TAKING THE HIGH GROUND:
A CASE FOR DEPARTMENT OF DEFENSE APPLICATION OF PUBLIC CLOUD COMPUTING

GRADUATE RESEARCH PROJECT

Kris E. Barcomb, Major, USAF

AFIT/ICW/ENG/11-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

The views expressed in this report are those of the author and do not reflect the official policy or position of the United States Air Force, the Department of Defense, or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT/ICW/ENG/11-01

Taking the High Ground: A Case For Department Of Defense Application Of Public Cloud Computing

GRADUATE RESEARCH PROJECT

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Cyber Warfare

Kris E. Barcomb

Major, USAF

June 2011

AFIT/ICW/ENG/11-01

Taking the High Ground:
A Case For Department Of Defense Application Of Public Cloud Computing

Kris E. Barcomb
Major, USAF

Approved:

/signed/ 12 May 2011

Robert F. Mills, PhD (Chairman) date

/signed/ 12 May 2011

Jeffrey W. Humphries, Lt Col, PhD (Member) date

AFIT/ICW/ENG/11-01

Abstract

Cloud computing offers tremendous opportunities for private industry, governments, and even individuals to access massive amounts of computational resources on-demand at very low cost. Recent advancements in bandwidth availability, virtualization technologies, distributed programming paradigms, security services and general public awareness have contributed to this new business model for employing information technology (IT) resources. IT managers face tough decisions as they attempt to balance the pros and cons of integrating commercial cloud computing into their existing IT architectures. On one hand, cloud computing provides on-demand scalability, reduces capital and operational expenses, decreases barriers to entry, and enables organizations to refocus on core competencies rather than on IT expertise. In spite of the benefits, security concerns are still the dominant barriers to cloud service adoption. This research explores public cloud computing services from a Department of Defense (DoD) perspective. The objectives are to improve the general understanding of cloud computing; describe its potential benefits to the DoD; examine public cloud computing adoption from a risk management perspective; present threats specific to public cloud computing; and provide a set of recommendations to help foster public cloud computing adoption within the DoD. In addition to advocating for incorporating public cloud computing into the DoD enterprise, this research also presents how it could be used by our adversaries to launch sophisticated attacks.

Acknowledgements

First and foremost, I want to thank God for giving me a tremendous number of blessings in my life, especially my beautiful wife and two lovely daughters. They are my world and none of this would be possible without their unending love and support. I'd also like to recognize my parents and grandparents for instilling in me a spirit of dedication, fostering a lifetime of learning, and helping me recognize the importance of always doing my best. The impetus for this work wouldn't be possible without the support of my previous boss and great friend, Dr. Pete Rustan, who entrusted me with a tremendous amount of professional freedom and responsibility. A portion of that trust allowed me to establish programs related to cloud computing, which has since become a passion of mine. I want to thank Dr. Robert Mills and Lt Col Jeffrey Humphries for their instruction, research guidance and feedback throughout the process of generating this document. Finally, its hard to describe how much fun it's been to be a part of such a unique program surrounded by an incredible group of classmates. In the short time we've had here at AFIT, I'm amazed at the close bonds that we've created both professionally and personally. Thank you all very much!

Kris E. Barcomb

Table of Contents

Page

List of Figures

Taking the High Ground:
A Case For Department Of Defense Application Of Public Cloud Computing

I. Introduction

> "What happened to the generation of power a century ago is now happening to the processing of information. Private computer systems, built and operated by individual companies, are being supplanted by services provided over a common grid–the Internet–by centralized data-processing plants. Computing is turning into a utility, and once again the economic equations that determine the way we work and live are being rewritten."
> - Nicholas Carr

> "The Federal Government's current Information Technology environment is characterized by low asset utilization, a fragmented demand for resources, duplicative systems, environments which are difficult to manage, and long procurement lead times. These inefficiencies negatively impact the Federal Government's ability to serve the American public." - Vivek Kundra

1.1 Making the Case for Public Cloud Computing

Cloud computing offers tremendous opportunities for private industry, governments, and even individuals, to access massive amounts of computational resources on-demand at very low cost. Recent advancements in bandwidth availability, virtualization technologies, distributed programming paradigms, security services and general public awareness have contributed to this new business model for employing information technology (IT) resources. The industry has advanced to the point where companies are now offering processing, system memory, storage and bandwidth as services over the Internet. For many applications, organizations and individuals no longer need to procure, operate and maintain their own compute infrastructures. Instead, they can have more flexible solutions provisioned from public providers at a fraction of the cost of a private deployment.

The business model for cloud computing is similar to that of a public utility. Central providers assume responsibility for the substantial capital investment, and the operations and maintenance functions for large IT infrastructures. Consumers pay for access to those resources according to the amount they use. The resources are provided, usually over the public Internet, to consumers on-demand. Providers deliver low-cost solutions by distributing their own costs across the large volume of consumers who benefit from their services.

Some have compared the current state of public cloud computing to the transition from localized electricity generation to centralized power stations that began in the late 1800s. [1] In many ways, this is a good analogy. Prior to Thomas Edison and Nikola Tesla inventing efficient mechanisms to transfer power over large distances, organizations and individuals relied on local means of generating electricity such as gas generators or water wheels. This required that the operator not only understand their business, but also know how to keep the power generation systems running. In the case of the water wheel, it also restricted where these businesses could operate. Despite the promise of lower costs and increased flexibility, evangelizing the benefits of centralized power was not without its difficulty. Obviously, without power a businesses could not operate and many expressed concern over outsourcing such a critical component of their infrastructure to an external provider. Centralization raised concerns over single points of failure that could lead to outages. The safety and security of transmitting high voltage power across long distances was also debated. Now, fast forward a hundred years and imagine what life would be like if we had not overcome those fears and found ways to mitigate the risks of centralized power while still being able to reap its benefits.

Unfortunately, the DoD has been slow to adopt public cloud computing as a viable business model. This stems from some of the DoD's inherent characteristics. First, it is a military organization chartered to fight to defend our nation's freedoms. That mission and its history in executing it, have caused the DoD to structure itself into a strict hierarchy designed to develop soldiers capable of successfully prosecuting warfare in hostile, denied, or degraded environments. Second, system designers face a fundamental tension between building functionality or increasing security. The grave subject matter the DoD often deals with tends to foster a culture that pushes that balance toward security over functionality.

Like many information technologies, cloud computing is distributed. It promotes access to information, and by extension decision making ability, at the lowest levels. Whereas the DoD is designed to maintain a rigorous chain-of-command, cloud computing can foster creativity and innovation right down to the individual. This level of power, at the "edge," as Alberts and Hayes put it, [2] is more difficult to manage from the top down. The ultimate benefits of incorporating modern IT can often be difficult to quantify and hard to predict. This is a fundamental problem for all government organizations seeking to capitalize on what they observe in the commercial IT market, but are bound by a responsibility to manage tax payer dollars responsibly. In the military, most members are classically trained to be soldiers, even those whose job is primarily to manage information technologies. That training has allowed them to see the benefits that modern IT could have in the battlefield, but it doesn't necessarily help them understand how to acquire, manage and operate those technologies in the most effective way.

Many programs the DoD executes are indeed extremely sensitive and require special considerations. Unfortunately, the mindset that necessarily governs these unique programs, also pervades the rest of its IT infrastructure. The DoD is often risk averse and tends to look at program acquisition and management only from the perspective of the impact that a security incident would have on the organization. This culture often fails to consider the real likelihood of that security incident actually occurring. Fiscal pressures and increasing demand on our nation's forces will drive a more widespread adoption of risk management over risk aversion. Otherwise, our adversaries will outmaneuver us on the digital battlefield that is increasingly relevant to military operations.

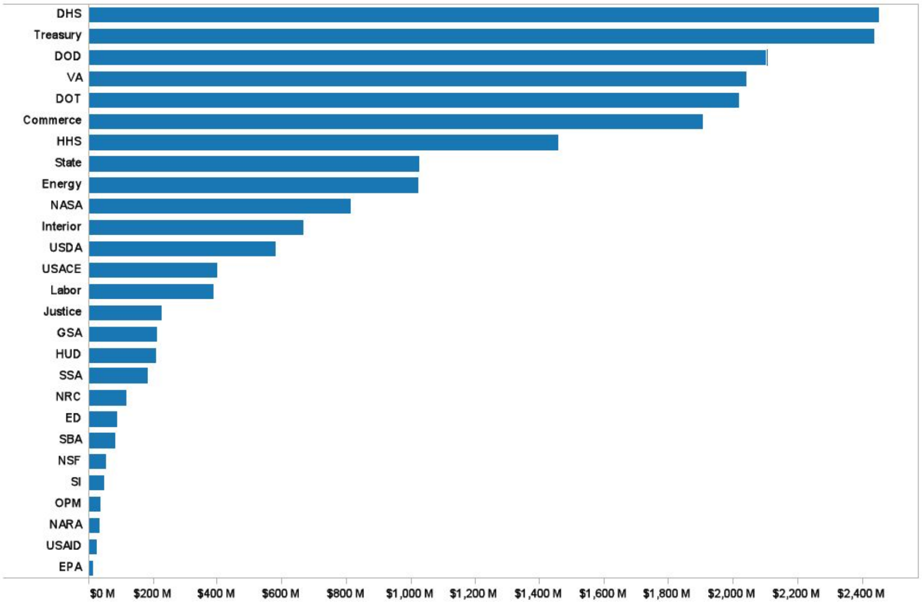

Figure 1: Potential Spending on Cloud Computing by Agency

Cloud computing offers an important opportunity for the DoD to increase its level of functionality at a lower cost with more design flexibility. Many federal agencies are actively incorporating commercial cloud computing into their own architectures. Their own budgetary realities, combined with an increased understanding of the promises of modern IT, have helped spur them down this path. Another driving factor is that the Federal Chief Information Officer (CIO), Vivek Kundra, and the National Institute of Standards and Technology (NIST) have been actively articulating the business case for cloud computing adoption, developing security standards for the technology and making cloud technologies more accessible within the federal government.

A recent report from the Federal CIO shows a number of examples depicting Federal agencies successfully employing cloud computing technologies. [3] The DoD can realize similar gains and must approach similar technologies with an open mind. The same report also highlights the DoD as the federal agency with the third largest potential spending on cloud computing. Figure 1 shows agency estimates for cloud computing spending as reported to the Office of Management and Budget (OMB).

The first step in successfully adopting cloud computing is to have a common understanding of what exactly it is. In this regard, NIST has made great strides in helping to clarify the terminology surrounding cloud computing. NIST partitions cloud computing into three service models: Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). It also defines four different deployment methods for those service models: Public, Private, Hybrid and Community.

This research is primarily concerned with public cloud computing offerings for two reasons. First, public providers benefit from the tremendous economies of scale of the commercial market place that they leverage to drive down their infrastructure costs. Second, competition drives innovation and continuous improvement. Government agencies deploying in-house solutions are unlikely to achieve comparable costs or maintain pace with advances in commercial IT. As Waxer observes, private infrastructures can rarely match the service levels offered by public cloud providers. [4]

In addition to defining cloud computing, NIST is taking a leading role in defining steps that government agencies can take to successfully manage the risk of adopting cloud computing while still reaping the benefits. They are actively producing standards and guidelines aimed at facilitating secure implementations of the technologies and business models involved in this area. As a result of their preeminent role, this research leans heavily on the documentation being produced by NIST and attempts to extrapolate their innovative look at cloud computing as it might apply to the DoD.

Other important themes of this research are education and trust. DoD employees must become more comfortable with modern information technologies in general. Only with a comprehensive understanding of the technology and its associated business models can the DoD successfully innovate in the digital world. Furthermore, commercial partnerships and the integration of off-the-shelf-technologies will continue to be increasingly important factors in maintaining a state-of-the-art Information Age military; therefore, the DoD must work closely with IT industry leaders and especially those commercial companies that are not traditionally associated with defense. Those companies are providing the rest of the nation with a continuous stream of innovation and the DoD must find ways to leverage capabilities coming from that sector.

1.2 Objectives

In addition to providing a general overview of cloud computing, this research explores the federal and DoD standards for security surrounding information technology programs and emphasizes the need for employing sound risk management processes in the decision to include public cloud computing in DoD information architectures. It also examines specific threats to public cloud computing so that program managers can understand root causes of risk and devise strategies to mitigate them. Trust will be a critical factor in the success or failure of public cloud computing and so examples of how the federal government and the commercial industry are working together to meet federal regulations and standards is also presented. The DoD must also recognize that the potential benefits of cloud computing can also be leveraged by our adversaries to execute sophisticated attacks. These attacks would not have been possible without access to these kinds of services, and so an overview of specific threats enabled by public cloud computing is also included. The research concludes with a set of recommendations for the DoD to help foster the adoption of public cloud computing services.

1.3 Organization

- *Background:* This research begins with a general overview of cloud computing as well as some of its key technological underpinnings. The information presented is largely based on the documentation being produced by NIST, since their framework for describing cloud computing has become the de facto standard in not only government circles, but also in the commercial sector.
- *Benefits of Public Cloud Computing:* After defining what cloud computing is, the benefits of cloud computing are described. Five particular benefits are

presented that are particularly relevant to the DoD. The chapter concludes with a set of examples to help the reader recognize these benefits.

- *Assessing Public Cloud Computing Security from a Risk Management Perspective:* This chapter focuses on the legal and policy considerations facing DoD IT managers who may wish to implement a portion of their enterprises by using public cloud computing providers. While there are a considerable number of rules that must be followed when protecting DoD information and implementing DoD information systems, the primary focus of this research is on the Federal Information Security Management Act (FISMA) and the thread of documentation that flows from it. NIST authoritative standards are presented along with guidelines from relevant special publications. Finally, applicable DoD Directives and Instructions are analyzed to show how public cloud computing could be used while maintaining the necessary safeguards.

  These laws and regulations exist to reduce the risks associated with information technologies. Ultimately, the value proposition of any solution must exceed the risk associated with its use. Within these boundaries, DoD IT managers must implement architectures that carefully balance functionality, cost, and security.

- *Threats to Public Cloud Computing:* As with any new technology, cloud computing carries with it a set of security challenges and potential threats. This chapter covers the natures of those threats that either present themselves primarily against cloud computing environments or threats that are exacerbated by the nature of cloud computing.

- *Threats Enabled by Public Cloud Computing:* Finally, whether or not the DoD decides to integrate the capabilities of public cloud computing into its IT enterprise, our adversaries will. Access to large-scale compute infrastructures has traditionally been reserved for nation-states, major corporations and sizable academic institutions. Cloud computing changes that by largely eliminating the accessibility barriers for massive computational power. Now, practically anyone with an Internet connection and a credit card can launch attacks that were infeasible only a few years ago. This chapter explores the kinds of attacks that public cloud computing enables so that the DoD can better prepare itself to defend against this relatively new threat.

Maintaining a technological edge is critical to the future of our fighting forces. Public cloud computing offers significant potential to assist the military in maintaining the state-of-the-art across its IT infrastructure. Many steps need to be taken, but the benefits of fully integrating public cloud computing cannot be ignored.

II. Background

> "Our industry is going through quite a wave of innovation and it's being powered by a phenomenon which is referred to as the cloud." - Steve Balmer

> "When it comes to Cloud Computing, it's important to understand that it is a 'disruptive technology' that requires a new foundation of knowledge to understand. With that said, we are just now beginning to understand this next-generation phenomena." - John Shea

The term "Cloud Computing" has been used in a variety of ways to describe many different concepts. [5] The multiplicity of definitions and perceptions of cloud computing confuse and detract from our ability to employ it effectively. Its origins come from the graphical representation of complicated network interactions as a "cloud" to assist in describing the flow of information at a high level without having to present all of the associated technical details. The diversity of definitions stems from the many ways individuals and groups approach describing those complex interactions. Often the definitions convey a specific meaning that only applies in the context of the individual using the term. Businesses have contributed to the confusion by appending the word "cloud" to an array of products as they try to capitalize on the expected market growth in the sector.

The best broad definition to date comes from NIST which describes cloud computing as a "model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models and four deployment models." [6]

2.1 Cloud Computing Essential Characteristics

The essential characteristics describe the nature of a true cloud computing implementation and help scope the requirements for implementing a cloud service offering. They also provide a framework of basic criteria for evaluating the quality of one implementation over another.

2.1.1 On-Demand Self-Service. This characteristic of cloud computing describes how resources are dynamically allocated to consumers upon request. The requests are driven by the consumers, and there is generally little-to-no human interaction required after the service level agreement (SLA) between the consumer and the provider is established and agreed upon.

2.1.2 Broad Network Access. Sufficient network availability and bandwidth are essential for employing cloud computing technologies. While similar computing models, such as grid and utility computing, have been around for some time, the ubiquitous nature of high-speed, high-availability connectivity has led to the rapid expansion of cloud computing as a mainstream capability.

2.1.3 Resource Pooling. Many cloud service providers make large quantities of resources available to consumers at extremely low costs. This is possible because they leverage the economies of scale associated with resource pooling. Providers employ technologies such as virtualization, dynamic provisioning, and load balancing to spread multiple customer requests across the entirety of their data centers. By leveraging the statistical nature of consumer demand over a large user population, providers can keep their data centers operating at high utilization rates.

2.1.4 Rapid Elasticity. Rapid elasticity describes how resources can quickly be added to, or removed from, the consumer's pool of allocated resources. It goes hand-in-hand with the on-demand self-service characteristic. The elastic nature of cloud-based resources allows consumers to operate with the exact amount of resources required to perform a given task. It helps streamline the acquisition processes and helps consumers avoid procuring physical assets that will likely go underutilized.

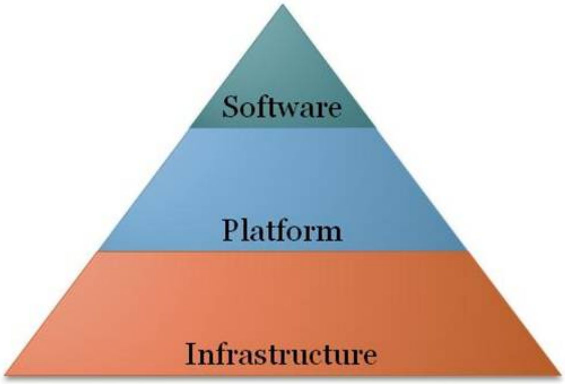

Figure 2: Cloud Computing Service Models

2.1.5 Measured Service. The final characteristic of cloud computing underpins its analogy to a public utility. Measured service enables providers to quantify the costs of allocated resources per unit time and subsequently charge a consumer for only what is used. The SLA between the consumer and the provider establish the length of the discrete time intervals and the nature of the services provided (e.g., 1 central processing unit (CPU) with 2 gigabytes (GB) RAM and 1 terabyte (TB) of storage = $1/hour).

2.2 Cloud Computing Service Models

The cloud computing service model describes the level of functionality offered by the provider. Figure 2 shows the relationships between the models represented as a layered pyramid. At higher levels, the consumer receives more functionality and knows less about the implementation details. The higher layers can also be built upon services received from lower levels. At lower levels, the consumer receives more granularity and higher levels of control. They are also increasingly responsible for the implementation details. The lowest levels can provide services that look and feel very much like locally owned and operated hardware.

2.2.1 Software as a Service (SaaS). At this highest layer, consumers gain access to fully functional applications without having to install or manage software on local systems. The SaaS provider is responsible for the implementation details as well as the compute resources required to execute the service. The provider also retains responsibility for maintenance and configuration management of the application. The consumer generally has little to no knowledge of how, or even where, the operations run. The strength of this model is that consumers can often operate these services from client platforms ranging from thick clients down to web-enabled mobile devices. Also, the state of the application can be seamlessly maintained from one device to another. Representative SaaS offerings include OnLive (`http://www.onlive.com`), an online gaming service; Google Docs (`http://docs.google.com`), a web-based document editing and management service; and Bing Maps (`http://www.bing.com/maps`), an online mapping service.

2.2.2 Platform as a Service (PaaS). PaaS providers offer a balance of functionality and control that falls in-between SaaS and IaaS. PaaS offerings are analogous to an application programming interface (API) where the complexities and redundancy associated with generating low-level code is abstracted away from the user. Consumers use PaaS to generate custom applications using software development languages and tools offered by the provider. The provider then hosts the resulting application on its own hardware provisioned to the client according to their needs. Examples include Google App Engine (`http://code.google.com/appengine`) and Microsoft Azure (`http://www.microsoft.com/windowsazure`).

2.2.3 Infrastructure as a Service (IaaS). The lowest level of service provides consumers with the most control over their environment. Amazon's Elastic Compute Cloud (EC2) (`http://aws.amazon.com/EC2`), Rackspace (`http://www.rackspace.com`), and GoGrid (`http://www.gogrid.com`) are all examples of public IaaS providers. In its purest form, IaaS provisions hardware resources directly and leaves most of the details of what to put on that hardware, including the operating system (OS), in the hands of the user. In practice, IaaS is often implemented using virtualization technologies serving to abstract the underlying hardware from the applications. The OS and applications are stored in a software bundle often referred to as a Virtual Machine Image (VMI). The consumer manages the contents of the VMI, while the provider manages which physical machines that execute the VMI as well as the specific characteristics of those machines. The user generally does not know the details of the underlying hardware. The description of the consumer's hardware performance needs is laid out in the SLA between the consumer and provider. The provider can meet the SLA by provisioning the VMI onto a portion of a resource (e.g. a single core of a multi-core system), an entire resource, or across multiple physical resources so long as they meet their obligations under the SLA.

2.2.4 X as a Service. The three tiers of cloud service types have been a boon to the cloud computing industry since NIST first formally presented them in 2009. Despite their help in demystifying the cloud, more details are needed to fully articulate the range of available services. In [7], the Federal CIO states that "these definitions need to be expanded to more comprehensively define a reference architecture and taxonomy to provide a common frame of reference for communication." To that end, other authors have attempted to further refine the categories of services offered in the cloud. In a recent book, Linthicum differentiates 11 different types of services. [8] His categories and their definitions are summarized here:

- *Storage as a Service*: Remotely accessible storage capacity.
- *Database as a Service*: Remotely managed and accessible database services.
- *Information as a Service*: Remotely hosted information (e.g. stocks, maps, and product prices) accessible through a well-defined interface.
- *Process as a Service*: Ability to bind many resources together, either hosted within the same cloud computing resource or remotely, to create business processes.
- *Application as a Service*: Another name for "Software as a Service" described by NIST.

- *Platform as a Service*: Equivalent to NIST's definition.
- *Integration as a Service*: Delivering a complete integration stack from the cloud including interfacing with applications, "smart" data and service discovery, process and information flow control and design.
- *Security as a Service*: Delivering core security services remotely, such as remote anti-virus or e-mail spam filtering.
- *Management/Governance as a Service*: Remotely administered cloud management functions and information, such as network topology, resource utilization, virtualization, and metrics gathering.
- *Testing as a Service*: Ability to test local or cloud-delivered systems using remotely hosted testing software and services.
- *Infrastructure as a Service*: Equivalent to NIST's definition.

While at first glance, having a number of additional cloud-based service categories can seem confusing, they actually help clarify and differentiate cloud service providers. A key aspect of migrating to a full or partial cloud computing solution will require that IT managers have a detailed understanding of the functionality of their existing architectures. The different functions will likely fall into similar classes as those described by Linthicum. With these definitions in mind, managers will be able to make more tailored decisions about which aspects of their architectures they can transition to a public cloud computing service provider.

2.2.5 Virtualization. At this point, it is beneficial to discuss virtualization in a bit more detail because it provides the technical underpinning for many of cloud computing's key benefits. Essentially, virtualization is a technique that decouples "guest" operating systems and applications from the underlying "host" operating system and hardware. Ideally, guests should have no knowledge that they are operating in a virtualized environment and should provide an equivalent set of functionality as if the guest were operating in a traditional installation environment. It is implemented using something called a hypervisor, which is also known as a Virtual Machine Monitor (VMM). The hypervisor is responsible for presenting the guests with a platform on which they can operate. It manages the guest execution by allocating the resources of the physical system to each guest in the environment. The hypervisor also serves to isolate each guest so that operations from one cannot impact, or obtain knowledge of, the operational state of another.

There are two primary flavors of hypervisors. A Type 1 hypervisor (See Figure 3) runs directly on the host's hardware and guest operating systems are installed one layer above it. *VMWare ESXi* and *Microsoft Hyper-V* are examples of Type 1 hypervisors. A Type 2 hypervisor (See Figure 4) is installed within a traditional host operating system and guest operating systems typically run in an application window. Examples of Type 2 hypervisors are *VMWare Workstation* and *Oracle VirtualBox*. The primary difference between the two types is whether or not there is a host operating system in between the hypervisor and the physical hardware. In general, Type 1 hypervisors have greater control of the underlying hardware as they do not have to exist within the context of a host operating system. Type 1 hypervisors can also be extremely compact and efficient allowing them to be built in a manner potentially more trustworthy from a security perspective than might be expected from a complete host operating system. This is important because if a host operating system can be compromised, then each guest operating system can potentially be compromised as well. Having a streamlined Type 1 hypervisor makes compromising guest operating systems much more difficult.

Virtualization has been around for many years, but relatively recent advances in the technology and general knowledge of the capabilities have made it more attractive to modern IT managers. Virtualization within a cloud computing provider offers a number of other key features that are advantageous to both consumers and providers. Virtualized implementations enable the provider to manage the underlying hardware more efficiently than if applications were implemented through "bare-metal" installations. This helps reduce the overall cost to both the consumer and provider.

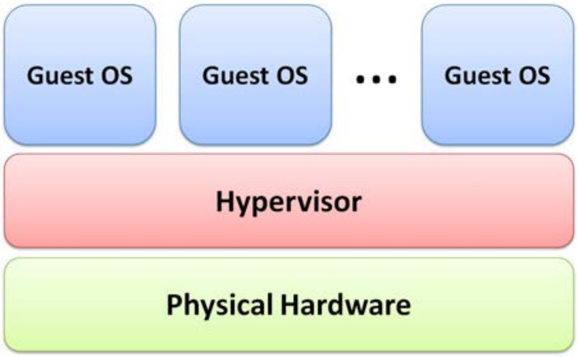

Figure 3: Type 1 Hypervisor

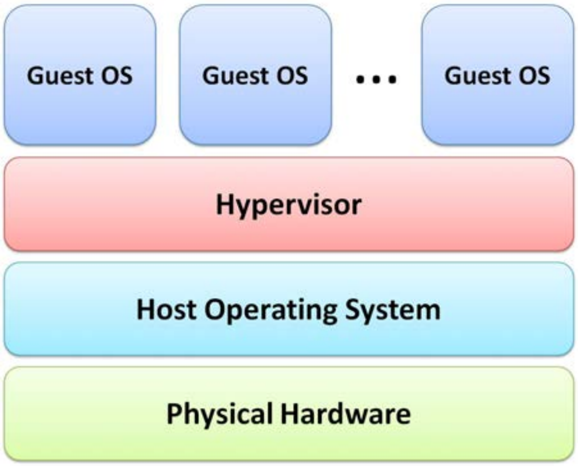

Figure 4: Type 2 Hypervisor

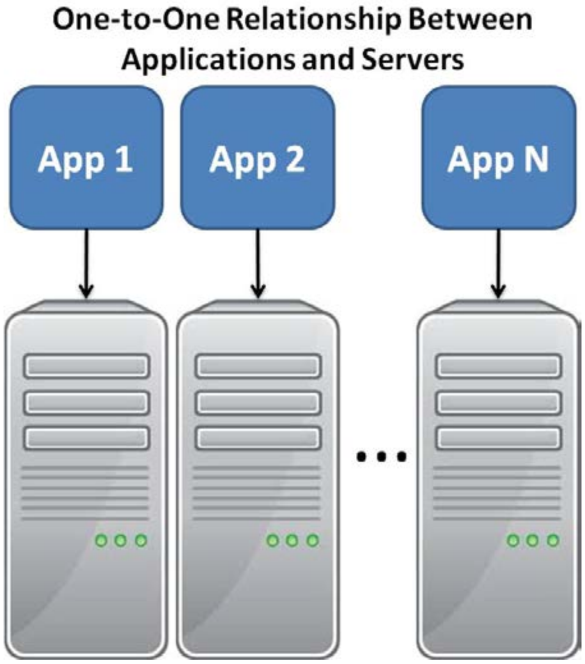

Figure 5: Traditional Data Center Deployment Model

As will be discussed in more detail in Chapter III, many data centers suffer from the problem of underutilization. This problem stems from the fact that the traditional approach to application deployment is based on “one application, one server.” This model predates widespread use of virtualization and attempts to ensure that if any one application fails it cannot affect other applications because of the physical separation between them. Figure 5 shows this one-to-one relationship between applications and physical resources in a data center. In general, the system is only performing its given function a limited percentage of time while taking up precious power, space and cooling all of the time. The utilization problem is exacerbated by managers typically deploying one server to run the application, another to serve as a backup to the first, and a third reserved for development purposes.

Figure 6 presents a graphical representation of this unfortunate reality. It shows the utilization of a typical server as a function of time. In this example, the server is sized to meet the peak load of some theoretical application with a load profile shown in gray. The server also includes some additional capacity in reserve to serve

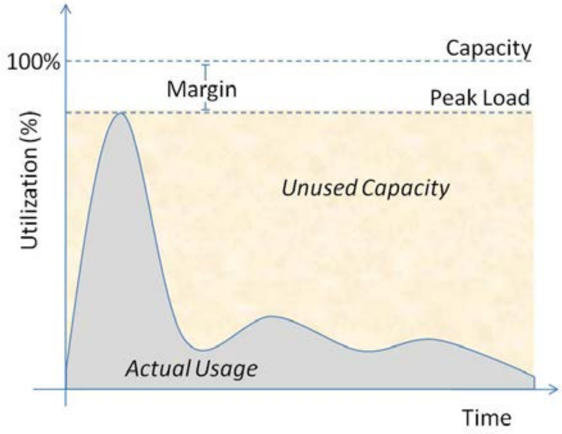

Figure 6: Example of an Underutilized Server

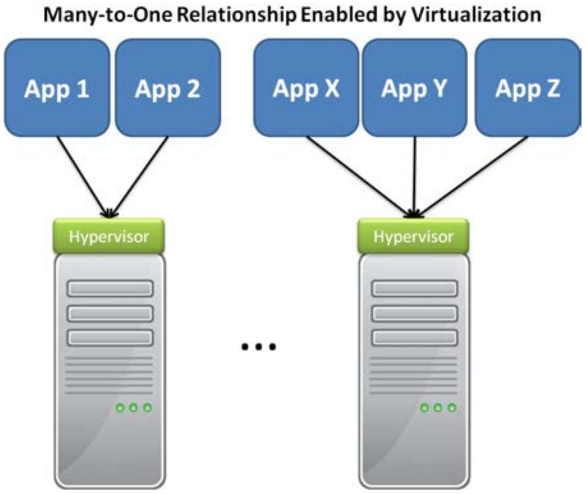

Figure 7: Consolidation Through Virtualization

as margin. The gray area of the curve represents the utilization as a percentage of total capacity at a given point in time. Like many applications this one only requires the full allocated capacity of the server on a periodic basis.

Many data center managers are rapidly moving away from the traditional one-to-one deployment model shown in Figure 5 because it results in the underutilization shown in Figure 6. Instead they are installing hypervisors on their physical servers so they can consolidate many applications on individual machines. A virtualized, “many applications, one server” deployment is shown in Figure 7.

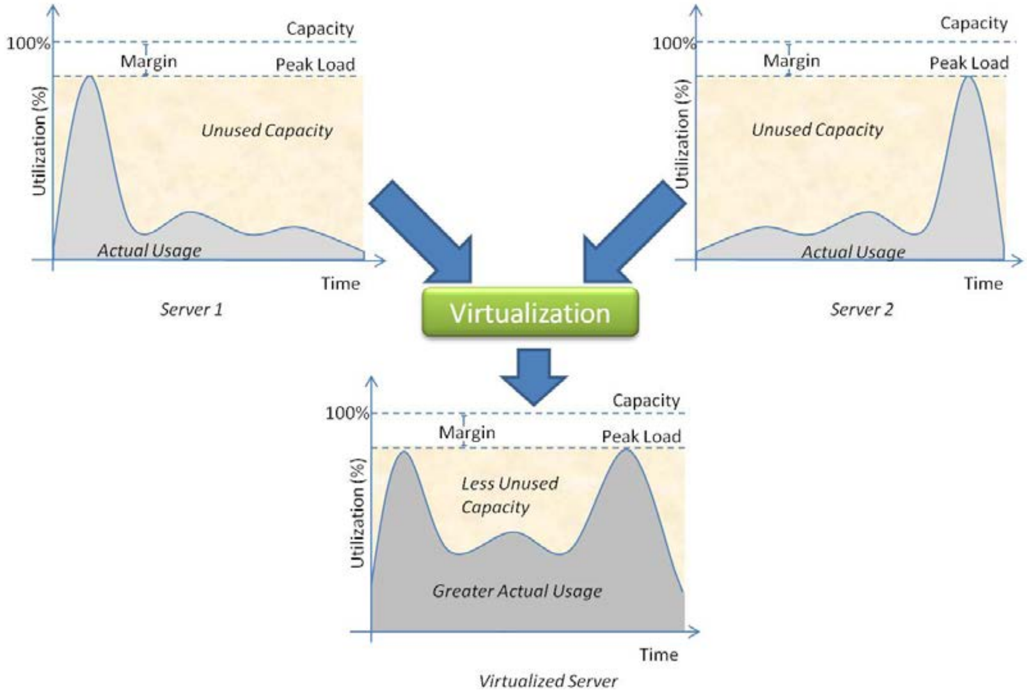

Figure 8: Increased Utilization Through Virtualization

Figure 8 represents the utilization benefits of such a deployment model. The top of the figure shows two servers with different functions and different utilization profiles. If these two functions are consolidated onto a single physical server through virtualization, then a single server would likely be able to carry both loads without impact to the user. The reduction in total servers within a given data center combined with an increased utilization per server can equate to significant cost savings.

Ultimately, this simple example demonstrates the variable nature of resource demand and how aggregation of that demand can improve resource utilization. When lots of different demand profiles are taken together, efficient cloud computing providers can leverage the statistical nature of that demand to produce *economies of scale*. The economies of scale help drive down costs as the provider dynamically allocates demand across a large pool of virtualized resouces. In effect, the cloud computing model represents a “many applications, many servers” relationship.

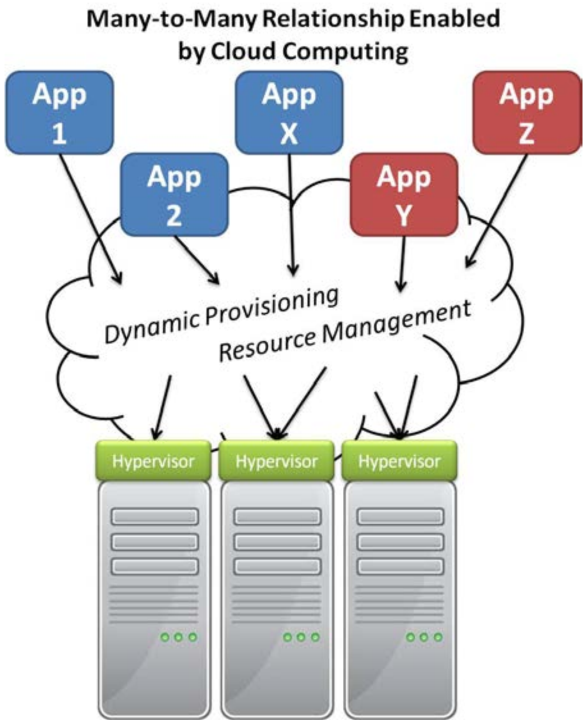

Figure 9: Multitenant Public Cloud Computing Deployment Model

Whereas Figures 5 and 7 represented private deployments, Figure 9 shows how a public cloud computing provider offers their services to multiple customers. This is a key aspect of public cloud computing, called *multitenancy*, which contributes to even higher utilization rates. Figure 9 highlights multitenancy by showing the applications in different colors to represent their ownership by different entities. The graphic also depicts the abstraction provided by cloud computing that separates applications from physical resources. Cloud consumers put applications into the cloud and the assignment to physical resources is handled exclusively by the cloud provider. By carefully balancing the load of all users across the resources of the entire data center, they can dramatically improve overall data center utilization and offer significantly reduced costs over a traditional "one application, one server" configuration.

Another key benefit of virtualization is "isolation." This serves to separate one consumer from another even when multiple consumer applications reside on the same physical resources. Isolation reduces the risk of errors in one application from having adverse affects on other user's applications. When properly configured, it also provides a security barrier between consumers by encapsulating information within a set of logically separated resources ensuring that different VMIs cannot access one another's applications or data. [9]

Finally, virtualization can also streamline configuration management and enable efficient scalability. For example, in an IaaS cloud, consumers manage the security configuration, patch level, software versions and implementation details in a single location inside their VMI. Then, the provider dynamically replicates that instance across as many resources as necessary to meet the customer's needs.

2.3 Cloud Computing Deployment Models

There are four deployment models for cloud computing: Public, Private, Hybrid and Community.

2.3.1 Public Clouds. A public cloud is implemented by a cloud service provider who makes those services available over the Internet to the general public. The provider is responsible for all of the capital and operating expense of the underlying infrastructure. It spreads that cost across all of its consumers either through a direct fee or through revenue generated from advertisements. The provider generally establishes a set of broadly applicable levels of service to attract the widest audience possible.

2.3.2 Private Clouds. A private cloud operates within an organization and provides services only to its own members. The organization can tailor the services it provides to meet its specific needs. Organizations often use private clouds when they have security concerns over moving applications or information onto resources operated outside of the control of the organization. Private cloud implementations often help increase the utilization rates of an organization's infrastructure by consolidating IT and dynamically provisioning organizational needs more efficiently. The primary drawback of a private cloud is that the organization must bear all of the costs associated with acquiring and managing the associated IT resources.

2.3.3 Hybrid Clouds. Hybrid clouds are architectures that mix both public and private cloud implementations. In a hybrid cloud, an organization implements some of its functionality within its own private cloud and outsources the rest to a public cloud provider. Often organizations implement hybrid clouds to obtain some of the cost savings associated with public clouds while also benefiting from the security of retaining sensitive information within the control of an organization's private cloud.

2.3.4 Community Clouds. Community clouds distribute the cost and management burden of a cloud architecture across different organizations with similar objectives or interests. The services provided in a community cloud are tailored to meet the needs of that community.

2.4 Cloud Computing Reference Architecture

Modern IT systems are often designed using an object oriented paradigm. The object oriented model helps to break complex systems down into their constituent parts to make them more manageable. Each part, or object, is well defined and performs specific functions. These objects work together to produce the overall functionality of the system. The interactions between objects are designed to be "loosely coupled" to ensure that changes within one module do not obviate changes in another. The concept of a Service Oriented Architecture (SOA) grew out of the object oriented paradigm. It was as a result of the trend for objects to interact outside of the bounds of a single system and instead interact across networks between systems. These models have proven very beneficial to designers and security professionals alike.

Cloud computing architectures are similarly complex and so NIST recently released the NIST Cloud Computing Reference Architecture to help clarify the relationships that exist in the cloud computing milieu. [10] It defines a neutral reference

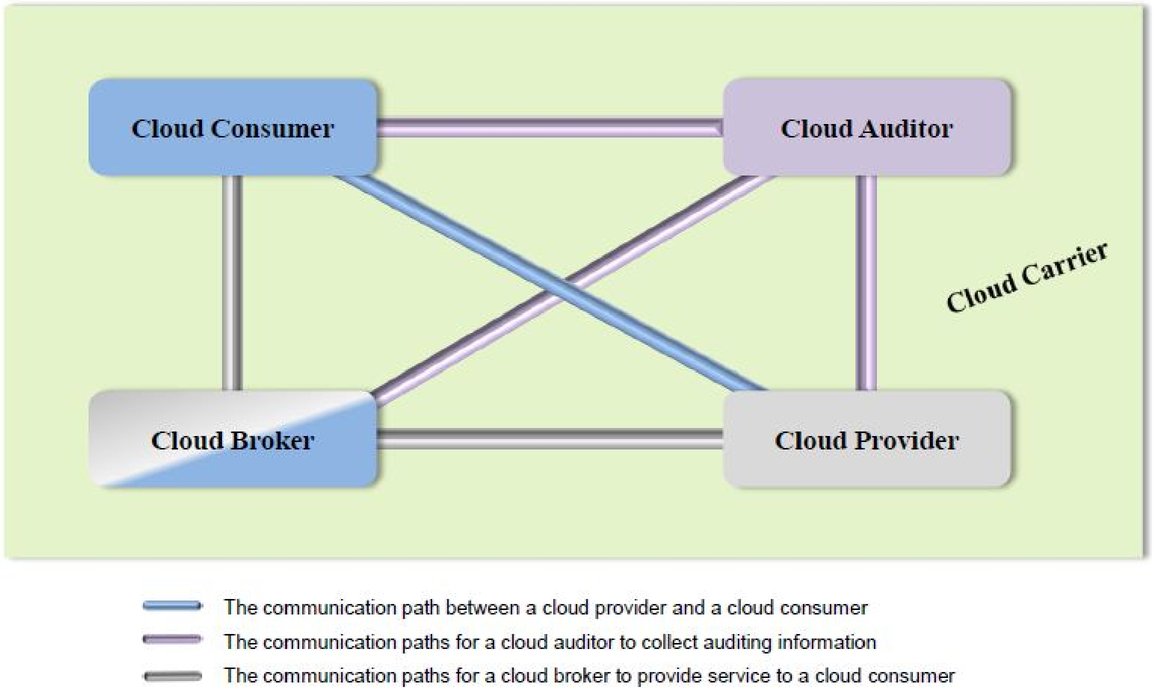
Cloud Consumer
Cloud Auditor
Cloud Carrier
Cloud Broker
Cloud Provider
The communication path between a cloud provider and a cloud consumer
The communication paths for a cloud auditor to collect auditing information
The communication paths for a cloud broker to provide service to a cloud consumer

Figure 10: Relationships Between Cloud Computing Actors

architecture consistent with NIST's definitions of cloud computing defined above. That architecture helps relate different cloud services and maps them in the context of an overall model. The document hopes to serve as a roadmap for IT professionals to understand, design and operate cloud computing-based infrastructures appropriately. This section summarizes the contents of the document.

Like the object oriented and service oriented models described above, the NIST Cloud Computing Reference Architecture breaks the complexity of cloud computing into its constituent parts. It starts by defining five major cloud *actors*: Consumers, Providers, Auditors, Brokers, and Carriers. Each actor has a *role* and performs a set of *activities* and *functions*. Figure 10 shows the relationships that exists between the five actors. [10]

2.4.1 Cloud Consumer. The *Cloud Consumer* is the person or organization that maintains a business relationship with, and uses services from *Cloud Providers*. Consumers are further broken into one of three different types depending on which

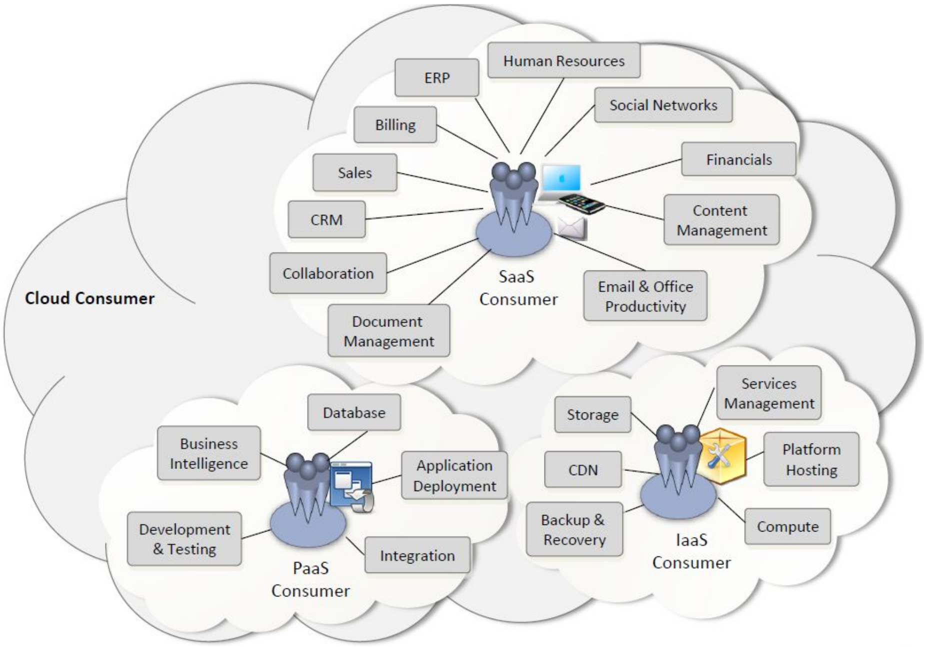

Figure 11: Services Available to Cloud Consumers

cloud service model they consume. A SaaS consumer's activities include using applications and services for business processes. PaaS consumers develop, test and manage applications hosted in a cloud environment. Finally, IaaS consumers create, install and monitor services for IT infrastructure operations. The breadth of services available to *Cloud Consumers* is shown in Figure 11. [10]

2.4.2 Cloud Provider. The *Cloud Provider* is the person, organization or entity responsible for making a service available to *Cloud Consumers*. The role of the *Cloud Provider* is to make five service categories available to consumers: service deployment, service orchestration, cloud service management, security and privacy.

- *Service Deployment:* This function provides for the type of deployment model (public, private, hybrid or community) required by the consumer.
- *Service Orchestration:* The infrastructure of the cloud environment must be arranged, coordinated and managed so that they can meet IT and functionality requirements. The orchestration function spans software, resource abstraction and control, and physical resource management.

- *Service Management:* This function refers to the service-related functions that enable *Cloud Consumers* to manage and operate their applications and data. These types of functions include: business support (e.g., billing, contracts, customer management), provisioning/configuration (e.g., resource modification, monitoring, metering), and portability/interoperability.
- *Security:* The *Cloud Provider* is responsible for providing many aspects of security to *Cloud Consumers.* Typically, these services include authentication and authorization; availability; confidentiality; identity management; integrity; security monitoring and incident response; and security policy management. Chapter IV describes the specific laws and regulation governing IT systems for the federal government and the DoD. The security responsibilities and relationships between government consumers and public providers will be discussed in more detail there.
- *Privacy:* This function goes hand-in-hand with the *Security* function. *Privacy* ensures that confidentiality is maintained.

2.4.3 Cloud Auditor. The *Cloud Auditor* is an entity that can conduct independent assessment of cloud services, information system operations, performance and security of the cloud implementation. They will provide an increasingly important role in developing the trust that is so critical to the relationship between *Cloud Providers* and *Cloud Consumers* especially when the provider uses a public cloud deployment model. Government agencies can enhance their security and functionality posture in a cloud environment by incorporating *Cloud Auditors* into the contracts governing federal and DoD IT systems.

2.4.4 Cloud Broker. The *Cloud Broker* is an entity that manages the use, performance and delivery of cloud services, and negotiates relationships between *Cloud Providers* and *Cloud Consumers.* A broker is especially helpful when the consumer does not have the knowledge or capability to effectively employ cloud services to their

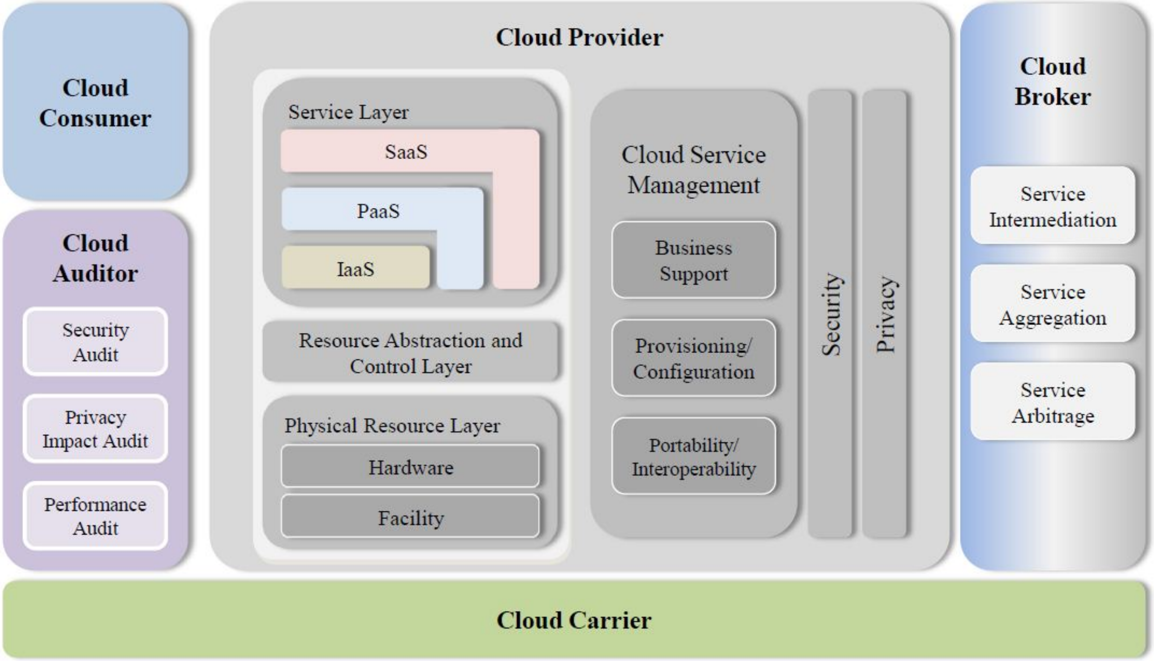

Figure 12: Cloud Architecture Reference Diagram

maximum potential. Brokers can provide service intermediation, aggregation and arbitrage. Service intermediation involves enhancing the offering of a *Cloud Provider* through "value-added" services. Service aggregation integrates multiple services into one or more new services. Finally, service arbitrage enhances flexibility by allowing the broker to evaluate the best service mix at the right time and provide a tailored response to meet consumer requests. The *Cloud Broker* is optional, since consumers have the option of working directly with a *Cloud Provider*.

2.4.5 Cloud Carrier. The final actor is the *Cloud Carrier*. The carrier is the intermediary that provides connectivity and transport of cloud services between all of the cloud computing actors through networking, telecommunication or other connectivity methods.

Figure 12 ties this entire section together by presenting a combined view of all of the actors and their major functions. [10]

2.5 Summary

This chapter covered the overall qualities of cloud computing and highlighted both the technologies as well as the business model associated with it. NIST and the Federal CIO have been very active in helping define the basics of cloud computing and articulating its key features. Having precise definitions and a clear understanding of the relationships between the various pieces is extremely important for DoD IT managers. With this information, the DoD can move past some of the early confusion in this space and implement sound acquisition strategies to taking advantage of the many potential benefits cloud computing. These benefits are presented in the following chapter.

III. Benefits Of Public Cloud Computing

> “There was a time when every household, town, farm or village had its own water well. Today, shared public utilities give us access to clean water by simply turning on the tap; cloud computing works in a similar fashion. Just like water from the tap in your kitchen, cloud computing services can be turned on or off quickly as needed. Like at the water company, there is a team of dedicated professionals making sure the service provided is safe, secure and available on a 24/7 basis. When the tap isn’t on, not only are you saving water, but you aren’t paying for resources you don’t currently need.” - Vivek Kundra

> “It is exciting to hear [federal] CIOs talk about how cloud computing can help the Federal Government focus on those activities that can really deliver real value for its citizens... Since the launch of those first AWS services... we have seen companies of every size... able to focus more and more on delivering value to their customers because of the use of our cloud services. We are excited and looking forward to counting the Federal Government among our customers and helping them achieve their goals.”
> - Werner Vogels

Public cloud computing can provide many advantages to DoD managers willing to incorporate its services into their IT architectures. While some authors have written on the subject of the basic benefits [11] [12] [13], this chapter focuses on five aspects of public cloud computing that are especially relevant to the DoD: *Continuous Refresh, Rapid Elasticity, Lower Cost, Improved Mission Focus* and *Lower Barriers to Entry.*

3.1 Continuous Refresh

DoD acquisition managers are under constant pressure to maintain currency across their enterprises and meet ever changing requirements over an increasingly complex infrastructure. It is extremely difficult to achieve such lofty goals under layers of bureaucracy and a six-year Planning, Programming, Budgeting, and Execution (PPBE) cycle. Information technologies are driven by the dynamic market trends and can rarely, if ever, be predicted years in advance through a formal requirements process.

Market forces and competition between IaaS providers serve to benefit the DoD as those pressures drive them to maintain the typical 12-24 month commercial refresh cycles for information technologies. This is also one of the key advantages of a cloud-based, virtualized solution. Virtualization abstracts the hardware from the application stack. It allows applications to migrate more easily from one hardware platform to another and facilitates transitions between upgrades.

3.2 Rapid Elasticity

One of NIST's essential elements of cloud computing is also one of the cloud computing's predominant benefits. It goes hand-in-hand with overcoming the dynamic nature of IT requirements mentioned in Section 3.1. Elasticity describes how cloud computing resources can grow or shrink depending on the demand and the consumer is only charged for what they use. This capability translates into much more relaxed constraints on early estimates of resource requirements since the available resource pool will adapt to meet the needs as they are presented.

The traditional DoD acquisition model can sometimes take months or more to augment existing resources. Not only must the physical system be purchased, installed and tested, a significant amount of documentation must be generated and authorities must be granted for those new systems to operate. By contrast, the rapid elasticity of public cloud computing makes the availability of new resources essentially instantaneous. Security configurations can also be uniformly distributed across those new resources as quickly as they come online.

3.3 Lower Cost

The historically stovepiped nature of DoD acquisition programs have lead to tremendous infrastructure redundancy and underutilization. A 2010 memorandum from the Federal CIO highlighted the DoD as having 772 individual data centers; the largest number as compared to all other federal agencies (see Figure 13). [14] In a keynote speech to the Brookings Institution in April 2007, he also stated that utilization rates for government data centers are typically only around seven percent.

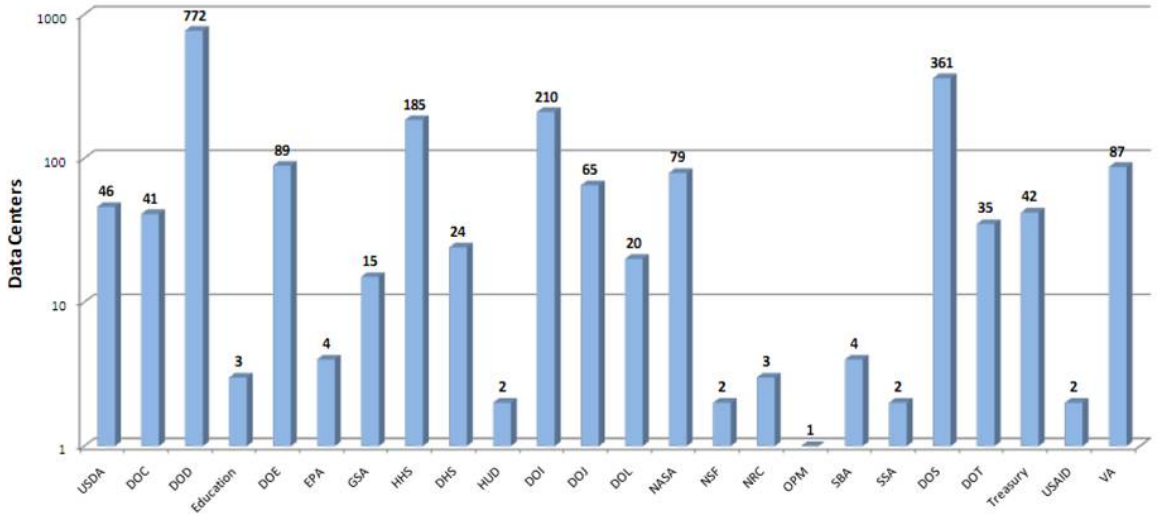

Figure 13: Number of Data Centers by Agency (as of 7/30/10)

Data centers go underutilized because they were designed to handle a given application's peak load, which only occurs over a limited period of time. For the majority of the time, these assets go unused. The separated nature of the DoD compute infrastructure compounds the problem because it prohibits one system from sharing its processing or storage load with other systems.

Public cloud service providers dynamically distribute application loads from multiple users across multiple systems. Any single user or application would generally have a cyclical demand for resources that will spike at certain times, but go unused at others. As stated in Section 2.2.5, public providers often host many different customers in their infrastructures simultaneously. Providers capitalize on the statistical nature of the loads generated by several consumers over time. As one user's demand drops, the provider reclaims resources from that user and reprovisions them to another user whose demand is rising. This is how public cloud computing providers maintain high utilization rates across their infrastructures. The utilization rates have a direct relationship to overall costs.

3.4 Improved Mission Focus

One useful way of framing the types of organizations on the Internet is to consider the difference between Application Service Providers (ASPs) and Internet Service Providers (ISPs). The two functions are separable, but are highly dependent on each other for providing a complete product to consumers. In its most basic form, the ISP provides consumers with connectivity to the Internet. The ASP creates applications that ride on top of that infrastructure and add value to it. For example, Comcast, Verizon, AT&T and Time Warner are all examples of ISPs. Facebook, Bing, eBay, and Wikipedia are examples of ASPs. The separability of these two functions allows each organization to focus on its core competencies without having to be experts in all aspects of the complicated, end-to-end solution required to deliver applications to users over the web.

By following the ASP/ISP model, the DoD can more effectively focus on its mission. Decoupling the mission applications and information from the underlying IT infrastructure needed to process and deliver them to the warfighter will allow the DoD to provide better mission services. Public cloud computing providers can become the ISP for many DoD applications. The DoD can leverage the provider's expertise and ultimately produce more agile and lower-cost solutions.

3.5 Lower Barriers to Entry

This last benefit gets to the heart of the current DoD industry base. Mergers and acquisitions, budget cuts and short-term focused planning have lead to a competition pool that consists of only a handful of massive prime contractors. This not only dilutes the economic market forces that help lower acquisition costs, it also stifles innovation.

The processing, storage and connectivity requirements of many applications keep many small business and non-traditional defense contractors out of the available pool of bidders. Often, these companies have ideas and talent that could provide the

DoD with tremendous benefit, but they cannot overcome contractual barriers such as access to Government Furnished Equipment (GFE) or meeting the myriad security requirements for connectivity.

A small business with an innovative idea may not be able to host the necessary infrastructure to handle their ideas even if that infrastructure were to be funded under a contract. For example, imagine a small business with an idea for an automated image processing algorithm requiring a data center with a minimum of 1000 CPUs to achieve timeliness requirements. Not only would the DoD be reluctant to fund a private data center for the company, the small company may not have the expertise, facility space or power required to operate it. The DoD can overcome this scenario by allowing contractors to leverage public cloud computing providers to develop and host their applications. Kaufman draws a similar conclusion by describing the benefits of cloud computing to small businesses as a"major selling point." [15]

3.6 General Security Benefits of Cloud Computing

Public cloud computing can even offer security advantages over private deployments. In [16], NIST presents the following areas where public cloud computing can provide superior security to traditional models. These areas are complimentary to those outlined above and should be carefully reviewed by the DoD so that the risks can be effectively balanced against the opportunities associated with public cloud computing:

- *Staff Specialization:* Similar to *Improved Mission Focus* described earlier, NIST emphasizes that cloud providers have a unique capacity for hiring and maintaining specialized staff that can focus solely on security and privacy.
- *Platform Strength:* Public cloud providers typically offer a more uniform solution than most traditional computing centers and this increased uniformity helps security professionals harden their environments. The reduction in complexity provided by standardization also helps improve manageability.

- *Resource Availability:* Despite the perception that external providers will not achieve availability requirements, public cloud computing environments can often foster even greater availability than a private deployment since redundancy and disaster recovery are often inherent in most public offering. The on-demand nature of public cloud computing also provides resilience against unexpected demand or denial of service (DoS) attacks.

- *Backup and Recovery:* The back up and recovery policies and procedures of public cloud providers may be superior to private deployments, since data and services are generally maintained across multiple geographically dispersed resources.

- *Mobile Endpoints:* The processing and storage model for public cloud computing eliminates much of the need to have large capacity devices at the consumption end-points for data and services. Often cloud data and services are accessed through the Internet with no more software than a web browser, which is available on everything from desktop computers to cell phones. This model makes data and service much more dynamic and accessible.

- *Data Concentration:* The centralization provided by public cloud offerings reduced the risk of loss associated with a compromised access device. For example, many organizations store and processes important data on a distribution of laptops or other devices, which opens up the possibility of data loss if those devices are lost or stolen. In a public cloud model, those devices would store only a minimal amount of data locally and therefore the impact of their loss would be much less.

3.7 Examples of Public Cloud Computing Applications

Over the last few years, a number of cloud service providers have made themselves available to the general public catapulting the “public cloud” deployment model into the forefront of e-services. Most notably, Amazon.com, Google, and Microsoft, have all started IaaS offerings. These offerings are Amazon Elastic Compute Cloud (EC2), Google AppEngine, and Microsoft Azure, respectively. Lesser known to the general public, but substantial in the industry are SalesForce, RackSpace, Parabon, GoGrid, and FlexiScale. All a user needs to do is sign up for an account and provide a credit card. Within minutes, they are up and running on servers hosted by any of these providers. Other companies also sell various software components that ease the setup, employment and distribution of workload across the resources obtained from these providers.

Notable examples of employing public cloud computing resources come from Animoto, New York Times, and MySpace. The stories of Animoto and the New York Times are two of the more famous case studies that leverage Amazon EC2. They highlight scalability and the on-demand capability of cloud computing. First, Animoto is a company that offers software for automatically creating video montages from user provided photos and music. In 2008, they launched their service as an application on Facebook. The demand spiked, forcing them to grow from 50 processing servers to over 3,500 in just 3 days. Fortunately, the on-demand, scalability provided by Amazon allowed them to maintain their service despite the tremendous increase in usage. [17]

In another example, the New York Times gave writer Derek Gottfird the task of organizing and converting 11 million articles from the news archives into PDF format. Rather than wait for days or weeks for the job to complete on the four available servers at the Times, he instead leveraged 100 servers at Amazon and completed the job in less than 24 hours. [18]

The popular social media website, MySpace, successfully used cloud computing to test its capacity to meet consumer demand. Prior to introducing a new streaming music video service in New Zealand, they worked with SOASTA, a company that specializes in web-based testing, to use the cloud as a load generation platform for simulating, measuring, and analyzing expected user traffic. SOASTA was able to simulate 1 million concurrent virtual users by rapidly deploying its testing software across 780 servers. [19] In all of these examples, neither company was required to purchase, install, or maintain large amounts of private infrastructure for performing the aforementioned tasks and they were able to access computing and storage resources precisely when and where they were needed.

The federal government is also actively leveraging public cloud computing. One example is `http://www.recovery.gov`, which grew from the passage of the American Recovery and Reinvestment Act. The Recovery Accountability and Transparency Board created Recovery.gov to enable citizens to see how their tax dollars are being spent. The website is hosted on Amazon EC2 and it is "the first government-wide system to migrate to a cloud-based environment." The move to a public cloud computing solution allows Recovery.gov to be fully scalable and efficiently manage variable demand for the service. It also saves the government money. The Board estimates that it will save $334,800 in FY2010 and $420,000 in FY11, which represents approximately 4% of its $18 million budget. [3]

3.8 Summary

As this chapter demonstrates, there are many potential benefits of public cloud computing. Five specific benefits were offered that are of particularly applicable to the DoD: *Continuous Refresh, Rapid Elasticity, Lower Cost, Improved Mission Focus,* and *Lower Barriers to Entry.* Many commercial businesses and other government organizations have implemented programs that have demonstrated the power of moving to a public cloud solution. The benefits are important, but they must be balanced against the responsibilities that the DoD has for ensuring the security of the information it handles. This responsibility is unique and it often requires special consideration that differentiates, but does not necessarily exclude, DoD application of public cloud computing from comparable application by commercial entities. Legal requirements and policy directives govern how the federal government and the DoD can safely implement IT solutions. The following chapter emphasizes the need for implementing a sound risk management process when incorporating public cloud computing and it puts that process in the context of current regulations and guidelines.

IV. Assessing Public Cloud Computing Security from a Risk Management Perspective

> "Cloud computing is nothing to be afraid of. It is a fundamental evolution of the way we do computing. At the very least, it will provide a well-defined architectural option for those who build enterprise architectures... The core idea of cloud computing is to open up your mind and your architecture to any sort of technology that will allow you to better serve the business." - David Linthicum

> "People have trouble estimating risks for anything not exactly like their normal situation." - Bruce Schneier

International Data Corporation (IDC) conducted a survey in 2009 asking IT executives and CIOs to "rate the challenges/issues of the cloud/on-demand model." Not surprisingly, security topped the list at 87.5% of respondents rating it as a concern. [20] The answer would almost assuredly be equal or even higher if the same poll was conducted within the DoD. Unfortunately, the current business model for managing DoD IT cannot be sustained in a declining budget environment with users demanding better services.

Wyld captures the essence of much of the problem for government IT managers by correlating the perception in government circles that *having control* is the same as *being secure*. He backs up this premise with quotes from Arun Gupta, a partner at Columbia Capital, and another from Linda Cureton, CIO of NASAs Goddard Space flight Center. Gupta stated that in order to succeed you have to have the confidence to say, "I don't need to control everything." Cureton said that it is imperative when considering cloud computing not to "confuse control and ownership with security and viability." [11] Golden amplifies these statements by observing that those who characterize cloud computing as too risky often have an overly optimistic view of their current risk management practices. He goes on to say, "this attitude reflects a common human condition: underestimating the risks associated with current conditions while overestimating the risks of something new." [21]

Successful adoption of public cloud computing requires that security concerns be broken out into their constituent parts and evaluated from a risk management perspective rather than one of risk avoidance. Many authors have created frameworks for categorizing these risks to help start this process. [11] [13] [22] [23] In addition, IT managers must resist the urge to equate their level of control with their level of security.

The DoD defines *risk* as a "measure of future uncertainties in achieving program performance goals and objectives within defined cost, schedule and performance constraints." [24] It is comprised of three components:

- A future *root cause* (yet to happen), which, if eliminated or corrected, would prevent a potential consequence from occurring.
- A *likelihood* assessed at the present time of that future root cause occurring.
- The *consequence* of that future occurrence.

Unfortunately, it seems that human nature is to approach the unknown by tending to focus on the consequence of a particular risk rather than accurately evaluating all three components in their entirety. The complexity of modern information technology combined with the rapid pace of its change makes much of it a mystery to many managers. When faced with this uncertainty, managers often react by maintaining the status quo, since it is something that they understand, not necessarily because it is the best solution. If we are to successfully incorporate modern IT into the DoD enterprise and maintain a competitive advantage over our adversaries, we must make an honest and informed assessment of the risks involved in adopting technologies like cloud computing and compare them against the benefits.

Once the risk have been accurately defined, then the risk mitigation process can begin. There are four options that managers can use to help reduce the likelihood of a risk becoming and issue (which is when a root cause actually occurs):

- *Avoiding* risk by eliminating the root cause and/or the consequence.
- *Controlling* the cause or consequence.
- *Transferring* the risk.

- *Assuming* the level of risk and continuing on the current program plan.

There are several statutory requirements governing Federal IT systems. The Clinger-Cohen Act of 1996 assigns responsibilities for the efficiency, security, and privacy of computer systems within the federal government. It also seeks to improve the acquisition and management process of information systems within federal agencies. Clinger-Cohen assigned a number of federal IT related responsibilities to the Office of Management and Budget (OMB). Under that authority, OMB issued a number of circulars to federal agencies. One of the most prominent is Circular A-130. It establishes policy for management of federal information resources, including procedural and analytic guidelines for their implementation. Appendix III of that circular requires that agencies provide adequate security for all of the information they collect, process, transmit, store, or disseminate. The Privacy Act of 1974 governs the collection, maintenance, use and dissemination of personally identifiable information maintained by federal agencies.

The most current regulation is the Federal Information Security Management Act of 2002 (FISMA) and the associated federal standards for information management that spawned from it. FISMA is the primary focus of this research because it is the most modern and its regulations provide an appropriate framework for evaluating the risk of public cloud computing. FISMA, and the guidance generated from it, define the approach for securing federal IT systems and the goal here is to understand how cloud computing adoption can be integrated appropriately within the context of these laws, standards, directives and instructions.

The federal regulations primarily serve as a mechanism for identifying the consequences associated with IT risks and are largely concerned with classifying government IT systems according to the impact that a security related event would have should it occur. The guidance, directives and instructions that flow from the regulations identify the general security areas where root causes can occur as well as presenting mitigation strategies that can be applied to manage the risks according to program needs. Chapter V dives more deeply into the specific kinds of root causes that are likely to affect public cloud computing. Finally, likelihood must be determined by the program manager's risk management team in the context of their specific system and their specific implementation.

4.1 Federal Information Security Management Act of 2002

FISMA was enacted under Title III of the E-Government Act of 2002 (Public Law 107-347). It requires each federal agency to "develop, document, and implement an agency-wide program... to provide information security for the information and information systems that support the operations and assets of the agency, including those provided or managed by another agency, contractor, or other source." Under FISMA, NIST was tasked with the responsibility for developing security standards and guidelines for the federal government, with the exception of those information systems designated as "national security systems." Specifically, NIST was tasked with developing:

- Standards for categorizing information and information systems collected or maintained by or on behalf of each federal agency based on the objectives of providing appropriate levels of information security according to a range of risk levels.
- Guidelines recommending the types of information and information systems to be included in each category.
- Minimum information security requirements for information and information systems in each such category.

From this tasking, NIST generated Federal Information Processing Standards (FIPS) Publication 199 and 200, and Special Publication (SP) 800-53.

At this point, it is important to differentiate a typical Federal IT system from one that is classified as a "national security system" to delineate when NIST guidelines should be applied. FISMA designates that national security systems are those systems (including telecommunications systems) that are used or operated by an agency, contractor of that agency or another organization on behalf of the agency that:

- Involves intelligence activities.
- Involves cryptologic activities related to national security.
- Involves command and control of military forces.
- Involves equipment that is an integral part of a weapon or weapons system.
- Is critical to the direct fulfillment of military or intelligence missions (with the exception of systems used for routine administrative and business applications (including payroll, finance, logistics, and personnel management applications).

IT systems also fall under the category of a national security system if the information they manage is to be kept classified in the interest of national defense or foreign policy as established by either an Executive Order or an Act of Congress. Even though national security systems are not bound by the NIST standards described later, those standards have been developed, from a technical perspective, to complement similar standards for national security systems. [25]

The current state of public cloud computing, along with the current level of trust between DoD IT managers and public cloud computing providers, precludes incorporating public cloud solutions into IT systems designated as national security systems at this time. Ultimately, all measures taken to provide security over IT systems must be commensurate with the risk and magnitude of harm resulting from the unauthorized access, use, disclosure, disruption, modification, or destruction of the information contained in these types of systems. The series of recommendations detailed in Chapter VII, do not advocate that public cloud computing should be included as part of the operational solution for national security systems, although those responsible for determining whether or not a system falls under this category should still resist overstating security risks at the detriment of system functionality and affordability.

To determine that level of risk, and therefore establish a basis for securing IT systems, FISMA defines three security objectives: *Confidentiality*, *Integrity*, and *Availability*. *Confidentiality* means "preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information." *Integrity* means "guarding against improper information modification or destruction, and includes ensuring information non-repudiation and authenticity." Finally, *Availability* means "ensuring timely and reliable access to and use of information."

These three objectives serve as a starting point for NIST standards and provide the contextual framework for how NIST accomplishes its mandate under FISMA.

4.2 Applicable NIST Guidance

4.2.1 FIPS 199: Standards for Security Categorization of Federal Information and Information Systems.

FIPS 199 addresses NISTs responsibility to develop standards for categorizing information and information systems as described in Section 4.1. It begins by describing the nature of the impact associated with the loss of one of the three security objectives defined by FISMA. Those definitions are as follows:

- A loss of *confidentiality* is the unauthorized disclosure of information.
- A loss of *integrity* is the unauthorized modification or destruction of information.
- A loss of *availability* is the disruption of access to or use of information or an information system.

The potential impact is broken into either a LOW, MODERATE or HIGH category depending on the severity. The following list describes how each of the impact categories are defined:

- *LOW:* loss of confidentiality, integrity or availability that could be expected to have a *limited* adverse effect on organizational operations, organizational assets, or individuals. A LOW impact event might cause degradation in mission capability to an extent and duration that the organization is able to perform its primary functions, but the effectiveness of the functions is noticeably reduced. It may also result in minor damage to assets, minor financial loss or minor harm to individuals.

- *MODERATE:* loss of confidentiality, integrity or availability that could be expected to have a *serious* adverse effect. It might cause a significant degradation in mission capability to an extent and duration that the organization is able to perform its primary functions, but the effectiveness of the functions is significantly reduced. It may also result in significant damage to assets, result in significant financial loss or significant harm to individuals that does not involve loss of life or serious life threatening injuries.
- *HIGH:* loss of confidentiality, integrity or availability that could be expected to have a *severe or catastrophic* adverse effect. It might cause a severe degradation in or a loss of mission capability to an extent and duration that it is not able to perform one or more of its primary functions. It may also result in major damage to organizational assets, result in major financial loss or in severe or catastrophic harm to individuals involving loss of life or serious life threatening injuries.

With this framework in mind, NIST presents a generalized equation, shown as Equation 1, for agencies to express the Security Category (SC) of their information and information systems. Establishing an appropriate SC requires IT managers to determine the potential impact for each security objective associated with a particular information type. The overall SC is based on the “high water mark concept” meaning that an information system’s SC will be no less than the highest impact category over all security objectives. Security categorization of federal information and information systems is the first step in the risk management process. [26] The resulting value helps determine which security controls are necessary for protecting the system.

$$SC_{(informationsystem)} = \{(\mathbf{confidentiality}, \mathit{impact}), (\mathbf{integrity}, \mathit{impact}), (\mathbf{availability}, \mathit{impact})\} \qquad (1)$$

4.2.2 FIPS 200: Minimum Security Requirements for Federal Information and Information Systems . FIPS 200 is the second of the mandatory IT security standards required by FISMA. It specifies the minimum security requirements covering seventeen security related areas for protecting confidentiality, integrity and availability of federal information systems and the information processed, stored, and transmitted by those systems. The security-related areas are *Access Control; Awareness and Training; Audit and Accountability; Certification, Accreditation and Authentication; Incident Response; Maintenance; Media Protection; Physical and Environmental Protection; Planning; Personnel Security; Risk Assessment; Systems and Services Acquisition; System and Communications Protection;* and *System and Information Integrity.*

These areas are grouped into one of three classes: *management, operational* or *technical.* The classes help to identify the overall approach used to maintain confidentiality, integrity and availability within a given security area.

- *Management Controls* focus on the management of risk and the management of information system security.
- *Operational Controls* are primarily implemented and executed by people as opposed to systems.
- *Technical Controls* are primarily implemented and executed by the information system through mechanisms contained in the hardware, software or firmware of the system.

The areas outlined here set the stage for the specific security controls described in detail in both NIST SP 800-53 and for the "Proposed Security Assessment & Authorization for U.S. Government Cloud Computing" document discussed in Section 4.2.3. The acronyms that follow the topic area are used to help organize and reference specific control measures outlined in those documents. The class of each security control area is also included.

Access Control (AC) - Technical. Access to IT systems must be limited to authorized users, processes and devices. This area covers one sense of the traditional security principle of "least privilege." According to this principle, users, processes and devices are only granted permissions commensurate with the functions they are required to perform and no more. Successfully implementing this control minimizes the "attack surface" that unauthorized agents can use to infiltrate a system.

Awareness and Training (AT) - Operational. Awareness and training are crucial aspects for ensuring that designers, developers and managers alike understand the security environment. It should cover not only the technical details of the security risks facing IT systems, but also the applicable laws, policies and regulations governing those systems. Finally, awareness and training helps to ensure that personnel are capable of adequately carrying out their assigned duties and responsibilities.

Audit and Accountability (AU) - Technical. This area covers two separate, but related activities. Auditing is the need to monitor, analyze, investigate and report any unauthorized, unlawful, or inappropriate activity on an information system. Accountability ensures that access to information systems can be ascribed to individuals, processes or devices in a non-repudiable[1] way. The combination of auditing and accountability provides security analysts the ability to determine malicious or erroneous activity and identify the source of its cause.

Certification, Accreditation and Authentication (CA) - Management. Certification, accreditation and authentication is the process by which organizations periodically assess the security controls of their information systems and make a determination concerning their effectiveness. It includes developing plans to correct any deficiencies in the information systems' security posture found during those reviews.

[1] Non-repudiation ensures that the originator of something cannot challenge the fact that they are, in fact, the originator. It essentially binds communication to the sender in a provable fashion.

The output of this process is a determination on whether or not to authorize the information system to connect to other systems and operate.

Configuration Management (CM) - Operational. The complexity of information systems requires that organizations have a well structured plan for maintaining detailed knowledge of their system and ensure that any changes to that system are implemented in a controlled fashion. Configuration management starts by establishing a baseline of the system to include the state of hardware, software, firmware and documentation. When a change is required, due to the need to fix discrepancies or to update functionality, the proposed change should flow through the configuration management process before being implemented in the operational system. This is a critical part of the systems engineering process and is vital for security analysts to gain a clear picture of the potential vulnerabilities of the information system.

Contingency Planning (CP) - Operational. In the event that a security related incident occurs, organizations must have response plans in place to handle the situation. Contingency planning includes emergency response procedures, backup operations and post-disaster recovery. Having appropriate plans in place to handle problems is critical to ensuring availability of services and continuity of operations.

Identification and Authentication (IA) - Technical. Identification and authentication provides a process for determining the identity of information system users, processes and devices prior to allowing them access to the information system. This area is a precursor to the Audit and Accountability area described earlier in that it ties the physical entity to their digital identity so that accountability can be traced and maintained.

Incident Response (IR) - Operational. Incident response implements contingency plans when a security incident occurs. It covers training, reporting, detection, analysis, containment, recovery and user response operations required to recover or maintain confidentiality, integrity and availability.

Maintenance (MA) - Operational. All information systems must undergo periodic maintenance to keep them up to date, operating efficiently and tuned for performance. System maintenance must work within the boundaries of configuration management to ensure a well documented baseline.

Media Protection (MP) - Operational. This area is largely concerned with protecting the physical aspects of stored information. Information systems employ both digital media (hard drives, USB devices, disks, etc.) and non-digital media (paper, microfilm, etc) to store information. Therefore, appropriate steps must be in place to protect those media from unauthorized access.

Physical and Environmental Protection (PE) - Operational. In conjunction with technical measures, such as passwords and digital ID cards, physical security must be in place to limit unauthorized physical access and guard against environmental hazards to systems, equipment and their operational area. Physical security is also responsible for protecting and maintaining the support infrastructure, such as power and cooling, critical to information systems.

Planning (PL) - Management. This area requires that organizations develop, disseminate and periodically update formal documentation concerning security policies and procedures. These policies are managerial in nature and address the purpose, scope, roles, responsibilities, management commitment, coordination and compliance. The documented procedures facilitate the implementation of the security policies and controls.

Personnel Security (PS) - Operational. Personnel security is concerned with ensuring that people who work in positions of responsibility over information systems are trustworthy and meet the established standards specific to their roles. This control extends to not only the employees of the agency, but also any third-party providers. Personnel security is also concerned with protecting information and information systems during and after personnel are either terminated from employment or transition to a different position.

Risk Assessment (RA) - Management. Risk assessment is a formal systems engineering process that helps an organization evaluate the pros and cons of employing particular solutions. This process strives to quantify the risk in measurable ways to assist management in making sound decisions. As mentioned earlier in this chapter, risk is measured according to both the *likelihood* of a particular *root cause* actually impacting the system and the *consequence* of that event actually occurring. When a particular security threat or vulnerability is discovered, the root cause, likelihood and consequence must be taken together so that managers prioritize and mitigate the risks with the resources available to them.

Systems and Services Acquisition (SA) - Management. Acquiring information systems involves a careful balance of overall system cost, delivery schedule and level of performance. In the context of security, managers must ensure that sufficient resources are allocated to provide for implementing the security controls necessary to protect confidentiality, integrity and availability. Agency managers must also ensure that third-party providers do the same.

System and Communications Protection (SC) - Technical. Information systems are useless without connectivity. Without it, information cannot be disseminated to those who need it. Therefore, organizations must monitor, control and protect those lines of communication at both the external and key internal boundaries of the information system.

System and Information Integrity (SI) - Operational. Organizations must protect information and information systems against malicious code by implementing measures such as firewalls, virus scanners and intrusion prevention and detection systems. With those systems in place, they must also identify, report, and correct any security discrepancies in a timely manner so that incident responders can take appropriate action quickly.

4.2.3 Special Publication 800-53 . After completing the security categorization process described in FIPS 199, organizations select an appropriate set of security controls for their information systems that satisfy the minimum security requirements set forth in FIPS 200. NIST SP 800-53 lays out specific security controls tailored for both the security area and for the impact level determined during the security categorization process. [27] Ultimately, it is the responsibility of the organization to select security controls appropriate to their specific systems.

SP 800-53 contains an extraordinary amount of detail covering each type of control that should be included in information systems to account for each of the 17 security areas. Therefore, each control will not be reproduced in their entirety here. Instead, it will suffice to present a representative example to give the reader an idea of what supplemental controls are necessary. The example also provides a sense of the taxonomy of the controls. The reader is strongly encouraged to review SP 800-53 to see all of the families of controls and their associated details.

For example, under Certification, Accreditation and Authentication family of controls, SP 800-53 defines seven specific types of controls:

1. Security Assessment and Authorization Policies and Procedures (CA-1)
2. Security Assessments (CA-2)
3. Information System Connections (CA-3)
4. Security Certification (CA-4)
5. Plan of Action and Milestones (CA-5)
6. Security Authorization (CA-6)
7. Continuous Monitoring (CA-7)

Then, within each of those controls is a description of the control itself and in many cases a list of *control enhancements*. The control enhancements can be either mandatory, depending on SC level, or implemented voluntarily if the SC level does not require it, but the agency feels that it is necessary for their particular system. For example, CA-2 has two control enhancements:

1. The organization employs an independent assessor or assessment team to conduct an assessment of the security controls in the information system
2. The organization includes as part of security control assessments [Assignment:organization-defined frequency], [Selection: announced; unannounced], [Selection: in-depth monitoring; malicious user testing; penetration testing; red team exercise; [Assignment: organization-defined other forms of security testing]]

For this particular family of controls, if the SC is LOW, then only the baseline CA-2 must be applied. If the SC is MODERATE, then baseline CA-2 plus enhancement (1) must be used. Finally, SP 800-53 states that if the SC is HIGH, then CA-2 and both control enhancements must be included. Each family of controls is laid out in a similar fashion.

SP 800-53 recognizes the challenge and importance associated with selecting the most cost-effective controls. It also presents a series of challenges when applying security controls in external environments including:

- Defining the types of external services provided to the organization.
- Describing how the external services are protected in accordance with the security requirements of the organization.
- Obtaining necessary assurances that the risk to the organization's operations and assets, and to individuals, arising from the use of the external services is at an acceptable level.

The primary consideration for overcoming these challenges is trust. Trust provides assurance or confidence that the risk to the organization's operations, assets and individuals is at an acceptable level. Federal agencies can manage the level of trust by exerting a certain amount of direct control over the external provider with contracts or SLAs. Trust can also develop from external providers convincing federal agencies they have credible security controls in place. This may be accomplished through third party audits or authoritative certifications of a provider's security control mechanisms.

4.2.3.1 Proposed Security Assessment & Authorization for U.S. Government Cloud Computing. More recently, the Federal CIO's office released a the Proposed Security Assessment & Authorization for U.S. Government Cloud Computing, Draft version 0.96. [28] That document takes the security controls established in SP 800-53 and augments them with tailored controls specifically for use with cloud computing.

Going back to the example in Section 4.2.3, recall that an information system with a SC of LOW, would only be required to implement the baseline control for CA-2. This document calls for both CA-2 and control enhancement CA-2 (1) when the information system incorporates cloud computing. Each security related area and set of controls has similar tailored security control requirements. The goal of this document is to help standardize the Assessment and Authorization of government systems that incorporate cloud computing, such that once a particular solution has been accredited, that accreditation can be passed on to other organizations and systems. This will hopefully reduce the burden and redundancy of the current process for accrediting government systems and accelerate adoption of cloud computing solutions.

4.3 Applicable DoD Direction

The Department of Defense Information Enterprise Strategic Plan 2010-2012 highlights the importance of information sharing, IT innovation and cultural change. It lays out six goals for establishing "a robust, reliable, rapidly scalable and interoperable infrastructure; and achieving synchronized and responsive operation of the DoD Information Enterprise." [29]

1. Information as a Strategic Asset
2. Interoperable Infrastructure
3. Synchronized & Responsive Operations
4. Identity & Information Assurance

5. Optimized Investments
6. Agile Information Management/Information Technology/Information Assurance Workforce

While cloud computing can help the DoD achieve all of these strategic goals, it is specifically identified in the strategic plan as key components of goals 1 and 2. Just as Linthicum does in [8], this plan emphasizes the relationship between cloud computing and SOA (see Section 2.4) for bringing enterprise services together efficiently. In striving to make information a strategic asset, the plan states that “cloud computing centers enable data and service transparency, and provide the foundation to run enterprise services securely and consistently across the DoD.” Developing and implementing cloud computing techniques will also help make current, accessible, secure and reliable information available to all authorized users, both known and unanticipated.

In describing how best to achieve an interoperable infrastrcture, the plan specifies that the “[DoD] must transform its infrastructure concept to support new service-oriented approaches, such as cloud computing and virtualization, for sharing, storing, processing and transporting information.” It goes on to describe the benefits of this approach in a similar fashion to those presented in Chapter III. Of particular interest, is how cloud computing can decrease the physical footprint, logistical trail and electrical consumption of IT resources, as well as how it can reduce the required amount of skilled touch-labor. Ultimately, these savings can be used to increase mission effectiveness for the warfighter.

The Department of Defense Information Enterprise Strategic Plan 2010-2012 does a good job describing the benefits, but achieving the stated goals still requires that the solutions meet the applicable regulations and guidance that governs information technologies. In this regard, DoD Directive 8500.01E “Information Assurance” and DoD Instruction 8500.2 “Information Assurance Implementation” are the most applicable documents for analyzing DoD specific IT security requirements impacting public cloud computing adoption. Analogous to FISMA regulation and the NIST guidance, 8500.01E lays out an overarching framework and 8500.2 provides specific controls required to secure DoD information systems.

4.3.1 DoD Directive 8500.01E: Information Assurance. The stated purpose of 8500.01E is to establish policy and assign responsibility to achieve DoD IA through a defense-in-depth approach integrating the capabilities of personnel, operations, and technology to support network-centric warfare (NCW). [30] It parallels FISMA regulation by declaring that DoD information systems, must to be categorized according to their required levels of confidentiality, integrity, and availability. It also recognizes the balance between the importance and sensitivity of information systems, the threats and vulnerabilities of those systems, the trustworthiness of users and connections and cost effectiveness of their implementation. The document goes on to define DoD information systems as a "set of information resources organized for the collection, storage, processing, maintenance, use sharing, dissemination, disposition, display, or transmission of information." DoD information systems can be categorized into four categories:

- Automated information system (AIS) applications
- Enclaves (which include networks)
- Outsourced IT-based processes
- Platform IT interconnections

The most interesting category from the perspective of public cloud computing is "Outsourced IT-based processes," since this category covers business processes supported by the private sector. Unfortunately, 8500.01E does not elaborate much on the particular pros and cons of designing DoD information systems using outsourced IT. The only specific guidance is that outsourced IT processes should perform clearly defined functions and have readily identifiable security considerations that are addressed in both acquisition and operations. This is one area where collaboration between the

DoD, the Federal CIO and NIST would be very beneficial, since the federal guidance outlined in Section 4.2.3 is especially relevant here.

In a similar manner to FIPS 199, 8500.01E also presents a mandate to classify the security posture for information systems. Although, instead of defining a security category according to Equation 1 it calls for DoD information systems to be assigned one of three Mission Assurance Categories (MAC). MACs focus primarily on integrity and availability. Levels governing the third security objective, confidentiality, are defined in DoD Instruction 8500.2 (See Section 4.3.2). The following list defines the criteria for MAC classification:

- *MAC I:* Systems handling information that is determined to be vital to the operational readiness or mission effectiveness of deployed and contingency forces in terms of both content and timeliness. The consequences of loss of integrity or availability of a MAC I system are unacceptable and could include the immediate and sustained loss of mission effectiveness. MAC I systems require the most stringent protection measures.
- *MAC II:* Systems handling information that is important to the support of deployed and contingency forces. The consequences of loss of integrity are unacceptable. Loss of availability is difficult to deal with and can only be tolerated for a short time. The consequences could include delay or degradation in providing important support services or commodities that may seriously impact mission effectiveness or operational readiness. MAC II systems require additional safeguards beyond best practices to ensure adequate assurance.
- *MAC III:* Systems handling information that is necessary for the conduct of day-to-day business, but does not materially affect support to deployed or contingency forces in the short-term. The consequences of loss of integrity or availability can be tolerated or overcome without significant impacts on mission effectiveness or operational readiness. The consequences could include the delay or degradation of services or commodities enabling routine activities. MAC III

systems require protective measures, techniques or procedures generally commensurate with commercial best practices.

4.3.2 DoD Instruction 8500.2: Information Assurance Implementation. 8500.2 augments the MAC concept, by defining the appropriate posture for DoD information systems with respect their required level of confidentiality. The confidentiality level is primarily used to establish acceptable access factors, such as personnel security and network controls. It breaks confidentiality into three levels: *classified, sensitive,* and *public.* [31]

The Instruction goes on to present the three MAC and three confidentiality levels in nine possible combinations and establishes the IA controls required for each possibility. As opposed to 17 security related areas defined by NIST, 8500.2 establishes eight IA Control Subject Areas:

- Security Design & Configuration (DC)
- Identification and Authentication (IA)
- Enclave and Computing Environment (EC)
- Enclave Boundary Defense (EB)
- Physical and Environmental (PE)
- Personnel (PR)
- Continuity (CO)
- Vulnerability and Incident Management (VI)

The controls associated with each MAC and confidentiality level are outlined in a series of appendices. Again, the reader is encouraged to review the appendices to obtain a comprehensive understanding of security control requirements that govern DoD information systems. The specific IA controls established in 8500.2 should be used as a basis for the SLAs and contracts between the program manager (PM) and the cloud service provider.

4.4 Bridging the Gap Between the DoD and Public Cloud Computing

The framework established by FISMA, NIST, the Federal CIO and the DoD provide federal agencies with a comprehensive risk management approach for securing information systems. It acknowledges the trend toward increased dependence on service offerings from external providers and establishes a basis for evaluating the security posture of public cloud computing providers with respect to federal agency requirements.

As highlighted in SP 800-53, trust is a primary consideration and it will continue to develop over time. Fortunately, external providers are making strides to help establish that trust. For example, In November 2009, Amazon Web Services (AWS) successfully completed a Statement on Auditing Standards (SAS) 70 Type II audit, and in November 2010 they were awarded an International Organization for Standardization (ISO) 20001 certification. In addition, a Department of Education application, with a security category of LOW under FISMA regulation, received the necessary Authority to Operate using AWS. [32] Google has also achieved similar certifications. [33]

Government agencies concerned about meeting FISMA regulations and their own agency's policies often find multitenancy to a considerable stumbling block. So, public providers are also increasingly making segmented versions of their offerings available in an effort to help organizations weary of commingling their data with others in a multitenant environment. In these cases, rather than allocating resources from a single large pool of hardware, the provider establishes a completely separate pool of resources for a specific organization or user. This has the benefit of ensuring that no other customer will share a physical resource. Also, this process helps the organization meet compliance requirements as well as eliminating some of the threats in Chapter V, especially those listed in Section 5.3. On the downside, customers will probably bear an increased cost since these systems cannot leverage the same economies of scale as those in multitenant environments. Fortunately, the total system costs are still likely to be less they would otherwise be in a traditional deployment. Also, other key benefits of public cloud computing, such as on-demand scalability and continuous refresh will still be realized.

Two recent examples are Google Apps for Government and Amazon EC2 Dedicated Instances. Google Apps for Government offers separate systems that are reserved for US government only use. The government has the ability to leverage all of the technology solutions, scalability and redundancy typical of Google while having the peace of mind that their data and services are not intertwined with others. Google reserves these systems in U.S. only locations and only makes them available to federal, state, or local governments in the U.S. [33]

Amazon EC2 offers users the option of requesting dedicated instances. The entire data center is not completely isolated as it is in Google Apps for Government, but only a single customer will be assigned to a dedicated physical server. For now, EC2 Dedicated Instances are only available in either Virginia or Ireland, which essentially restricts U.S. government use to a single region. This could potentially open up the possibility of data loss if the Virgina facility were to fail. Amazon will likely continue to expand its offerings to other regions in the U.S. as demand for this type of service increases. EC2 Dedicated Instances have two pricing components. One is an hourly per instance usage fee that is slightly higher than a standard instance charge. The other is flat rate fee per hour, and also per region, for using the service regardless of the total number of running dedicated instances. [34]

The federal government is also taking proactive steps to strengthen the trust between external providers and federal agencies. The Federal CIOs office recently launched the Federal Risk and Authorization Management Program (FedRAMP). The program was established to "provide a standard approach to Assessing and Authorizing (A&A) cloud computing services and products. FedRAMP allows joint authorizations and continuous security monitoring services for Government and Commercial cloud computing systems intended for multi-agency use." It also aims to provide a risk model enabling the government to “approve once, and use often.” [35] This program should help streamline the approval process for public cloud computing services and enable IT managers to capitalize on the benefits that public cloud computing offers.

DoD managers should also understand that commercial cloud service providers are highly incentivized to take a proactive stance toward security. In June 2010, IDC forecast that worldwide revenue from public IT cloud services would reach $55.5 billion in 2014. [36] Providers recognize that despite some technical challenges, there are not significant barriers preventing consumers from switching from one provider to another. A sufficient security breech could result in a tremendous loss of revenue as consumers make that change. These financial realities make security a top priority, which ultimately make public offerings more attractive.

Much work has been done to help address the minimum security requirements in the regulations and policies outlined in this chapter. Cloud computing providers have built their business models around common security best practices. Physical security, personnel training and monitoring, audit logging, patch configuration management, data encryption, secure authentication, network firewall management and intrusion detection are well established and routinely audited by independent reviewers. [37] [38] [39] Researchers are also constantly investigating ways to make the cloud more palatable for both the government and private industry. Examples include *FABRIC*, a platform for securing distributed workloads; [40] *TERRA*, a trusted computing platform for virtual machines; [41] the Trusted Cloud Computing Platform (TCCP) for guaranteeing confidential execution of guest virtual machines in IaaS environments; [42] and the cloud security solution based on virtual-machine introspection presented in [43].

4.5 Summary

With these facts in mind, DoD managers should resist the urge to fall back on the status quo. In addition, they must not draw incorrect conclusions that FISMA compliance necessarily excludes public cloud computing as some government publications have in the past. [44] As NIST and Federal CIO guidance shows, security measures can be put in place that allow external providers to meet federal regulations. [16] [27] [28] DoD managers can successfully leverage the technology coming from the commercial sector by evaluating their necessary security posture and then implementing security controls tailored to meet their needs. The result will be a public cloud computing solution that is both security and effective.

V. Threats to Public Cloud Computing

"The man who trades freedom for security does not deserve nor will he ever receive either." - Benjamin Franklin

"We will bankrupt ourselves in the vain search for absolute security." - Dwight D. Eisenhower

So far, this research has presented a framework for implementing secure architectures and for establishing trust between cloud consumers and cloud providers. The intent of this chapter is to outline specific threats to IT solutions that incorporate public cloud computing and to focus primarily on those threats that leverage the unique characteristics of cloud computing to implement their malicious behavior. Therefore, generic security attacks are not covered, since those attacks threaten all IT systems.

Information about specific threats to cloud computing is a rapidly evolving area of study; for both good and bad actors. The constant game of "chicken and egg" will continue to play out for the foreseeable future. Countermeasures are implemented almost as fast as these styles of vulnerabilities emerge, so any discussion of specific attacks is likely to be out-dated rather quickly. The following examples aim to help the reader gain a deeper appreciation for what kinds of threats can present themselves as well as foster a better understanding of some of the more unique aspects of cloud computing security. The threats presented here represent *root causes* to be used in risk analysis. Managers should also determine the *likelihood* and *consequence* of each threat occurring in their enterprises through their formal risk management process.

5.1 Undermining Authentication

Recall from Chapter II that one of cloud computing's essential characteristics is "Broad Network Access." Consumers of cloud computing must connect to the provider through a network to obtain access the information and services that provider hosts. In the case of public cloud computing, the connection flows across the Internet and therefore strong encryption and authentication methods are required to safeguard the data.

Weak passwords have always posed a security threat, but the increased reliance on remote connections associated with public cloud computing makes the attack surface for a malicious user much broader than it would otherwise be against isolated applications. Developers who use public cloud computing often access the services through Virtual Private Networks (VPNs), remote desktop clients, Secure Shell (SSH), or web consoles using Secure Sockets Layer (SSL) and/or Transport Layer Security (TLS). While each of these applications can be implemented securely, their security features can be easily undone if weak passwords are employed. Once authentication or identity management mechanisms have been compromised an attacker can potentially violate any or all of confidentiality, integrity or availability.

Cox states that "weak authentication, due to poor credentials, is the main exposure on the Internet, period." [45] He goes on to provide a series of mitigating strategies to alleviate the problem. The most effective strategy is to use multi-factor authentication where users would not only employ a password, but also some other form of credential. Identity management often relies on authenticated through "something you know," "something you have," or "something you are." Passwords are a part of the first category. Other options for authentication include token generators that provide the user with a random number that is only valid for a specific period of time. To authenticate, the user must not only provide their password, but must also provide the number present on their token generator. The DoD Common Access Card is another example of multi-factor authentication since the user must know the secret PIN, but also be in possession of the card itself. "Something you are" refers to biometric devices that check for unique physical characteristics of the user such as fingerprint or iris scans. Using more than one factor to authenticate to a cloud service provider greatly enhances the security of the connectivity between the consumer and provider.

5.2 Undermining Virtualization

Providers of virtualization solutions often highlight the security benefits of employing a virtualized environment as opposed to a traditional, bare-metal implementation. One of the key benefits of virtualization is isolation (see Section 2.2.5). Unfortunately, creators of malicious code are actively attempting to break the security barrier offered by virtualization thereby invalidating one of the fundamental security assumptions about virtualized environments.

Joanna Rutkowska is one of the researchers actively involved in this area. Two well known attack vectors discovered by Rutkowska are referred to as "Red Pill" and "Blue Pill." The names of the attacks are a nod to the popular science fiction movie, "The Matrix," where the lead character Neo is offered two different colored pills that allow him to escape his alternate reality. If the reader is familiar with the movie, then parallel between it and virtualization should be obvious.

In 2004, Rutkowska revealed an extraordinarily compact piece of code (5 lines) that could detect the presence of a hypervisor in between the operating system and the hardware. She called her code the Red Pill. [46] In theory, guest operating systems should not be able to determine if they are running on bare-metal or on a hypervisor. Detecting the presence of the hypervisor is the first step in escaping its confines. This is especially important for malware researchers who often employ virtual machines as a safe way of monitoring the execution of malicious code. An alarming trend in malware is to employ techniques like the Red Pill and modify their functionality if they detect the presence of a virtual machine. In this fashion, their real behavior cannot be determined making it much more difficult to create detection signatures.

In 2006, Rutkowska briefed the Blue Pill at a Black Hat conference. [47] The basic idea behind the Blue Pill is to create a malicious virtual machine monitor that inserts itself between an operating system and its hardware. If successful, the malware would be extremely difficult to detect and would have total access to practically everything on the system. Fortunately, her Red Pill technique provides an effective way of actually detecting the presence of such a nefarious piece of code. The original attack relied on specific virtualization enhancements provided by AMD-V chipset and was later ported to the Intel VT-x chipset. Both of these chipset technologies include functionality essential for efficient operation of virtual machines.

Cloud computing providers employ chipsets with similar virtualization enhanced hardware in their resource pools. While some of the claimed capability of the Red Pill was disputed by researchers, it definitely highlights that virtualization technologies are potentially vulnerable to attacks. If a Red Pill style attack could be executed against a running hypervisor, as opposed to simply a host operating system, that would put all running guest operating systems at risk and invalidate the isolation between them. This would be particularly devastating to public cloud computing, which relies on multitenancy to drive down costs.

5.3 Undermining Multitenancy

Ristenpart et al. leveraged the multitenant nature of public cloud computing to launch a number of attacks against confidentiality in public clouds. [48] Their analysis sought to answer the following four questions about the predictability of cloud-based resource allocation and the potential for information leakage:

1. Can one determine where in the cloud infrastructure an instance is located?
2. Can one easily determine if two instances are co-resident on the same physical machine?
3. Can an adversary launch instances that will be co-resident with other user's instances?
4. Can an adversary exploit cross-virtual machine information leakage once co-resident?

The attacks defined in [48] follow two basic steps, *placement* and *extraction*. The goal of placement is to get an attacker's malicious VMI on the same physical server as a victim's VMI. They refer to this result as "co-residence." They successfully demonstrated that they could achieve co-residence approximately 40% of the time employing only the tools available to standard users and without any help or inside knowledge of the cloud provider's internal operations. Factors contributing to the success of their attack included matching their own machine instance type to that of the victim, choosing the same availability zone[1] as the victim and launching their malicious VMI at roughly the same time as the victim's launches. The first two contributing factors are relatively straight-forward to determine. The researchers found that the Internet Protocol (IP) addresses assigned to different VMIs and different availability zones was highly predictable. They were able to use network probing techniques, such as *nmap*, to determine the instance type and availability zone of target machines by inspecting their IP addresses.

The third factor in achieving co-residency seems much more difficult accomplish. At first glance, one would expect that it would be next to impossible for an attacker to determine exactly when a victim launches a VMI. A key insight of the research was to leverage the dynamically scalable nature of cloud computing resources to help achieve success. They were able to demonstrate that VMIs would often be assigned the same IP address if it was terminated and then relaunched. This could occur if demand for the service fell and then reappeared at a later time. By continuously probing the victim's machine, with any technique similar to a *ping* request, the attacker could see a VMI stop and then restart. As soon as a restart is detected, the attacker could flood the cloud provider with requests for resources and have a reasonable chance of achieving co-residency. A more insideous version of the attack involves the attacker actually causing a victim's VMI to launch by flooding the victim's service with artificial demand. The cloud provider would respond to the additional load and dynamically scale the victim's service by starting additional instances. The attacker would time

[1] Availability zones describe the geographic area where an instance will be assigned. Users have the ability to select availability zones to support distributed backup strategies as well as meet legal requirements.

the launch of their own instances to coincide with their artificially generated demand and have a high-likely hood of instigating co-residence.

Once placement is achieved, Ristenpart et al. showed how side-channel attacks could be used to infer information about other VMIs running on the same server. The primary technique involved using time-shared CPU caches to measure when other instances assigned to the same physical server are under computational load. While this attack does not allow direct access to data residing on the victim instances, it does allow the attacker to gain insight into the level of activity the victim is experiencing. This information might reveal that some particularly important event is occurring or allow the attacker to gather information about the overall level of business activity of the victim. A worse case scenario would allow the attacker to infer user entered passwords for remote logins. If the victim's machine is sufficiently idle, then each keystroke entered would correspond to some small spike in CPU utilization. In theory, by timing these spikes, the attacker could infer the layout of the victim's password on the keyboard by measuring the time difference between keystroke presses. Fortunately, such an attack requires measurement granularity and extremely "quiet" environments that make attacker's success in using such a technique in a public cloud extremely unlikely.

Finally, the authors provide a series of recommendations to counter this style of attack. The first approach is for cloud providers to obfuscate the internal resource allocation mechanisms to confuse the adversary's ability to achieve co-residence. Second, users can employ "blinding" techniques that mask utilization. For instance, a user could stuff communication channels with fake data to mask when actual data, such as key strokes, are traversing their connections. Ultimately, the most effective solution is for providers to offer consumers the ability to request dedicated physical resources or to only allow co-residence with other trusted entities. In this case, the user would be required to bear the costs associated with the underutilization that would occur with respect to a more efficient, pure multitenant environment. Fortunately, cloud providers are well on their way to implementing this solution. As described in Section 4.4, both Amazon EC2 and Google App Engine have established mechanisms for isolating physical resources while still realizing the benefit of employing cloud computing.

5.4 Undermining Community Trust

Marco Slaviero presented a series of attacks against IaaS providers at BlackHat in 2009. [49] One particular attack leveraged the trust relationship that often exists between cloud developers. The attack involved how public IaaS providers often offer consumers prebuilt VMIs to help get them started. For example, there may be a VMI that contains everything the user needs to get a website up and running. That image might come preloaded with Linux, Apache, MySQL and Perl preinstalled saving the user a considerable amount of setup time. Often IaaS providers take advantage of the developer community by allowing them to load custom VMIs into a public directory where other users can download them and put them to their own use. This developer environment is a tremendously powerful way to propagate good ideas.

Unfortunately, Slaviero was able to demonstrate how a malicious user could undermine the trust inherent in this distribution mechanism. First, he created a VMI with a script that would "call home" when another user in the community loaded an instance of his VMI. He then scripted the machine image registration process for an IaaS provider in such a way that his VMI would show up toward the top of the list of available VMIs. Slaviero recognized that the IaaS provider was assigning a random unique identifier to the newly uploaded VMI and was using that identifier to sort the list of available images. So, he created a script that would repeatedly register the image until it was assigned an identification number that put it at the top of the list. As most people are aware, search rankings are extremely important for generating traffic and Slaviero assumed that if his VMI ranked high on the list that users would be more likely to use it. He was correct.

The process took approximately 12 hours to achieve its goal, but eventually the image showed up in the 5th spot on the list of machines. Once it was listed, Slaviero changed the name to something that looked useful (in this case, he used "fedora_core_11"). As luck would have it, his machine was the highest ranked Fedora image on the list and so anyone who was looking for a machine pre-loaded with Fedora would see his image first. By monitoring the call-back mechanism that had been implanted in the VMI, Slaviero was able to show that his scheme had worked. Within four hours he received notification through the call-back script that his image had been loaded by another user.

This style of attack undermines the trust that is often created in these types of communities. The problem is exacerbated if the IaaS provider is a reputable source, since users might feel that anything offered on the site may also share the same level of trustworthiness. Obviously, the best way to prevent such an attack is for users to build their own VMIs. This would be more time consuming, but would provide the user with a better understanding of the security pedigree of the loaded software. Another method would be to only use images offered by the provider themselves. Presumably, these images would actually be as trustworthy as the provider. Finally, third party curators or a community review/rating mechanism could be used to establish VMI trust.

5.5 Undermining Virtual Machine Image Integrity

One of the key technical challenges associated with successfully implementing a cloud-based business model is accurately measuring the amount of usage of a particular resource. While CPU cycles, storage and memory can be allotted with exceptional granularity, it is much more difficult to measure software usage with similar accuracy. Software vendors are having to come to grips with a deployment architecture that turns the traditional software licensing model on its head. For example, the traditional user might purchase a single copy of a piece of software where they would be authorized to run that software on a single physical machine. When a copy of a piece of software is loaded into a VMI, then that VMI can easily migrate from one server to another, or even be copied across multiple machines with relative ease.

The scalability of cloud computing is heavily dependent on software licensing agreements catching up with the technology. The rise of open source software under General Public License (GPL) agreements has enabled much of the growth in the cloud sector. The Linux community, with its vast array of powerful and free software is a critical component of this space. That doesn't mean that traditional vendors, such as Microsoft, are left out. Many consumers still require the special functionality that sometimes only comes from purchased software.

As a result of this, commercial software vendors are quickly developing methods for accurately metering software usage and providing licensing agreements that allow consumers to leverage the full power of the cloud. One particular method for managing the usage of software is to allow consumers to "check out" VMIs from cloud providers at a usage cost commensurate with the capability installed within the VMI. For example, Amazon charges $0.085/hr for a small instance running Linux, while it charges $0.12/hr for a small instance running Windows. [50] Microsoft will receive a commensurate portion of the proceeds when consumers use an instance running their products. Other vendors also offer bundled commercial software for additional fees.

With that background in mind, Slaviero also demonstrated in [49] a mechanism that he called "AMI Stealing."[2] He found that he could subvert the process that Amazon EC2 uses to associate pay-for-use AMIs with their owners so that the owners can be paid when consumers deploy their AMI. Essentially, he showed that it was possible to violate the integrity of the instances by purchasing a single AMI instance, stripping the product code and ownership information, repackaging the instance using Amazon's tools and then uploading the modified AMI back into the cloud for subsequent use. In this manner, it was possible to only pay for an AMI once and then never have to pay the usage fee again.

[2] AMI stands for Amazon Machine Image and is Amazon's version of a VMI.

5.6 Potential Hazards of Third Party Involvement

Most of the attacks presented in this chapter are technical in nature, but technical threats are not the only aspects of cloud computing that need to be accounted for. Chapter IV emphasizes the criticality of establishing solid contracts and SLAs between the consumer and provider. Multiple federal documents that cover security in a cloud computing environment back up that assertion. [7] [27] [3] [16] When establishing these contracts, consumers must take into account any relationships between the primary provider and any third parties and that the security controls established by the contract/SLA with the primary provider also apply to their subcontractors. The following two examples highlight failures originating from third party relationships. Both cases resulted in a loss of availability and significantly impacted consumers.

The first case involves a widely publicized loss of data for users of Microsoft's Sidekick mobile device operating on the T-Mobile cellular network. On October 2$^{\text{nd}}$ 2009, a server failure at Danger, a subsidiary of Microsoft providing cloud based storage for the mobile device, resulted in customers not being able to access their calendars, address books and other personal information. The failure occurred in both the core database managing the service and the backup. Over the next two weeks, Microsoft frantically rebuilt the service and eventually recovered most, if not all, of the lost data. Unfortunately, by the time the data were recovered, the damaging publicity was extensive and it called into question the level of trust that users should have in cloud service providers. In this case, they were drawn to the perceived security of using a Microsoft product, but behind the scenes they were getting a service from a newly acquired subsidiary without appropriate safeguards in place. [51]

The second example involves another cloud storage provider. TheLinkup was meant to be a social network for file sharing that would allow users to share data over the web. Behind the scenes, though, the business relationships that were involved in making TheLinkup a reality were in serious trouble. The company was in the process of rebranding itself and was working to migrate user data from an older system to the new TheLinkup social environment. Another company, Nirvanix, was tasked with to managing the storage of customer data. Unfortunately, the task of migrating user information was technically infeasible. The convoluted business arrangements and a serious technical error by a system administrator caused a widespread data loss for many of the company's 20,000 paying customers. Eventually, TheLinkup shut down its site and many users were never able to recover their data. Both entities publicly blamed the other for the fiasco and customers were left out in the cold. [52]

5.7 Undermining Resource Availability

A third example from Marco Slaviero presented at the same BlackHat conference in 2009 demonstrated how consumers could potentially perform a DoS attack against a cloud resource provider. [49] Typically, cloud providers limit the total number of resources that any single account can access at the same time. Slaviero circumvented this safeguard by creating a VMI who's sole function was to create new accounts on the system, request the maximum number of resources allowed from each new account, replicate itself, and then start the process all over. By recursively registering accounts and loading them with self-replicating instances, an attacker could, if left unchecked, eat up all available resources inside the cloud. Once all of the resources have been taken, then other users would be effectively locked out of obtaining any new resources from the provider, which would severely limit their scalability and prevent any on-demand usage. This demonstration was possible because the provider did not take steps to prevent automatic registrations and didn't check for multiple uses of the same credit card across the different accounts.

While this was definitely not part of Slaviero's original demonstration, one could envision a scenario where the functionality inside the self replicating VMI is augmented with a network traffic generator designed to flood the provider's network with bogus traffic. If such a VMI were created, and allowed to propagate, the entire service of the cloud provider could potentially be stopped.

There are other threats to availability that are not necessarily malicious. While the focus tends to be on the technical aspects of information technology, environmental safeguards and fiscal solvency should also be a key factors in evaluating public cloud computing providers. In a private deployment, the agency responsible for the hardware is also directly responsible for ensuring that hardware is safe from fires, floods, severe weather and other forms of natural disasters. In a public deployment, the cloud computing provider bears the responsibility for ensuring these events do not impact their customers. Fortunately, most providers inherently offer data and service replication capabilities. That replication is often geographically dispersed to help avoid single points of failure that might lead to loss of availability due to a local emergency or natural disaster. Government managers looking to work with public providers may need to obtain additional insight into the security controls offered by the provider to ensure adequate safeguards are in place.

Also, the fiscal solvency of a public provider can be a serious issue. If a provider files for bankruptcy, it may shut down its services without notice leaving customers without access to their data or services. Not only should government managers carefully craft service agreements with individual providers, they may also wish to develop their infrastructures by using multiple public cloud service providers as a hedge against any one of them shutting off their services unexpectedly.

5.8 The Insider Threat

Probably the most feared aspect of public cloud computing is the insider threat. While a threat from insiders exists to some degree regardless of whether the solution uses public providers or not, the lack of direct control over personnel in a public cloud solution tends to heighten the level of fear of this particular threat. By using public cloud computing, some amount of control over one's data must necessarily be delegated to the provider. So, it is natural to worry that the provider won't live up to that responsibility. Many providers acknowledge this risk and are putting steps in place to counter it. Physical security, multi-factor authentication, least-privilege security policies and functionality separation are all actively employed by public providers to help develop the trust necessary to successfully operate their businesses. [37] [39] While these security mechanisms reduce the risk, consumers should still employ strong encryption and keep access control lists locked down to the minimum set of necessary personnel.

5.9 Summary

This chapter aimed to help the reader better understand some of the specific threats facing public cloud computing providers and their customers. Unfortunately, malicious actors are constantly looking for new ways to compromise information. As such, there is no way to provide a truly comprehensive overview of the threats because the threats are constantly changing. While the focus here was on security threats against public cloud computing providers, the reader should remember that a similar security struggle exists for private IT enterprises albeit with a somewhat different mix of threats.

When performing risk management, these topics can be evaluated as potential root causes of risk and they should be mitigated according to their likelihood and consequence in the context of the design goals of the architecture. IT security will continue to be top priority for the foreseeable future and as always, IT managers should carefully balance the benefits against the risks of any proposed solution. As discussed in Section 4.4, cloud computing providers are highly incentivized to ensure the security of their offerings. As the business model for cloud computing matures and its foundational technologies advance in capability, the security of these services will continue to increase.

VI. Threats Enabled by Public Cloud Computing

"Adversaries in cyberspace are exploiting low-entry costs, widely available resources, and minimal required technological investment to inflict serious harm, resulting in an increasingly complex and distributed environment."
- Air Force Doctrine Document 3-12: Cyberspace Operations

"Disruptive challenges may come from adversaries who develop and use breakthrough technologies to negate current U.S. advantages in key operational domains." - National Defense Strategy, 2005

Cyberspace is a contested domain, and the DoD has established that warfare exists both in and through cyberspace. Like all other domains where warfare exists, combatants strive to dominate their opponent with various capabilities to ensure their superiority in the domain. The objectives of the United States Air Force are no different in cyberspace as outlined by Air Force Doctrine Document (AFDD) 3-12 "Cyberspace Operations." [53] *Cyberspace Superiority* is the operational advantage in, through, and from cyberspace to conduct operations at a given time and in a given domain without prohibitive interference. To now, this research has focused on DoD application of public cloud computing. This chapter shifts the focus to public cloud computing as a potential force-equalizer that adversaries may employ to reduce US superiority in the cyber domain.

Until recently, the costs associated with building and maintaining large IT infrastructures have kept massive scale processing out of the hands of many would-be adversaries. Now, cloud computing is poised to revolutionize the ability for governments, private industry, non-state actors, and private individuals to access tremendous computing power at a fraction of the traditional cost and effort. Unfortunately, technological innovations foster malevolent applications as readily as the beneficial ones. As a result, each of the benevolent use cases highlighted in Section 3.7 has such a malevolent equivalent. It does not require much imagination to realize that cloud computing could be a key enabler of our adversaries. The following section focuses on cloud computing as a weapon that could be employed by our adversaries against our interests.

6.1 Using Cloud Computing for Denial of Service

Just as MySpace was able to leverage a public cloud to perform large-scale, simulated load testing on its servers, other organizations or individuals may wish to employ similar applications to overload target networks with a flood of simultaneous requests or unsolicited communications. The result, for an unprepared host, would be a complete shutdown of its service. This is referred to as a *Distributed Denial of Service (DDoS)* attack.

In the past, these types of attacks were promulgated by infecting unwitting computers (known as *zombies*) with viruses, trojan horses and other forms of malware to form a *botnet* of machines. The network of infected computers generally lay hidden and dormant until called into coordinated and directed action by a control node. Now, the task of executing a DDoS attack may no longer require an attacker to subvert security measures and choreograph distributed malware across a heterogeneous mix of machines they don't own. Instead, an attacker looking to incapacitate a network or host node could call up, for example, 1000 servers from an IaaS provider, then install and launch its attack from a dedicated pool of on-demand resources. For reference, 1000 "small instances" from Amazon EC2 could cost as little as $85/hr. [50]

In an actual demonstration of this style of attack, two security researchers presenting at DEF CON 18 showed how they could take down a client's web page through a DDoS attack launched from Amazon EC2. They called their platform "Thunder Clap," which basically used packet flooding originating from within the servers provisioned by Amazon to completely saturate the bandwidth of their target website. They performed the attack for two hours with only three cloud-based servers at a cost of only $6. This proof of concept not only demonstrated the power of cloud computing as a potential attack platform, but also showed how few constraints and how little oversight there is against malicious activity originating in the cloud. During their demonstration they did not encounter any bandwidth restrictions imposed by the provider nor did their activity seem to trigger any kind of alerting mechanisms.

One of the presenters, David Bryan, even went so far as to say that, “with the help of the cloud, taking down small and midsize companies’ networks is easy.” [54]

6.2 Using Cloud Computing to Crack Passwords

Instead of launching DDoS attacks, that same 1000 machines at $85/hr described earlier could be employed in an attempt to break encryption mechanisms used for secure communication. This would lead to adversaries having access to sensitive information, without the victim even knowing that their connections were not secure. Legitimate companies and organizations provide private enterprises and law enforcement with software tools that enable the distribution of decryption workload across thousands of machines simultaneously. AccessData teamed up with Parabon to create *Distributed Network Attack with Frontier* for just such a purpose. [55]

An example of this style of attack is publicly available for as little as $17! An attacker can upload captured WPA-PSK network traffic to WPACracker.com to launch a 135-million word dictionary attack against a wireless network’s password. The service leverages 400 CPU instances from Amazon EC2, which provide an average run-time of only 20 minutes. The website will even expand the dictionary it uses to perform the cracking operation to 284 million words for only $40. WPACracker.com advertises that a similar attack would take five days running on a dual-core PC. Since it is a dictionary attack, there is no guarantee that this particular approach will be successful, but this does provide an early example of cloud based decryption services. [56]

Fortunately, there are currently no known techniques for breaking strong encryption techniques such as the Advanced Encryption Standard (AES), Triple Data Encryption Standard (3DES) or the standard founded by Rivest, Shamir and Adleman, called RSA, in a reasonable amount of time. To provide a little background for the reader, encryption is often based on the difficulty of factoring the products of two large prime numbers. If these numbers could be factored, then it would be trivial to decode the contents of the message. Even with strong encryption, an attacker with enough computing power and enough time could factor the key and break the encryption using a brute force approach. To date, the largest key to be successfully factored is a 768-bit RSA key (a number with 232 digits) and it took the equivalent of 2000 2.2GHz-Opteron-CPU years to perform the necessary calculations. To break a more representative 1024-bit RSA key would require approximately 1000 times more effort. The research team that broke the 768-bit RSA key estimated that the growth in processing capacity will enable the same feet to be accomplished on 1024-bit key before the end of the decade. [57]

In November of 2010, Amazon announced that it would offer clustered graphics processor units (GPUs) as a new type of cloud instance. GPUs are known for their ability to perform massive amounts of parallel mathematical calculations to render complex scenery in gaming environments and to enable other high-performance applications. Now that these devices are being offered in the cloud, the computational power available to a would-be attacker has increased dramatically at the current cost of only $2.10/hr. [50]

Shortly after Amazon released its new instances, Thomas Roth, a German researcher, took advantage of their tremendous power and demonstrated how one could conduct efficient brute force attacks against the SHA-1 encrypted passwords and even WPA-PSK wireless infrastructures. The difference between this and other brute force attempts was the availability of low-cost, extremely fast hardware that enabled the passwords to be cracked in a relatively short amount of time. In his blog, he highlights how SHA-1 was never intended for cryptography. SHA-1 was intended to output fast "digital fingerprints" for validating the integrity of electronic files. The speed at which SHA-1 operates contributed to Roth's ability to crack the password hashes quickly. His technique enabled him to crack 14 SHA-1 encrypted passwords in only 49 minutes. Each password was 1-6 characters long and consisted of a set of 95 possible characters. In conducting this research, Roth advocated the use of more cryptographically secure hashing functions, such as *scrypt* or *PBKDF2*, which are intentionally slow to complicate similar cracking attempts. [58]

Then in February 2011, Roth expanded his research toward cracking WPA-PSK encrypted wireless communication. Using the same processing architecture on Amazon, he was able to generate approximately 400,000 cracking attempts per second. He believes that he can achieve 1,000,000 attempts per second in the near future by optimizing his algorithm. While many headlines indicated that WPA had been "broken," the fact was that the attack style employed a brute force approach using extremely fast, cheap, and readily available compute resources from the cloud. [59]

6.3 Using Cloud Computing For Information Warfare

Cloud computing can provide tremendous technical capability to organized criminals, state or non-state adversaries, or simply rogue actors. Unfortunately from a national security perspective, cloud computing providers can offer even more basic capabilities that are more accessible to the masses since they don't require detailed technical knowledge to make them operate. Numerous chat rooms, blogs, websites, file sharing services, and wiki pages operate on high-level SaaS offerings. Recall from Section 2.2 that there are many cloud service models. The previous examples were primarily carried out using IaaS, which puts very low-level control in the hands of the user. SaaS instead offers predefined functionality in well defined packages. It is generally simple to use and often comes free of charge.

SaaS services are being used to conduct information warfare against our national interests. Terrorists and criminals alike can post information on blogs, propagandize on websites, coordinate activity through social media and communicate through e-mail. Twitter, Facebook, Gmail, Blogger, and Google Sites, just to name a few, allow users to freely access these types of services. Setting up an account requires no credentials and no credit card for basic services making them extremely fast to set up and as well as effectively anonymizing the user. Government entities trying to combat these types of attacks have to play the proverbial "Whack-a-Mole," because as soon as one service or account is shut down, the user can simply recreate their activity on a different site.

Washington Post reporters Coll and Glaser highlight how “al-Qaeda has become the first guerrilla movement in history to migrate from physical space to cyberspace.” They go on to give examples of how al-Qeada uses transient message boards and chat rooms to communicate about subjects like mixing ricin poison, making bombs from commercial chemicals, shooting U.S. soldiers and celestial navigation in the desert at night. Since these channels can be vulnerable to eavsedropping, terrorists sometimes turn to more advanced covert techniques. Khalid Sheik Mohammed, a key planner of the 9/11 attacks, used a digital “dead drop” to evade eavesdropping by the U.S. and other governments. To do so, he and his operatives opened random accounts on free, web-based e-mail hosting services. They would write messages as e-mails, but instead of sending them (which would open them up to possible interception), they saved the message as a draft. Then, they would post the account login information on a message board so that the person they wished to communicate with could use the credentials to open the draft message and review its content. [60] These examples show how public cloud computing can be a powerful tool for maintaining anonymity and confidentiality.

Google Earth is a SaaS platform that provides commercial satellite and airborne imagery spanning the entire globe to its users for free. In addition to the imagery, users have access to many layers of geospatial information compiled from content generated by Google, publicly available sources and an avid user community who regularly contribute to the project. The service has been around since 2005 and after the initial fascination with the product wore off, users began to see how access to such a wealth of information could be potentially damaging in the wrong hands. Some have even petitioned Google to blur the imagery of select locations for security reasons.

Unfortunately, terrorist movements have been quick to adopt the service as a cheap and efficient intelligence gathering platform. In 2007, the Daily Telegraph reported that Google Earth imagery of British bases in Basra, Iraq had been found inside the homes of insurgents. They were using the tool to plan attacks against allied forces. The high quality of the imagery, along with its precise geopositioning, gave the insurgents an information advantage that they could not have had without the power of the Internet. [61]

The concern over such services has gotten so great that a court in India has even been called to ban on Google earth because it "aids terrorists in plotting attacks." Advocates are asking the court to direct Google to blur images of sensitive areas until the case can be decided. The petition comes in light of the attacks that took place in Mumbai in 2008, where technically savvy gunman used GPS, satellite phones and satellite imagery, found through services such as Google Earth, to carefully plan and execute the deadly attack. [62]

Then, there is the case of Colleen LaRose. LaRose is an American citizen who used the public cloud services, most notably YouTube and Dailymotion, to recruit jihadists and also to spread anti-U.S. propaganda. She posted videos on the two services under the aliases "Jihad Jane" and "Fatima LaRose." The FBI tracked her online activities for several months after being tipped off by a vigilant group of volunteers who scour the web looking for extremists. They eventually arrested her in Ireland where she had gone to assassinate a cartoonist named Lars Vilk. Eventually, she pleaded guilty to providing material support to terrorists, conspiracy to kill in a foreign country, and attempted identity theft. [63] [64]

The story of Jihad Jane is an example of the power that public cloud services can have in spreading propaganda. Their low cost, ease of use and widespread accessibility are all contributing factors. Her case gained a significant amount of notoriety because of her background. Despite her upbringing in a small Pennsylvania town and her Christian origins, she was lured into the world of violent extremists by their successful use of Internet technologies to proliferate their message. After succumbing to their ideology, she also employed these tools to communicate her own message and organize an assassination plot. Fortunately, thanks to a set of watchful citizens, she was thwarted before she could carry out her plan.

A more recent, high-profile example of a nefarious actor employing cloud-based services to host information attacks occurred in the fall of 2010. Hackers launched a number of DoS attacks against the infamous Wikileaks website in response to the information that it was presenting on its pages. The attacks forced Wikileaks to abandon their original hosting servers and migrate to Amazon's cloud. The robustness of Amazon's architecture and its simple interface enabled the controversial site to rapidly re-host itself. To its credit, after discovering that its infrastructure was being used in this way, AWS quickly removed the offending website from its cloud. This action demonstrated an early, high-profile example of how providers can police themselves against bad actors. There is still a long way to go, but examples of this type of responsible behavior is promising for the future of effective cloud computing. [65]

6.4 Summary

Public cloud computing services that manage information provide society with tremendous benefits, but unfortunately these same services can also be used for insidious purposes. Our nation and our allies must carefully make decisions concerning the best way to maintain the appropriate use of these information sharing technologies without also infringing on the personal freedom and liberty of law abiding citizens.

The examples in this chapter were included to demonstrate the power that large-scale compute infrastructures can have in formulating attacks in cyberspace. Until recently, processing on this scale has only been available to large corporations, universities and state-sponsored organizations. This will not be the case for long, as public cloud computing service mature and grow in capacity. When they do, this power will be readily available to anyone with an Internet connection for both good and evil purposes.

The cloud services at Amazon.com play a part in many of the preceding examples. The intention here was not to single them out, but the facts are that the bulk of the currently publicized examples of how the cloud can be used maliciously involve their services in one form or another. Their technological leadership, brand recognition and efficient API have made them a de facto standard in the public cloud computing arena for both good and bad applications. As a result, they tend to receive far more attention than any other cloud service provider. Hopefully, the reader will not take away a negative view of this particular provider. Instead the reader should recognize that the first, and arguably the best, large-scale public provider of cloud services has produced a platform, not unlike the Internet at large, that provides a capability both tremendously beneficial, but also potentially dangerous. Also, it should be noted that each example described above could just as easily have been carried out using another IaaS provider with similar results.

From a doctrinal perspective, AFDD 3-12 does an excellent job of highlighting the threat of low-cost access to modern compute power by stating that “adversaries in cyberspace are exploiting low-entry costs, widely available resources, and minimal required technological investment to inflict serious harm, resulting in an increasingly complex and distributed environment. The expanded availability of COTS technology provides adversaries with increasingly flexible and affordable technology to adapt to military purposes. Low barriers to entry significantly decrease the traditional capability gap between the US and our adversaries. Adversaries are fielding sophisticated cyberspace systems and experimenting with advanced warfighting concepts.” As this research demonstrates, cloud computing can make massive-scale compute infrastructures available to friends and adversaries alike at extremely low costs. The preceding sections have identified the potential for public cloud computing to be used as a weapon in the cyber warfare domain.

Dr. David G. Ullman cautioned that when sufficient uncertainty exists decision makers can get stuck in the *Observe* and *Orient* phases of Col John Boyd’s OODA loop. [66] This chapter aimed to expand Ullman’s concept by acknowledging that many of our adversaries have moved past the *Observe* and *Orient* phases and are instead stuck at the *Decision* phase. Prior to widespread availability of public cloud computing services, many adversaries had the will, but not the capacity, to act on the decisions that they have made. The expansion of access to massive amounts of low-cost computational resources provided by public cloud computing service may enable adversaries to close their OODA loops and threaten U.S. efforts to achieve superiority in cyberspace.

VII. Recommendations

> "I cannot help fearing that men may reach a point where they look on every new theory as a danger, every innovation as a toilsome trouble, every social advance as a first step toward revolution, and that they may absolutely refuse to move at all." - Alexis de Tocqueville

> "As DoD moves further along the net-centric operations path, the Department must transform its infrastructure concept to support new service-oriented approaches, such as cloud computing and virtualization, for sharing, storing, processing and transporting information." - DoD Information Enterprise Strategic Plan 2010-2012

The benefits and challenges of adopting public cloud computing have been presented to the reader for consideration. The focus will now turn toward a series of recommendations to help the DoD integrate public cloud computing into its IT enterprise.

7.1 Education and Awareness

DoD IT managers must be educated on the benefits and challenges of using public cloud computing. Many managers are unaware that these services even exist. Once a general awareness is established, then those managers will be able to assess the pros and cons of the technologies and business models associated with public cloud computing in a more thorough manner.

The DoD must also tighten up the terminology it uses to describe cloud computing. There can often be a significant gap between what senior program managers say when they discuss the cloud and what IT experts hear if they are tasked to implement a cloud-based solution. The NIST definition of cloud computing should be used throughout the DoD and the Cloud Computing Reference Architecture should be used as a basis for training. Employing a consistent vocabulary will help managers generate better requirements and help the DoD communicate with public service providers through more precise SLAs and contracts.

7.2 Adopt Risk Management Over Risk Avoidance

Risk management is a critical aspect of all acquisition. It helps program managers evaluate the risk versus reward of one option over another. Unfortunately, DoD managers are often reluctant to allow external IT services into their option space because of the perception that their level of control is proportional to the security of their system. Hopefully, with a better understanding of cloud computing, a determined Federal CIO behind them, and continuous improvements in the security posture of public cloud providers, managers will take a closer look at employing public cloud computing into their IT architectures.

DoD managers must also ensure that when they discuss security risks that they take into account not only the impact of a security related event, but also the likelihood and root cause of that event. Two characteristics of modern IT exacerbate the problem: complexity and ubiquitousness. Stated simply, IT is everywhere and it permeates essentially every program in the DoD; yet very few have an understanding of how it actually works. The combination of these two factors lead many to a miscalculate risk. The complexity of modern IT often negates people's ability to comprehend the real root causes of security events, never mind estimate likelihoods of their occurrence. The ubiquitous nature of modern IT also serves to amplify our perception of what the consequences would be if a security event occurred.

When implementing their risk management process, DoD managers must strive to understand the nature of public cloud computing, develop trust relationships with external providers and gain an appreciation of the security architectures that public providers employ. Not doing so will restrict the DoD to the status quo and we will fail to capitalize on the benefits and innovations coming from the commercial sector.

7.3 Publish DoD Specific Guidance on How to Incorporate Public Cloud Computing

The Federal CIO and NIST are making great progress in defining and developing cloud computing for use in the federal government. The Federal Cloud Computing Strategy and the Proposed Security Assessment & Authorization for U.S. Government Cloud Computing have presented the foundation for moving the government toward cloud computing while also establishing the commensurate security controls necessary for its success.

The DoD should take similar steps to emphasize the need to integrate cloud computing into its information enterprise as well as establishing the appropriate methodology for its employment. The Department of Defense Information Enterprise Strategic Plan 2010-2012 is a step in the right direction, but additional guidance should be documented that focuses specifically on a DoD cloud computing strategy. The next step following the publication of a DoD cloud computing strategy would be to publish a complementary document to the Proposed Security Assessment & Authorization for U.S. Government Cloud Computing. The focus here would be to expand on the IA requirements outlined by the DoD 8500 series documents with a specific emphasis on the security controls necessary to accredit DoD information systems that leverage public cloud computing. Both the DoD cloud computing strategy and the IA controls for public cloud computing should should be developed in close coordination with the Federal CIO and NIST.

7.4 Engage Public Cloud Providers More Directly

Ultimately, security is based on trust. Managers have to trust the personnel, hardware, software, and business model along with a host of other factors when balancing risk mitigation strategies against their cost of implementation. The level of trust between the DoD and public cloud computing providers will be a crucial component of public cloud computing adoption. The DoD must engage public cloud providers directly so that it can gain a better appreciation for the security posture of those services as well as the business model required for their proper employment.

The primary vehicle for establishing and maintaining that trust will be SLAs. [16] states that the "SLA represents the understanding between the cloud subscriber and cloud provider about the expected level of service to be delivered and, in the event that the provider fails to deliver the service at the level specified, the compensation available to the cloud subscriber." There are two types of SLAs: predefined non-negotiable agreements and negotiated agreements. The former represent the agreements that providers have with the general public. They are unilateral in nature and are immediately available for use by anyone who desires to employ their services. These are likely to present the consumer with the maximum level of benefits of cloud computing since these SLAs are specifically crafted to leverage the greatest economies of scale for the provider. The downside of non-negotiable SLAs is that they can often change without notice and it can be difficult to obtain useful compensation if breaches occur.

Negotiated SLAs are more closely related to traditional contracts. The DoD can tailor the level of required performance, necessary security controls, articulate ownership and exit rights, and mandate compliance with federal and DoD regulations and policies. These SLAs are likely to be less cost effective, but can provide the DoD with additional controls over its relationship with an external provider. The DoD must be extremely cautious of overly restricting these agreements so as not to undermine the benefits of cloud computing. Only those measures absolutely necessary to achieve an appropriate security posture should be included.

By starting now, the DoD can begin to develop the appropriate balance of control versus provider flexibility to optimize the value proposition of public cloud computing. This will be a learning process for both the DoD and the public providers as they both strive to maintain the integrity of their offerings while providing the best possible services to their customers. Program managers, strategic leadership, lawyers, technical consultants, corporate entities and cloud brokerage agents must all come together to bridge the gap between the current state-of-practice and the art-of-the-possible. Eventually, what will likely be a difficult journey at first, will become the foundation of trust necessary for realizing the benefits of cloud computing.

7.5 Expand the Role of Organizational CIOs To Include More Mission Oriented Services

In both the DoD and in the federal government at large, the CIO is often perceived as being primarily responsible for administrative IT services for the organization. Mission applications are still largely the purview of PMs with cost, schedule and performance objectives specific to their individual programs. Unfortunately, this mindset has produced to the current state of vertically integrated, stove-piped DoD systems. The model works well from an accountability perspective, since there is a single PM who can be either acknowledge or admonished based on the success or failure of their charge. While it is true that the CIO has to be concerned with IT functions such as e-mail, data storage, and other administrative services, they should share an equal role in acquiring IT infrastructure related to mission applications so that the benefits of standardized IT can be realized across the entirety of the DoD.

PMs and CIOs should work toward developing a similar relationship to that found in the ASP/ISP model of the commercial sector (see Section 3.4). In this fashion, PMs could focus on delivering mission applications without also having to provide expertise associated with the lower-level IT functionality. CIOs could focus on providing the underlying support infrastructure as services to the PMs. This model would provide the CIO with the ability to homogenize the IT infrastructure of the organization, making it more secure, predictable and reliable. Standardizing the supporting IT infrastructure would also facilitate information sharing and interoperability since having a single point of contact for IT would be much more likely to produce commonality than with multiple PMs acquiring one-off solutions.

Initially, adopting this model might be difficult. PMs are still responsible for the end-to-end solutions and are unlikely to willingly yield control over a critical portion of their systems. Also, measures of performance are still mostly focused on individual programs and not the programs in their aggregate. Therefore, there are no real incentives for an individual PM to give up a portion of their own budget or performance specifications for the greater good of other programs. DoD senior leadership will have to strongly support efforts by their CIOs to standardize the IT enterprise so that individual PMs do not create unsustainable, tailored solutions that have only benefit their own programs and are likely to have long-term negative consequences.

This relationship model will require a great deal of trust, not only between PMs and CIOs, but also in the theory that a standardized approach to IT will yield greater good for the organization as a whole even if individual programs initially feel constrained. Standardized IT solutions generally yield greater agility, flexibility and ease of integration, but these qualities are not as easily measured as metrics like cost and performance. Senior leadership often cites, the iPhone, Facebook, or Google, as exemplars of the benefits of commercial IT. None of these examples would be possible without the Internet having a widely adopted set of standards which facilitate these modern capabilities. Similarly, senior leadership must strive to create IT enterprises where similar functionality can flourish within the government. Enhancing the role of the CIO in the acquisition of mission related programs will be a crucial step in achieving similar effects across DoD programs.

Once the IT enterprise within an organization is aligned under the management of a single responsible entity, then commonality can emerge and long-term focused planning will become more stable. Both of these characteristics are important from the perspective of integrating public cloud computing. First, the likelihood of crafting successful SLAs with public cloud providers would go up dramatically. The SLAs would be established by personnel dedicated to acquiring IT who have the appropriate focus and knowledge. Commonality within the SLAs would also help public providers realize greater efficiencies within their own enterprises and pass greater savings and performance back to the government customer. Second, standardization would enable better security planning and reuse of accredited systems. As with the FedRAMP program, a common IT enterprise would promote an “approve once, and use often” security posture improving the responsiveness of IT resources across all programs. Initial attempts at incorporating public cloud computing will have to overcome considerable technical, legal and policy obstacles. Once those obstacles are overcome the same difficulties should not be repeated as they would if individual programs attempted to accredit their own unique solutions. Frameworks for successful integration of public cloud computing should be established and reused by subsequent programs. Finally, commercial providers are much more likely to produce products and services that meet the needs of the government when stable IT strategies and roadmaps are documented and adhered to. Consolidating all organizational IT, to include mission applications, will enable commercial providers to invest their own research and development money appropriately and provide a more predictable stream of innovation back to the government.

7.6 Quantify the Value Proposition of Public Cloud Computing

Equation 2 presents a formula for managers to use when evaluating the total cost of ownership of moving toward a public cloud computing solution. The on-demand, scalable nature of public cloud computing enables this cost equation to center around actual usage instead of a more traditional cost estimation method based on procuring a data center sized to meet peak operating load.

The left-hand side represents the cost of a public cloud implementation. The first term is the $Usage$ (in units of time) multiplied by C_{Unit}, the unit cost of the service per unit of time. $C_{Transition}$ represents the transition costs, such as software recoding, of porting existing applications to the cloud. Next, $C_{Connectivity}$ represents the cost of the connectivity between the point of service consumption and the cloud platform. The equation accounts for the estimated reduction in labor and facilities costs that would have been spent to implement a traditional private solution by subtracting C_{Labor} and $C_{Facilities}$. Finally, C_{Risk} is a term for capturing the risk, in dollars, associated with a cloud based solution. The value of this term would flow from performing the process outlined by the governing documents described in Chapter IV. This term is analogous to management reserve that could be used to mitigate risks and deal with issue as they emerge. The right-hand side, $C_{Traditional}$, represents the estimated costs of a traditional acquisition as determined through normal cost estimation techniques. All other factors being equal, managers should choose the public cloud computing solution if the cost on the left-hand side is less than the right hand side.

$$Usage * C_{Unit} + C_{Transition} + C_{Connectivity} - C_{Labor} - C_{Facilities} + C_{Risk} < C_{Traditional} \quad (2)$$

7.7 Pilot Programs

Pilot programs will help develop a practical understanding of how DoD applications can best leverage public cloud computing. They are also essential for determining appropriate performance metrics. Cost is obviously a big factor, but benefits such as rapid elasticity, improved mission focus and lower barriers to entry are more difficult to quantify. Through practice, the DoD can develop more tangible metrics to quantify those kinds of benefits.

At Gartner's 22nd Annual Application Architecture, Development & Integration Summit in 2009, David Cearley and Gene Phifer presented a keynote speech outlining six candidate application types likely to provide early successes in the cloud:

- Prototyping/Proof of Concept
- Development/Test & Projects
- Web Application Serving
- SaaS, E-Mail, Collaboration
- Departmental & Workgroup Apps

- Simple Parallelized Workloads

The DoD should start by looking for these classes of IT programs within its enterprise. Then, it should sort out those that are categorized as MAC II or III that process publicly releasable information as defined by DoD Directive 8500.01E (See Section 4.3.1). This should produce a set of low-risk, high-reward pilot candidates. In addition, public IaaS offerings are likely to be the best starting point for cloud service adoption. IaaS provides the consumer with the most control over the environment and as such, the DoD has more opportunities to customize it to match unique security requirements and application designs.

One excellent example of a DoD pilot program related to cloud computing is the Defense Information Systems Agency's (DISA) Rapid Access Computing Environment (RACE). In an effort to help facilitate the adoption of cloud computing into the DoD information enterprise, DISA created RACE, which is essentially a private IaaS offering. Rather than acquiring their own independent IT resources, customers of RACE can be provisioned resources from DISA within 24 hours of their request. All the user needs is network connectivity to the self-service portal and a government credit card. The standard services available to RACE customers include a CPU, one GB of memory, 60 GB disk storage, and Internet Information Services (IIS) for Windows-based servers or LAMP (Linux, Apache, MySQL and PHP) for Red Hat servers. The server configuration can be customized as needed with up to four CPUs and eight GB of system memory. [67]

The program started as a platform for development and testing prior to fielding operational solutions and had success in hundreds of military applications ranging from satellite programs to convoy management. [7] Recently, RACE has moved beyond development and test and is now offering production environments where mission oriented services can be hosted to address warfighter's operational needs on both the Unclassified but Sensitive Internet Protocol (IP) Router Network (NIPRNet) and the Secret Internet Protocol Router Network (SIPRNet). RACE offers something called

"Path to Production" to facilitate meeting the regulatory and policy requirements outlined in Chapter IV. It provides a "Host-Tenant accreditation model, standardized system configurations, the Vulnerability Management System (VMS) and the Enterprise Mission Assurance Support Service (eMASS) to ensure compliance with the DoD Information Assurance Certification and Accreditation Process (DIACAP) and the DISA Security Technical Implementation Guides (STIGs)." [67]

Hopefully, by incorporating RACE into their IT solutions, DoD managers will obtain a better understanding of the power of cloud computing services and become more comfortable with the benefits and the issues surrounding this business model. Despite this necessary step in the right direction, RACE still has some weaknesses when compared to public cloud service providers. First, the services are provisioned on a monthly basis and the costs of those resources, while likely to be substantially lower than independently procured resources, is still almost an order of magnitude greater than commercial prices. For example, a 1 CPU/1 GB Memory configuration in RACE cost $467/month in 2011, which equates to roughly $0.65/hr. A comparable configuration from Amazon EC2 cost only $0.085/hr. Also, server demand that can be broken up by the hour is likely to achieve a higher utilization efficiency over a monthly allocation. The reaction time for RACE, in terms of on-demand provisioning is around 24 hours as compared to single digit minutes in the commercial sector.

So, where possible, DoD IT managers should still strive to put appropriate services into public clouds to benefit from the best that cloud computing technology has to offer. That being said, for those programs that require additional security, especially those that are classified, RACE is an excellent way to achieve a more agile and cost effective IT infrastructure.

7.8 Invest in Cloud-Based Research and Development

In conjunction with developing pilot projects to help program managers get a better feel for the technology, the DoD should invest in cloud-based research and development projects to help sort out the intricacies of both the technologies and the business model of cloud computing. Metrics development and security should be two key focus areas.

Cost and performance will be the most quantifiable metrics at the early stages of public cloud computing adoption. Of the two, cost will be the most easily accessible metric for determining the success of a particular cloud computing solution. It's relatively straightforward to measure and everyone understands it. Research projects would help validate value-based analysis models like the one for total cost of ownership presented in Equation 2. Once the terms of this model are properly understood and verified, then they can be put to more precise use. Many performance metrics will also be readily quantifiable, such as processing time or total storage capacity. These can also be included in overall assessment of the benefits of cloud computing solutions.

While these two metrics are more easily quantifiable they are by no means the only ways that success should be measured. Scalability, flexibility, improvement in mission focus, transition-to-operations time, availability, overall security posture, and technology insertion rates are some other factors that the research and development community could make strides in helping the acquisition community define appropriately. Some of these terms will necessarily have to be measured in their aggregate across multiple programs to adequately quantify their benefit. For example, scalability and agility may not be as important to all programs to the same degree, but portfolio managers and senior leadership should see overall gains to their enterprise as a result of many programs having access to such capabilities.

Three methods are offered for starting a cloud computing research and development portfolio to help achieve these objectives. First, the DoD has a number of laboratory data centers that offer computational resources to multiple organizations. Organizations come to these labs to take advantage of their large infrastructures without having to acquire their own. In this regard, these data centers offer a similar business model to an IaaS provider. It would be very beneficial to help capture the benefits of public cloud computing by porting a number of these programs to public cloud computing providers. Each program would be run on both the public cloud and the DoD data centers and then direct comparison of the relative pros and cons of each approach could be performed. Again, cost would dominate the initial focus, but this would also provide key insight into how the other metrics could be formalized more precisely.

Second, researchers could help find ways to bridge the gap between public cloud computing and a private solution that will likely be more palatable to DoD program managers in the short term. One way to do this would be to research ways to convert existing portions of the DoD laboratory data centers into model versions of public providers. An open source project, called *Eucalyptus* (`http://www.eucalyptus.com`), would be an excellent starting point for such a project. Eucalyptus is a software platform for implementing private cloud computing solutions on top of an existing data center. The software was designed to be compatible with the Amazon EC2 application programming interface. In this sense, employing software like Eucalyptus on existing data centers, would provide at least two benefits. First, application developers who are already familiar with Amazon's public cloud offering would be able to produce functional software quickly. Second, application frameworks could be developed quickly in Amazon's cloud without putting any sensitive information in its infrastructure. Then, when the application framework is complete, it could be ported back into the Eucalyptus-based private cloud where the sensitive data resides with a minimal amount of difficulty. Since Eucalyptus is open source, DoD researchers could experiment with the code base to meet DoD specific requirements. Even if programs choose to remain in private data centers, research like this would still provide benefits. The data centers themselves could work toward adopting similar streamlined, service-based approaches to public providers, which could ultimately speed up the request-to-provisioning time, automate the request process, and achieve better utilization of their own infrastructures.

Finally, many industry partnership opportunities are emerging that could help DoD researchers begin experimenting with cloud. Amazon recently announced its

“Free Usage Tier” (http://aws.amazon.com/free) aimed at helping people get started with public cloud computing. New customers to the service get a year’s worth of a Linux Micro Instance which includes 613 MB of memory, 15 GB of storage, 30 GB of data transfer as well as access to a number of benefits such as load balancing, messaging and database services. Google also created the “Google Exacycle for Visiting Faculty Grant Program” (http://research.google.com/university/exacycle_program.html), which allows researchers access to the Google’s computing infrastructure for developing high-performance, CPU-intensive application related to science and engineering.

Focused research would go a long way toward helping the DoD get a firm grasp of the benefits and challenges of public cloud computing. Research would enhance education and awareness, support the development of tailored security controls, and help foster innovation. These efforts would also help improve the trust relationship between the DoD and public providers that is so crucial to the adoption of public cloud computing.

7.9 Move Toward Distributed Software Development

Recent technology trends are pushing software development away from algorithms that operate in serial to those capable of distributing their workload across multiple processing units. In the early 2000’s CPU clock frequencies began to plateau as chip designers pushed the limits of power consumption and heat dissipation within individual processing cores. Rather than continue to increase the speed of a single core, designers made a fundamental architectural change toward multiple cores on a single chip. Figure 14 depicts this phenomenon. There has also been explosive growth in graphics processor units driven by demand from high-performance gaming. These processors can have hundreds of cores. From here, the next logical step is to expand beyond individual chips toward developing applications that run across multiple machines. Cloud computing is the represents the realization of this future.

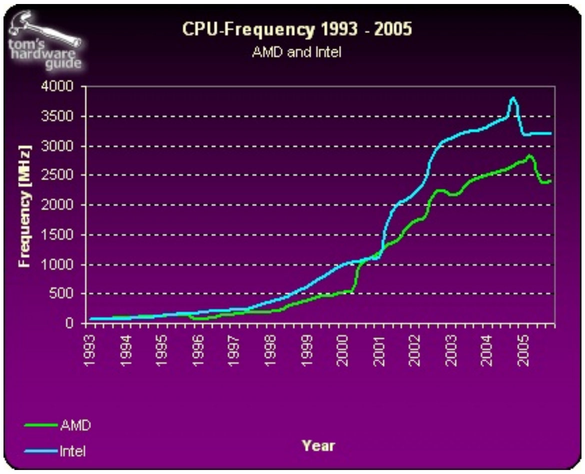

Figure 14: CPU Frequency by Year

The original trend afforded software developers "free" increases in performance. Software applications would automatically see an increase in performance as the underlying hardware got faster. This is no longer the case since the clock rates have essentially stopped growing. In a distributed architecture, managing challenging concepts like timing, synchronization, choreography, thread management and parallelism, is critical. Fortunately, software developers are increasingly expanding their expertise with techniques for taking advantage of distributed systems with multiple cores. They are demonstrating how complex, processing-intensive algorithms can be executed on commodity hardware by making readily available technologies perform functions that used to be reserved for specialized systems.

In general, public cloud computing is focused on providing commodity style processing resources. Providers offer these resources for extremely low cost and in potentially mass quantities. Therefore, developers who are adept at distributed programming can take advantage of public cloud to produce extraordinary performance at low cost. The DoD will need to support software development efforts that can take advantage of the power of distributed commodity hardware resources. To do otherwise would eliminate the scalability benefits that cloud offers and would keep the DoD hardware baseline primarily focused on costly, specialized systems.

7.10 Formally Establish Cloud Computing in the Acquisition Process

One of the key tenants of the Federal Cloud Computing Strategy is to harness the benefits of cloud computing by instituting a “Cloud First” policy. The policy is intended to “accelerate the pace at which the government will realize the value of cloud computing by requiring agencies to evaluate safe, secure cloud computing options before making any new investments.” [7] The document calls for all federal agencies to re-evaluate their technology sourcing strategy to include consideration for, and the application of, cloud computing solutions as a part of their budget processes. When opportunities are found, agencies are called to modify their portfolios to take full advantage of the benefits of cloud computing to maximize resource utilization, and improve flexibility and responsiveness, while minimizing cost.

In concert with the “Cloud First” policy, the Federal CIO offers the simple diagram shown in Figure 15 to help agencies plan for cloud migration. The chart has two axes: *Value* and *Readiness*. The *value* dimension measures the realized benefits from cloud computing. For the DoD, this relatively subjective measure would be a combination of the five benefits described in Chapter III. The *readiness* dimension measures how likely services are to be successfully transitioned to cloud computing in the short term. This measurement is a combination of program security posture, market characteristics, management willingness and program life-cycle stage. The need for pilot programs emphasized in Section 7.7 would be a part of the “First Movers” category.

Hopefully, the DoD will also recognize the benefits of leveraging public cloud computing as it implements these recommendations. In concert with the Federal CIO’s Cloud First policy, the DoD should formally establish cloud computing in its acquisition process. Program managers should be required to justify decisions to procure private IT solutions over those of external providers at each of the acquisition milestones described in the DoD 5000 series documents that govern acquisition. The DoD CIO should lead this effort. This step will help reinforce the need to consolidate

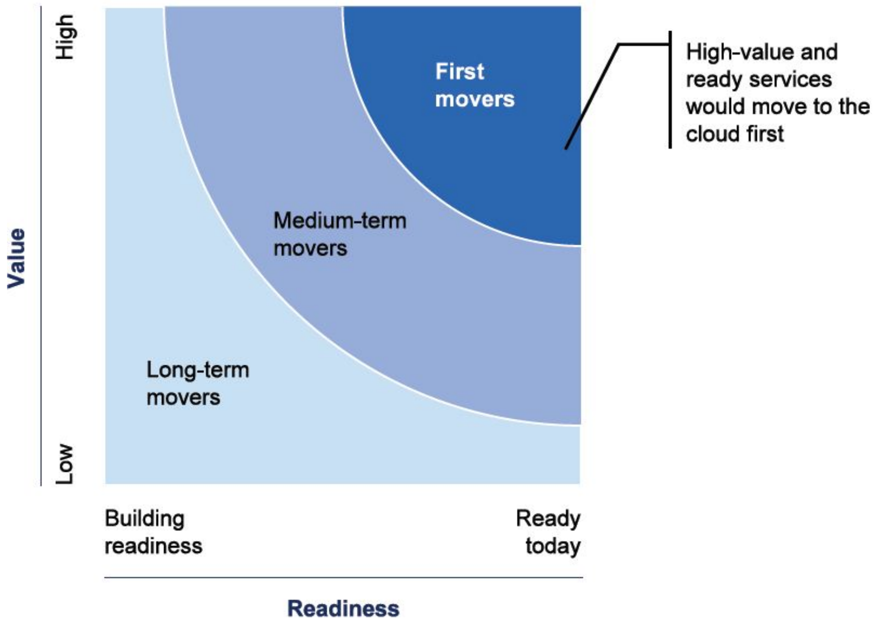

Figure 15: Selecting Services for Cloud Migration

the DoD IT infrastructures, encourage standardization across the enterprise, allow for more agile responses to changing warfighter needs, and promote faster acquisitions timelines.

7.11 Summary

By implementing the recommendations contained in this chapter, the DoD will be better postured to maintain a state-of-the-art information enterprise and continue to lead in the field of information technology. That enterprise will foster innovation, improve mission capability, and make the force more agile all within the constraints of increasingly tighter budgets. Implementing public cloud computing effectively within appropriate portions of the DoD architecture will require close coordination and partnerships with commercial companies, the Federal CIO, NIST and others. Ultimately, those partnerships will help develop the trust that is so critical to achieving success.

VIII. Conclusion

"Innovation is the ability to see change as an opportunity - not a threat."
- Anonymous

"All of us have a place in history. Mine is clouds." - Richard Brautigan

The time is right for DoD program mangers to seriously consider incorporating IT services from public cloud computing providers into their architectures. Growth in this sector is increasing rapidly as more and more companies, organizations, and even individuals look to cloud computing to provide low-cost, on-demand access to tremendous amounts of computational resources. As the number of users expands, increased revenue and competition are driving providers to continuously innovate and make their offerings more capable, secure, efficient, and reliable. There benefits and risks must be considered carefully, but future of this sector is bright.

This research has presented a number of topics concerning cloud computing. It began with an overview of what it is and how it has been formally been defined by thought leaders in this space. Emphasis was placed on NIST and the Federal CIO who have gone to great lengths to overcome the confusion that dominated only a few years ago and replace it with a solid framework for understanding cloud computing services. NIST's formal definitions have been very beneficial to both the government and commercial worlds alike and their baseline definition of five essential characteristics, four deployment models and three service models will continue to be refined as the capabilities grow and become more modular.

NIST's reference architecture for cloud computing also helps consumers, especially new ones, understand that cloud computing is not just a collection of technologies, it is equally a business model. IT managers must understand that there is no such thing as "Cloud-in-a-Box" that can be purchased from a vendor with magically beneficial results. While technologies like virtualization and high-bandwidth connectivity provide a technological foundation for cloud computing, those technologies must be employed properly to realize the benefits that cloud computing promises. One example is budgeting. Whereas traditional IT infrastructures would require substantial capital expenditures early in a program's life cycle followed by operations and maintenance phases. Public cloud computing eliminates the need for consumers to outlay considerable amounts of capital expenses to get a program going. Instead, the cost burden is spread over actual usage during the operations phase. Maintenance charges are also subsumed by the usage fee, since they become the responsibility of the cloud provider. Budget planners must adapt to meet this shift in resource phasing.

Another example, is the increased emphasis on standardization when cloud computing services are employed. When programs ask for modifications to provider's offerings to support some non-standard solution, they will erode the economies of scale that drive so many of cloud computing's benefits. Senior leaders and CIOs need to endorse those standards while recognizing that having a consistent IT framework will provide the most good for the most programs. Metcalfe's Law states that the value of a network is proportional to the square of the number of connected users of the system. What is implied, but not directly stated, is that the connected users must adhere to the common standards of the network. The exponential increase in value would be invalidated if each user decided to use their own standards when trying to communicate. Unfortunately, unique implementations in many DoD IT programs has lead to a high-degree of fragmentation in its networks. As a result the DoD is not achieving the n^2 value that its IT enterprise should otherwise make possible.

After defining the terminology and business relationship associated with cloud computing, the focus shifted to cover its key benefits. There are many potential benefits, but five were presented that are of particular interest to the DoD: *Continuous Refresh, Rapid Elasticity, Lower Cost, Improved Mission Focus* and *Lower Barriers to Entry.* Continuous refresh and lower barriers to entry will help the DoD maintain a posture of cutting-edge innovation. Rapid elasticity focuses on alleviating some of the constraints on early design requirements because the overall system is flexible and responsive to changing demand. Finally, lower cost and improved mission focus help either maintain the current functional baseline or even potentially improve performance even on a declining budget.

Of course, the benefits of cloud computing cannot be taken in isolation from the risks that it presents. While, the associated risks have many similarities to traditional IT deployments, the relative level of exposure in each risk area may be different. Federal law and agency doctrine mandate many aspects of how managers must mitigating the security risks associated with federal IT. This research covered the legal requirements that will impact public cloud computing adoption, primarily from the perspective of FISMA, which mandates that all federal IT systems be accountable in terms of their *confidentiality, integrity,* and *availability.* In response to direction originating in FISMA, NIST defined standards, guidelines and minimum security requirements for federal IT. The DoD also created a series of directives and instructions that provide for information assurance within DoD specific systems. At the lowest level within both the NIST and DoD documentation is a list of specific security controls that can be used by program managers to ensure confidentiality, integrity and availability within their respective systems. When implementing public cloud computing solutions within the DoD, these security controls can be tailored to overcome risks specific to cloud computing while still meeting regulatory and policy requirements.

Risk management was a key theme throughout this research. A sound risk management process emphasizes that risks be assessed according to a particular *root cause*, a *likelihood* of that root cause occurring, and finally the *consequence* that would result should that root cause occur. Too often, when faced with something new or something not well understood, people tend to focus on the consequence that a security related event would have, without an appropriate understanding of the likelihood or real root cause of that event. Since cloud computing is relatively new, it is still easier to focus on what can go wrong, when mangers should be striving to understand how to quantify the pros and cons in terms that can be analyzed systematically. Ultimately, through a sound risk management process covering public cloud computing, managers will be able to develop reliable metrics for evaluating whether the benefits exceed the risks. To help with this process, a chapter was devoted to covering some of the specific threats against public cloud computing to help managers understand real root causes. With this information, they should be able to make better assessments of the likelihood of these threats being realized in their solutions.

Up to that point, the focus was on how the DoD could implement public cloud computing and also realize its many benefits. Unfortunately, the low barriers to entry can not only empower the DoD, it can also be used by our adversaries against us. A series of topics were covered to highlight how state and non-state actors, organized criminals or simply rogue individuals now have access to massive amounts of computational resources at extraordinarily low cost. Access to computational resources at this scale was once reserved to a select few, which help limit the potential for their use in malicious ways. Now, extremely fast password cracking; large-scale, coordinated and distributed DoS attacks; and mass propagandizing are all possible by anyone with connectivity to the web. While realizing the benefits of public cloud computing in its own infrastructure, the DoD must also safeguard itself against a new wave of extremely powerful attacks that were not possible prior to widespread access to public cloud computing services.

Finally, a series of recommendations were presented to help foster the adoption of public cloud computing throughout the DoD IT enterprise. These recommendations cover a wide array of topics ranging from simply providing a well-defined education and awareness program to integrating an emphasis on public cloud computing directly into acquisition decision criteria. Some of the recommendations will be relatively straightforward to implement, especially in the research and development community. Others will require that organizational and cultural barriers be broken down so that public cloud computing can move beyond the labs and into mainstream programs.

Ultimately, trust will be the deciding factor in how successful public cloud computing will be within the DoD enterprise. Trust must exist between the DoD and the commercial providers; between program mangers and CIOs; and senior leadership and the technical, managerial and operational security controls implemented to protect systems that employ public cloud computing. These trust relationships are essential and must be actively developed by all parties to achieve real success. Once that trust is established, then these solutions can help realize the exponential value inherent in the DoD IT enterprise. Information technologies work best when they are constrained as little as possible. They exist to facilitate information sharing and when they become overly restricted, either out of fear or lack of trust, the very benefits that were originally expected never materialize. Therefore, the most important part of the trust relationship will be a faith in the emergent properties of modern IT; the ones that Metcalfe's Law predicts. Those properties are often stifled by hierarchical control, and instead thrive when enabled to flourish; especially at the edges where innovation happens. Public cloud computing is a key component of that innovation.

If you love your data, set it free!

Bibliography

1. N. Carr, *The big switch: Rewiring the world, from Edison to Google*, WW Norton & Company, 2008.
2. D. Alberts and R. Hayes, "Power to the Edge: Command... control... in the Iinformation Age," *US DOD Command and Control Research Center Publications* , 2003.
3. V. Kundra, "State of Public Sector Cloud Computing," May 2010. urlhttp://www.cio.gov/documents/StateOfCloudComputingReport-FINALv3_508.pdf.
4. C. Waxer, "Is the cloud reliable enough for your business?," 2009. http://www.networkworld.com/news/2009/060109-is-the-cloud-reliable-enough.html.
5. L. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," *ACM SIGCOMM Computer Communication Review* **39**(1), pp. 50–55, 2008.
6. P. Mell and T. Grance, "The NIST Definition of Cloud Computing. Version 15," *National Institute of Standards and Technology* , 2009.
7. V. Kundra, "Federal Cloud Computing Strategy," February 2011. http://www.cio.gov/documents/Federal-Cloud-Computing-Strategy.pdf.
8. D. Linthicum, *Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide*, Addison-Wesley Professional, 2009.
9. B. Golden and C. Scheffy, "Virtualization for Dummies-Sun and AMD Special Edition," 2008.
10. NIST, "NIST Cloud Computing Reference Architecture, Version 1," March 2011. http://collaborate.nist.gov/twiki-cloud-computing/pub/CloudComputing/ReferenceArchitectureTaxonomy/NIST_CC_Reference_Architecture_v1_March_30_2011.pdf.
11. D. Wyld, "Moving to the cloud: An introduction to cloud computing in government," *IBM Center for the Business of Government E-Government Series* , 2009.
12. A. Michael, F. Armando, G. Rean, D. Anthony, K. Randy, K. Andy, L. Gunho, P. David, R. Ariel, S. Ion, *et al.*, "Above the clouds: A berkeley view of cloud computing," *EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2009-28* , 2009.
13. P. Mell and T. Grance, "Effectively and securely using the cloud computing paradigm," *Natl Inst. of Standards and Technology* , 2009.

14. V. Kundra, “Memorandum for Chief Information Officers: Update on the Federal Data Center Consolidation Initiative,” October 2010. http://www.cio.gov/Documents/Update-Federal-Data-Center-Consolidation-Initiative.pdf.

15. L. Kaufman, “Data security in the world of cloud computing,” *Security & Privacy, IEEE* **7**(4), pp. 61–64, 2009.

16. NIST, “DRAFT Guidelines on Security and Privacy in Public Cloud Computing,” January 2011. http://csrc.nist.gov/publications/nistpubs/800-53-Rev2/sp800-53-rev2-final.pdf.

17. J. Bezos, “Jeff Bezos talks about Animoto at startup school 2008,” 2008. http://www.youtube.com/watch?v=uIc-VB-ke9o.

18. D. Gottfrid, “Self-Service, Prorated Supercomputing Fun!,” 2007. http://open.blogs.nytimes.com/2007/11/01/self-service-prorated-super-computing-fun/.

19. S. P. Release, “SOASTA Leverages the Cloud to Test 1,000,000 Users on MySpace,” November 2009. http://www.soasta.com/info-center/press-releases/soasta-leverages-the-cloud-to-test-1000000-users-on-myspace/.

20. International Data Corporation, “New IDC IT Cloud Services Survey: Top Benefits and Challenges,” 2009. http://blogs.idc.com/ie/?p=730.

21. B. Golden, “The case against cloud computing,” 2009. http://www.cio.com/article/478419/The_Case_Against_Cloud_Computing%_Part_Two.

22. J. Heiser and M. Nicolett, “Assessing the security risks of cloud computing,” *Gartner Report* , 2008.

23. K. Popovic and Z. Hocenski, “Cloud computing security issues and challenges,” in *MIPRO, 2010 Proceedings of the 33rd International Convention*, pp. 344–349, IEEE, 2010.

24. Department of Defense, “Risk Management Guide for DoD Acquisition, 6th Edition,” tech. rep., DEFENSE SYSTEMS MANAGEMENT COLL FORT BELVOIR VA, August 2006.

25. NIST, “FIPS PUB 200, Minimum Security Requirements for Federal Information and Information Systems,” 2006. http://csrc.nist.gov/publications/fips/fips200/FIPS-200-final-march.pdf.

26. NIST, “FIPS PUB 199, Standards for Security Categorization of Federal Information and Information Systems,” 2004. http://csrc.nist.gov/publications/fips/fips199/FIPS-PUB-199-final.pdf.

27. NIST, “Special Publication 800-53, Information Security,” 2007. http://csrc.nist.gov/publications/nistpubs/800-53-Rev2/sp800-53-rev2-final.pdf.

28. CIO Council, "Proposed Security Assessment & Authorization for U.S. Government Cloud Computing, Draft version 0.96," November 2010. https://info.apps.gov/sites/default/files/Proposed-Security-Assessment-and-Authorization-for-Cloud-Computing.pdf.

29. Department of Defense, "Department of defense information enterprise strategic plan 2010-2012," May 2010. http://cio-nii.defense.gov/docs/DodIESP-r16.pdf.

30. Department of Defense, "DoD Directive 8500.01E: Information Assurance," April 2007.

31. Department of Defense, "DoD Directive 8500.2: Information Assurance Implememtation," February 2003.

32. J. Bezos, "AWS Receives ISO 27001 Certification," November 2010. http://aws.typepad.com/aws/2010/11/aws-receives-iso-27001-certification.html.

33. Google, "Secure applications to meet the needs of government," 2011. www.google.com/apps/intl/en/government/trust.html.

34. A. W. Services, "Amazon EC2 Dedicated Instances," 2011. https://aws.amazon.com/dedicated-instances/.

35. U. S. C. I. O. Council, "Federal Risk and Authorization Management Program (FedRAMP)." http://www.cio.gov/pages.cfm/page/Federal-Risk-and-Authorization-Management-Program-FedRAMP.

36. International Data Corporation, "Through 2014 Public IT Cloud Services Will Grow at More Than Five Times the Rate of Traditional IT Products, New IDC Research Finds." http://www.idc.com/about/viewpressrelease.jsp?containerId=prUS22393210.

37. A. W. Services, "Amazon Web Services: Overview of Security Processes," 2010. http://media.amazonwebservices.com/pdf/AWS_Security_Whitepaper.pdf.

38. Rackspace, "Rackspace Security: Triple-strength Security Backed by Fanatical Support," 2009. http://broadcast.rackspace.com/downloads/pdfs/RackspaceSecurityApproach.pdf.

39. M. G. F. Services, "Securing Microsoft's Cloud Infrastructure," 2009. http://www.globalfoundationservices.com/security/documents/SecuringtheMSCloudMay09.pdf.

40. J. Liu, M. George, K. Vikram, X. Qi, L. Waye, and A. Myers, "Fabric: A platform for secure distributed computation and storage," in *Proceedings of the ACM SIGOPS 22nd symposium on Operating systems principles*, pp. 321–334, ACM, 2009.

41. T. Garfinkel, B. Pfaff, J. Chow, M. Rosenblum, and D. Boneh, "Terra: A virtual machine-based platform for trusted computing," *ACM SIGOPS Operating Systems Review* **37**(5), pp. 193–206, 2003.

42. N. Santos, K. Gummadi, and R. Rodrigues, "Towards trusted cloud computing," in *Proceedings of the 2009 conference on Hot topics in cloud computing*, USENIX Association, 2009.

43. M. Christodorescu, R. Sailer, D. Schales, D. Sgandurra, and D. Zamboni, "Cloud security is not (just) virtualization security: a short paper," in *Proceedings of the 2009 ACM workshop on Cloud computing security*, pp. 97–102, ACM, 2009.

44. S. Paquette, P. Jaeger, and S. Wilson, "Identifying the security risks associated with governmental use of cloud computing," *Government Information Quarterly* , p. 250, 2010.

45. P. Cox, "Iaas threats in the cloud," 2010. `http://www.systemexperts.com/assets/pdf/SystemExperts-IaaSThreatsInTheCloudPt2.pdf`.

46. J. Rutkowska, "Red pill... or how to detect vmm using (almost) one cpu instruction," 2004. `http://invisiblethings.org/papers/redpill.htm`.

47. J. Rutkowska, "Subverting vista kernel for fun and profit," 2006. `http://www.blackhat.com/presentations/bh-usa-06/BH-US-06-Rutkowska.pdf`.

48. T. Ristenpart, E. Tromer, H. Shacham, and S. Savage, "Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds," in *Proceedings of the 16th ACM conference on Computer and communications security*, pp. 199–212, ACM, 2009.

49. M. Slaviero, "Blackhat presentation demo vids: Amazon," August 2009.

50. Amazon Web Services, "Amazon ec2 pricing," 2009. `http://aws.amazon.com/ec2/pricing/`.

51. Microsoft, "Microsoft Confirms Data Recovery for Sidekick Users," October 2009. `http://www.microsoft.com/presspass/press/2009/oct09/10-15sidekick.mspx`.

52. DEMO, "Streamload/MediaMax/TheLinkup death spiral dogs Nirvanix," July 2008. `http://www.demo.com/community/?q=node/160512`.

53. USAF, "Air Force Doctrine Document 3-12, Cyberspace Operations," July 2010.

54. D. Bryan and M. Anderson, "Cloud Computing, a Weapon of Mass Destruction?," 2010. `http://defcon.org/html/links/dc-archives/dc-18-archive.html#Bryan1`.

55. P. P. Release, "AccessData Selects Parabon's Frontier Grid Platform to Power Extreme-Scale Decryption across the Enterprise," February 2009. `http://www.parabon.com/news-events/accessdata.html`.

56. WPACracker.com, “Questions About WPA Cracker,” 2011. http://www.wpacracker.com/faq.html.

57. T. Kleinjung, K. Aoki, J. Franke, A. Lenstra, E. Thomé, J. Bos, P. Gaudry, A. Kruppa, P. Montgomery, D. Osvik, *et al.*, “Factorization of a 768-bit RSA modulus,” *Advances in Cryptology–CRYPTO 2010* , pp. 333–350, 2010.

58. T. Roth, “Cracking Passwords In The Cloud: Amazon’s New EC2 GPU Instances,” 2010. http://stacksmashing.net/2010/11/15/cracking-in-the-cloud-amazons-new-ec2-gpu-instances/.

59. T. Roth, “Upcoming Black Hat Talk,” 2011. http://stacksmashing.net/2011/01/12/upcoming-black-hat-talk/.

60. S. Coll and S. Glaser, “Terrorists Turn to the Web as a Base of Operations,” August 2005. http://www.washingtonpost.com/wp-dyn/content/article/2005/08/05/AR2005080501138.html.

61. T. Harding, “Terrorists ’use google maps to hit uk troops’,” April 2007. http://www.telegraph.co.uk/news/worldnews/1539401/Terrorists-use-Google-maps-to-hit-UK-troops.html.

62. R. Blakely, “Google earth accused of aiding terrorists,” December 2008. http://technology.timesonline.co.uk/tol/news/tech_and_web/the_web/article5311241.ece.

63. V. O’Connell, A. Efrati, and E. Perez, “’Jihad Jane’ Had Troubled Past,” *The Wall Street Journal* , March 2010. http://online.wsj.com/article/SB10001424052748704655004575114071680745664.html.

64. C. Johnson, “’Jihad Jane’ Pleads Guilty In Terrorism Plot,” *National Public Radio* , February 2011. http://www.npr.org/2011/02/01/133404223/jihad-jane-pleads-guilty-in-terrorism-plot.

65. D. Gross, “WikiLeaks cut off from Amazon servers,” December 2010. http://edition.cnn.com/2010/US/12/01/wikileaks.amazon/index.html?eref=edition.

66. D. Ullman, “OO-OO-OO! the sound of a broken OODA loop,” *CrossTalk–J. Def. Software Eng* , pp. 22–25, 2007.

67. Defense Information Systems Agency, “Rapid Access Compute Environment,” 2011. http://www.disa.mil/computing/cloud/race.html.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704–0188*

1. REPORT DATE *(DD–MM–YYYY)* 16–06–2011	**2. REPORT TYPE** Graduate Research Project	**3. DATES COVERED** *(From — To)* 18 June 2010 — 16 June 2011

4. TITLE AND SUBTITLE Taking the High Ground: A Case For Department Of Defense Application Of Public Cloud Computing	**5a. CONTRACT NUMBER** N/A **5b. GRANT NUMBER** N/A **5c. PROGRAM ELEMENT NUMBER** N/A
6. AUTHOR(S) Kris Barcomb, Maj, USAF	**5d. PROJECT NUMBER** N/A **5e. TASK NUMBER** N/A **5f. WORK UNIT NUMBER** N/A
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology Graduate School of Engineering and Management 2950 Hobson Way WPAFB OH 45433-7765	**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT/ICW/ENG/11-01
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Not Disclosed	**10. SPONSOR/MONITOR'S ACRONYM(S)** Not Disclosed **11. SPONSOR/MONITOR'S REPORT NUMBER(S)** N/A

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approval for public release; distribution unlimited. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Cloud computing offers tremendous opportunities for private industry, governments, and even individuals to access massive amounts of computational resources on-demand at very low cost. Recent advancements in bandwidth availability, virtualization technologies, distributed programming paradigms, security services and general public awareness have contributed to this new business model for employing information technology (IT) resources. IT managers face tough decisions as they attempt to balance the pros and cons of integrating commercial cloud computing into their existing IT architectures. On one hand, cloud computing provides on-demand scalability, reduces capital and operational expenses, decreases barriers to entry, and enables organizations to refocus on core competencies rather than on IT expertise. In spite of the benefits, security concerns are still the dominant barriers to cloud service adoption. This research explores public cloud computing services from a Department of Defense (DoD) perspective. The objectives are to improve the general understanding of cloud computing; describe its potential benefits to the DoD; examine public cloud computing adoption from a risk management perspective; present threats specific to public cloud computing; and provide a set of recommendations to help foster public cloud computing adoption within the DoD. In addition to advocating for incorporating public cloud computing into the DoD enterprise, this research also presents how it could be used by our adversaries to launch sophisticated attacks.

15. SUBJECT TERMS
Cloud Computing, Virtualization, Infrastructure as a Service, Information Technology, Data Centers

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr. Robert F Mills
U	U	U	UU	119	**19b. TELEPHONE NUMBER** *(include area code)* (937) 255–3636, ext 4527 robert.mills@afit.edu