring, and paying for the services, and IT providing the necessary support to ensure that the service operates effectively in the overall context of the company's IT environment.

Discussions about responsibility for specific IT processes and how specific risks are dealt with help to construct the overall IT organization. Who are the process owners? Where do they report? How are they measured? Who makes what decision?

Frameworks for IT governance and service management are already prepared and handle many of the complicated IT management aspects proposed by adoption of cloud computing. Implementation of complete service life cycle management practices as presented in ITIL v3 is rare given the recent date of introduction in 2008. However, many individual

processes and functions have been adopted and IT organizations are transforming to a process orientation. The ability of an IT organization to deal with cloud computing will depend to a large extent on the adoption of a service management type of approach to IT.

Giving up IT responsibility for certain decisions to business counterparts, and working with business units to determine specific requirements and concerns about security and privacy, provides a mechanism for IT to respond to IT's changing role in the corporation. The risk is that the cost of managing the CSP is not a value add and becomes a bureaucracy unto itself, standing in the way of responsiveness to business needs. The roles of the IS governance group are to direct and monitor the use of IT, acquire new technology, and assess the competence of the staff managing the business and technical aspects of IT. As with any transformational change, a clear vision of the nature of the change and the ability to sponsor the change will be required from company leadership.

In summary, the new IT function will spend fewer resources on building and maintaining commoditized functions and more on differentiated functions. Skills become more oriented around governance, IT service management, architecture, procurement, accreditation, a common vocabulary, and inspection and monitoring. In many cases, these are new skills in an IT department that need to be acquired or may come from outside the IT department. Given this, IT function skills will migrate to CSPs as new skills are acquired in organizational IT departments.

Without this type of IT change, there is a risk of going back to the stovepipes, duplicating inconsistent data as happened before deployment of ERP systems. The balance between control to achieve effective integration and speed of response will be the art and skill of the new IT group function.

6.4 Governance Factors to Consider When Using Cloud Computing

The CSP and its customer have a number of processes to manage. As explained in earlier chapters, such processes include:

- Managing identity
- Provisioning access
- Defining data storage requirements
- Managing key management
- Monitoring and managing service levels
- Monitoring and maintaining availability
- Providing assurance on internal controls
- Providing secure connectivity
- Providing for data governance
- Managing for problem management and incident response
- Developing, maintaining, and, when necessary, executing a business continuity program

It is important for the CSP and the customer to understand the various levels of responsibilities as they relate to the aforementioned processes. Some of these processes may be achieved by the CSP itself, or by the customer itself. In some cases, they may be jointly performed. As a basic rule, during early adoption of the CSP it would be prudent for customers to take on as many of the processes as they can. As the customer gets more familiar with the services of the CSP the customer may slowly transition some of the processes to the CSP where the CSP shows competency and a proven track record to take on these processes. Clear metrics, boundaries of responsibilities between the customer and the CSP, as well as adequate policies need to be in place for the processes to be well managed.

A critical question to address is which part of an organization should manage these processes and which function and executive officer should own the relationship of the CSP? Should this be an operational concern where the chief operating officer becomes the owner? Should this be a technology concern where the chief information officer becomes the owner? Should this be a business-led concern where the most relevant business executive (e.g., finance, sales, legal, human resources, logistics, etc.) becomes the owner?

To answer these questions, look at the nature of the service being provided by the CSP, the culture of the customer, the competency and skills of the customer and the CSP, and the level of executive sponsorship at the customer organization. It is clear that the role of IT may change with the adoption of cloud computing from pure implementation and maintenance of technology to integration of cloud computing into the organization.

Look at a software company that has a traditional IT group. This IT group would manage IT resources to help sustain, grow, and manage business needs. This would typically involve providing back office support in managing the network, application development, and help desk support. Often the software development of the company's own products would be managed by engineering, and in most cases, the related infrastructure would also be managed by engineering. However, with the advent of cloud computing there is a potential for traditional IT groups to play a larger role. For example, the IT group could be seen as a facilitator with the adoption of CSPs to benefit both business initiatives as well as product development. IT groups can promote more common standards and ensure that appropriate measures are in place to mitigate and manage enterprise risks. This will free up resources in other areas to focus more on their core competencies and thereby increase the productivity of the company. Business units and engineering should not be concerned with confidentiality, integrity, and availability concepts as IT professionals are well rehearsed in these areas and are best positioned to define and manage these principles.

6.5 Summary

Just as outsourcing, collocation facilities, and application service providers (ASPs) have had an impact on corporate IT, cloud computing will do the same. In many respects, that cloud computing impact will be an extension or continuation of the trend that the other factors listed have had as well: more work formerly done in-house by corporate IT will shift to outside the organization. Corporate IT departments will become more like managers of the IT services provided than the actual providers of those services.

Particular attention should also be paid to the economics of cloud providers. It is inevitable that many start-ups will go out of business, but the growing competition that is being seen creates business models with razor-thin margins that create challenges to invest

in appropriate support and quality. Ultimately, this becomes a major governance issue for internal IT, who now become consultants and business analysts. They need to take steps to ensure that their cloud provider is a healthy business and can provide a sustainable solution for the long term.

However, there are two differences worth noting. First, delineation of responsibilities between providers and customers is much more nebulous than that between customers and outsourcers, collocation facilities, or ASPs. Given the newness of the cloud computing business model and its nascent stage of development, this lack of clarity on responsibilities is to be expected. With time, more maturity in defining those responsibilities between CSPs and customers will occur. The other difference between cloud computing and other, earlier trends in shifting IT services is that cloud computing is likely to involve much more direct business unit interaction with CSPs than with other providers previously. In fact, there will be many instances where business units go directly to CSPs without even consulting corporate IT departments. Essentially, IT is being cut out of the business loop—deliberately. Many business units see a diminishing value of corporate IT departments to provide for the services that business units need—not only in direct IT skills, but even in management or oversight of IT operations. This diminished view of IT's value directly to business units is reinforced by cloud computing's "pay as you go" business model and the shift from capital expenditures (CapEx) to operational expenditures (OpEx).

References

1 http://www.opencloudmanifesto.org/.
2 http://en.wikipedia.org/wiki/ISO_38500#cite_note-0#cite_note-0
3 http://en.wikipedia.org/wiki/ISO_38500#cite_note-0#cite_note-1

Chapter 7. The Future of the Cloud

Cloud Computing has the potential to be a disruptive force by affecting the deployment and use of technology. The cloud could be the next evolution in the history of computing, following in the footsteps of mainframes, minicomputers, PCs, servers, smart phones, and so on, and radically changing the way enterprises manage IT. Yes, plenty of questions are still left to be answered regarding security within the cloud and how customers and cloud service providers (CSPs) will manage issues and expectations, but it would be a severe understatement to say simply that cloud computing has generated interest in the marketplace.

The hype regarding cloud computing is unavoidable. It has caught the imagination of consumers, businesses, financial analysts, and of course, the CSPs themselves. Search for "cloud computing" on the Internet and you will uncover thousands of articles defining it, praising it, ridiculing it, and selling it.

So powerful is the term cloud computing that according to some, just the mere mention of it may help to drive additional attention and revenues for providers. Take, for example, the case of Salesforce.com. According to Marc Benioff, CEO of Salesforce.com, his software-as-a-service (SaaS) organization did not embrace the use of the term until he read an article that referred to Google and Amazon as cloud computing leaders in December 2007. Soon afterward, Salesforce.com started to leverage the term in its marketing efforts and collateral. In the full fiscal year since Salesforce.com started using the term cloud computing, its revenues grew 44%. "I think it's the most powerful term in the industry," said Benioff.

OK, what does all this mean? Does this mean enterprise adoption of the cloud is a "sure thing"? Maybe. Or maybe not. As we noted in the previous chapters, a number of key drivers for cloud computing may make the move compelling for enterprises, including low levels of initial investment and ongoing costs, economies of scale, open standards, and sustainability. Additionally, there are some potential barriers to adoption that we have discussed, including concerns regarding security, privacy, and compliance and governance.

At the end of the day, you are probably going to make your own assessment regarding the future of the cloud and whether the use cases and associated value propositions are appropriate for you and/or your organization. The objective of this chapter is to give you a snapshot of analyst and IT leadership thoughts regarding the potential of the cloud and our thoughts regarding the future of the cloud.

7.1 Analyst Predictions

Most financial analysts feel that cloud computing will be a huge growth area in terms of IT spending and revenue streams over the next few years, but the estimates vary.

According to a May 2008 forecast by Merrill Lynch, the volume of the cloud computing market opportunity will amount to $160 billion by 2011, including $95 billion in business and productivity applications and $65 billion in online advertising[1].

According to a March 2009 forecast by Gartner, worldwide cloud services are on pace to surpass $56.3 billion in 2009, a 21.3% increase from 2008 revenues of $46.4 billion. The market is expected to reach $150.1 billion in 2013[2].

“Cloud computing is a broad and diverse phenomenon. Much of the growth represents a transfer of traditional IT services to the new cloud model, but there is also scope for creation of substantial new businesses and revenue streams,” said Ben Pring, research vice president for Gartner. “Cloud computing enables a shift in IT provision from direct purchase and payment for services to provision of services which are free at point of use and where revenue is derived from advertising. Services supported by advertising are currently, and will remain, the largest component of the overall cloud services market through 2013.”

According to an IDC October 2008 forecast, spending on IT cloud services is growing at five times the rate of traditional, on-premises IT. Also according to IDC, even more striking than this high growth rate is the contribution that the cloud offering’s growth will soon make to the “The Internet Industry Is on a Cloud—Whatever That May Mean,” by Geoffrey A. Fowler and Ben Worthen. The Wall Street Journal, March 2009.

IT market’s overall growth. As illustrated in **Figure 31**, cloud computing services will generateapproximately one-third of the net new growth within the industry.

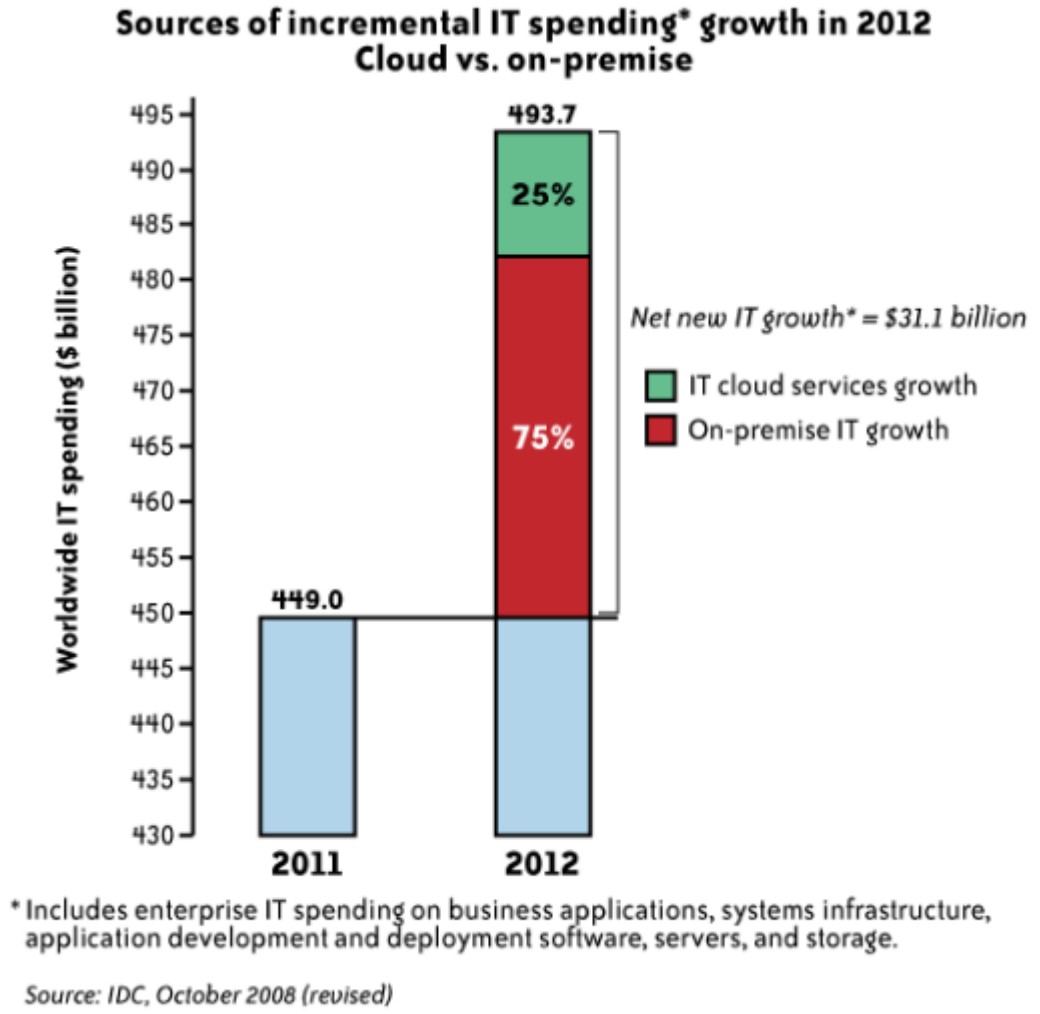


Figure 31. Sources of incremental IT spending growth

Additionally, according to IDC and illustrated in **Figure 32**, projected spending on cloud services will nearly triple by 2012, and will continue to be dominated by SaaS offerings over this period of time. But as you can see from the sheer scale of the increase in overall cloud services spending, platform-as-a-service (PaaS) and infrastructure-as-a-service (IaaS) services will also experience strong growth[3]. So, the analysts seem to be sold on the growth potential of cloud computing. What about IT and business leaders?

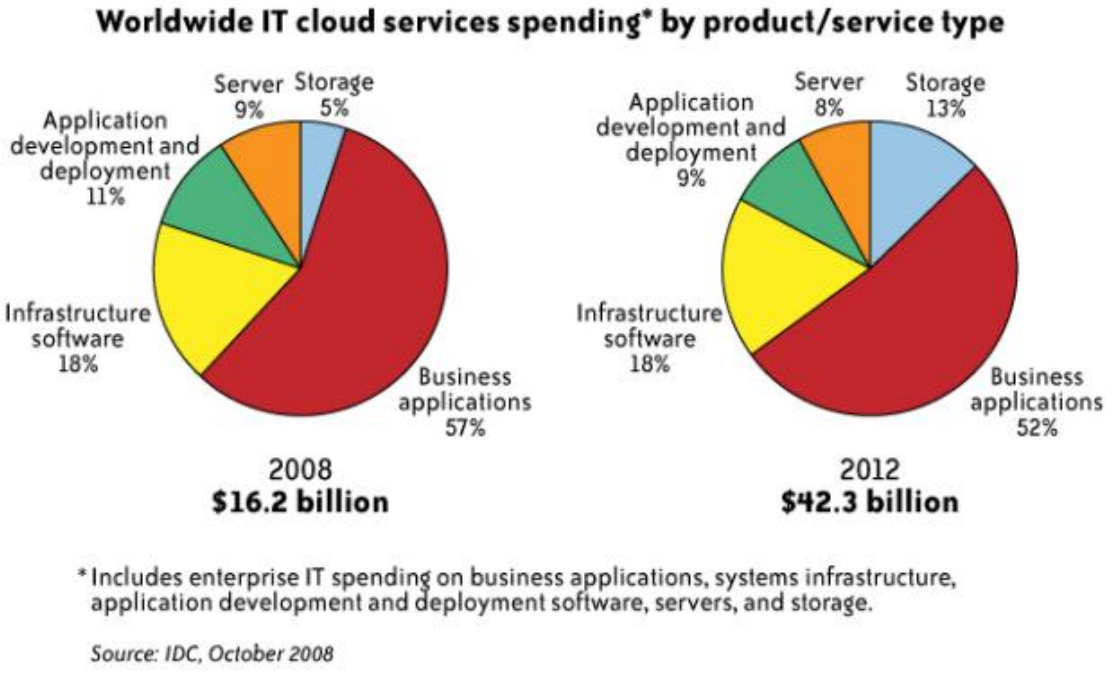


Figure 32. Worldwide IT cloud services spending

7.2 Security in Cloud Computing

Since the premise of our book is that security is a concern when discussing cloud computing, let's revisit the security considerations we previously discussed and conclude with our thoughts on the current and future states of these considerations for the cloud:

- Infrastructure security
- Data security and storage
- Identity and access management
- Security management
- Privacy
- Audit and compliance
- Security-as-a-[cloud] service
- Impact of cloud computing on the role of corporate IT

Infrastructure Security

During our discussion of infrastructure security, we looked at network-, host-, and applicationlevel security and the issues surrounding each level with specific regard to cloud computing. At the network level, although there are definitely security challenges with cloud computing, none of those challenges are caused specifically by cloud computing. All of the network-level security challenges associated with cloud computing are instead exacerbated by cloud computing, not specifically caused by it. Likewise, security issues at the host level, such as an increased need for host perimeter security (as opposed to organizational entity perimeter security) and secured virtualized environments, are exacerbated by cloud computing but not specifically caused by it. And the same holds true for the application level. Certainly, there is an increased need for secure software development life cycles due to the public-facing nature of (public) cloud applications and the need to ensure that APIs have been thoroughly tested for security, but those application-level security requirements are again exacerbated by cloud computing and not specifically caused by it.

Therefore, the issues of infrastructure security and cloud computing are about understanding which party provides which aspects of security (i.e., does the customer provide it or does the CSP provide it)—in other words, defining trust boundaries.

With regard to infrastructure security, an undeniable conclusion is that trust boundaries between customers and CSPs have moved. When we see poll after poll of information executives (e.g., CIOs) and information security professionals (e.g., CISOs) indicating that security is their number one concern with cloud computing, the primary cause for that concern is really over moved trust boundaries. To be more specific, the issue is not so much that the boundaries have moved, but more importantly that customers are unsure where those trust boundaries have moved to. Many CSPs have not clearly articulated those trust boundaries (e.g., what security is provided by the CSP versus what security still needs to be provided by the customer), nor are those new trust boundaries reinforced in operational obligations such as service-level agreements (SLAs).

Although the CSPs have the primary responsibility for articulating these new trust boundaries, some current confusion about this is also the fault of information security personnel. There are some information security professionals who, either fearing something new or not fully understanding cloud computing, are engaging in FUD (fear, uncertainty, and doubt) with their business customers.

Similar to confusion over moved trust boundaries is the fact that the established model of network tiers or zones no longer exists. That model has been replaced with domains, which are less precise and afford less protection than the old model. (Domain names are used in various networking contexts and application-specific naming and addressing purposes based on DNS.) If we can no longer trust the network (organizational) perimeter to provide sufficient protection and are now reliant on host perimeter security, what is the trust model between hosts?

An analogy of this problem already exists and was dealt with 20 years ago—STU- (Secure Telephone Unit) IIIs used by the U.S. Department of Defense and the intelligence community. In that model, each STU-III unit (a host) was responsible for its own “perimeter security” (i.e., the device’s electronic components were tamper-resistant), and each device had a secure authentication mechanism (i.e., a dongle with an identity written to it, protected and verified by asymmetric encryption and Public Key Infrastructure or PKI). Additionally, each device would negotiate a common level of authorization (classification level) based on an attribute included with the identity in the dongle.

Today, we have no such model in cloud computing. The STU-III model simply is not viable for cloud computing, and there is no trusted computing platform for virtual machine (VM) environments. Therefore, host-to-host authentication and authorization is problematic in cloud computing since much of it uses virtualization. Today the use of federated identity management is focused on trust, identity, and authentication of people. The identity

management solutions of today do assist in managing host-level access; however, there is no viable solution today that addresses the issue of host-to-host trust. The host-to-host trust issue is exacerbated in cloud computing because of the sheer number of resources available.

Conceptually similar to the trust boundary problem at the application level is ensuring that one customer's data is not inadvertently provided to another, unauthorized customer. Data has to be securely labeled to ensure that it remains separated among customers in a multitenancy environment. Today, data separation in cloud computing is logical, not physical, as was done previously, and there are valid concerns about the adequacy of that logical separation.

Data Security and Storage

During our discussion of data security and storage, we looked at several aspects of data security and the storage of data. If cloud computing customers are concerned about the security afforded by infrastructure security and are counting on data security to provide compensating controls, those customers will be disappointed. A major reason for the lack of effective data security is simply the limitations of current encryption capabilities. However, efforts to adequately detail data lineage (mapping) are simply not possible in today's cloud computing offerings. The amount of effort (and cost) to provide such mapping runs counter to the economic incentives of cloud computing. Another major problem with current cloud computing offerings is a lack of serious attention (effective action) to customers' concerns about data remanence (i.e., data residue left behind and possibly becoming available to unauthorized parties).

These concerns with data security do not negate the capabilities or advantages of utilizing storage-as-a-service in the cloud—for non-sensitive, non-regulated data. If customers do want to (simply) store organizational data in the cloud, they must take explicit actions, or at least verify that the provider will and can adequately provide such services, to protect their data stored in the cloud.

We know how to effectively encrypt data-in-transit, and we know how to effectively encrypt data-at-rest. But because encrypted data cannot be processed, indexed, or sorted, to do any of those important activities requires that the data be unencrypted—hence, a security concern, especially if that data is in the cloud and is beyond the data owner's direct control.

Even efforts to effectively manage data that is encrypted are extremely complex and troublesome due to the current inadequate capabilities of key management products. Key management in an intra-organizational context is difficult enough; trying to do effective key management in the cloud is frankly beyond current capabilities and will require significant advances in both encryption and key management capabilities to be viable. Claims of key management products being effective currently are naïve at best.

Identity and Access Management

Managing access control and governance within identity and access management (IAM) to meet today's business needs in the cloud remains one of the major hurdles for enterprise adoption of cloud services. IAM support for business needs ranges from secure collaboration with global partners to secure access for global employees consuming sensitive information from any location and any device at any time. Thanks to the proliferation of consumer technologies (e.g., Apple iPhone) into the enterprise (consumerization of IT) and the steady dissolution of the network perimeter, enterprises are faced with greater risks in protecting their intellectual property and sensitive information as well as sustaining compliance. Easily accessible, user-friendly Web 2.0 technologies delivered via browsers is one other catalyst that is accelerating the trend of "consumerization of identity and access management" services (e.g., consumer-based identity services such as OpenID). In short, IT is constantly challenged to support today's business needs with yesterday's technologies and static processes. And the information protection challenges are exacerbated by increasingly mobile,

dynamic, replicated, and scattered data on a variety of media ranging from USB memory sticks to storage-as-aservice.

On the other hand, IT is grappling with user access management dissatisfaction issues among business users who are increasingly frustrated with today's "user-unfriendly" IAM techniques (e.g., carrying a token card that performs two-factor authentication, remembering a variety of user IDs and passwords for various services, and forcing users to choose a strong password that they write down and carry in a wallet). And it is no secret that users will do anything to sidestep identity or any other security controls that slow their productivity and business agility. Hence, IAM solutions need to strike a balance and act as enablers of security controls to increase user adoption and compliance.

Although the basic technology building blocks (trusted identity stores, provisioning processes, authorization and authentication methods, federation) for IAM exist today, the migration and extension of those technologies into cloud services in their current form will not yield the purported IAM benefits of efficiency, efficacy, and business agility. The sheer volume of dynamic cloud compute resources (compute nodes, storage, network policies) combined with the magnitude of users and services accessing those resources are challenging the scalability, automation, and availability requirements of today's directory and identity infrastructure services. The primary reason is that today's IAM solutions deployed in the enterprise are complex, require extensive customization, are expensive, and are not easily extendable to cloud services. Furthermore, the trusted source of identity in the cloud is still an issue and needs to be addressed. On the other hand, support for IAM practices and standards by CSPs is sparse and is not adequate for most enterprises. Although large SaaS cloud services are showing signs of support for federation standards such as the Security Assertion Markup Language (SAML), they are largely absent from PaaS and IaaS services. A word of caution: viral adoption of cloud services driven by business units that don't leverage your own federated identity management infrastructure and IAM processes risks repeating the mistakes (e.g., provisioning of multiple credentials per user) that caused you to implement enterprise identity management solutions in the first place.

Today's early adopters—small and medium-size businesses (SMBs)—who are driven by the economic advantages of cloud computing have silently embraced the basic low-assurance authentication methods, leaving the enterprises waiting on the sidelines. Enterprises are hoping that the CSPs will offer IAM capabilities that are standard within their enterprise, and have come to expect this in any new service.

Enterprise cloud adoption barriers include lack of support for federation (single sign-on or SSO), integration with corporate directories, risk-based authentication, scalable identity services, and the extension of the IAM practice to the CSP. Hence, IAM solution design for cloud services will require careful consideration of cloud use cases, investment in processes and architecture that address cloud user access provisioning (including privileged users), service-to-service authentication and user-to-service authentication, and management of the user and access life cycle.

A small set of CSPs (mostly large SaaS service providers, such as Salesforce.com) are beginning to pay attention to enterprise IAM requirements, including support for standards such as SAML that facilitate SSO using federation. However, given the early adoption cycle by large enterprises, from an enterprise perspective IAM capabilities are primitive at best. Customers should continue to demand IAM features, including support for SAML, user provisioning using the Service Provisioning Markup Language (SPML) standard, and an open application programming interface (API) to support various user and access automation requirements. This IAM capability chasm has given birth to a new breed of cloud-based identity services; for example, identity services and frameworks such as secure token services (STSs) from Microsoft's Azure support basic federation from Active Directory to Microsoft's cloud services and facilitate user SSO from on-premises Active Directory to

Microsoft's cloud services. Although these cloud-based identity services are lowering the barriers to entry for SMBs, they are deemed inadequate to meet most enterprise requirements such as custom reporting and compliance management. Trust and user data management are other barriers, and most enterprises are not willing to store their trusted source for identity outside controlled enterprise boundaries. This issue is further exacerbated by use cases in which attribute data associated with identities is either copied or stored in the cloud service. Synchronizing multiple identity repositories remains a key challenge for enterprises. Working with cloud-based services and addressing synchronization issues by way of federation, virtual directories, and an open API will reduce these barriers. To avoid costly retrofits and integration with aftermarket products, organizations looking to adopt cloud-based services should embed an IAM strategy into the cloud service strategy road map. Organizations that have been investing in directories, IAM capabilities, and practices should therefore stand to gain by leveraging an optimized internal IAM strategy and practice in the cloud. The most important success factor for an enterprise to effectively manage identities and access control in the cloud is the presence of a robust directory and federated identity management capability within the organization (an internal or cloud-based identity service)—for instance, architecture and systems, user and access life cycle management processes, and audit and compliance capabilities. When it comes to authenticating users and services to the cloud, organizations need to pay attention to simplicity and ease of use in addition to risk-based authentication methods (e.g., look up when sensitive data is accessed). Another premise to keep in mind is that "all clouds are not created equal," so enterprises need to have a strategy for employing risk-based IAM methods, including strong authentication, automated provisioning, deprovisioning, auditing, and monitoring to address risks that are specific to a CSP.

Although identification and authentication challenges can be overcome (when those capabilities are made available by the service provider) with a well-architected IAM infrastructure and IT processes, authorization services in the cloud are very basic and evolving. Cloud users should be aware that granular application authorization is immature at this point. Where it does exist, it is usually implemented using CSP proprietary profiles and primitive roles—often CSPs offer primitive roles such as "user" and "administrator." As a long-term strategy, customers should be advocating for greater support of eXtensible Access Control Markup Language (XACML)-compliant entitlement management on the part of cloud providers, even if XACML has not been implemented internally. XACML provides a standardized language and method of access control and policy enforcement across all applications that enforce a common authorization standard. At the very least, CISOs should be thinking about authorization standards and avoid any temptation to customize a solution based on the provider's capability.

Business and IT stakeholders should also be advocating standardization of enterprise roles within the enterprise—in other words, roles mapped to user business functions (e.g., accounts payable manager, people manager, and purchase order approver). In the future, well-defined enterprise roles should be mapped to the cloud service roles or profiles supported by the CSPs. We believe SPML and XACML will play a role in that regard. (Currently, we are not aware of any effort to standardize the naming conventions of enterprise roles.)

IT architects should be advocating externalization of authentication and authorization components from applications (loosely coupled) as this can aid in the rapid adoption of cloudbased services including cloud identity services, policy-based authentication, centralized logging, and auditing (e.g., OpenSSO from Sun Microsystems and Microsoft's Geneva claimsbased authentication framework can help externalize authentication).

Security Management

With the adoption of cloud services, a large part of your network, system, applications, and data will move to a third-party provider's control. The cloud services delivery model brings new challenges to the IT operations and management staff in the area of availability, access control, vulnerability, security patching, and configuration management. As a first step, cloud customers will have to understand all the layers they own, touch, or interface with—network, host, application, database, storage, and web services, including identity services. To tackle these challenges, you will need to understand the interfaces and the scope of IT system management responsibilities, including your responsibilities for access, change, configuration, patch, and vulnerability management.

Although you may be transferring some of the operational responsibilities to the provider, you may still own some of the responsibilities whose scope will depend on a variety of factors, including the type of cloud service. Major factors to consider are the SLA, monitoring capability, and provider-specific security management capabilities to support the extension of your internal operations management processes and tools.

Today, customers largely rely on CSPs for the service instrumentation to measure and manage the security, availability, and performance of their services in the cloud. Most CSPs are sharing the overall service metrics via a dashboard (e.g., Amazon's service health dashboard at http:// status.aws.amazon.com/). Although a CSP may be publishing the most up-to-the-minute information of its overall system status across all customers, the onus is on you to keep abreast of the service status. To manage the availability of your application you will need to measure, monitor, and manage service levels from your perspective (i.e., for your virtual environment). Unfortunately, the lack of standards and weak capabilities from CSPs to help customers place probes into their virtualized environment have exacerbated cloud service management. Hence, as a tenant of a aaS service, you will have to understand what instrumentation and dashboards are made available to you by the service provider to help manage service levels to your users.

From a security management perspective, a key issue is the lack of enterprise-grade access management features. Since access control features will vary with the service delivery model and provider, customers will have to understand what access control features are available (strong authentication, user provisioning) and what their responsibilities are in managing the life cycle of user access to the cloud service. Some service providers are making an effort to keep their customers informed of new threats and educating them on ways to protect the information hosted in their cloud (e.g., Salesforce.com publishing threat and security practice information via http://trust.salesforce.com/).

In a virtualized environment where infrastructure is shared across multiple tenants, your data is commingled with that of other customers at every phase of the life cycle—during transit, processing, and storage. Even if you are able to install monitoring probes at infrastructure layers available to you, the resource bottlenecks that are visible to your instrumentation may not be able to give the necessary information to perform root-cause analysis (e.g., latency of packets between your system nodes in the cloud). Outages that impact the entire population will be visible to all users. Another dimension in cloud computing is the issue of monitoring and measuring disruptions across your users—depending on the cloud service architecture, failures of the infrastructure components may impact only a subset of the population and it would be hard to detect the service disruption unless the affected users report it (e.g., Google mail disruption events that impact only a subset of users). Hence, it is important to understand the location of the service, service-level guarantees such as internode communication, and storage access (read and write) latency.

The scope of security management of cloud services will vary with the service delivery model, provider capabilities, and maturity. Customers will have to make trade-offs with respect to the flexibility and control offered by the SPI services. The more flexible the service (i.e., the lower the service abstraction), the more control you can exercise on the

service, and with that come additional security management responsibilities. Given that most cloud service offerings lack transparency in the area of SLA, provider management capabilities, and security responsibilities, the management functions will continue to challenge enterprises that have established IT governance, tools, and processes. Those frameworks, processes, and tools that address systemic qualities including reliability, availability, and security may not be extensible to the CSP. If you have adopted standard IT frameworks including the Information Technology Infrastructure Library (ITIL) and ISO 27002 in your organization, they should be reviewed and continuously adjusted based on the cloud service capabilities, sensitivity of information, and SLA that govern various management functions.

Privacy

Cloud computing offers significant challenges for global organizations that are facing multiple global and sometimes conflicting privacy rules, regulations, and guidance. Organizations need to adopt a systematic approach to addressing privacy in the cloud. Cloud computing is facing a challenge that has existed for many years: how to deal with crossborder data flows. Since this involves a number of foreign jurisdictions, complexities start to develop due to conflicting rules among foreign governments (or even among various states within the United States). The nature of, and one of the major benefits of, cloud computing just expands this challenge. It is worth noting that an organization can define to the CSP in which country it would like to have its data stored and processed. However, determining which specific server or storage device will be used is difficult to ascertain due to the dynamic nature of cloud computing.

We further explored the impact of cloud computing on Organization for Economic Cooperation and Development (OECD) and other privacy principles, and we concluded that:

• The CSP requires strong data governance (managing the entire life cycle of the data from creation to destruction) to enable client organizations to respond to requests for government disclosure of data.

• Care should be taken to delete storage devices, especially as it relates to virtual storage devices where storage is constantly being reused.

• Transferring data to third parties will require consent from the data owner.

• Multiple privacy laws and regulations, such as the European Union and U.S. Safe Harbor Program, require knowledge of where data is stored at all times. This will encourage CSPs to store data on servers located in specific jurisdictions that minimize legal risk (potentially outside Europe and the United States).

• Data protection and privacy policies should be applied to data and should follow through the data's life cycle to ensure that original commitments are met and to create accountability and knowledge of what happens to data.

Organizations are expected to be responsible for knowing and managing how data is being handled and stored at all times. This becomes difficult in a cloud computing environment since IT resources are often shared and used on demand. There are a few steps that a CSP can take to improve data privacy and security. This includes improving security solutions such as IAM (restricting access), key management (encrypting data), secure event and incident monitoring (monitoring for security breaches), and data loss prevention solutions (monitoring for data breaches). The organization's privacy commitments (legal, regulatory, and contractual) should be attached to the data elements across their life cycle. There are many debates regarding who should be responsible for privacy—perhaps the CSPs?

However, it is a commonly held belief that the accountability for privacy protection falls on the organization that collected the information in the first place. To fulfill this role, it is essential for these organizations to understand the privacy and security policies and security architecture of the service the CSP is delivering, to have the right contractual arrangements in place, and to monitor the CSP's compliance. The various reporting standards satisfy the

multiple requests that organizations will have from the CSP. However, these reports tend to be generic and may not explain the specific nature of the processes and controls associated with the specific data in mind. There is a need for a globally consistent privacy standard that the CSPs will adopt and independent third parties will monitor for compliance.

It is worth noting that payroll processing has been around for a long time and data is regularly sent to payroll bureaus for processing. Such data is sensitive and contains a lot of personally identifiable information (PII). Most organizations have relied on SAS 70 reports to gain comfort regarding the processes and controls supporting the payroll process. These payroll processors have multiple customers and process a number of payrolls at the same time. The current SAS 70s, however, don't provide user organizations with comfort regarding the privacy of the data.

The risks and issues around payroll processing are very similar to concepts being introduced by cloud computing. However, since payroll processing has been around for a longer time, organizations have gotten used to relying on it for security. Granted, organizations can recalculate the accuracy of the processing, but the payroll service provider is still responsible for securing the data.

Audit and Compliance

It is clear that the CSP will face a large number of requests from its customers to prove that the CSP is secure and reliable. There a number of audit and compliance considerations for both the CSP and the customer to consider in cloud computing. First, which compliance framework should a CSP adopt to satisfy its customers and manage its own risks? The customer base will largely determine the framework that the CSP would choose. Most IT service providers are adopting a combination of ITIL, ISO 27001, and specific industry standards such as the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the Payment Card Industry (PCI). It is expected that the CSP will adopt the same approach. It is possible that the adoption of cloud computing may impact an organization's SarbanesOxley (SOX) program. At the moment, most organizations are resisting putting any data relating to financial reporting systems into cloud computing. However, email is often used as the means for communicating the authorization or approval of a control activity, and this may end up in the cloud. Alternatively, smaller organizations with finite resources may use PaaS and thereby bring software development life cycle controls into scope.

Many forms of reporting are available to satisfy these requests. The most relevant standard for the CSP to adopt would be SysTrust, or eventually, SAS 70, once new changes are made to this standard.

For the CSP to be successful it will be necessary to establish an appropriate framework of processes and controls. This framework needs to be comprehensive and globally accepted to meet the challenges of the various industry verticals. Imagine if the CSP customer is a health care provider or a bank. The requirements will be different for each and it can be expensive for a CSP to meet the various industry requirements.

A growing concept in the industry is the development of an IT Governance, Risk, andCompliance (GRC) program. The intent of such a program is to develop an IT uniformedcompliance framework. A number of tools are available today that can automate this process. Such tools have:

• A library of controls covering standards such as ISO 27001, PCI, Control Objectives for Information and Related Technology (COBIT), ITIL, the National Institute of Standards and Technology (NIST), and many others

• Built-in connectors to leverage existing security tools deployed in the IT environment

• A flexible, real-time reporting engine that can report on various standards and organizational units

Figure 33 illustrates an overview of the capabilities of an IT GRC program and its relationship to the broader enterprise GRC. A large proportion of its function relates to security, and such programs result in the adoption of compliance dashboards that can be configured to various levels of management and show real-time compliance and an indicator of where risk exists.

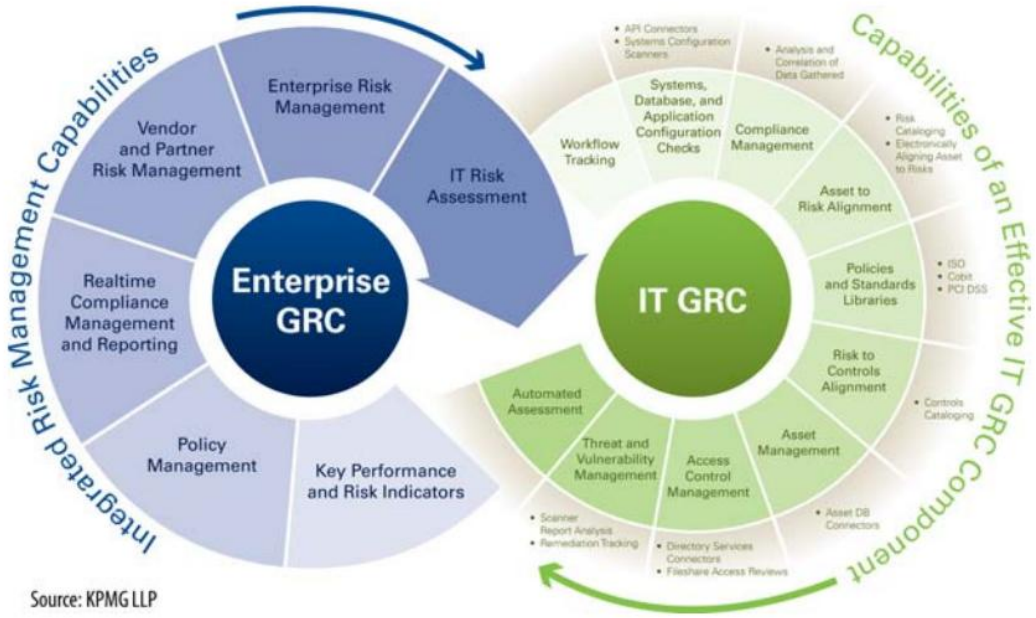


Figure 33. An overview of IT GRC

Such tools can report on specific organizational units against a specific standard, or a combination of standards against a tailored framework. This would allow a CSP to reduce its cost of compliance and create a more sustainable solution. The adoption of IT GRC will allow the CSP to deliver more custom reports to reflect the standards relevant to the customer and in a timelier manner.

Security-As-a-[Cloud]-Service

Security-as-a-service is already well established in the nascent cloud computing space. In fact, it is likely to continue to grow both in terms of market share against traditionally delivered security capabilities and in terms of depth of offerings. For example, not only is the relatively new identity-as-a-service (IDaaS) a needed alternative for individual organizations, but IDaaS will become even more desirable for growing organizational types, such as increasingly multistatus organizations (i.e., employees, contractors, interns, other companies' employees, and vendors, all working in the same shared workspace), co-opetition (cooperative competition), and virtual organizations. Additionally, other important security services could be outsourced and provided in a cloud environment, such as logging, auditing, and security incident and event management (SIEM).

Security-as-a-service is likely to see significant future growth for two reasons. First, it is likely that a continuing shift in information security work from in-house to outsourced will continue. What started with email filtering and managed security services will continue and expand as organizations look to reduce capital expenditures (CapEx) further and increasingly concentrate on their core capabilities. Second, several other information security needs are present for organizations currently, but they will accelerate in need and complexity with the growing adoption of cloud computing. That growing complexity will further fuel the growth of SaaS. Specifically, we are referring to two preventive (proactive) controls and two detective (reactive) controls. The two proactive controls are also important to the growth of cloud computing: identity management that is intercloud and scalable to the cloud size, and (encryption) key management. Significant improvement in both is needed for cloud

computing, and that will make potential solutions very valuable. The two reactive controls are needed for audit and compliance purposes as well: scalable and effective SIEM, and data leakage prevention (DLP). Trying to provide solutions to each of these controls will be difficult and requires significant complexity that must be hugely scalable and yet easy to use. However, all of these needs also pose significant and growing opportunities for vendors as cloud computing continues to grow in adoption.

Impact of Cloud Computing on the Role of Corporate IT

Almost certainly, many corporate IT departments will continue to be redefined by this latest model of outsourcing. As with earlier outsourcing (e.g., to large IT services firms such as CSC, EDS, and IBM Global Services, or application development to China or India), use of collocation facilities or application service providers (ASPs) and IT functions previously done in-house are moving outside corporate IT departments. With growing IT needs at the cost of growing complexity, many organizations are deciding that IT is not a core competency for their organizations and much of the IT work required to run today's organizations is being turned over to specialist companies. Cloud computing is a further example of this.

However, cloud computing is in some respects also a repudiation of traditional corporate IT departments. Business units are tired of hearing CIOs and IT departments telling them that the costs of their desired projects are excessively high and that there will be an excessive time delay until those projects can be implemented. Part of cloud computing's appeal is the speed with which business units can be up and running on their desired platform or application, along with the perceived lower costs of "pay as you go" and lack of upfront capital expenditures. As such, it really should be no surprise that the push for the use of cloud computing in most organizations is coming from business units and not from within IT. The long-standing tech mantra of better, faster, cheaper has come home to roost for corporate IT departments.

7.3 Program Guidance for CSP Customers

It is important for customers of CSPs to develop a strategy to manage the security issues mentioned earlier. We suggest that the strategy be based on developing capabilities in the manner illustrated in **Figure 34**.

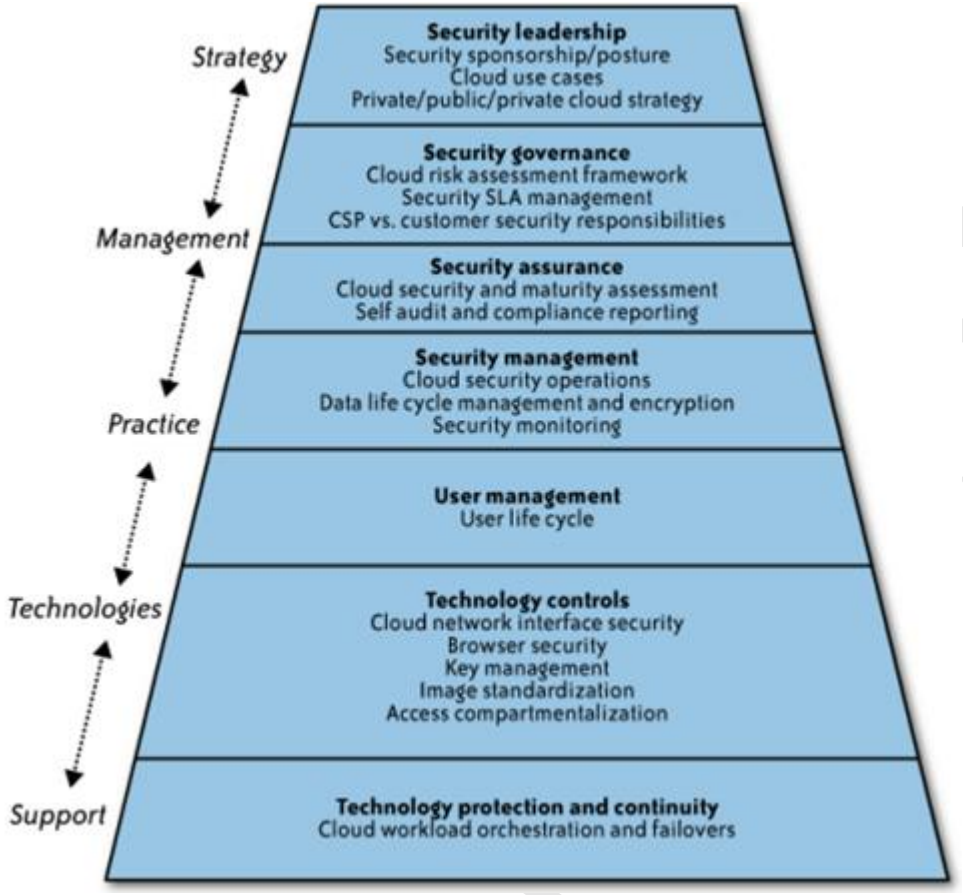


Figure 34. An enterprise security architecture for cloud computing

Let's briefly examine each component of this enterprise security architecture.

Security Leadership

Appropriate leadership needs to be involved with any strategy involving cloud computing. This applies to both CSPs and their customers. Customers are likely to have a decentralized approach as each business unit adopts its own plans for addressing the CSP. It is important to centralize this planning to ensure that consistent practices are adopted and that the maximum purchasing power is leveraged. Considerations of using the public, private, or hybrid clouds need to be standardized. Today, some customers of CSPs have IT departments whose staff members have little knowledge of how the CSPs are being used.

It is important for management to have a deep understanding of the issues around cloud computing and it is vital that they are educated on the latest solutions and challenges with cloud computing. The traditional security paradigm is different for cloud computing, so it is essential for leadership to fully understand the complexities and capabilities of solutions in the cloud. Applying traditional security techniques is not sufficient. For CSPs, it is important to have adequate senior leadership involved in all security matters to ensure that they are appropriately addressed.

Security Governance

Another critical success factor is that appropriate governance needs to be in place. That is, is an appropriate organizational structure in place to manage the organization facing the cloud computing solution? A risk assessment framework should be adopted to ensure that consistent and reasonable practices are applied. Defining security metrics will be key to both the CSP and the customer. Each will have different perspectives and it is important to ensure

that both understand their responsibilities well and none rely on each other. Key security policies that would become critical would be the handling of data, storage of data, communication policy, vendor management (including external connections), trust reporting (i.e., how to give assurance to third parties and customers of the reliability and security of thesolution), and awareness policy (both for customers and for internal users to the boundaries of responsibilities around security).

Security Assurance

Another key aspect to overcome is for the CSP to provide assurance to its customers that their operations are secure and reliable. SAS 70 is not an adequate reporting format and CSPs will need to develop a more transparent means of gaining the confidence of their customers. Customers of CSPs need to perform their own audits and ensure that they have the right to audit for key operations. Clearly, this will become burdensome for CSPs, so they will need to develop more assurance by becoming compliant with standards such as ITIL, ISO 27001, and others to build up confidence from the market.

Security Management

Information governance, that is, the need to have controls over the life cycle of data, is crucial for both the CSP and its customers. One of the biggest issues is the difficulty in tracking the exact location of data during processing times; therefore, having control over its creation, storage, use, and destruction is important. Leveraging data mining tools and providing sound IT operational practices will be key to managing data.

Developing capabilities around information asset security will be challenging to CSPs. As we discussed earlier, although host-level security can be addressed, host-to-host communication and its integrity are much harder to secure due to the volume and dynamic nature of how data travels through the infrastructure. Although traditional security scanners can be deployed, it is critical to have real-time reporting around them; therefore, an IT GRC solution would assist in providing management with a “dashboard” of key metrics to provide oversight of site security and reliability.

User Management

Identity access management can be leveraged to assist the CSP in providing access more seamlessly to its customers. However, IAM solutions today need to be enhanced to deal with having multiple CSPs providing access to the same customer. Also, these solutions need to provide the ability for self-provisioning in such a multitenancy environment. User awareness will be key, and more education is needed for the customers of CSPs to understand how the security posture will be changed with the CSP.

Technology Controls

A number of new and exciting technologies can be applied to both the CSP and its customers. A central question to ponder is who should manage the keys as they relate to the encryption of data? Can the CSP be trusted and does it have the expertise to hold the keys? Other factors to address would be browser security, image stabilization, and how access can be controlled.

Technology Protection and Continuity

CSPs provide for a resilient system; however, there will be times, perhaps due to a failure by the ISP or telecommunications carrier, when the customer may not be able to access the CSP’s environment. Although most CSPs will build resiliency and redundancy into the design of their services, it is inevitable that there will be some outages. It is essential for both CSPs and their customers to have robust business and disaster recovery plans. The responsibilities for certain tasks will not be clear, so it is important for both parties to recognize who will be responsible for which part of the business continuity plan and/or the disaster recovery plan. The testing of each plan will be critical here to ensure that the right level of coordination between the CSP, ISP, and customer as well as others exists.

Overall Guidance

Overall, both customers and CSPs need to work together to mutually agree on what aspects of security will be provided and monitored by both parties to manage the risks of leveraging cloud computing, and the traditional models of security have to be retooled to address the risks of cloud computing. It has been shown that the security around the CSP can be monitored and controlled; however, it will ultimately be the responsibility of the customers of the CSP to ensure that the appropriate measures are taken and that they cannot rely on the CSP to provide a secure and reliable environment without consultation and advice from the CSP's customers.

7.4 The Future of Security in Cloud Computing

Over the course of this book, we have discussed key drivers for adoption and potential barriers, including the inherent security concerns associated with cloud computing. Let's look forward and see what potential direction security may take in the areas we just discussed.

Infrastructure Security

There is without question a need for greater transparency regarding which party (customer or CSP) provides which security capabilities, as well as greater assurance over the CSP's capabilities and efforts. It is likely that there will be increased agreement on what security capabilities each party is to provide, as well as some level of standardization across CSPs regarding CSP security capabilities with respect to specific offerings in the SPI service delivery model. It is also likely that this standardization and agreement will be reflected in operational SLAs.

In the future, identity management should be adopted to address the interrelationships between systems, services, and people. As intercloud (i.e., cloud-to-cloud) communications come into existence, due to customer demands these interrelationships will take on even greater urgency.

Data Security and Storage

Due to the nature of cloud computing (e.g., multitenancy) and the volume of data likely to be put in the cloud, data security capabilities are important for the future of cloud computing. Because of that, coupled with today's inadequate encryption and key management capabilities, cryptographic research efforts, such as predicate encryption[5], are underway to limit the amount of data that can be decrypted for processing in the cloud. Recently announced capabilities of fully homomorphic encryption to process encrypted data should be a huge benefit to cloud computing[6]. Future commercial viability of such capabilities would be a huge benefit to cloud computing. Similar research into large-scale, multi-entity key management should also be encouraged, as it would be of enormous benefit to cloud computing.

Identity and Access Management

Today, access governance within the enterprise is a constant struggle and requires constant customization. This is compounded by the fact that no single monolithic IAM solution is available to meet the basic use cases, such as SSO, within an enterprise. Although enterprises are deploying IAM solutions to address yesterday's problems, today your business units may be adopting cloud services in an ad hoc, viral fashion. Although user-provisioning project cost overruns and failures have reduced customer expectations,

federation is viewed positively and web access management, enterprise SSO, audit, and compliance have become IAM drivers. Hence, enterprises will have to rapidly reevaluate the IAM strategy approach to address IAM use cases for cloud services. With the advent of cloud-based identity services, enterprises may adopt a hybrid IAM strategy where some aspects of IAM that require architectural change migrate to cloud services while the trusted source and processes stay within the enterprise trust boundary.

When it comes to the trusted source of identities, the standard practice within enterprises is to rely on a well-established, trusted source of identity registries (e.g., an enterprise HR database for managing the identities of employees, contractors, and partners). That practice and process architecture will be challenged by new enterprises that grow with cloud services and come to rely on "everything as a service." The trusted source model will be disrupted when HR services move from controlled enterprise boundaries to cloud services (e.g., Workday for HR services). In that IT delivery model, there are a few issues to ponder: how will the trusted source manifest when the HR service is delivered from the cloud? Can we trust those services to be the authority of identities? And what new connector services will be required to manage access control and compliance in the cloud?

The "identity-aware cloud service" is another thing to watch for. When identity becomes pervasive and portable across clouds (e.g., cross-domain authentication) a new level of granular access control can be deployed across the cloud. The cross-cloud security policies should be able to map sophisticated policies that go beyond a single cloud or domain (e.g., "user x can connect to service y that connects to service z ").

Today's cloud APIs are squarely focused on cloud service deployment and management, including provisioning and managing the life cycle of cloud resources (computing, storage, network). In the future, we'll see.

APIs encompass cloud user access management and role life cycle management functions leveraging industry standards including SAML, SPML, and XACML.

In addition to user-to-service authentication, service-to-service authentication and authorization frameworks will emerge. These frameworks will aid in delegated authorization without disclosing credentials. We are witnessing the genesis of flexible frameworks such as Microsoft claims-based authentication and the mash-up of OpenID and OAuth—a hybrid model where OpenID is used for federated login with the OAuth authorization process. In that loosely coupled model, authenticated users can be assigned a more granular artifact for authorization—a claim. This model helps developers to design applications and services so that they aren't tied to a particular credential type or to a particular set of roles. This will allow developers to externalize authentication and authorization from the application. Claims-based authentication also gives users more control as it allows users to reveal an appropriate level of user attributes based on user consent.

In this era of business consolidation where mergers and acquisitions are the norm, identity and access management solutions will become dynamic and flexible to meet the needs of a merged corporate entity or divested entities. In this scenario, the agile cloud-based identity and access processes anchored on "trusted relationships between domains" will obviate the need for any major architectural or costly implementations to reflect the changed access landscape and support new entitlement requirements.

Security Management

Today, given that a large segment of early adopters (SMBs) are solely focused on cloud service business benefits such as reduction of operational expense, elasticity, and on-demand service delivery, CSPs do not have the necessary market impetus to compete on service management support differentiators and capabilities. Enterprise customer adoption and standardization of application delivery models (compute and storage) will drive the need for fine granular instrumentation that offers a customer-specific view for services in the cloud.

To achieve a consistent service quality coupled with repeatability and predictability, customers will have no choice but to turn to automation and standardization on service management frameworks. The purported benefits of scalability and elasticity of cloud services can only be accomplished with strong management capabilities including centralized monitoring, provisioning, and configuration management practices. These practices have proven to deliver quality service for enterprise users and will continue to play a role in the cloud.

Although CSPs may not be able to offer a comprehensive set of management features and services, we believe that independent service providers (including start-ups) will be able to exploit market opportunities to deliver new cloud management services. We are witnessing the early stages of these services (e.g., Amazon's cloud watch services that offer visibility into resource utilization, operational performance, and overall demand patterns, including metrics such as CPU utilization, disk reads and writes, and network traffic). Driven by customer demand, more of these types of services will emerge, offered by either the CSP or certified third-party specialists who customize their offerings on the service provider platform. Hence, we will witness the emergence of a new breed of security-as-a-service offering that addresses security management issues including logging, security event management, vulnerability management, and incident response (e.g., Qualys's service offering of vulnerability management as a service).

Similar to the management standards that were established during the client/server computing era (e.g., Common Information Model, Java Management Extensions, Simple Network Management Protocol, and WS-Management), we will see the emergence of cloud management standards that facilitate unified management functions across CSPs. An example is a recent initiative from the Distributed Management Task Force standards body, called the Open Cloud Standards Incubator; the objective of the group is to standardize interactions between cloud environments by developing cloud resource management protocols, packaging formats, and security mechanisms to facilitate interoperability. (The scope of this activity is limited to the cloud resource management aspects of IaaS with some work touching on PaaS, including SLAs, quality of service, utilization, provisioning, and accounting and billing.) Another effort is driven by the Open Cloud Computing Interface workgroup, which is governed by the Open Grid Forum. The group's objective is to deliver an API specification for remote management of a cloud computing infrastructure, allowing for the development of interoperable tools for common tasks including deployment, autonomic scaling, and monitoring. The scope of the specification will be the high-level functionality required for the life cycle management of VMs (or workloads) running on virtualization technologies (or containers) supporting service elasticity.

In the future, we might see other standards organizations, such as ISO, the World Wide Web Consortium (W3C), the Organization for the Advancement of Structured Information Standards (OASIS), and the Internet Engineering Task Force (IETF) initiate new efforts to standardize management protocols that interoperate with many clouds. To accelerate enterprise cloud adoption, it is imperative that cloud management standards are created that will be supported by CSPs and that facilitate seamless interoperability across disparate clouds. Similar to the client/server era, standards will help to create an ecosystem of ISVs and service providers that provide customers with choice, flexibility, and greater agility by way of automation.

Privacy

It will be essential for the CSP to understand international privacy laws to comprehend how data can be transferred from one part of the world to the other. This was a challenge during the globalization of the world economy. It is unlikely that this will be resolved without some form of government intervention or the creation of a global privacy standard that will provide

consistency across jurisdictions. Such standards will help define the way businesses can leverage cloud computing.

Once cloud computing becomes more mainstream, the standard audit reports (e.g., SAS 70 Type II and SysTrust) augmented by specific requirements around privacy and security (such as the AICPA/CICA Generally Accepted Privacy Principles—GAPP) may suffice the audit concerns regarding handling of data and its privacy concerns. In the meantime, most organizations will have to rely on on-site audits, physical inspections, and reviews of security architectures until cloud computing becomes an accepted practice.

Audit and Compliance

It is likely that each CSP will define its own processes and controls (i.e., compliance), and in the short term this does not present a problem. However, as CSPs start to connect to each other and provide cross-CSP solutions, a uniform compliance framework will become more important to ensure that appropriate security measures are being consistently applied. The adoption of the IT GRC program would a good starting point to gain agreement on the adequacy of security measures since the discussion will be based on standards relevant to the CSP and its customers. Given the volume and multitenancy of cloud computing, the compliance program for CSPsneeds to be more real-time and have greater coverage than most traditional compliance programs.

Impact of Cloud Computing on the Role of Corporate IT

As adoption of cloud computing continues to grow, there will be a greater shift of IT functions and jobs from traditional corporate IT departments to CSPs. This will result not only in a downsizing of corporate IT departments, but also in a commoditization of IT functions (e.g., which CSP provides the best of service x) and jobs. For organizations, this will likely mean hiring fewer specialized IT personnel. Those IT personnel who are hired will likely not be actual practitioners, but managers or supervisors of the IT services provided by CSPs. It is likely that organizational costs spent on IT will decrease, as falling hardware costs will have to be passed on to customers at least partially by CSPs because of competition and fewer in-house IT personnel with skills demanding higher compensation than many other jobs. In addition, a shift in organizational payment for computing services from a centralized IT budget to businessunit budgets will lead to greater efficiencies in computing services used.

This will affect the IT profession itself. Custom applications will be developed less frequently,and only in very specialized cases (i.e., narrow or niche markets). Similarly, applications will likely be less customized. (However, there will be an increased demand for and increased competition from CSPs to provide greater personalization of applications offered by CSPs.) This will lead to fewer application developer positions. It is also likely that strong pressure by customers for open systems will result in fewer proprietary systems and fewer systems using proprietary languages, such as today's use of Apex by Salesforce.com or ABAP (Advanced Business Application Program) by SAP. Similarly, corporate IT departments are likely to hire far fewer system administrators, and such responsibility will shift to CSPs. And the growing number of servers maintained by CSPs will require a greater number of system administrators to be hired, in spite of increasing use of automated tools for configuration management. (Google alone is rumored to operate about 500,000 servers. And think about how many servers can fit into the 8 million square feet of data center space in which IBM Global Services operates?) There will also be a decrease in the number of network engineers needed by corporate IT departments, and again many of those jobs will shift to CSPs.

So, the future of corporate IT departments will see significant changes. First, the relationship between business units and corporate IT departments vis-à-vis CSPs will shift. Greater power will shift to business units from IT. Second, a number of functions performed today by corporate IT departments will shift to CSPs, along with corresponding job positions. Third, the functions performed by corporate IT departments will shift from those who do (i.e.,

practitioners who build or operate) to those who define and manage. And fourth, IT itself will become more of a commodity as practices and skills are standardized and automated. Dan Geer famously warned of a Microsoft monoculture. That has not occurred, but there will be less proprietary technology and less computing diversity simultaneously going forward.

Coclusion

Looking at cloud computing, it is important to step back and keep the big picture in view. What is really new here, and what changes impact security and privacy? Remember that cloud computing is a change in business models, and not a new technology. From an information security perspective, the single biggest change with cloud computing is the use of shared resources, or multitenancy. The impact of that change is that trust boundaries have moved. The real source of concern for information security practitioners is that it is not clear where those trust boundaries are now. With each level of the SPI delivery model the trust boundaries are different, and even within each level the trust boundaries change from provider to provider.

It is also important when looking at the security afforded by CSPs to keep your current (information security practitioner) perspective in mind. For information security professionals from large enterprises looking at the security afforded by CSPs, that security may very well look weak and even unacceptable in comparison with their current (large enterprise) security posture. However, for many information security professionals from SMBs looking at the security afforded by CSPs, that security may look acceptable and even better in comparison with their current (SMB) security posture. Where you “sit” may have a significant influence on your view of the security provided by CSPs.

That being said, going forward there is definitely a need for greater transparency by CSPs regarding their security practices, and to document those efforts through auditing for the benefit of their customers. To that end, the existing, commonly used auditing framework, SAS 70, is really no longer adequate for cloud computing audit purposes. Thankfully for all, that framework is now being updated to better reflect changed needs.

However, greater transparency alone will not be sufficient for improving the levels of security that are needed in cloud computing. There need to be significant improvements in security technology as well. Those improvements are needed in both preventive (proactive) controls and detective (reactive) controls.

IAM technology is really not acceptable in today’s non-cloud computing environments. IAM today fails to provide an adequate solution in enterprise environments. Significant IAM improvements are needed for cloud computing, and hopefully the IDaaS business model will spur those changes. Failure to do so will hamper the growth of public cloud computing.

Similarly, today’s (encryption) key management capabilities cannot even meet today’s enterprise requirements. Expecting those same technologies to scale to the cloud, and to provide easy-to-use management of complex needs, is simply wishful thinking. A radical improvement in key management capabilities is needed to meet cloud computing demands. Failure to do so will hamper the growth of cloud computing.

With regard to security monitoring, SIEM technologies are barely able to meet today’s large enterprise needs. It simply is not realistic to expect that today’s SIEM solutions will be able to scale to the cloud level. Additionally, the whole approach to SIEM probably needs to be revisited with regard to its ability to handle intercloud monitoring. Cloud customers are already demanding cloud portability through an open cloud API, and will shortly be demanding to use multiple clouds simultaneously. That demand for multiple cloud use simultaneously will break today’s approach to SIEM.

As more and more data is put into public clouds, customers will demand greater efforts by CSPs to protect their data. Those customers who happen to be large enterprises will be looking at their own DLP efforts, and demanding the same of their CSPs. With far

greater data volume transfers than their current gateways handle, and an increasing volume of encrypted network traffic, it is doubtful that today's DLP solutions will prove effective in cloud computing.

Do these deficiencies in current information security technologies mean the demise of the technologies or of cloud computing itself? No, definitely not. However, these deficiencies do mean that many customers are likely to be unsatisfied with CSP security efforts in the short term. These current deficiencies also mean an opportunity for information security vendors— and for new information security start-ups looking to shake up the current approaches that today's technologies provide.

A great part of the concern today about cloud computing security and privacy is based on unfamiliarity—it's new, and not enough people understand it well enough to make informed judgments. Real security issues for and by CSPs absolutely exist today. However, better understanding, greater transparency, and better security technology capabilities going forward mean that the hue and cry of today over the cloud's lack of security will soon fade, and will become yesterday's concern.

References

1. "The Cloud Wars: $100+ billion at stake," by Merrill Lynch, May 2008.
2. http://www.gartner.com/it/page.jsp?id=920712.
3. http://blogs.idc.com/ie/?p=224.
4. "Forecast: Cloud Computing Looms Big on the Horizon," by CIO Focus, October 2008.
5. Predicate encryption is a form of asymmetric encryption where encrypted data can be selectively decrypted by different individuals (or groups) without having to decrypt all of the encrypted data. "Predicate Encryption Supporting Disjunctions, Polynomial Equations, and Inner Products," by JonathanKatz, Amit Sahai, and Brent Waters, at http://eprint.iacr.org/2007/404.pdf.
6. http://www-03.ibm.com/press/us/en/pressrelease/27840.wss.

XACML Engine

In XACML, core logic related to policy evaluation resides in a software component called "XACML Engine". "WSO2 Identity Server" has a XACML (Sun – XACML) engine embedded.

XACML engine takes two inputs and gives a single output. Inputs it is taking are "XACML Policies" and a so called "XACML Request". Output per request can be one of the following;

- **Permit** – Request is evaluated against all applicable policies given to XACML engine and request is authorized to carry on the operation/actions
- **Deny**– Request is evaluated against all applicable policies given to XACML engine and request is not authorized to carry on the operation/actions
- **Not Applicable** – XACML engine didn't find any applicable policy for given request and request is not evaluated in XACML engine

XACML Policies

XACML engine can take multiple policies and evaluate request against all those policies. XACML engine evaluates policies independently. Result of evaluating a single policy is same as the results stated in previous section (Permit, Deny, and Not Applicable). So now the problem is how output from multiple policies is combined together. For that we have something know as "Policy Combining Algorithm". This algorithm is responsible for combining outcome from multiple policy evaluations.

- **permit-override** – If at least one policy is evaluated as "permit", the integrated output will also be "permit".
- **deny-override** – If at least one policy is evaluated as "deny", the integrated output will also be "deny".

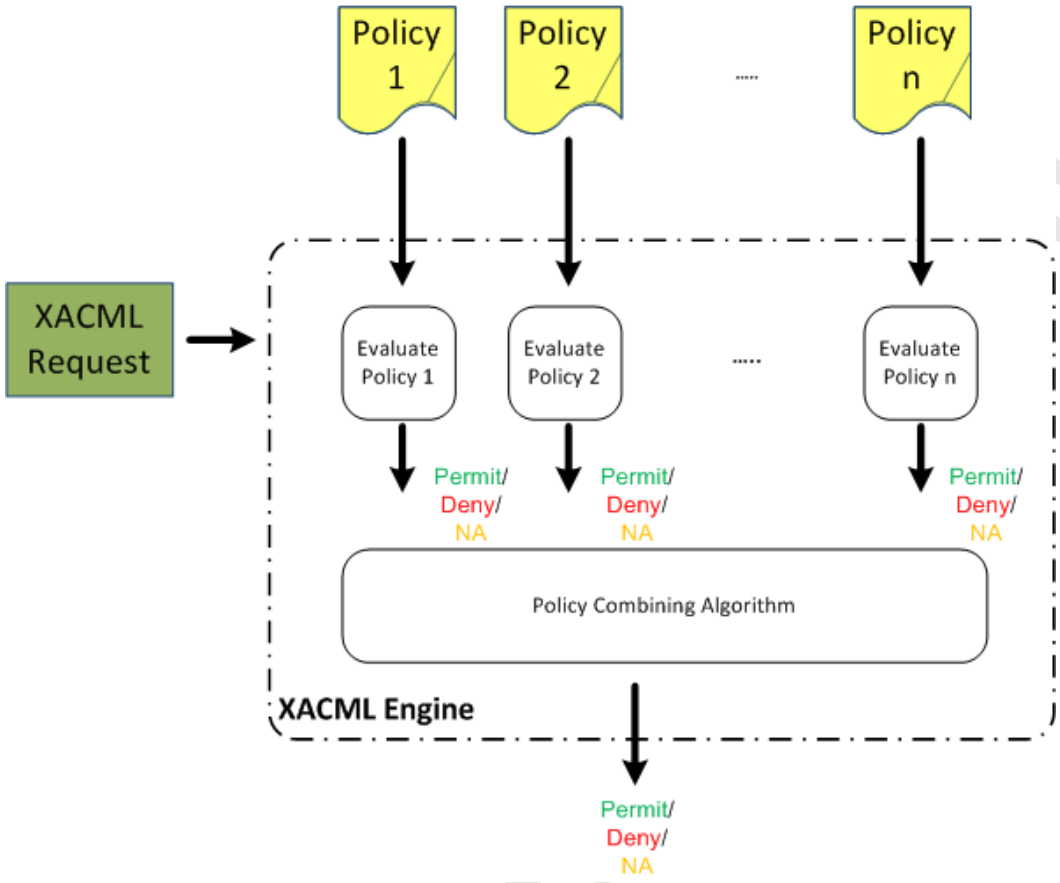


Figure 1, depicts the above mentioned behavior.

XACML Request

A XACML request contains information necessary to take authorization decision. Basically it contains attribute names and attributes values. When evaluating the policy, attribute names and attribute values are compared according to criteria defined in the policy.

Applicability of XACML Policies

XACML engine will get all policies as inputs. But it might not be necessary to evaluate all of them for a given request. **So how does XACML engine filter applicable policies for a given request?**

The logic which says whether a given policy is applicable to a request is defined in the policy itself. Each policy has a XML element known as "**Target**". The Target element's attribute values are matched with the incoming request attribute values. If those are matched with each other, XACML engine decides that the request should be evaluated against the complete policy.

XACML Policy Elements

Now you should have a basic idea how XACML engine processes XACML policies. Now let's focus on how actual evaluation is taking place.

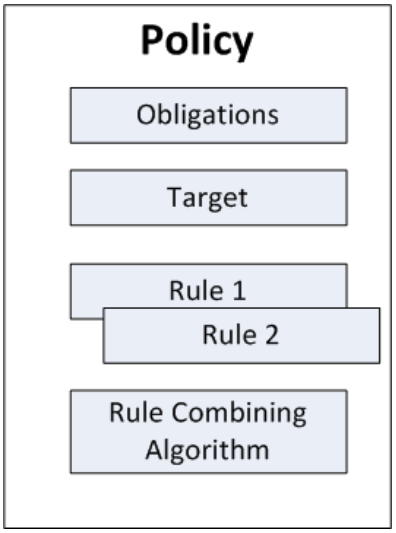


Figure 2, Mainly a XACML policy consists of 4 sub-elements

Target Element

Every XACML policy should have a "Target" element. As described in "Applicability of XACML Policies" section, "Target" element decides whether a particular policy is applicable to a given request. The "Target" element contains 4 sub-elements. They are depicted in Figure 3.

Figure 3

The sub-element values are compared against the XACML request values. When comparing, all sub-elements are taken into account. If the request attributes match with "Target's attributes, the policy will be further evaluated else XACML engine decides that given XACML request is not applicable to the policy. Let's look at what each of sub-elements represent.

Subjects – This is a parent element which contains 1 or more "**Subject**" elements. A **Subject** element usually represents the identity of the entity, which performs an action. Within the "**Subject**" element we need to define matching criteria for policy. For that we are using another sub-element called "**SubjectMatch**". Within the "**SubjectMatch**" we define the logic to match elements. A **SubjectMatch** element contains 2 parameters;

- Value of the subject attribute

- Name of the subject attribute

We can directly specify comparing attribute value with the relevant data type (in the policy). When XACML needs to retrieve attribute values from incoming Subject (in request), it uses "**SubjectAttributeDesignator**". Similar to SubjectAttributeDesignator, there are "**ResourceAttributeDesignator**"s, "**EnvironmentAttributeDesignator**"s and "**ActionAttributeDesignator**"s. In the attribute designator we specify the fully qualified name of the attribute and its type. Though, this looks complicated it is essential retrieving attribute values from the request.

So what are attribute designators?

Attribute designators instruct XACML run environment to look for values from the XACML request. As described in an earlier section, there are 4 types of attribute designators. They are "SubjectAttributeDesignators", "ResourceAttributeDesignators", "EnvironmentAttributeDesignators" and "ActionAttributeDesignators". All attribute designators instruct XACML engine to look for values from the XACML request. "SubjectAttributeDesignators" instructs XACML engine to look only for values within "Subject" element. Similarly "ResourceAttributeDesignators" will only look for "Resource" and "ActionAttributeDesignators" will only look for "Action" in the XACML request. Same pattern is applied to "EnvironmentAttributeDesignators".

Rule Element/s

A policy can have one or more rules. While target element evaluates the applicability of a policy, rule elements implement the actual authorization logic. Structure of a rule is depicted in Figure 4

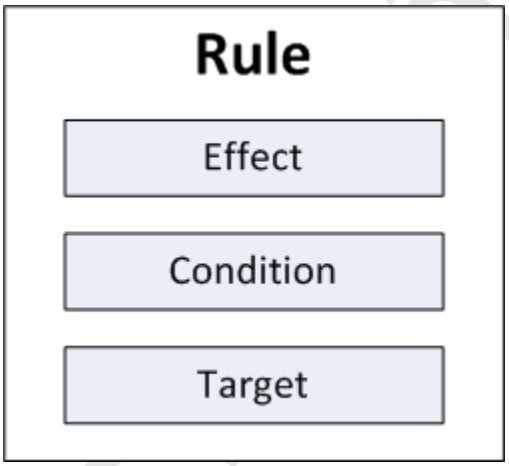


Figure 4

Target – A policy can have multiple rules. But it is not necessary to evaluate all such rules for a given request. A rule has a target element similar to policy's target element. Role of this target element is to decide whether a rule should be evaluated or not for a given request. If there isn't a target, the rule will be evaluated for all requests applicable to policy.

Condition – You can treat the condition as the core element of a rule. Within the condition we specify the exact authorization logic which always contains a Boolean expression. Based on the outcome of Boolean expression, the rule will be evaluated as true or false. Within the condition element we can use certain functions to implement authorization logic.

Rule Combing Algorithm

In a policy we can have many rules. Each rule says whether a request is permitted/denied or "not applicable". When we have multiple rules, how can we decide the final outcome of a policy? For that policy defines a "**rule combining algorithm**". Rule combining algorithm is quite similar to "policy combining algorithm" described above (In XACML Policies section). In fact algorithms are also same as in "Policy combining algorithms". i.e. "permit overrides", "deny overrides" etc …

Obligations

This is the least used element in a policy. This is an optional element and it is used to invoke certain actions upon a policy evaluation.

E.g.:- If policy is evaluated to "permit" send a mail to head of the department. This can be used to audit certain authorization logic evaluations. But due to its performance reasons this is not widely used in practical scenarios.

Use Case

I have a web service called "BankService". "BankService" has 2 operations.

1. **deposit (int accountNumber, double amount)**
2. **withdrew (int accountNumber, double amount)**

The deposit operation is allowed to "managers" and "executives". But withdrew operation is only allowed to "managers". Thus the "BankService" web service is deployed in "wso2.org" domain. "Manager" and "Executive" is defined as roles in the system.

Implementation

Steps 1 - Let's first try to understand the attributes and resources in above use case.

The use case says that above authorization is only applied to "wso2.org" domain. So we can define that in policy's target.

Resources in use case are;

1. **BankService/deposit** – Deposit operation
2. **BankService/withdrew** – Withdrew operation

Since we are always invoking a web service we can treat action as "**execute**".

Subjects are "**Manager**" and "**Executive**".

Steps 2 - Define the XACML request.

We will build the XACML request based on data gathered in Step 1.

```
<Request>
    <Subject SubjectCategory="urn:oasis:names:tc:xacml:1.0:subject-category:access-subject">
        <Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:subject:subject-id"
                   DataType="http://www.w3.org/2001/XMLSchema#string">
            <AttributeValue>manager</AttributeValue>
        </Attribute>
    </Subject>
    <Resource>
        <Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:resource:resource-id"
                   DataType="http://www.w3.org/2001/XMLSchema#string">
            <AttributeValue>BankService/withdrew</AttributeValue>
        </Attribute>
    </Resource>
    <Action>
        <Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:action:action-id"
                   DataType="http://www.w3.org/2001/XMLSchema#string">
            <AttributeValue>execute</AttributeValue>
        </Attribute>
    </Action>
    <Environment>
        <Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:environment:environment-id"
                   DataType="http://www.w3.org/2001/XMLSchema#string">
            <AttributeValue>wso2.org</AttributeValue>
        </Attribute>
    </Environment>
</Request>
```

Figure 5

A sample XACML request which suits above use case is defined in Figure 5. The request in Figure 5 will arrive in to the system when a “manager” tries to “withdraw” money. In a complete setup there will be a component which will craft above request. In XACML terms we call that component, “**Policy Enforcement Point** ”.

Steps 3 – Define policy target.

According to use case description, the authorization logic should be evaluated only if request’s domain is identical to “**wso2.org** ”. Therefore our policy should be evaluated only if request’s environment-id is equal to “wso2.org”. We need to define our policy target to cater this.

```
<Target>
 <Environments>
  <Environment>
   <EnvironmentMatch MatchId="urn:oasis:names:tc:xacml:1.0:function:string-equal">
    <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">wso2.org</AttributeValue>
    <EnvironmentAttributeDesignator
         AttributeId="urn:oasis:names:tc:xacml:1.0:environment:environment-id"
         DataType="http://www.w3.org/2001/XMLSchema#string"/>
   </EnvironmentMatch>
  </Environment>
 </Environments>
</Target>
```

Figure 6

Figure 6 depicts the policy target, which we defined for the use case. As you can see we are comparing value "**wso2.org** " with the value in request. To retrieve XACML request value we are using an "**EnvironmentAttributeDesignator** ".

Steps 4 – Define Rules

As per, use-case description, we can come up with 3 rules.

Rule 1 - If user is a **manager** or an **executive** allow access to resource "**BankService/deposit** ".

Rule 2 - If user is a **manager** only allow access to resource "**BankService/withdrew** ".

Rule 3 - If user is neither a **manager** nor an **executive** do not allow access to any of the resources.

We can specify rule 1 as following Boolean expression.

And (isResourceEqual("**BankService/deposit** "), Or(isManager(), isExecutive())

The XACML rule which depicts above Boolean expression is shown in Figure 7. Here we are using in-built functions in XACML engine. If the rule is evaluated to true we should "permit" the operation.

```
<Rule Effect="Permit" RuleId="Rule_On_Deposit">
 <Condition>
  <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:and">
   <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-is-in">
    <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">BankService/deposit</
AttributeValue>
    <ResourceAttributeDesignator AttributeId="urn:oasis:names:tc:xacml:1.0:resource:resource-id"
                                 DataType="http://www.w3.org/2001/XMLSchema#string"/>
   </Apply>
   <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:or">
     <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-is-in">
      <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">manager</AttributeValue>
      <SubjectAttributeDesignator AttributeId="urn:oasis:names:tc:xacml:1.0:subject:subject-id"
                           DataType="http://www.w3.org/2001/XMLSchema#string"/>
     </Apply>
     <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-is-in">
      <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">executive</AttributeValue>
      <SubjectAttributeDesignator AttributeId="urn:oasis:names:tc:xacml:1.0:subject:subject-id"
                                  DataType="http://www.w3.org/2001/XMLSchema#string"/>
     </Apply>
    </Apply>
  </Apply>
 </Condition>
</Rule>
```

Figure 7

Equivalent Boolean expression for Rule 2 is as follows;

And (isResourceEqual(“**BankService/withdrew** ”), isManager())

Figure 8 depicts relevant rule for above Boolean expression.

```
<Rule Effect="Permit" RuleId="Rule_On_Withdrew">
 <Condition>
  <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:and">
   <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-is-in">
    <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">BankService/withdrew</
AttributeValue>
    <ResourceAttributeDesignator AttributeId="urn:oasis:names:tc:xacml:1.0:resource:resource-id"
                                 DataType="http://www.w3.org/2001/XMLSchema#string"/>
   </Apply>
   <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-is-in">
    <AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">manager</AttributeValue>
    <SubjectAttributeDesignator AttributeId="urn:oasis:names:tc:xacml:1.0:subject:subject-id"
                                DataType="http://www.w3.org/2001/XMLSchema#string"/>
   </Apply>
  </Apply>
 </Condition>
</Rule>
```

Figure 8

To implement Rule 3, we can use an empty deny rule and a rule combining algorithm. Since we set effect of other rules to be “permit” we can set rule combining algorithm to “permit-override” (urn:oasis:names:tc:xacml:1.0:rule-combining-algorithm:permit-overrides). Empty “deny” rule would be as follows;

<Rule Effect="Deny" RuleId="Deny_Rule"/>

Steps 5 – Upload policy to “WSO2 Identity Server”

Save policy to a file.

Start “WSO2 Identity Server” and navigate to Entitlement -> Administration. Then select “**Import New Entitlement Policy**” from the appearing screen.

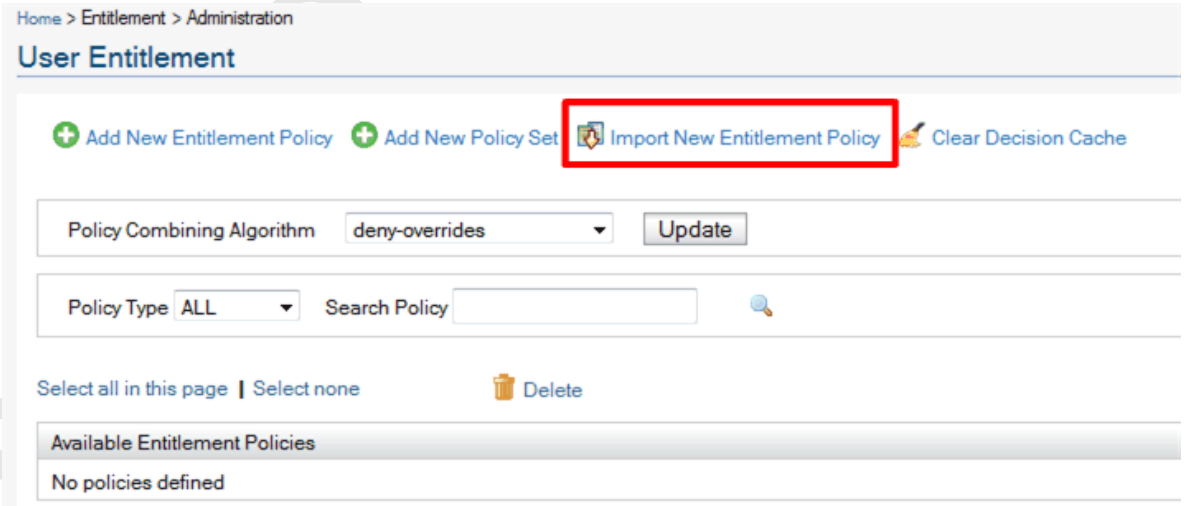


Figure 9

From next screen select "**Import Entitlement Policy from FileSystem**" and browse and give the file location of the policy file. Then upload the policy.

Once policy is uploaded, it will be in **disable** state. **Therefore, you need to enable the policy as soon as you upload.**

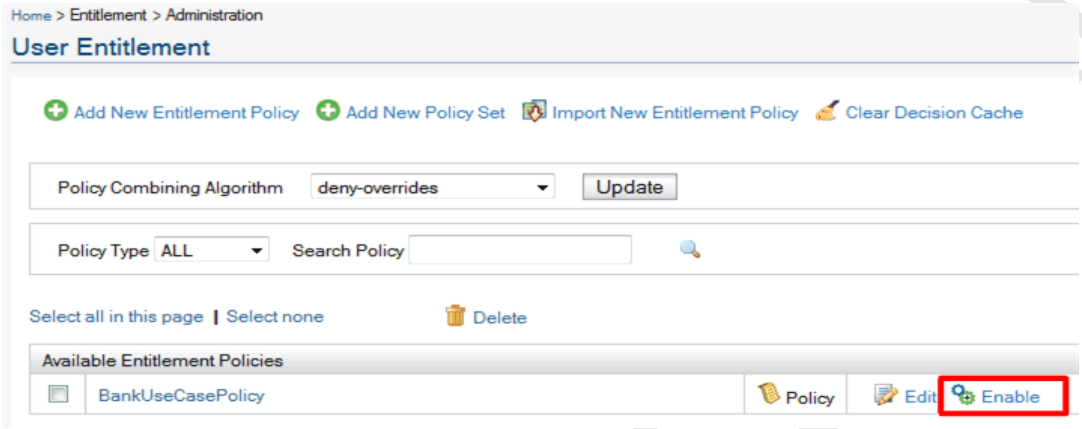


Figure 10

Step 6 – Testing

To test XACML policy you can use the "**TryIt**" functionality in "WSO2 Identity Server". Navigate to "Entitlement -> TryIt".

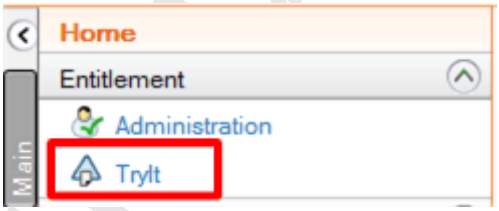


Figure 11

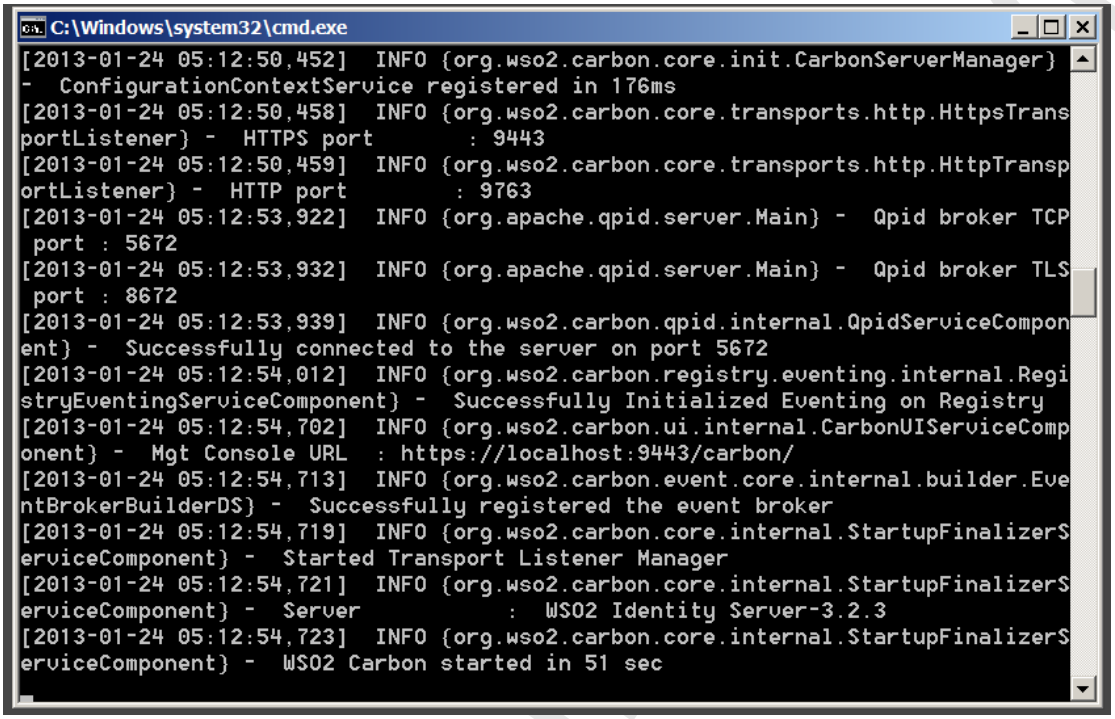
In the request editor you can create the request, evaluate and validate the policy accuracy.

For this use case we will use 4 requests to test.

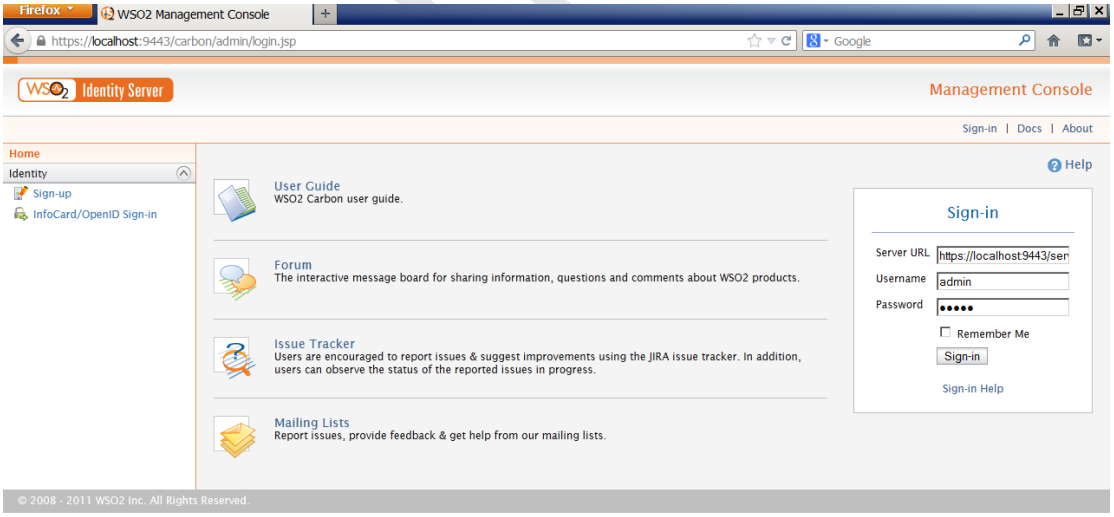
1. Test Deposit Rule – For this you can use **request-deposit.xml**. Copy and paste the content of request-deposit.xml into request editor and evaluate. The evaluation should **permit** the action.
2. Test Withdrew Rule – **Use request-withdrew.xml**. This test is similar to deposit test. Here we are specifying only the manager. Request **request-withdrew-deny.xml** should be **denied** as we are using "executive" as the subject and executive is not allowed to withdraw money.
3. Negative scenario – Use **request-negative.xml**. A cashier is trying to authorize. The output should be "**deny** ".
4. Not Applicable – Use **request-na.xml**. Here we are specifying a different domain (test.org) as oppose to wso2.org. So the output should be "**Not Applicable** ".

Screenshots

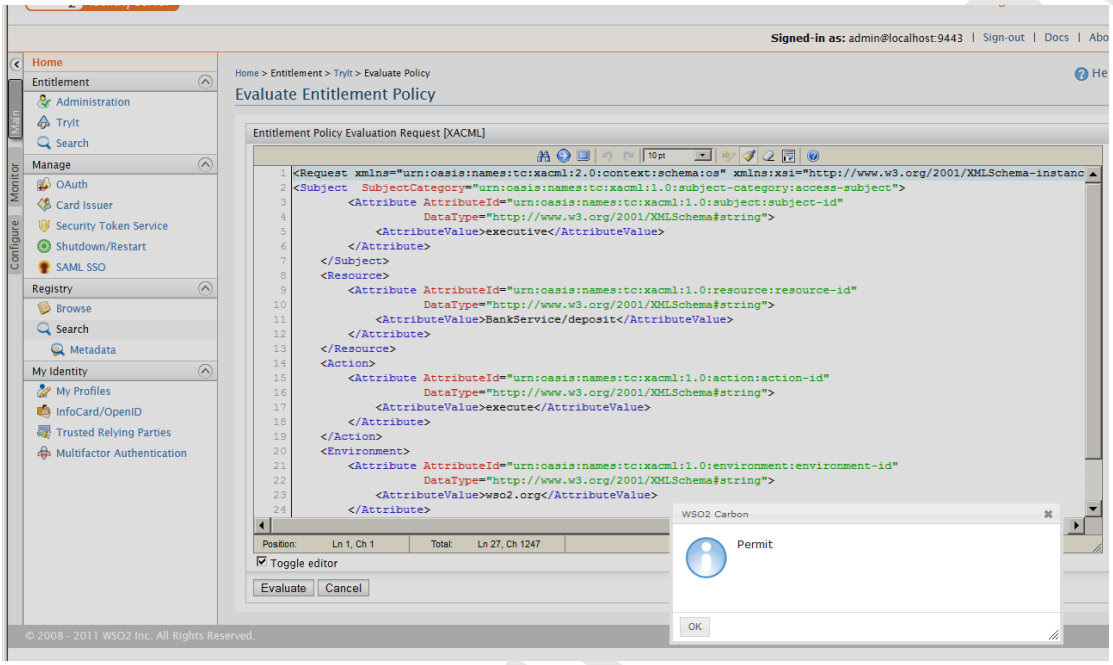
Initialize wso2 Identity Server

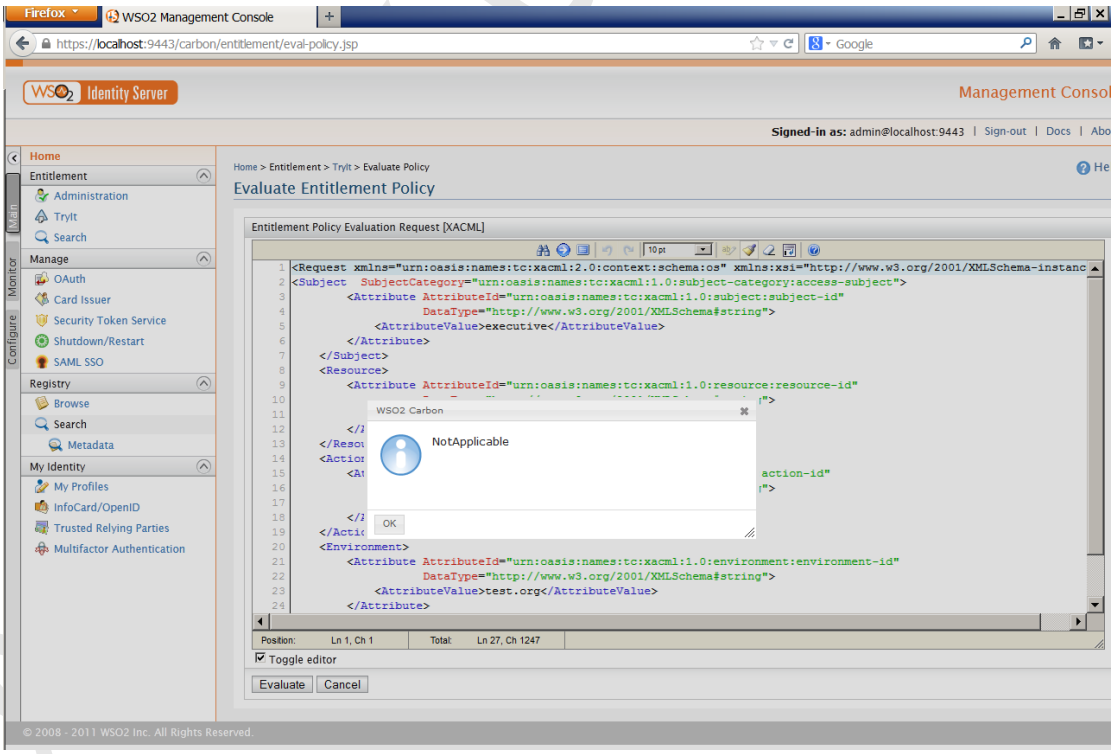


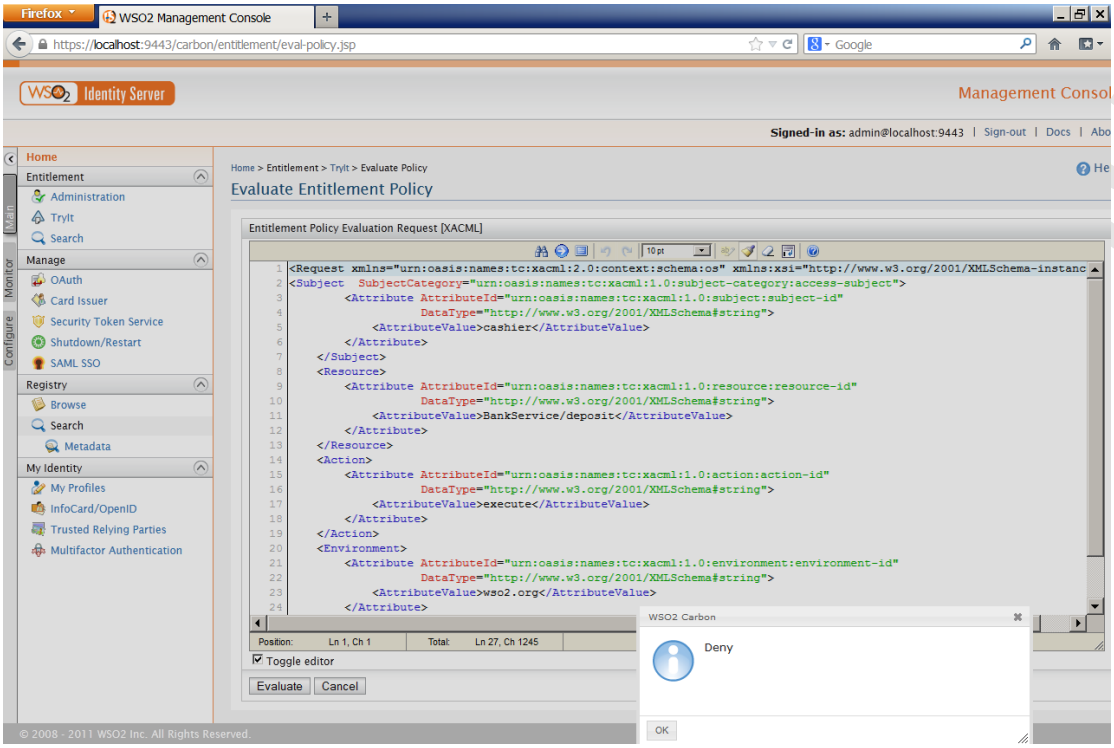
Log in to the Management url of the server

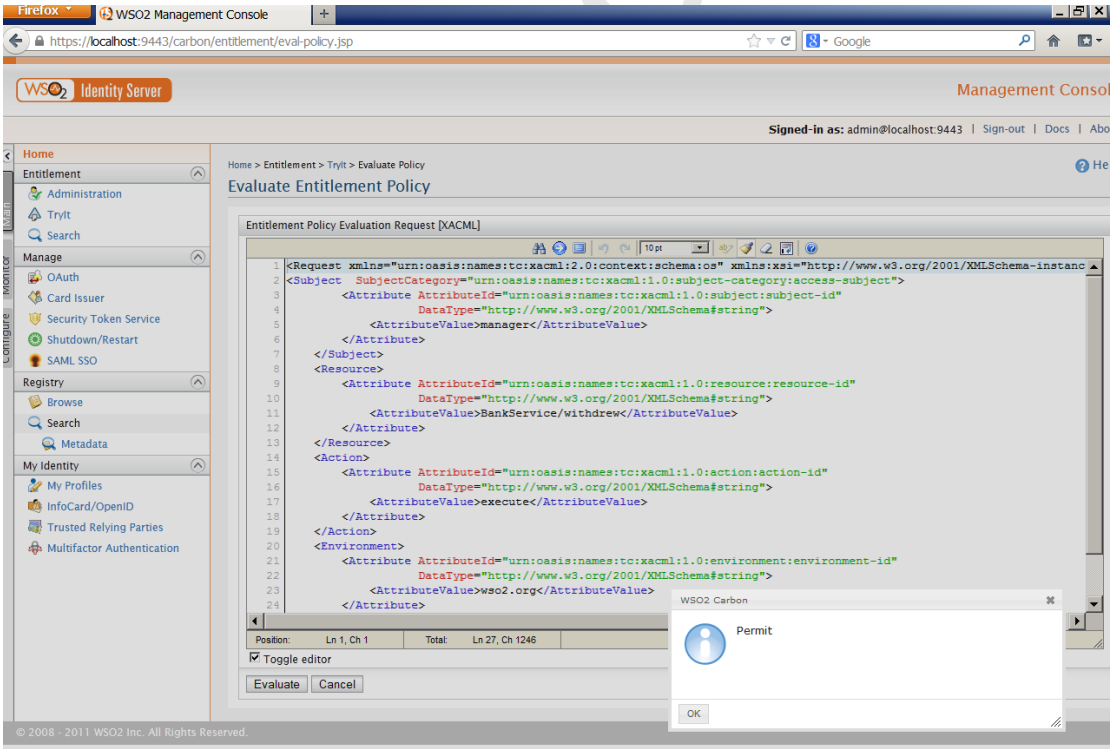


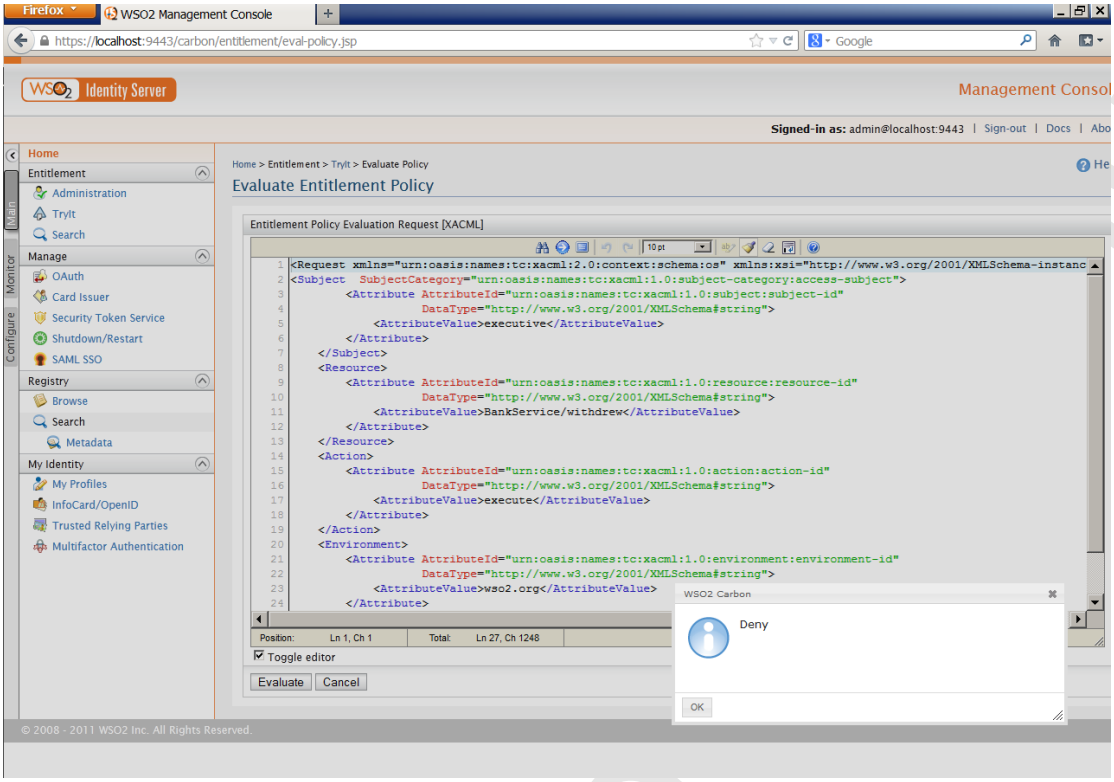
Testing the implemented policy for the different requests according to the different scenarios






WSO2 Management Console
https://localhost:9443/carbon/entitlement/eval-policy.jsp
Identity Server
Evaluate Entitlement Policy
Entitlement Policy Evaluation Request [XACML]
<Request xmlns="urn:oasis:names:tc:xacml:2.0:context:schema:os" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instanc
<Subject SubjectCategory="urn:oasis:names:tc:xacml:1.0:subject-category:access-subject">
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:subject:subject-id"
DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>cashier</AttributeValue>
</Attribute>
</Subject>
<Resource>
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:resource:resource-id"
DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>BankService/deposit</AttributeValue>
</Attribute>
</Resource>
<Action>
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:action:action-id"
DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>execute</AttributeValue>
</Attribute>
</Action>
<Environment>
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:environment:environment-id"
DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>wso2.org</AttributeValue>
</Attribute>
Toggle editor
Evaluate
Cancel
WSO2 Carbon
Deny
OK


WSO2 Management Console
https://localhost:9443/carbon/entitlement/eval-policy.jsp
Identity Server
Evaluate Entitlement Policy
Entitlement Policy Evaluation Request [XACML]
<Request xmlns="urn:oasis:names:tc:xacml:2.0:context:schema:os" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instanc
<Subject SubjectCategory="urn:oasis:names:tc:xacml:1.0:subject-category:access-subject">
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:subject:subject-id"
DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>manager</AttributeValue>
</Attribute>
</Subject>
<Resource>
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:resource:resource-id"
DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>BankService/withdrew</AttributeValue>
</Attribute>
</Resource>
<Action>
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:action:action-id"
DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>execute</AttributeValue>
</Attribute>
</Action>
<Environment>
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:environment:environment-id"
DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>wso2.org</AttributeValue>
</Attribute>
Toggle editor
Evaluate
Cancel
WSO2 Carbon
Permit
OK


WSO2 Management Console
https://localhost:9443/carbon/entitlement/eval-policy.jsp
Identity Server
Management Consol
Signed-in as: admin@localhost:9443 | Sign-out | Docs | Abo
Home
Entitlement
Administration
TryIt
Search
Manage
OAuth
Card Issuer
Security Token Service
Shutdown/Restart
SAML SSO
Registry
Browse
Search
Metadata
My Identity
My Profiles
InfoCard/OpenID
Trusted Relying Parties
Multifactor Authentication
Home > Entitlement > TryIt > Evaluate Policy
Evaluate Entitlement Policy
Entitlement Policy Evaluation Request [XACML]
<Request xmlns="urn:oasis:names:tc:xacml:2.0:context:schema:os" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instanc
<Subject SubjectCategory="urn:oasis:names:tc:xacml:1.0:subject-category:access-subject">
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:subject:subject-id"
DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>executive</AttributeValue>
</Attribute>
</Subject>
<Resource>
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:resource:resource-id"
DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>BankService/withdrew</AttributeValue>
</Attribute>
</Resource>
<Action>
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:action:action-id"
DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>execute</AttributeValue>
</Attribute>
</Action>
<Environment>
<Attribute AttributeId="urn:oasis:names:tc:xacml:1.0:environment:environment-id"
DataType="http://www.w3.org/2001/XMLSchema#string">
<AttributeValue>wso2.org</AttributeValue>
</Attribute>
Position: Ln 1, Ch 1
Total: Ln 27, Ch 1248
Toggle editor
Evaluate
Cancel
WSO2 Carbon
Deny
OK
© 2008 - 2011 WSO2 Inc. All Rights Reserved.