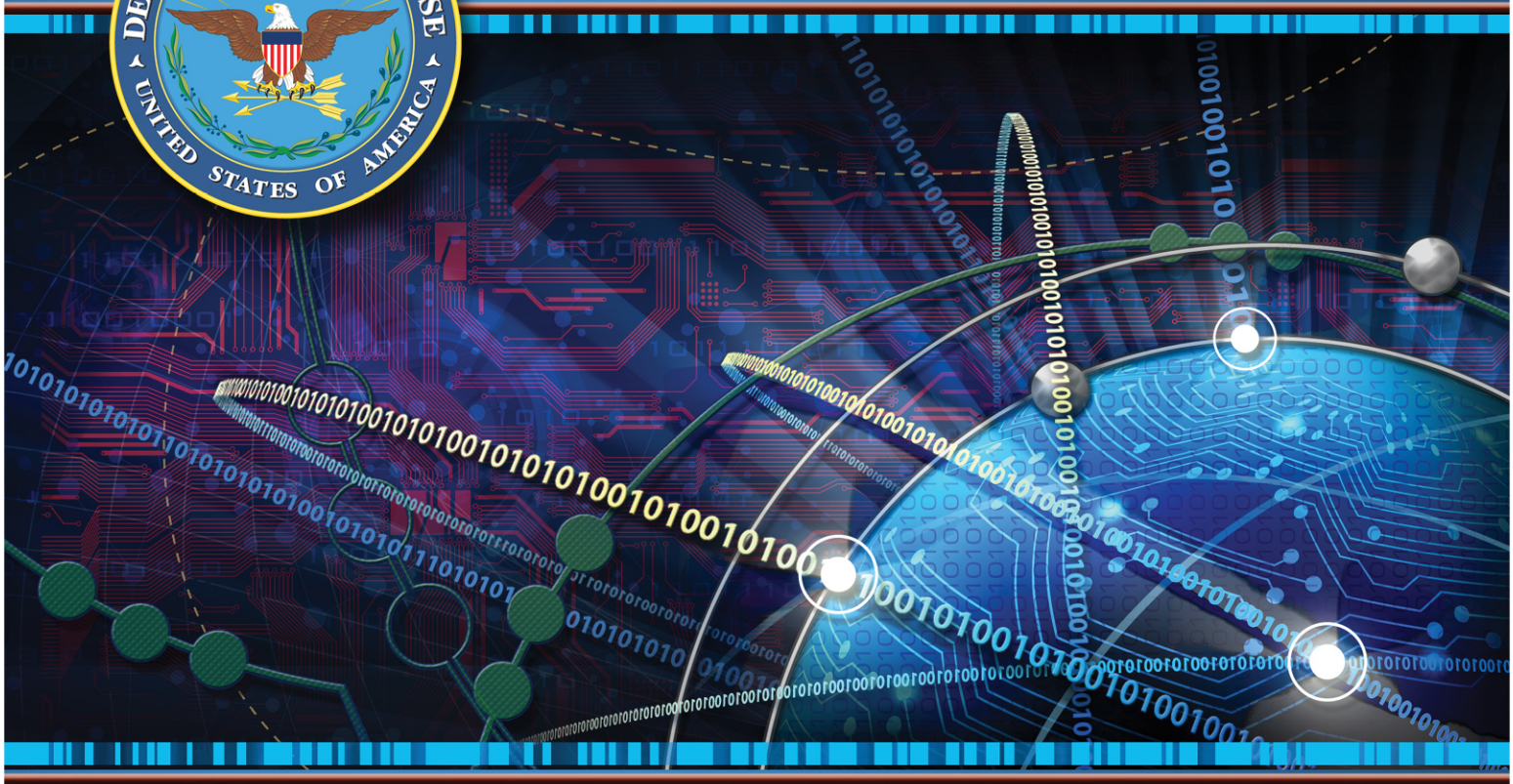
DEPARTMENT OF DEFENSE

DEFENSE SCIENCE BOARD

TASK FORCE REPORT:

Cyber Security and Reliability in a Digital Cloud



OFFICE OF THE UNDER SECRETARY OF DEFENSE FOR ACQUISITION, TECHNOLOGY, AND LOGISTICS
WASHINGTON, D.C. 20301-3140

JANUARY 2013

REPORT OF THE DEFENSE SCIENCE BOARD

TASK FORCE ON

Cyber Security and Reliability in a Digital Cloud

JANUARY 2013

Office of the Under Secretary of Defense
for Acquisition, Technology, and Logistics
Washington, D.C. 20301-3140

This report is a product of the Defense Science Board (DSB).

The DSB is a Federal Advisory Committee established to provide independent advice to the Secretary of Defense. Statements, opinions, conclusions, and recommendations in this report do not necessarily represent the official position of the Department of Defense (DoD). The Defense Science Board Task Force on Cyber Security and Reliability in a Digital Cloud completed its information-gathering in March 2012. The report was cleared for open publication by the DoD Office of Security Review on January 16, 2013.

This report is unclassified and cleared for public release.

OFFICE OF THE SECRETARY OF DEFENSE
3140 DEFENSE PENTAGON
WASHINGTON, DC 20301–3140

DEFENSE SCIENCE
BOARD

November 27, 2012

MEMORANDUM FOR UNDER SECRETARY OF DEFENSE FOR ACQUISITION, TECHNOLOGY & LOGISTICS

SUBJECT: Final Report of the Defense Science Board (DSB) Task Force on Cyber Security and Reliability in a Digital Cloud

I am pleased to forward the final report of the DSB Task Force on Cyber Security and Reliability in a Digital Cloud. This study comprises one part of a DSB Cyber Initiative. A study on Resilient Military Systems is the other component of the initiative.

The Task Force assessed the implications of using cloud computing resources and services for Department of Defense (DoD) mission needs. The report offers important recommendations for the DoD focused on: identification and application of cloud computing resources to DoD mission areas; improving DoD's implementation of cloud computing; enhancing cloud resiliency in degraded operations; and finally, areas requiring further research and development. Particular emphasis is given to improving cloud computing resilience for deployed forces.

I fully endorse all of the Task Force's recommendations contained in this report, and urge their careful consideration and soonest adoption.

Paul B. Kaminski

Dr. Paul Kaminski
Chairman

OFFICE OF THE SECRETARY OF DEFENSE
3140 DEFENSE PENTAGON
WASHINGTON, DC 20301–3140

DEFENSE SCIENCE BOARD

November 27, 2012

MEMORANDUM FOR UNDER SECRETARY OF DEFENSE FOR ACQUISITION, TECHNOLOGY, AND LOGISTICS

Subject: Report of the Defense Science Board Task Force on Cyber Security and Reliability in a Digital Cloud

The final report of the Defense Science Board Task Force on Cyber Security and Reliability in a Digital Cloud is attached. The Task Force conducted an independent assessment of the suitability of cloud computing architectures for DoD applications. Key factors in the assessment included DoD mission enhancements, cyber security benefits and risks, and potential cost savings associated with cloud computing.

The Task Force also investigated the benefits and risks of cloud computing for the needs of deployed forces. Several enhancements in cloud computing architectures and training and operational exercising are recommended to improve the access to important data and computing resources under degraded operational conditions.

The Task Force recommends that for sensitive, classified, or time-critical applications, the DoD should pursue private cloud computing to enhance mission capabilities, provided that strong security measures are in place. This report recommends several improvements in cloud computing implementations to strengthen cyber security and reliability.

Dr. Eric D. Evans
Co-Chairman

Dr. Robert L. Grossman
Co-Chairman

Table of Contents

Executive Summary

Cloud computing is viewed by many as the next major step in the evolution of computing infrastructure. Very large commercial cloud computing data centers have emerged around the world with petaflops of processing capacity, hundreds of petabytes of data storage, and wideband network access. Services, including electronic mail, data storage, database management, application hosting, very large dataset processing, and high performance computing, are globally available today from many cloud computing data centers. Cloud computing advocates promise on-demand delivery of these massive, warehouse-scale computing resources simply and easily through a network browser.

Much of the technology and computer architecture that enable modern cloud computing has roots in the mainframe, client-server, and early internet computing eras. What has emerged in recent years, however, differs from all of these in many attributes. Cloud computing data centers have different capabilities, risks, and security concerns than conventional networks, as well as different cost and efficiency models.

These differences are substantial, and have resulted in a wide variety of realistic and unrealistic claims for cloud computing, as well as a good deal of hype and confusion. With the proper implementation and operations, cloud computing data centers have demonstrated as good or better cyber security, capabilities, and cost than is currently available in Department of Defense (DoD) data centers. These improvements, however, are by no means guaranteed for every case and very much depend on the specific details of the implementation and operations.

Cloud computing offers the DoD new, agile computational capabilities to support increasingly multifaceted missions. Some DoD missions likely to benefit from cloud computing services will involve varying or unpredictable computing requirements, or the integration of many, high-capacity data feeds from sensor networks and other sources. Other missions may include the analysis of very large data sets or those that require the ability to move computational resources. An additional benefit is the productivity gained from a ubiquitous connection to common cloud-based services, such as email, shared calendars, unclassified training, or document preparation.

This study investigates the suitability of the cloud computing approach for addressing the DoD enterprise and operational computing needs. Over the past few years, DoD has transitioned some of its computing needs to cloud computing data centers. The main factors driving this transition include enhanced mission capabilities, potential reduction in data center costs, and potential improvement in cyber security. This study has investigated these factors in detail and has analyzed the characteristics that should be considered when DoD contemplates moving applications onto cloud computing data centers. The study also investigated ways for the DoD to manage the cyber security risks and benefits associated with cloud computing.

Important Cloud Computing Issues for the Defense Use

Types of cloud computing service configurations

An important issue is selecting an appropriate configuration of cloud services for DOD cloud computing applications:

- Cloud computing services may be provided by a company that provides similar services to the public, a defense-only contractor, or the DoD itself.
- Cloud computing resources may be shared among a number of customers, or only a single organization.
- The staff that manages the hardware, software, and services may be uncleared employees of a public company, cleared DoD contractors, or DoD employees.
- The cloud computing hardware resources may be located in shared space with other customers, in dedicated space in a building with other customers, at a dedicated facility, or on a military base.
- Cloud computing software resources may be based on a standard or modified software stack used by a public cloud computing services provider, standard or modified open source software stack, proprietary software stack, custom software stack, or some combination of these.

As is clear from this list, multiple dimensions distinguish how cloud computing services may be provisioned. Simply distinguishing between "public clouds"—commercial public companies operating their own data centers that are shared among many external customers using their own custom software and their own staff—and non-public or private clouds can cause confusion. In this report, the task force describes the specific aspects of the cloud computing configuration that are relevant to avoid the simple choice of public or private clouds.

National security concerns clearly preclude putting the computing resources of some sensitive DoD missions and capabilities in public shared clouds operated by non-cleared personnel. In general, however, the decision whether to host a particular application in a particular cloud computing data center depends upon the specific details of the application and the data center.

Detailed mandates for enhanced cyber security

An issue of importance to DoD is the development of a detailed approach for enhanced cyber security across both its conventional and cloud computing enterprise.

The hardware and software used in cloud computing, like all hardware and software, may have vulnerabilities that can be exploited by adversaries. Cloud computing processes, fortunately, offer the potential for improved cyber security through a number of attributes, primarily better traffic filtering and malware scanning, monitoring of usage

patterns and end-device configurations, varying provisioning of data resources, and improved management of systems operations. Whether allocating an existing application to a cloud computing data center increases or decreases cyber security depends upon the specific application, the specific characteristics of the configuration, and the specific implementation.

The cyber security of cloud computing needs additional attention when it is used to support mission-critical DoD applications. The task force found that, in many cases, deploying applications to cloud computing data centers increased cyber security, especially against less sophisticated threats. The task force also found that many risks can be managed with available hardware and software measures, but the DoD needs to carefully implement these measures before transitioning existing applications to cloud computing systems.

Research and development work within the Military Services, the Defense Advanced Research Projects Agency (DARPA), and the intelligence community offers technology that promises significant improvements for cloud computing cyber security in the long term, and this work should be better integrated with acquisition planning for DoD cloud computing data centers. In some DoD cloud computing implementations currently underway, a larger emphasis on cyber security measures is needed.

Control of cloud computing transition and sustainment costs

Realizing the potential cost savings associated with cloud computing is important to DoD. The transition of Federal government applications to cloud computing data centers have, in some cases, resulted in cost savings. The task force found the actual cost benefits to be highly case-dependent.

This cost savings for the transition from conventional enterprise computing to cloud computing has been achieved in a number of ways: through staffing, electric power usage, and computing efficiency. Conventional systems typically require one professional staff per tens to hundreds of servers, whereas most cloud computing data centers only require one professional staff for thousands of servers. Electric power is a large component of data center costs, and cloud computing data centers can be located where power is relatively less expensive. Finally, through virtualization and improved processing management, servers in cloud computing data centers can be more efficiently used, often achieving greater than five times the server efficiency as compared with conventional computing.

The required cost to enhance cyber security for any cloud computing implementation will need additional investigation. Some additional hardware and software will be required, and the cost for these components will need to be incorporated into the transition and sustainment costs when contemplating transition to a cloud computing data center.

DoD cloud computing data centers

Of particular importance to DoD will be finding ways to mitigate risk while achieving the capability benefits and potential cost reductions that cloud computing promises. An important aspect of cloud computing is the ability to operate infrastructure at a warehouse-scale data center and, thus, to provide new capabilities and enable cost savings. But warehouses are, by their nature, highly visible; having only a few, very large DoD data centers may create attractive targets for an adversary to attack. Further, the centralization implied by a "Fort Knox" approach—with a single, very large data center—cannot provide DoD with resilience or low-data transfer latencies required for global operations.

The task force therefore recommends that DoD design, implement, and deploy a set of geographically distributed data centers that could be could be operated as a single system. A few tens of such consolidated cloud computing data centers, established across the United States and around the world, seems like a good start at creating a sensible cloud capability for DoD. If appropriately designed, a collection of modular data centers would provide DoD with robust and elastic computing capacity.

Commercially available data centers, with servers embedded in modular units, offer DoD a relatively low cost and rapid way to develop a defense cloud computing infrastructure. The DoD could situate clusters of these modular data centers in physically secure areas. These may include military bases that have access to low cost and reliable power and wideband networks.

These modular data centers could be designed as a unit and purchased over time. In this way, standard best practices could be applied, such that one-third of the decentralized data center could be refreshed each year to ensure ongoing modernization. Such a design can also provide agility because computing infrastructure could be moved between geographic locations when needed.

Resilient cloud computing resources for deployed forces

A final issue of importance for the DoD is to provide resilient cloud computing resources at the warfighter "edge"—locations and times with scarce bandwidth. Deployed forces often execute their missions under degraded conditions and disadvantaged data links, and this limits a warfighter's access to the most current data. In these cases, thick clients—with enhanced data storage and redundant data links—could ensure limited access to data. When low-latency processing is needed, cloud computing data resources could be deployed in close proximity to the data streams.

The availability of secure, modular cloud computing resources could provide DoD with the capability to forward-deploy data and computing resources to meet warfighter needs.

Summary of Key Findings and Recommendations

The Significance and Impact of Cloud Computing

Finding 1: Although cloud computing is an overloaded term, cloud computing providers are offering services that are fundamentally new and useful, typically delivering the:

- ability for massive scale-up of storage and computing
- rapid, agile, elasticity with the ability to increase and decrease storage and computing capacity on-demand, when the community of tenants don't all require that capacity at the same time
- metered services where the user pays only for what is used
- self-service start-up and control

Finding 2: Modular data centers offer an approach to quickly set up cloud computing capacity, to add additional capability to existing cloud computing data centers, and to easily refresh or update existing capability. This concept is illustrated in Figure F-1.

Finding 3: Cloud computing services can scale to data centers or "warehouse-scale" computing. Elastic, warehouse-scale cloud computing is fundamentally new and can provide DoD with important new capabilities.

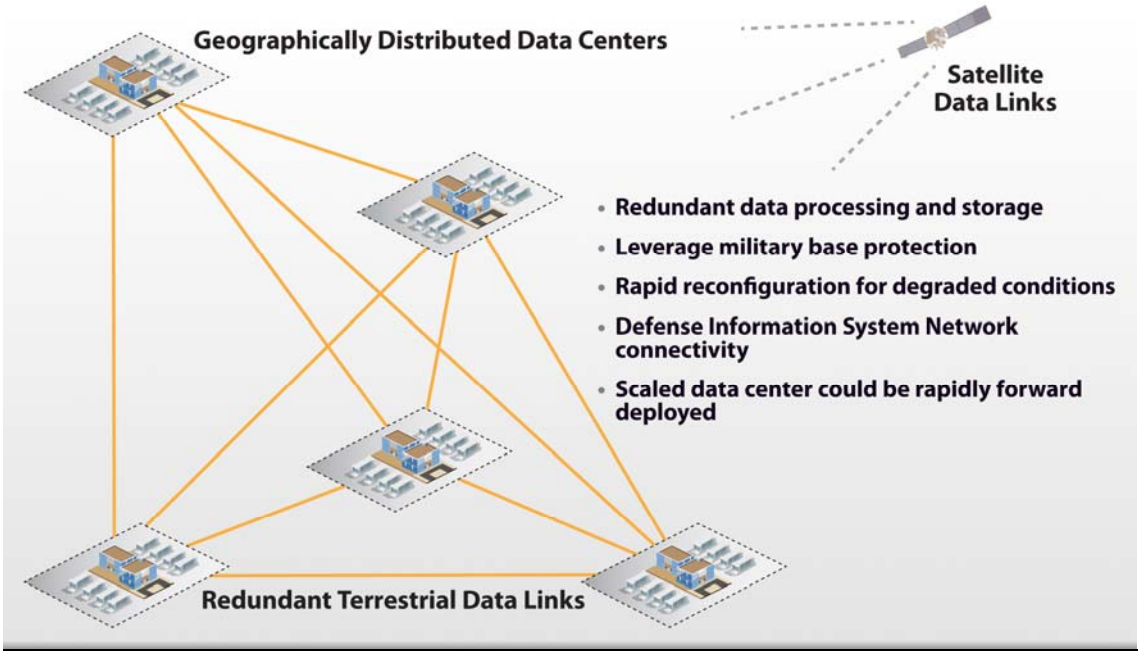


Figure F-1. Concept for a geographic distribution of DoD data centers

The Security of Cloud Computing

Finding 4: Cloud computing is not intrinsically more secure than other distributed computing approaches, but its scale and uniformity facilitate and enable the wholesale and consistent application of security practices. Secure aspects include large scale monitoring and analysis of data to detect attacks, and automated and persistent provisioning and re-provisioning to foil intrusions. For these reasons, well-operated cloud computing facilities can exhibit better security hygiene than conventional data centers. However, the centralization of resources in a huge data center also encourages more determined attacks, especially on critical components broadly affecting security. This is similar to conventional systems where attacks are observed to focus on central directories.

Finding 5: The scale of cloud computing enables the analysis of packet and log data that provides new capabilities for event forensics and real-time detection of malicious behavior. The ability to manage very large, diverse datasets facilitates a data-centric security model in which users are authorized to work with data based upon their security credentials and the security markings on the data rather than the conventional enclave-centric security model in which users are provided access to an enclave and can access all the data in the enclave.

Finding 6: No cloud computing deployment model is uniformly suitable for hosting all DoD applications. In general, sensitive, classified, and time-critical DoD applications should be deployed only in private clouds or conventional non-cloud approaches.

Finding 7: The case for transitioning a DoD application to a cloud computing data center must include a security assessment detailing the impact of the transition. Whether security will be improved by transitioning an application to a cloud computing data center will depend on factors specific to the application, to the cloud computing data center, and to the transition process.

Finding 8: The DoD has not established effective plans for cloud computing facility backup or for dealing with any anticipated degradation of communications between the cloud computing facilities and the end user.

The Costs Associated with Cloud Computing

Finding 9: Potential cost reductions or increases incurred during the transition to and sustainment of cloud computing infrastructure depend on the specifics of the implementation. Potential cost-reduction factors include a higher utilization of servers, lower professional support staff needs, economies of scale for the physical facility, and the flexibility to locate data centers in areas with lower-cost power.

Research and Development for Cloud Computing Technologies

Finding 10: The DoD has active research and development efforts in technology areas applicable to cloud computing performance and security. Sustained DoD investment in cloud computing security technology is critically important to allow DoD data centers to continue improving their defenses against evolving threats. Research and development in software stack protection, monitoring, and forensics of very large datasets, secure hypervisors, and advanced encryption offer significant possible security benefits.

Overarching Recommendations

Recommendation 1: For some sensitive, classified, and time-critical applications, the DoD should pursue private cloud computing, provided that strong security measures are in place.

In particular, cloud computing-based solutions should be considered for applications that require the agility, scale-out, and ability to integrate and analyze massive data that cloud computing can provide. Examples of such applications include: big data analysis and all-source intelligence integration; processing, exploitation, and dissemination of data gathered through intelligence, surveillance, and reconnaissance (ISR); large-scale modeling and simulation; open source data collection, storage, and assessment; and advanced decision support systems.

Recommendation 2: The DoD CIO in partnership with the military Services should deploy interconnected, modular cloud computing data centers located at secure locations, such as military bases.

The development of large, private community clouds in DoD will enable greater computing and storage elasticity and the improved ability to operate under degraded conditions. The DoD CIO should guide this development with an eye on both current and future DoD computing needs.

A DoD private community cloud may include in-house, in-sourced, or out-sourced private clouds. Implemented through interconnected, modular cloud computer data centers, this can be operated as an integrated unit to improve the potential reducing costs.

Because large data centers can also be attractive targets, geographically distributed modular data centers are recommended that are operated as a single, large-scale, distributed cloud. The design should include a distributed data center architecture that allows access by multiple Services and Agencies. Cost savings would be achieved through shared development, operations, and maintenance support.

These modular data centers could be located on military bases in order to provide good physical security. The location should also be influenced by the cost and availability of reliable electric power. It is anticipated this will be similar to the National Security Agency private

cloud models. Shared cyber security event response and rapid forensics would be an enhanced capability.

By designing and acquiring these data centers as a system, the DoD can achieve the economies of scale typically associated with large data centers.

Recommendation 3: The DoD CIO and DISA should establish clear security mandates for DoD cloud computing.

Security mandates should be aimed at reducing the number of cloud compromises and to mitigate those that occur. Some examples of potential mandates include:

- Hypervisors hosting DoD operating systems should have effective cryptographic sealing, attestation, and strong virtual machine isolation.
- Data at rest should be stored in encrypted form with keys protected using hardware attestation, such as a trusted platform module (TPM).
- Data in transit on communication lines should be encrypted with keys protected using hardware attestation, such as a TPM.
- Access to cloud computing systems should require multifactor authentication.

Recommendation 4: The DoD CIO should establish a central repository to fully document cloud computing transition and sustainment costs and best practices for programs underway or completed.

Because the cost savings to be gained through cloud computing are case-dependent, a central repository documenting DoD cloud computing programs is needed. The goal of this repository is to improve the understanding of the following:

- system costs before the switch to cloud computing, costs during transition, and sustainment costs
- enhanced functionality attributable to cloud computing architectures
- best practices for cloud computing security
- issues surrounding service license agreements
- metrics for availability and reliability

This repository will enable leveraging the lessons learned from several DoD cloud computing initiatives underway, including:

- NSA development and use of private clouds
- DISA Rapid Access Computing Environment (RACE)
- Army Enterprise Email

Recommendations to Improve DoD's Implementation of Cloud Computing

Recommendation 5: The DoD USD AT&L and the DoD CIO should establish a lean, rapid acquisition approach for information technology infrastructure, including cloud computing hardware and software.

Acquisition guidelines for all information technology—not only cloud computing hardware and software—should strive to create a lean, capabilities-based approach with strong, clear security mandates. Rapid certification and accreditation (C&A) and other characteristics to streamline acquisition of cloud computing hardware and software should be developed and implemented quickly.

Recommendation 6: The DoD CIO and DISA should establish standard service level agreements for private and public cloud computing.

Key attributes that should be included in service level agreements include availability, authentication and authorization approaches, data processing and storage locations, software and data back-up approaches, cyber attack event notification, required staff clearances and background checks, software and data disposition, risk disclosure requirements, and contingency plan. Transparency in all of these aspects for DoD service providers will help set standards for secure cloud computing across the economy.

Recommendation 7: The DoD CIO and DISA should participate in the public development of national and global standards and best practices for cloud computing.

A key outcome of this activity will be to inform the private sector and open source developers about the agility and auditability requirements for DoD cloud computing.

Recommendations to Improve Cloud Computing for Degraded Operations

Recommendation 8: The DoD and the intelligence community leadership should develop a unified approach for training and exercising with degraded information infrastructure, including cloud computing hardware and software.

Degraded operations in a realistic operational exercise must be implemented organically, *i.e.*, beyond simply holding up a white card to introduce a cyber event to an exercise. Advanced cyber security threats should be exercised, including a gradual ramp-up of threat and loss of disadvantaged communication and data links as well as primary capabilities. Enhanced red and blue teaming should be established along with operational exercises incorporating degraded cloud computing infrastructure. Participants should demonstrate a rapid forensics response and effective backup plans.

Recommendation 9: The Joint Chiefs of Staff and Combatant Commands should establish effective back-up plans for operations with degraded information infrastructure, including cloud computing hardware and software.

Candidate plan attributes include implementing thicker clients and forward caching of data as well as backup data networks, processors, and storage. Each organization should also develop operational contingencies for degraded networks. Potential strategies also include using local network connectivity for forward clients and narrowband, analog communication links for situational awareness and warning.

Recommendations for Investment

Recommendation 10: The DoD should continue investing significantly in information security research and development, including research and development for secure cloud computing technology.

To best leverage state-of-the-art cloud computing technologies for DoD, significant investment should continue for technology research and development activities in areas such as: efficient operations of cloud computing data centers; cloud security; secure, lean hypervisors; micro-virtualization; advanced TPMs; homomorphic computing; and cloud situational awareness software.

1. Scope of the Report

1.1 Terms of Reference

Over the past several years, cloud computing has had a major impact on commercial information processing. This report examines the suitability of cloud computing for DoD infrastructure, support applications, and mission applications.

The terms of reference for this study identified the following topics for investigation:

- Characterize the operational properties of clouds and the quality of service that can be delivered to connected users.
- Consider alternative designs and implementations of these technologies and evaluate their use for varied military and intelligence applications.
- Evaluate the vulnerability of a cloud infrastructure to various attacks, compared to alternative infrastructures.
- Determine how to avoid the danger of concentrating data and computation.
- Review and project the consequence of current trends in digital technology on cloud deployments.
- Comment on customer practices and modes of interaction with the cloud that may aid in increasing security.
- Make recommendations on what aspects of these technologies should be considered to increase reliability and to assure security as the military and intelligence communities evolve their digital infrastructure.
- Identify research opportunities and estimate the level of investment to achieve results consistent with DoD needs.

The full terms of reference can be found on page 68 of this report.

1.2 Task Force Approach

As shown in Figure 1, the task force investigated in detail cloud computing definitions, attributes, and service management models, as well as dimensions for implementation. Proposed motivation that were assessed for transitioning to cloud computing architectures included potential DoD mission capability enhancement, security improvements, and cost reductions.

The task force then developed examples for areas where cloud computing would benefit DoD missions. This resulted in a set of findings and recommendations for improving the DoD's ability to use cloud computing architectures effectively, with cost reductions and sufficient levels of security.

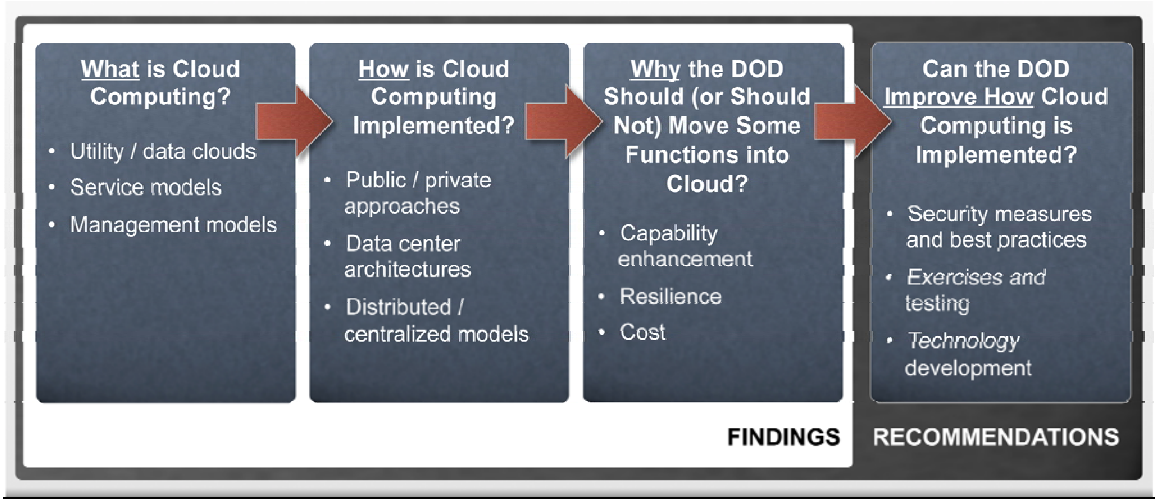


Figure 1. The task force approach

In a final phase, the task force discussed in several meetings how the DoD could improve the implementation of cloud computing systems for DoD missions and applications.

1.3 Organization of the Report

An overview of cloud computing is presented in Chapter 2. This chapter also defines terms and concepts used throughout the report. The National Institute of Standards and Technology (NIST) provided a consensus definition of cloud computing that was a useful starting point for discussions; however, the task force found places where a broader definition was also useful.

In Chapter 2, a variety of different service models and deployment models for cloud computing are described. The task force found it helpful to view a cloud computing facility as a warehouse-scale computing facility that supports computing applications and services for remote users connected using a network.

Some well-known examples of commercial cloud service providers include Google, Amazon, Yahoo!, and Microsoft, but these services can also be provided by defense agencies or defense-only contractors. Confusion regarding DoD use of cloud computing has arisen, in part, because of unstated assumptions on who provides the service.

Chapter 3 looks in some detail at cloud computing architectures and how cloud computing is implemented. A commercial cloud computing facility can contain hundreds of thousands of servers, with applications and services scaled to employing this capacity. Computing at this scale is a fundamentally new capability.

One way that commercial cloud computing facilities achieve efficiencies is through virtualization. With virtualization, operating systems and applications operate on independent virtual machines that share physical processors. By implementing many virtual machines entirely in software on a large physical machine, the arrangement more efficiently utilizes physical resources while providing computational isolation. Because virtual machines can be migrated between computers located in different geographically distributed data centers, the system experiences improved fault tolerance and load-balancing.

Chapter 4 looks at some of the benefits to DoD's mission that could be enabled by cloud computing. The mobility of computing infrastructure has important implications for DoD. The ability to move collections of virtual machines and the virtual networks that connect them will be critical for future DoD applications and missions.

Today, commercial cloud computing facilities offer an ability to self-provision computing infrastructure on demand and as needed, paying just for what the customer uses. This agility is extremely useful for settings where there is widely varying or unpredictable computing needs. The task force also observed that the wide availability of cloud computing leads to the reasonable assumption that adversaries of the United States may use cloud computing for both defensive and offensive missions.

Chapter 5 discusses security of cloud computing, which has been questioned in a number of strategies and studies.[1,2,3,4] The task force found this to be a complex subject where evolving objectives make analysis particularly difficult. The task force observed several subtleties that affect this analysis. These are highlighted here, and discussed in detail in Chapter 5.

The responsibility for security in most cases is shared between a cloud service provider and a cloud service client. Different cloud computing service and deployment models split this responsibility differently, with many models requiring that two or more parties be involved in managing the computing infrastructure and security measures. Such sharing can be a problem when the provider and client are different organizations without unrestricted two-way communication.

Security cannot be discussed independently of a defined threat. Protecting against high-level threats is extremely difficult; the safest course is to assume that any computing infrastructure might be compromised, to develop mechanisms that operate in the

1. L. Leong and N. MacDonald, "Mitigating Risks in Cloud Infrastructure as a Service" (Gartner Research G00235858, July 11, 2012). Available at time of press at http://goo.gl/oIeq5
2. United States Department of Defense, "Cloud Computing Strategy" (DoD Chief Information Officer, July 2012). Available at time of press at http://goo.gl/MfFQg
3. IBM. "X-Force 2011 Trend and Risk Report," IBM Security Collaboration (March 2012). Available at time of press at http://goo.gl/MW0qH
4. V. Winkler, "Securing the Cloud: Cloud Computer Security Techniques and Tactics" (April 2011). Available at time of press at http://goo.gl/AVEIO

presence of such compromise, and to design in a way that will mitigate the impact of compromises. Cloud computing differs little from conventional computing infrastructures in this regard.

The scale of cloud computing is vastly different from conventional computing systems. Such scale requires automation for provisioning and management of the computing infrastructure with humans out of the loop. For this reason, the security hygiene of cloud computing systems tends to be better than computing systems of comparable size. Thus, cloud computing can offer equivalent or better protection against low level threats that tend to exploit vulnerabilities caused by poor system hygiene.

Chapter 6 considers issues and circumstances in which cloud computing can be expected to lower the costs of computing infrastructure. By leveraging scale, commercial cloud computing suppliers can offer computing services and applications at lower cost than a company or organization can often achieve internally.

For example, because of the scale and the automation of provisioning and management of computing infrastructure, commercial cloud computing data centers generally require far fewer systems administrators. As an example, conventional enterprise-computing might require one system administrator per tens or hundreds of servers, while a commercial cloud service provider might only require one system administrator per thousands of servers.

These advantages must be considered against the higher costs that defense systems may incur. These may include DoD acquisition process requirements or specific certification and accreditation processes.

Chapter 7 suggests areas for research and development of technology that could be important to the DoD's use of cloud computing. An emphasis is placed on research that improves the security and capabilities of cloud computing systems. Payoffs for some investments will be seen in a few years; other problems, however, will be solved only with longer-term sustained research support.

Finally, Chapter 8 presents the study recommendations that flow from the assessments and findings in the first seven chapters. The chapter includes proposed DoD leads to take responsibility for the recommendations, and some additional detail is provided to clarify the intent of the recommendations..

2. Overview of Cloud Computing

The phrase "cloud computing" has evolved to have different meanings for different people. Rather than defining it, this chapter describes some historical background, various types of cloud computing platforms, and different characteristics of cloud computing architectures. The task force believes that, over time, cloud computing models will evolve, and this evolution may not be reflected in today's descriptions. In this report, standard definitions are used where they suffice and are expanded where necessary.

2.1 The Latest Step in an Evolutionary Process

Cloud computing can be viewed as the natural evolution of a variety of computing technologies, including virtualization, client-server architecture, the World Wide Web, and networking. The evolution of some computing platform precursors to cloud computing is shown in Figure 2.

As early as the 1960s, mainframe computers were shared among multiple users across an enterprise, while logically isolating their processing and data from each other. In the 1980s, standardized packet network protocols were developed and widely deployed, along with client-server architectures to utilize them. The ability to connect users to computing and data resources via standardized networks is a key enabler of cloud computing.

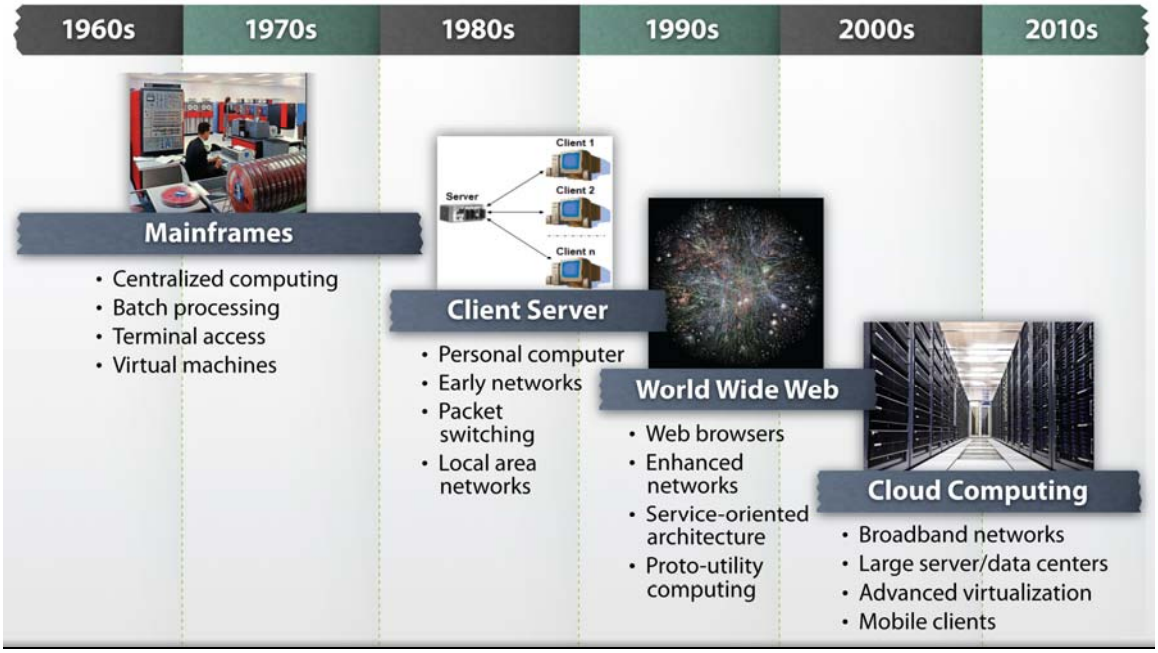


Figure 2. Historical precedents for cloud computing

The development of the World Wide Web in the 1990s, with its standard markup language, transfer protocol, and graphical browsers, made client-server computing ubiquitous. Business began to provide servers to deliver content and services at a truly global scale.

Seen in this historical context, the development of cloud computing is the next logical step in the evolution of computation. It has been enabled by the availability of broadband networks and inexpensive end-user devices, as well as commodity computing nodes that can be simply interconnected and controlled, and virtualization to provide the appearance of isolating processes that share computers.

2.2 What is Cloud Computing?

One well-known definition of cloud computing was provided by NIST.[5] It begins:

"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (*i.e.*, networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

The definition goes on to identify five essential characteristics of cloud computing. These are as follows:

- **On-demand self-service**. A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.
- **Broad network access**. The cloud's capabilities are available over the network from a wide variety of edge devices, including workstations, laptops, tablets, and mobile phones.
- **Resource pooling**. The cloud computing provider's resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. The customer (or tenant) generally has no control or knowledge about the exact location of allocated resources, but may be able to specify location at a higher level of abstraction (*e.g.*, country, state, or datacenter). Examples of resources include storage, processing, memory, and network bandwidth.
- **Rapid elasticity**. Cloud computing capabilities allocated to the customer can be elastically provisioned and released as required by demand, in some cases automatically. To the customer, the cloud capabilities available often appear to be unlimited and can be appropriated in any quantity at any time.

5. P. Mell and T. Grance, "The NIST Definition of Cloud Computing" (September 2011). Available at time of press at http://goo.gl/eBGBk

- **Measured service**. Cloud computing systems automatically control and optimize resource use by leveraging a metering capability appropriate to the type of service (*e.g.*, storage, processing, bandwidth, and active user accounts), typically on a pay-per-use basis. Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

Two aspects of cloud computing are of particular significance. The scale of processing and storage that becomes available through cloud computing is unprecedented, with up to hundreds of thousands of computers acting in concert. It is this large scale—sometimes called warehouse-scale or internet-scale computing—that enables the design of reliable computing services using less-than-reliable commodity computers. Solving this challenge has provided new capabilities.

Also new is the ease-of-use of cloud computing services. Many cloud computing service providers allow a user to configure a new computing infrastructure through a simple web form with instantaneous payment by credit card. The ability to remotely request hundreds of servers for a few hours and to have them available a few minutes later is another new capability. This capability has transformed the work of many scientists and engineers, as well as their information technology support personnel.

2.2.1 Data, utility, and other cloud computing services

Different types of cloud computing are provided from large, remotely located, interconnected data centers—hence, the common use of cloud computing to describe different uses. Cloud computing services are primarily categorized as utility- or data-intensive, and also include storage, high performance computing, and other specialized functions.

Utility computing is a label for cloud service providers that make computing resources available to consumers, much as electric companies and other utilities provide services to consumers. With an electric power utility, a homeowner can, within limits, request electricity simply by flipping a switch on a device, receive that power instantly from a distant generating facility, share the generating facility with thousands of other customers, use more or less power as needed, and pay only for the power actually used.

Utility computing customers include, for example, a retailer who chooses to purchase cloud services to host an internet-facing e-commerce web site. In this way, the retailer gains increased capacity and geographic presence over what could be obtained if the retailer had to buy, operate, maintain, and upgrade their own dedicated computing resources. Another example of a utility computing customer was demonstrated by the *New York Times* in 2008 to process more than a hundred years of digitized archived images, articles, and metadata in order to produce more web-friendly images and more

accessible JavaScript data files. By using Amazon Web Services, the *Times* completed this enormous task in less than 36 hours.[6] In these examples, cloud computing service providers enabled customers to perform compute-intensive processes as needed without a large investment in infrastructure.

Utility computing enables a cloud service provider to exploit economies of scale and uncorrelated customer demands to share computing capability among a collection of customers, at an attractive price. Individual consumers perceive that they are accessing an infinite resource on demand. They also perceive that their computing tasks are operating in isolation from those of other consumers.

Data-intensive cloud computing is a type of parallel processing applied to very large datasets. An example of data-intensive computing is the process by which search engines index the data available on the World Wide Web. The underlying computational steps required to index data are simple—sorting, counting, merging, and so on—but the amount of data to be processed is so large that it requires specially-adapted software for data ingestion, analysis, database operations, and file system storage.

Data centers will generally be designed and optimized for different requirements. Utility computing design focuses on sharing resources, lowering the cost of computing to the customer, and providing computing capacity on-demand. The utility computing customer trades capital costs for operating costs. Data-intensive computing focuses on performing rapid analysis of large datasets, and vast amounts of computing resources may be dedicated to a single user or task. A data-intensive architecture will be optimized for large scale parallelization.

2.2.2 Cloud computing software stack and virtualization

Today, cloud computing infrastructure usually consists of a large number of interconnected, inexpensive, commodity processors. The software running on each processor is modular and layered. Figure 3 shows a typical layered "stack" of software running on a single cloud computing node, with descriptions of each layer in the stack.

The hypervisor provides virtualization by providing an interface to the virtual machines (VMs) that gives each of them the illusion that they have complete, exclusive access to the underlying hardware resources. The ability to run multiple, isolated virtual machines on a single hardware node is fundamental to cloud computing because it enables resource pooling and rapid elasticity. Multiple users can use the same physical node without interfering with each other, and nodes can be rapidly assigned and reassigned as users' computing demands ebb and flow.

6. D. Gottfrid, "The New York Times Archives + Amazon Web Services = TimesMachine," *New York Times* (May 21, 2008). Available at time of press at http://goo.gl/G7uvG

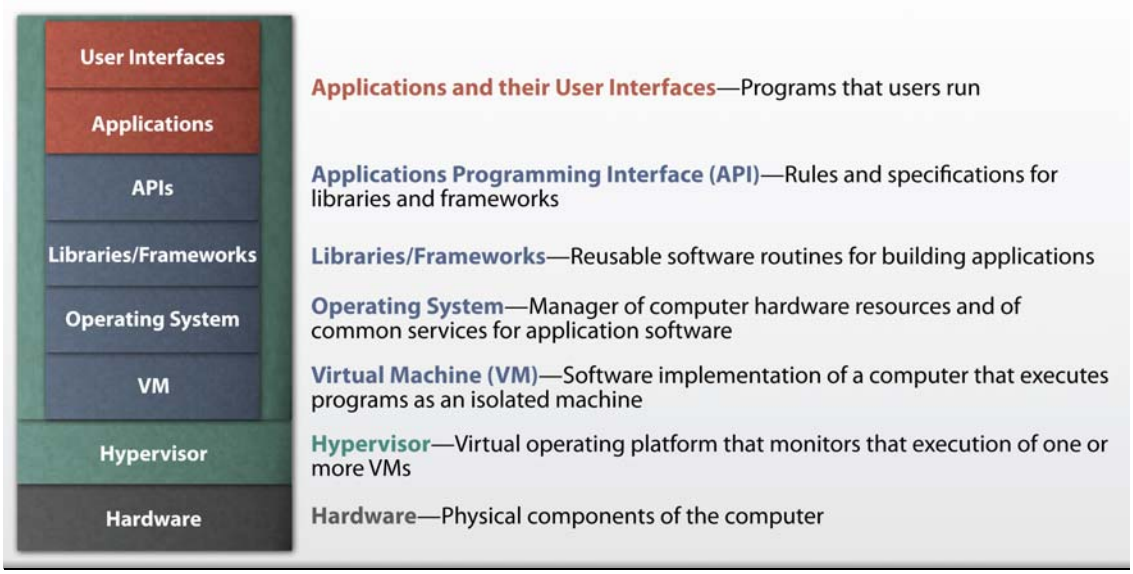


Figure 3. The main components of a cloud computing software stack

Although virtualization is one of many enablers of cloud computing, cloud computing is more than just virtualization, and there are applications of virtualization that are not instances of cloud computing. For example, a departmental data center may use virtualization to allow a single hardware server to run multiple VMs, with each VM configured to run only one specific service. Such implementations may offer limited resource pooling and no elasticity—the virtualization is used in this case merely as a convenient mechanism for ensuring adequate isolation between services that is more cost-effective than assigning one hardware server per service.

The various cloud software service models assign responsibility for managing the software stack differently. Figure 4 shows that the cloud service provider provides the underlying hardware and the hypervisor in all service models, and that the upper layers of the stack can be provided and managed either by the service provider or by the cloud computing customer.

2.2.3 Cloud computing service models

Different types of service are available to cloud computing customers, depending on how much control a customer requires. The NIST definition of cloud computing describes three service models (as reflected in Figure 4):

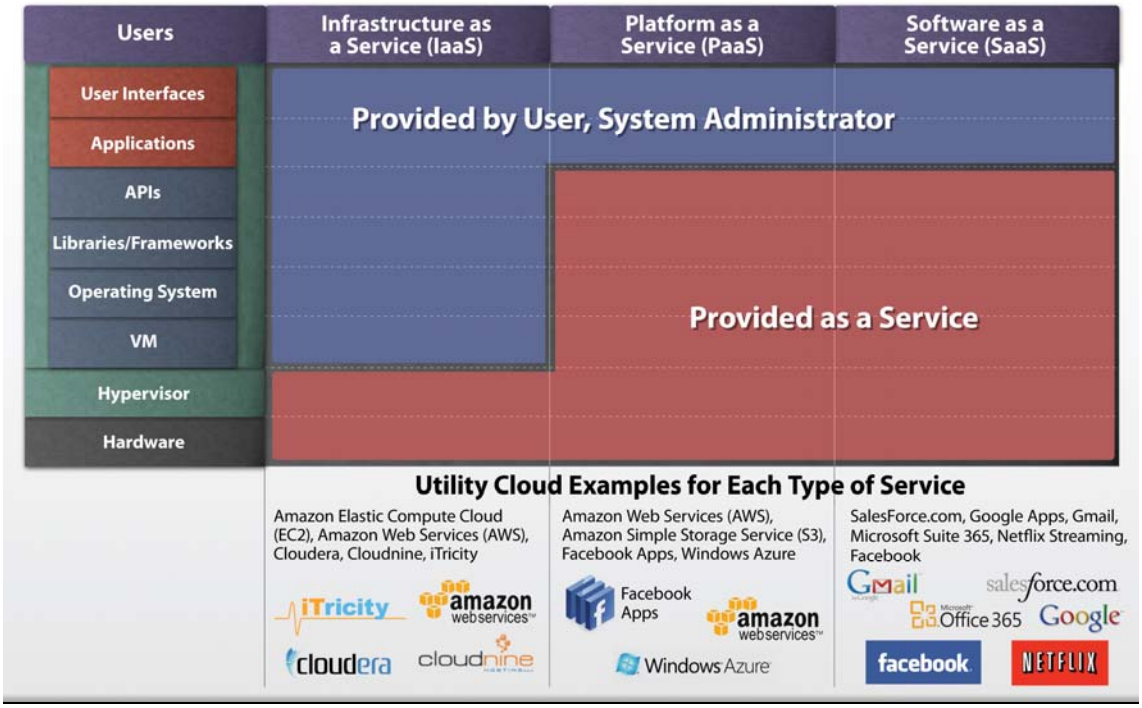


Figure 4. A cloud computing software stack responsibility as a function of service model

- **Software as a Service (SaaS)**. With SaaS, customers use software applications that are developed, managed, and operated by a provider. The applications are accessible from various client devices through either a thin client interface, such as a web browser (*i.e.*, web-based email), or a specifically developed programmatic interface. The customer does not manage or control the underlying cloud infrastructure, including network elements, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.
- **Platform as a Service (PaaS)**. With PaaS, customers create their applications using standardized programming languages, libraries, services, and tools supported by the provider. The customer does not manage or control the underlying cloud infrastructure (including networks, servers, operating systems, or storage) that execute the applications, but the customer has control over the deployed applications and possibly over configuration settings for the application-hosting environment.
- **Infrastructure as a Service (IaaS)**. With IaaS, the customer provisions processing, storage, networks, and other low-level computing resources. Also with IaaS, the customer can deploy and run arbitrary software, which can include operating systems and applications. The customer has some control over operating systems,

storage, and deployed applications; and possibly limited control of select networking components (*i.e.*, host firewalls or software defined networks).

In SaaS, customers have limited ability to make configuration changes, but cannot modify the application, such as a web-based email system. In PaaS, the consumer is able to build and upload his own software applications for running on the provider's computing resources, but is constrained to use the tools supported by the cloud provider. This gives the PaaS consumer more flexibility than SaaS without all the complexity of managing lower-level components (*i.e.*, the operating system). In IaaS, consumers have maximum control over the software running on the provider's hardware, with responsibility for many of the attendant management and security challenges.

These NIST-defined service models span a single dimension—the level of software control ceded by the provider to the consumer. Another important dimension is how the computing provided by the cloud is used, which leads to phrases like "data as a service" for cloud storage of data, and "security as a service" for security services provided via cloud computing, such as host-based antivirus and firewall software.

2.3 Managing Cloud Computing

The NIST definition lists four deployment models for sharing cloud resources:

- **Private cloud**. Provisioned for exclusive use by a single organization, the cloud infrastructure might be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.
- **Community cloud**. Provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (*i.e.*, mission, security requirements, policy, and/or compliance considerations), the cloud infrastructure can be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it might exist on or off premises of one of the organizations.
- **Public cloud**. Provisioned for use by the general public, the cloud infrastructure might be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.
- **Hybrid cloud**. The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (*i.e.*, cloud bursting for load balancing between clouds).

Consumers of services with the least tolerance for sharing resources and relinquishing control usually choose a private cloud deployment, while those with more tolerance will choose a community, hybrid, or public cloud.

2.3.1 Cloud management models

NIST defines the four deployment models, but users of cloud computing services face many more choices as they decide who will own, manage, operate, and support the site of the cloud computing infrastructure, who will own, manage, operate, and support the hardware, who will own, manage, operate, and support the various layers of software, and so on. For all cloud computing users, these configuration options must be weighed against economics, security, and other factors. This balance is evolving.

Figure 5 lays out a range of possibilities for management and operating cloud computing infrastructure. The far left of the table suggests that very sensitive applications are not suitable for deployment in cloud computing architectures. Such applications include nuclear weapon security systems, some command and control systems, and weapon system fire control.

Other sensitive applications are best reserved for in-house private cloud computing. For the in-house private approach, the DoD would host the cloud computing infrastructure, control the hardware and software implementation, and use DoD-employed staff support.

For the in-sourced private model, the DoD hosts the cloud computing infrastructure. The hardware and software stack and staff support might be provided by an external contractor. For the out-sourced private model, an external cloud service provider would host hardware to be used exclusively by the DoD. The control of the hardware and software stack and staff support might be provided by both the DoD and an external contractor. Some less sensitive applications may be appropriate for these approaches.

	Not Cloud	In-house Private	In-sourced Private	Out-sourced Private	Public
Hosting	DoD site	DoD site	DoD site	Commercial site	Commercial site
Hardware	DoD owned	DoD owned	DoD or Commercial owned	DoD or Commercial owned	Commercial owned
Software	DoD specialized	DoD stack	DoD or Commercial stack	DoD or Commercial stack	Commercial stack
System Administrator	DoD	DoD	DoD/Commercial	DoD/Commercial	Commercial
	Nuclear security Command and control Weapons fire control Highly classified needs	**More Sensitive** Geospatial image analysis Moving target radar analysis Geo-location processing Data integration Large simulations	**Less Sensitive** Email Calendars Logistics Purchasing Finance records		Public released data External web sites Public services Recruiting

Figure 5. Cloud computing management models

For the public cloud computing model, an external cloud service provider employs hardware at its site. The hardware may not be used exclusively by the DoD, and the cloud service provider controls the software stack and employs the staff support. For some applications, where data or processing has been publicly released and required latency and system availability is consistent with public cloud service providers, public cloud computing could be acceptable for the DoD.

The following cloud data center management models are described in more detail:

- **In-house private design**. DoD privately operates the data center with high physical security. The DoD directly controls the hardware and software configuration, and the IT operational support staff is employed by the DoD. The cloud data center may have as its tenants a single mission or multiple missions.
- **In-sourced private design**. DoD privately operates this data center with high physical security. DoD either directly or through contractors assembles the infrastructure, with the goal of maximizing the use of well-vetted infrastructure components. However, DoD (or its contractors) might themselves build some infrastructure or application components when assurance needs dictate. The cloud data center might have as its tenants a single mission or multiple, shared missions.
- **Out-sourced private design**. A DoD contractor operates the data center. All DoD applications are run in a DoD enclave, physically segregated from non-DoD tenants in the cloud data center. Data center personnel with access to the DoD enclave are U.S. citizens, meeting specified personnel security requirements. All security-critical components in the cloud data center are subject to DoD review. These components are likely to include those concerned with, data integrity, software integrity, key management, and key storage. DoD has access to incident and forensic information concerning all cloud tenants. Individual tenants are responsible for building or integrating applications, but those applications are subject to the data center's security requirements for operation and audit.
- **Public, with user provenance design**. A commercial contractor operates the data center. It provides services under commercially available terms. Security-critical components in the cloud data center are subject to DoD review, including components that affect data security, data integrity, software integrity, key management, and storage. DoD has access to incident and forensic information concerning all cloud tenants. Individual tenants are responsible for building and integrating applications, and those applications are subject to commercially established security requirements for operation and audit.

2.3.2 Cloud computing service providers and proprietary networks

Services offered by cloud computing providers may be connected to proprietary networks to provide services that users require. These services might be deployed according to any of the various deployment and management models described above.

A salient example is the Global Information Grid (GIG). The GIG is the DoD's globally interconnected end-to-end set of information capabilities, associated processes, and personnel for collecting, processing, storing, disseminating, and managing information on demand. Users include warfighters, policy makers, and support personnel. The GIG includes DoD-owned and leased communications, computing systems and services, software applications, data, security services, and other associated systems.

Figure 6 is a simplified diagram that shows a notional relationship of cloud computing to some traditional functional components of the GIG. As shown here, cloud computing brings a new capability to the GIG, but it does not replace the GIG.

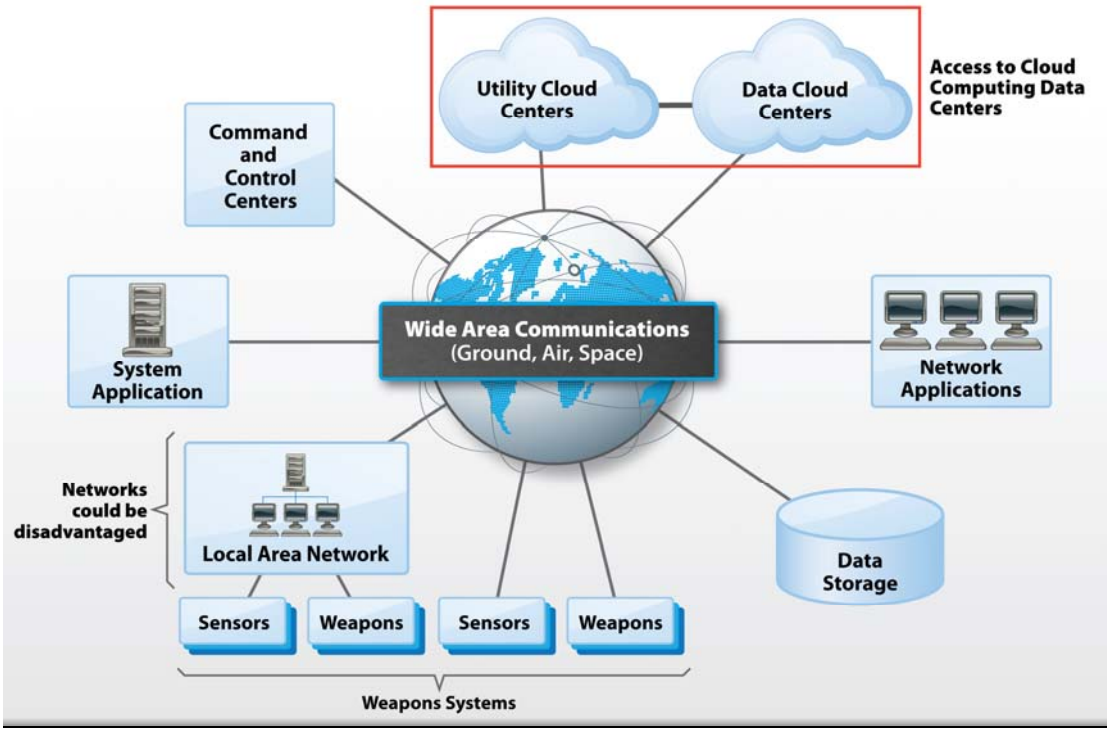


Figure 6. Cloud computing hardware and software as components of the GIG

3. Cloud Computing Architecture and Implementation

Cloud computing architectures have developed to support the key benefits of cloud computing: elasticity, economy of operation, massive scaling of computing resources to solve critical problems, and real-time responsiveness.

3.1 The Building Blocks of Cloud Computing

Technology in several key areas has driven cloud computing architecture development and made cloud computing possible. A few critical developments are described here:

- **Commoditization of microelectronics**. Digital devices have become cheaper and more capable over time and, as a result, are now widely available and deployed in a broad set of contexts. Computing devices that cost millions of dollars in the 1960s have equally powerful descendants costing hundreds or even tens of dollars. The explosion of the personal computing market from 1975 through 2005 represents one outcome of this revolution in capacity and cost.
- **Networking**. Fast, cheap, ubiquitous networking between previously unrelated parties is critical for the success of cloud computing. The Internet fulfills this role by imposing interoperability requirements on its end-hosts. These interoperability requirements, in turn, have led to standards for protocols and services that have facilitated the emergence of a low-cost and widespread network infrastructure.
- **Virtualization**. A hypervisor provides an interface to implement VMs, giving each VM the appearance of exclusive access to a physical processor. The ability to run multiple, isolated virtual machines on a single hardware processor is fundamental to cloud because it enables multiple users to use the same physical node without interference, and it enables nodes to be rapidly assigned and reassigned as user computing demands ebb and flow. Virtualization facilitates the resource pooling and rapid elasticity that characterize cloud computing.
- **Commodity hardware**. Software is increasingly targeted at commodity computer hardware. A noteworthy outcome of this trend shaped high performance computing (HPC), where performance equal to or better than custom processor designs (*i.e.*, supercomputers such as Cray machines) has been achieved at much lower cost by using clusters built from commodity processors and high-speed interconnects. These clusters are the forerunner of today's cloud computing data centers.
- **Open source software**. Unix was among the first operating systems with widely available source code. This availability fostered broad community participation in

the development of operating systems, applications, and tools that can be used as-is, or can be adapted by any developer. Open-source systems have, in many cases, outstripped expensive commercial software in function and quality, and many cloud computing systems, development tools, infrastructure, and applications are built from open-source components. Cloud providers routinely contribute labor and computing capacity to these open-source development efforts and the cloud community has several complete open-source frameworks.

3.2 The Scale of Cloud Computing

Combining the technology developments listed above would be a positive step but would yield only incremental improvements in cost and capacity. The signature characteristic of cloud computing is scale. Scale is the differentiator that brought a giant leap in computational and storage capacity in recent years. Search engines and other massive data applications were initial drivers for the evolution of cloud computing architectures, as exemplified by Google's mission "to organize the world's information and make it universally accessible and useful."[7] Today, cloud computing infrastructures support the large-scale storage and processing of many different types of data.

The scale of a modern cloud computing data center is sometimes difficult to comprehend. They are designed to support hundreds of thousands of central processing units, many petabytes of data on shared disk drives, and nearly a petabyte of dynamic storage of memory.

Thus, as is shown in Figure 7, cloud computing data centers are very large physical plants, sometimes with acres of computers. Even a small facility might consume several megawatts of electricity for cooling and powering the electronics, and some of the larger data centers today consume much more. A communications infrastructure is likely to support tens to hundreds of gigabytes per second of network ingress and egress; storage requirements dictate many thousands of disks.

3.2.1 Benefits and challenges of scale

One of the most attractive new capabilities of cloud computing is elasticity—the ability to rapidly and dramatically increase the computing resources assigned to solve a problem. Elasticity is achieved mainly by designing resilient infrastructure and applications and by deploying uniform hardware and standardized operations so that tasks can be redistributed and relocated within the computing substrate.

7. "About" Google. Available at time of press at http://www.google.com/about/company/

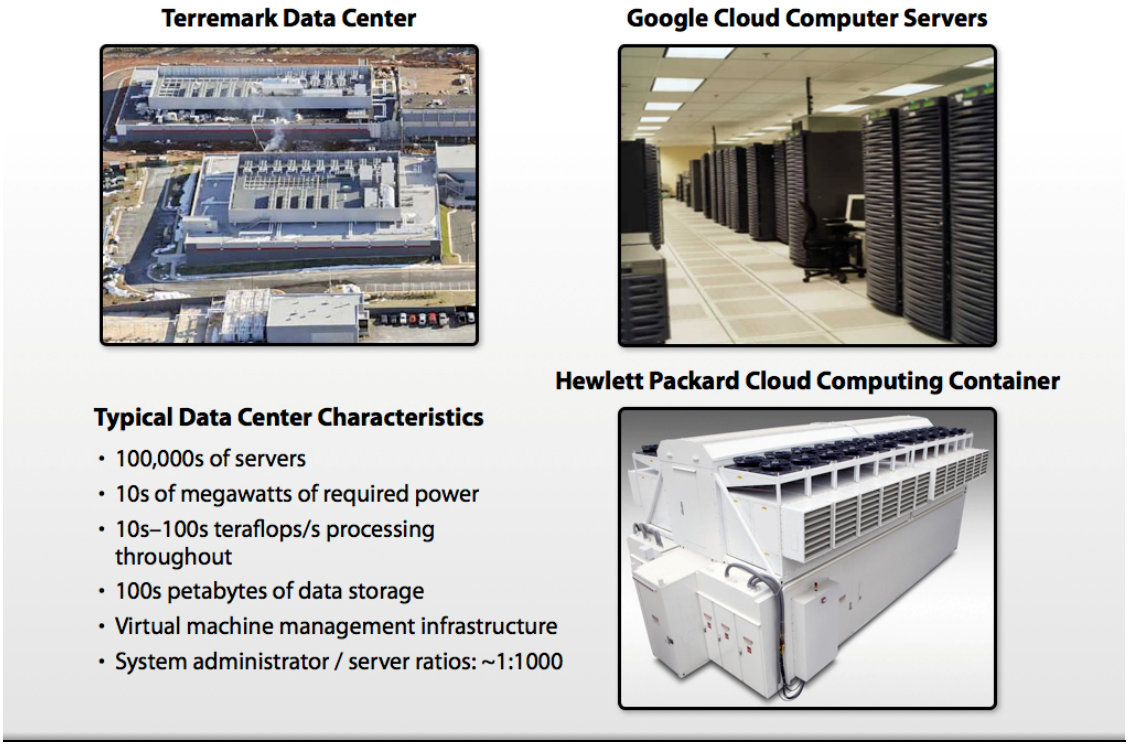


Figure 7. Cloud computing data center characteristics and examples

The ability to scale seamlessly can also enable rapid processing and analysis of very large datasets through using highly parallel operations. This is an important capability for DoD.

Using scale to achieve reliability as well as performance leads to profound changes in application design and application management. Many cloud computing applications are long-running and have thousands of software components that cooperatively and continuously execute across multiple data centers. In a typical cloud computing data center, even with high quality parts, hundreds of disks and electronic components fail every day. Homogeneity of hardware, infrastructure, service, deployment, and operation of applications is crucial for achieving the efficiencies of a cloud.

Assembling a large facility at an enormous capital expense dictates that different parties with different goals and objectives must be able to use the facility efficiently and securely, despite none having physical access to the facility. This style of sharing could be accomplished through strict physical isolation between tenants, or through strong but flexible access controls to support collaborative access to shared data.

Another potential benefit of scale is cost savings. Applications that may experience the most cost benefit from cloud computing architectures are those used in the same

way by large numbers of people, such as email. Many, if not most, DoD applications, however, are not standardized.

Designing more specialized applications to operate efficiently in a cloud computing environment can require large up-front development costs. Even slight customizations, such as the ability to find and erase any unintentionally transmitted classified material, could quickly and dramatically reduce potential savings of using an existing email application. However, even addressing all of the technical challenges, large cloud data centers have been shown to achieve a factor of ten cost savings over smaller installations. Cost savings are discussed in more detail in Chapter 6.

3.3 Specific Cloud Characteristics Affecting Architecture and Implementation

Each of the following characteristics informs cloud architecture design tradeoffs that materially affect performance and security. These tradeoffs will be explored further in the discussion on security in Chapter 5.

3.3.1 Automatic provisioning and infrastructure management

Commercial cloud computing data centers have developed an operational model that requires few visits by human operators to individual computers or nodes. Many operations can be automated and the physical configuration for all the machines in a data center is generally homogeneous. No application can rely on special setup of a particular machine or continuity of execution on the same computer. Some installations simply disable nodes as individual components fail, but many providers have excellent diagnostic software that can help to forecast and avoid hardware failures proactively.

Applications must be packaged to support automated data center operations. Typically this involves expressly specifying provisioning requirements—what resources are required to run the application—and designing software to tolerate the full range of resource assignment within the scope of the specified requirements. Data center personnel generally do not know application behavioral patterns; hence the operations staff cannot detect or fix anomalous behaviors *except* when problems are expressly registered with data center operations software. When such a report is made, a pre-specified automatic procedure is performed without human intervention.

3.3.2 Application development and scale out

Existing application software sometimes does not perform efficiently when it is simply deployed on a cloud computing system. Well-architected cloud computing applications must detect failures, incorporate failover alternatives, and provide sufficiently robust diagnostic support to allow remote analysis and debugging of problems as well as implementing automated contingency planning and

reconfiguration. Conventional computing, conversely, generally deals with component failure using lower-level mechanisms, such as redundant hardware, mirrored storage, and automatic failover. These are more costly and more time-consuming.

Many cloud applications are accessed through Internet browsers, which can be challenging to secure. For these applications, there is a performance premium on reducing data exchanged between the browser front-end and the cloud deployed back-end. Network round-trips must also be reduced, and if disconnected operation is required, provisions may be designed to cache data at clients.

When applications deployed in cloud computing share data and infrastructure, they must use standard protocols, which can limit flexibility and can make application development, debugging, and testing expensive. As a result, the expense of customizing some legacy applications for cloud deployment can be substantial.

Redesigning a legacy application so that it benefits from scaling can sometimes require significant effort. An application developer must carefully consider shared volatile state management and full system effects such as latency, network and storage failures, and correlated hardware failures.

3.3.3 Application centralization

Because cloud data centers involve replication of hardware, systems software, and application software elements, care must be exercised to avoid the risks that monocultures bring. Fortunately, the uniformity also means that changes and security updates can be installed very quickly. Moreover, one impediment to installing updates—backward compatibility—is less of a problem in cloud computing because data can be migrated at the same time as the software update is deployed.

Because cloud-based applications typically are partitioned between a client front-end and a cloud back-end server, security issues arise that do not occur when client and server are deployed within the same enclave. In particular, in cloud deployments, strong client and server authentication must be used.

3.3.4 Data collection and centralization

Cloud computing is a natural repository for large and complex datasets that cannot be easily managed or accessed using traditional database management tools. Indeed, cloud computing services, such as Facebook, Google, and Amazon, rely on such centralized data repositories. Central repositories are attractive attack targets—both by insiders and outsiders. For this reason, special attention must be paid to data provenance for damage control, forensics, accountability, and data-quality control.

3.3.5 Clients

Almost all cloud computing services are accessed through a client—an application or system that accesses a service made available remotely. Client design is thus an integral part of any cloud application. Many reported cloud security failures have been attributed to bad or compromised client machines.

3.4 Architecture of a Modern Cloud Data Center

When building a cloud data center, a prospective designer must specify the machine and cluster configurations, storage architecture, network connectivity and management, and physical infrastructure, such as power and cooling. Often data centers are built near hydroelectric facilities to exploit the cheap power and near major fiber links to facilitate high-bandwidth remote access to the cloud. Site characteristics conducive to cooling, as well as access to a trained support staff, are important. The expected frequency of natural disasters (*i.e.*, earthquakes, floods, or hurricanes) and proximity to transportation are also key factors in site selection. In addition, the buildings and campus themselves must be built to ensure physical security.

Designing data center software to manage and monitor machines and the network, as well as providing software for common tasks, is just as critical as physical construction details and hardware procurement choices. In fact, how DoD obtains, develops, maintains, and evaluates software will have a big impact on cloud security, economy, and performance. Key software elements include:

- storage systems software, including access control
- network management software
- software to help detect and correct malfunctions or malicious activity
- resource allocation software to assign tasks to hardware elements
- system software to isolate tenants, be they clients or clouds, so that a malicious tenant cannot affect any other tenant
- plant software to manage power and cooling in the data center
- software for load balancing within and between data centers

A notional design of such a data center is depicted in Figure 8. This data center uses virtualization technology, which is common in data centers, but is not required. Key elements in Figure 8 are:

- **Network head node**: These components provide external data center network access.
- **Network forensics analytics**: These components monitor network behavior to detect attacks and failures.

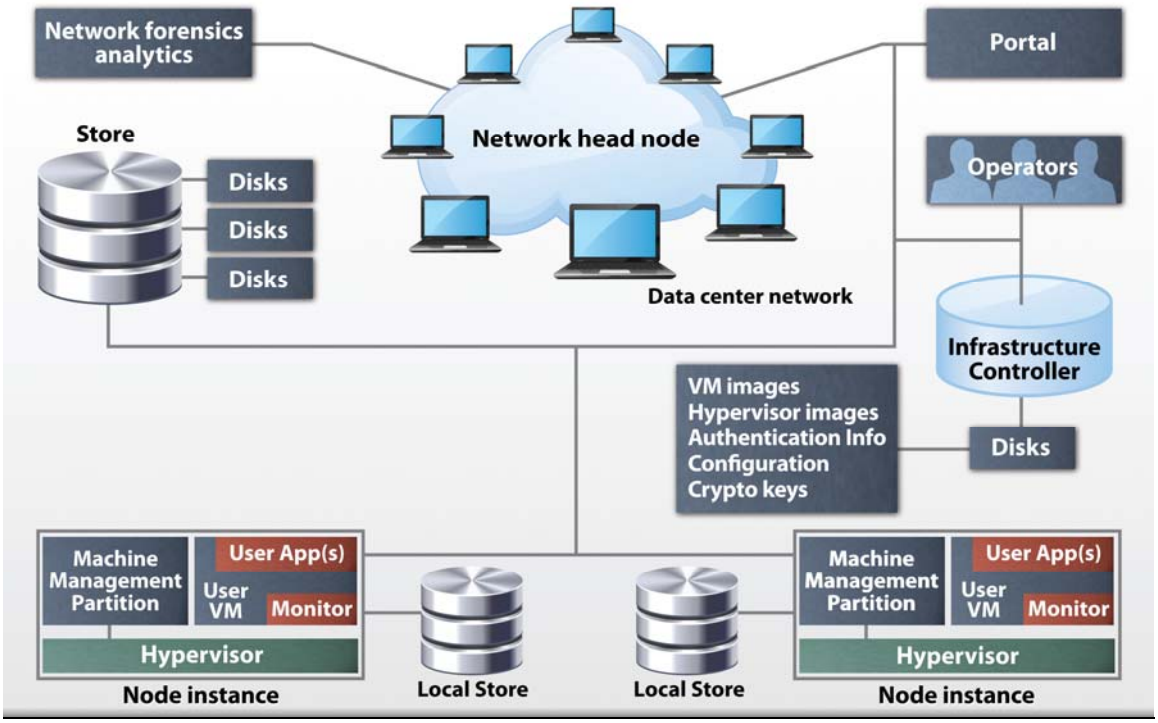


Figure 8. Example of a cloud computing data center architecture

- **Data center network**: This is a high performance network connecting all machines within a data center.
- **Portal**: This component registers new data center users, obtaining billing and authentication information. Tenants access the portal to transfer software and data to the cloud, negotiate resource assignments (*e.g.*, how many machines are needed, when, how much storage capacity, networking characteristics, and special requirements).
- **Storage**: A high speed, fault-tolerant storage system for the data center.
- **Infrastructure Controller**: This component allows data center operators to assign physical resources, monitor hardware and software health and operations, and detect and remedy attacks and failures as they arise. All data center software is deployed and managed through the infrastructure controller.
- **Node instances**: These machines run the applications for tenants. Tens or hundreds of thousands of nodes are in a typical cloud data center. Each node has a hypervisor to manage machine resources and to isolate and protect user software from other software sharing the node. A management partition obtains, configures, and starts user software on the machine, and monitoring software monitors node heath and operations.

Each of these elements represents design choices that affect cost and performance. For example, the network might allow one cluster within a data center to become partitioned from another but will not allow a partition within a cluster. Management software, in conjunction with user-supplied information, would then be knowledgeable about this clustering and allocate to each application only those elements located within the same cluster. Similarly, a data center may provide some heterogeneous computing elements—powerful processors that can perform computations such as fast Fourier transforms—much more quickly than normal computing units. This heterogeneity will also be visible to the management software so that appropriate resource allocations will be made.

3.4.1 Modular data centers

One innovation in the design of data centers is to use pre-assembled modular units that together create a data center of varying size, depending upon the number of units used.

Early versions used standardized shipping containers and contained racks of computers and all the associated power distribution and cooling units required. These containers were simply hooked up to power, chilled water, and networking cables to make them ready to be used. A simple concept for such a modular data center is shown in Figure 9.

Today, new variants of modular data centers include those that use custom racks for greater densities, separate containers for the associated cooling, and custom containers that are easier to maintain. New designs may also assemble modular units of different configurations that, as an aggregate, provided all the required computing, power distribution and cooling required. As an example, a single modular data center might contain 44 racks with 7,000 servers and require 1.3 megawatts of power.

Although it would be less expensive to build a full-size data center rather than construct it entirely from modular units, in practice, modular data centers are much faster to install. They can also make it much easier to add incremental or refreshed

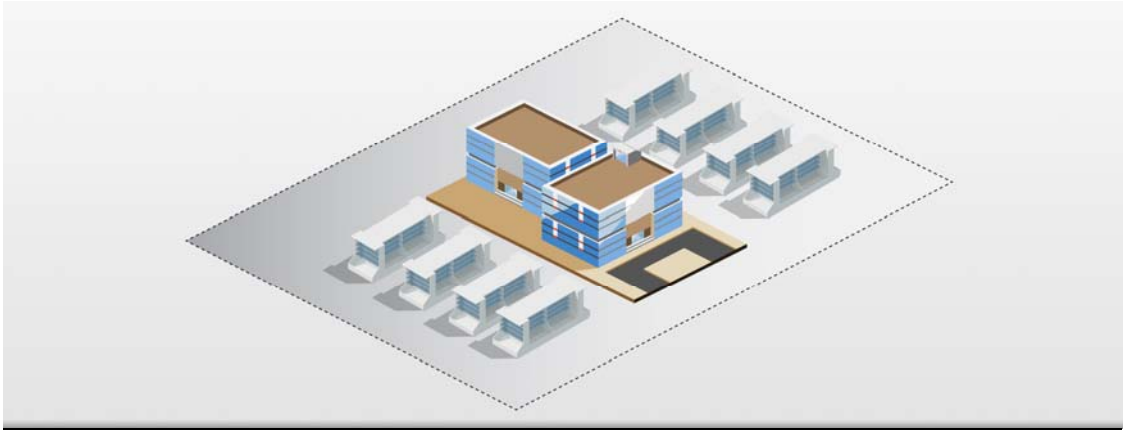


Figure 9. Concept for a modular DoD data center

computing capacity by building them a container at a time. Modular data centers can also be easily transported to where they are needed. For these reasons, modular data centers are often used in cloud computing.

Alternatives to modular data centers include data center designs in which entire rows of preconfigured racks can be quickly snapped into place by simply connecting the appropriate electrical cables and cooling hoses.

3.4.2 Critical cloud computing design choices

A well-designed cloud computing data center will reflect its projected uses. The degree of automation and flexibility in the management software of a cloud computing data center will depend on the applications that are run there.

For example, a single, large SaaS application (such as search) may be operated and used by a single organization, and as such will require only modest data security. The need for a hypervisor to provide isolation between tenants becomes less compelling, because only a single application is being run with no need to co-locate with different—and potentially adversarial—applications. By contrast, when an application involving highly sensitive data is deployed in a cloud computing data center running many other programs, the system design will include a hypervisor on each processor to assure isolation. If multiple tenants share a facility, it becomes important to manage resource usage quite strictly to assure responsiveness for all.

Social networking or search applications will interact almost exclusively with client machines through internet browsers. Data center and application design in this case would focus on protecting data leaks from one user to another, which may be achieved at the expense of availability.

In some special cases, a cloud data center may be used for software development. The design will need to allow access and control of running programs to facilitate debugging, with the understanding that speed may suffer. Other data centers may be designed to minimize latency, support high interactivity, or maximize physical security. These may benefit most from locating a small cloud computing data center near the user, either as the sole source of computing capabilities or as an intermediary.

Finding

Finding 1: Although cloud computing is an overloaded term, cloud computing providers are offering services that are fundamentally new and useful, typically encompassing the:

- ability for massive scale-up of storage and computing
- rapid, agile elasticity with the ability to increase and decrease storage and computing capacity on-demand, when the community of tenants don't all require that capacity at the same time
- metered services where the user pays only for what is used

- self-service start-up and control

Finding 2: Modular data centers offer an approach to quickly set up cloud computing capacity, add additional capability to existing cloud computing data centers, and easily refresh or update existing capability.

4. Cloud Computing Benefits to the DoD Mission

Cloud computing offers the DoD new ways to provide computational capabilities for missions. DoD missions most likely to benefit from cloud computing services will satisfy one or more of the following:

- **Scalable, on-demand computing.** The elasticity and resource-pooling provided by cloud computing is useful to applications that involve varying or unpredictable computing capacity. This model works well for applications that do not require highly correlated computing capacity, so it may not be useful for active missions or intensive exercises.
- **Integration of many, high-capacity data feeds**. The DoD collects high-capacity data from sensor networks and other sources, and data clouds have proven effective for the large-scale ingestion and integration of this kind of data. If cloud computing data centers are not used, custom-designed large-scale computers would be required to support these applications, and the construction of such machines is far more costly.
- **Analysis of very large datasets**. The DoD has the requirement to analyze large datasets. Over the past several years, a number of cloud computing applications have been developed, including Hadoop, Accumulo, Cassandra, and Hive, that scale to many thousands of processors and support easy-to-program parallel computing frameworks. These make big data analysis a practical enterprise.
- **Connections to common services.** Such applications as email, shared calendars, unclassified training, or document-preparation can benefit from SaaS, PaaS, or IaaS. Accessing these applications through cloud computing results in lower computation cost, lower software management costs, and enforced uniformity and interoperability. DoD has already begun to move some common services into private and public cloud computing architectures.

In this chapter, five examples of defense applications are discussed that have proven to be well-suited for cloud computing data centers.

4.1 Example: Communication and Networking

Email, calendars, and contact lists are applications found in many of today's commercial cloud-based computing services with millions of regular users. These applications rely on redundant storage to enable widespread availability, many identical processors for interactive performance, and a simple and uniform user interface across different internet browsers. The required bandwidth from client machines to the cloud computing data center is relative low, so the internet suffices. These services are also easily accessed from highly portable devices—cell phones and tablets—that are useful in many DoD scenarios.

Technologies for e-learning will also be increasingly important to the warfighter. As applications such as YouTube and Netflix have demonstrated, commercial cloud computing is a reliable, economical, and highly scalable way to provide video to users. The ability to access a YouTube-like system for virtual training sessions is an integral part of the Army Training Concept. New ways to deliver multi-media content will be needed in all locations that the DoD operates.

4.2 Example: Analysis of Large Datasets

A wide variety of cloud-based applications have been developed to support the analysis of extremely large datasets. Specialized clouds are being built to process, manage, and analyze signals, imagery, and other types of intelligence.

One example is Hadoop, which consists of a distributed file system and a simple to use parallel programming framework called MapReduce. Hadoop is widely used by the DoD and other U.S. government agencies, as well as in commercial applications. The Hadoop Distributed File System is designed to run on top of unreliable computer servers, and uses replication in order to safeguard the data it manages. One of the reasons for the popularity of Hadoop is that there is a rich ecosystem of applications based upon Hadoop, including Hive, a data warehouse infrastructure; HBase, a distributed database; and Pig, a high-level data flow language.

Accumulo, is a distributed database that is built over Hadoop, and includes cell-level security. Accumulo has proved effective for several DoD applications.

Another example of applications working with large datasets is sometimes called NoSQL databases. These applications relax some of the characteristics usually required for databases, such as transactions, in favor of scale. A widely used example of this type of application is Cassandra, which is a highly scalable key-value store.

Hadoop and NoSQL databases are currently used effectively to support intelligence analysis applications. These applications range from simple search to more sophisticated queries that look for patterns of interest in the data.

4.3 Example: Operational Support for the War Fighter

Cloud computing can offer great benefits to warfighters, especially when they have adequate connectivity back to a cloud computing data center. Useful services that can be provided by the cloud include translation, maps and navigation, searchable stored images, and specialized analyst applications. Smaller-scale, modular data centers can be used to support these types of applications in theater.

With the poor and uncertain communications that sometimes occur in the field, cloud architectures are needed that include thick clients and other components, such as local

caches, that enable continued, though degraded, operations when network connectivity is not available.

4.4 Example: Situational Awareness for Cyber Security

The DoD continuously monitors the health of their computing systems to ensure that these information systems are capable of supporting DoD missions. A part of that monitoring process involves analysis of data reporting adversary activities against DoD networks, both on the boundaries and inside them. A notional architecture for supporting this situational awareness would array analytic, foundational, and enterprise services with associated sensor applications, as they are made available to users. Data flows into the system from multiple sensors, including client and server logs and network traffic capture. Various services analyze and process the data, providing risk and mission readiness assessments, malicious activity analyses, and anomaly detection.

Some of the challenges faced by engineers building such situational awareness systems include the high rate of data ingest from a multiplicity and variety of data sources. Data is produced by each host, server, network device, and process on the network being monitored. In addition, raw network traffic is also collected at many points within the network and at network boundaries.

Some types of data analytics are compute-intensive. Further, the computing capacity required will vary, depending on the data ingest rates, which themselves may vary depending on time of day, day of the week, time of year, world events, and so on. A good example of compute-intense algorithms are graph-based algorithms that look for anomalies in graphs built from the cyber data. In one study, an enterprise with 3,500 users and 9,500 hosts on its internal network accessed more than 200,000 web servers during the course of one month, producing 7.5 million unique connections and more than 500 million proxy log entries.[8]

4.5 Example: Wide-area Persistence Surveillance

Over the past ten years, DoD has built a large inventory of airborne battlefield sensors. Examples of the kinds of data being collected include signals intelligence and full-motion video of moving targets. These sensors are capable of collecting large amounts of data, which require subsequent processing before they can be exploited and disseminated. Figure 10 shows that data rates for a variety of advanced sensors are

8. B.A. Miller, N.T. Bliss, and P.J. Wolfe, “Toward signal processing theory for graphs and non-Euclidean data, in Proc.,” IEEE Int. Conf. Acoust., Speech and Signal Process (pp. 5414–5417, 2010).

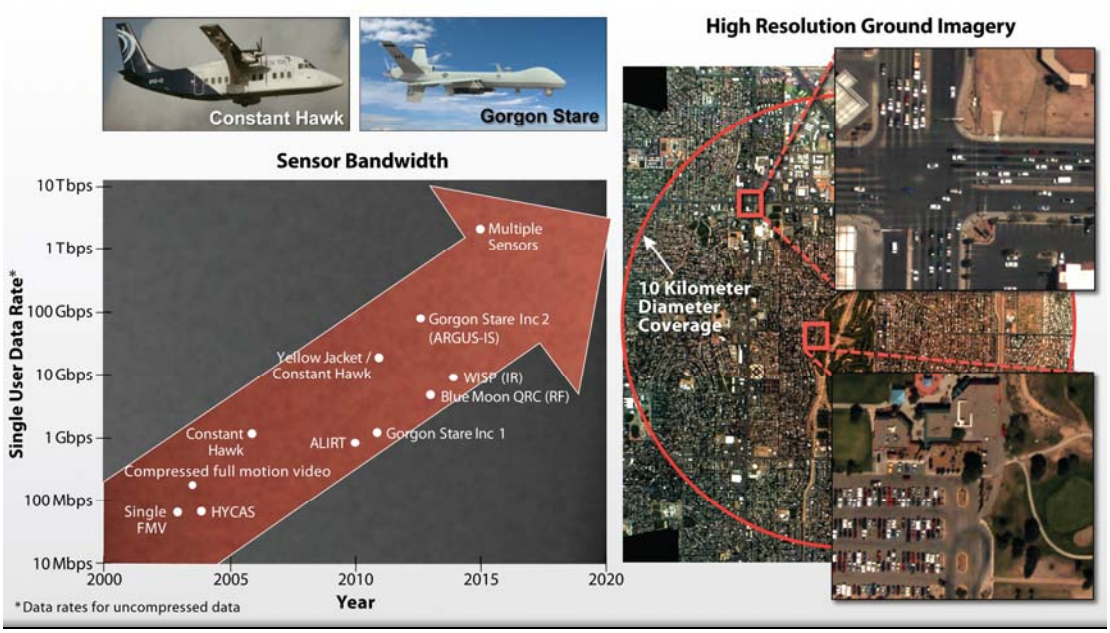


Figure 10. Exponential growth of data readouts from advanced DoD sensors

growing exponentially as a function of time. At times, the amount of data collected can exceed the DoD's ability to perform processing, exploitation, and dissemination.

One example of the new generation of airborne sensors is the Autonomous Real-Time Ground Ubiquitous Surveillance Imaging System (ARGUS-IS).[9] The ARGUS-IS video sensor produces 1.8 billion pixels per frame at 12 frames per second, providing continuous coverage of a large field of regard—up to a 100-square kilometer area. A single ARGUS-IS class sensor can produce more than a petabyte of data per day. Processing video data from sensors such as ARGUS-IS requires stitching the images from the individual cameras, rectifying the data according to known geographic references, removing the effects of motion by the airborne platform, modeling the unchanging background, detecting and tracking vehicle motion, selectively compressing the raw data, and archiving the results. This processing chain might require 100 operations per pixel.

Finding

Finding 3: Cloud computing services can scale to data centers or "warehouse-scale" computing. Elastic, warehouse-scale cloud computing is fundamentally new and can provide DoD with important new capabilities.

9. A. Heller, "Autonomous Real-Time Ground Ubiquitous Surveillance Imaging System," *Science & Technology Review* (Lawrence Livermore National Laboratory, April/May 2011). Available at time of press at http://goo.gl/w4H2D

5. Cloud Computing Security

Enhancing today's cloud computing security is emerging as an important priority for the DoD.[10] This chapter discusses needed enhancements and discusses assumptions about hardware, software, people, and physical plant that are part of the security operating picture for a cloud computing system.

Security experts evaluate real systems through careful analysis, accurate auditing, and vigorous red teaming. Most current software is poorly suited for forensics and auditing, even though these capabilities are extensively used in security analysis. As a result, most DoD software today is not amenable to accurate operational risk assessments. With current software, the defensive and offensive roles in systems are highly asymmetric, favoring the attacker who succeeds by finding a single vulnerability over the defender who must eliminate all vulnerabilities in a large system.

Despite uncertainties about security, new capabilities enabled by cloud computing could provide significant benefits to the DoD. In many cases, the DoD could achieve some cost savings when using cloud computing, even though secure cloud operations may involve additional work and expense.

With the requirements and risks in mind, three questions emerge:

- How should DoD build secure cloud environments?
- How can DoD take advantage of cloud benefits in the near future by careful risk management involving existing technology?
- What should DoD do to ensure access in the future to secure, efficient, and effective cloud technology?

5.1 Security Assessment

Security assessment provides the decision-maker with a basis for selecting a cost-effective computing system that provides sufficient assurance for completing assigned missions, despite the plausible actions of the anticipated adversaries. The assessment must be based on both the material and activity that must be protected, and the resources and access of the adversaries. A worthy long-term goal is to make all security-critical functions remotely verifiable by the user through simple technological means.

A security maxim is that the worst enemy of good security is the delay and expense caused by an insistence on perfect security. No activity is unconditionally secure against

10. A. Sternstein, "NSA Chief Endorses the Cloud for Classified Military Cyber Program," *Nextgov* (June 13, 2012). Available at time of press at http://goo.gl/bUrS5

an adversary with unlimited resources. In this light, risks and tradeoffs must be carefully examined and security decisions must be realistic.

To be effective, security must be described in terms of specific, concrete goals. None of the goals listed here can, unfortunately, be purchased off-the-shelf or simply marked off on a checklist:

- **Preserving confidentiality and integrity of data**. Confidentiality means the data should not be disclosed to unauthorized parties. The integrity requirement is concerned with ensuring that unauthorized parties cannot corrupt data.
- **Protecting the computational confidentiality and integrity of the software**. This requirement means that program execution must not be visible to adversaries. The integrity requirement means the software should operate in accordance with what its designers intended. Adversaries should not be able to tamper with a program and cause subsequent execution to produce incorrect output or undesired side effects. In some cases, it may be important to keep code confidential, because it discloses techniques and methods.
- **Ensuring resiliency, availability, reliability, and predictability of operation.** Under actual operating conditions, the target computer services must recover quickly from potential failures, operate reliably, be available when needed, execute in a predictable way, and deliver results in a predictable time frame.
- **Avoiding single points of failure in applications.** Code for many of today's applications is vulnerable to intentional or accidental modification by insiders. In one case, an unintentional incident crashed major data centers for hours because an error condition accidentally triggered a flaw in a single protocol that many data centers were running. Application diversity and formal fault analysis are promising approaches for mitigating these types of risks.
- **Retaining agility**. Security measures should not limit the rapid development of new programs or create unreasonable delay in the deployment of existing programs in other suitable environments.
- **Detecting system failures and enabling ongoing evaluation of system performance and safety**. This is sometimes called situational awareness, and it requires audit and forensics. It means detecting failures and, when they do occur, collecting information that helps assess ongoing operational security posture. Systems, processes, and skilled staff must be in place so that an intrusion will be detected quickly and its effects rapidly and thoroughly remediated.

5.1.1 Comparing cloud and conventional computing approaches

Cloud computing is not intrinsically more secure than other distributed computing approaches, but its scale and uniformity facilitate and enable the wholesale and consistent application of security practices. Secure aspects include large scale monitoring and analysis of data to detect attacks, and automated and persistent

provisioning and re-provisioning to foil intrusions. For these reasons, well-operated cloud computing facilities can exhibit better security hygiene than conventional data centers.

Cloud computing deployment can offer potential for improvements over the security found in some existing DoD systems. For many current DoD applications, moving to a well-implemented cloud computing data center would improve both the overall security and security of individual applications. On the other hand, for applications that were built with attention to security, a cloud deployment environment, in which security is not a priority, would likely decrease the security of the application.

Some threats are exacerbated by cloud computing deployments, absent further security provisioning. For example, cloud computing data centers are generally large in scale and, therefore, they present attractive targets. Insiders often have access to immense resources, and determined, well-funded adversaries could exploit even generally trustworthy insiders to compromise a cloud system. Cloud services often rely on connectivity; for many applications, a network outage—perhaps due to an attack—would render that application completely unusable. Poor development practices are also more damaging in a cloud environment, because tenants share resources, which means tenants execute in close proximity to other's programs and data. Note that these threats are not unique to cloud computing. They would be present in any large-scale DoD computing system that relies on network connectivity.

For a system to be built with today's technology—which is notoriously insecure whether in a cloud or conventional deployment—it may be more productive to compare the relative security of existing conventional and proposed cloud computing systems rather than the absolute security of either one.

5.1.2 Classifying threats

Threats refer to specific opportunities by identified adversaries to defeat security goals. It is useful when thinking of threats to separate them into the categories shown in Table 1.

5.1.3 Classifying adversaries

Adversaries are entities that may or may not be intentionally malicious, but either way an adversary is a person or group that undertakes actions that will cause one or more of the security goals to be violated. Under this definition, adversaries include well-intentioned operators who improperly configure systems, as well as outsiders who can access computer systems, and insiders who are entitled to access computer systems but in an unauthorized manner. Perhaps the most powerful adversaries are skilled and well-financed nation states that can exercise attacks ranging from exploiting software vulnerabilities to corrupting operational elements of a system either before or after deployment.

Table 1. Example Threat Classifications

Threats that apply to computing systems in general:

- Malicious insiders who take sensitive data or knowingly interfere with the proper operation of the system adversely.
- Benign insiders who accidentally leak sensitive data, improperly configure components, fail to carry out responsibilities, or commit unintentional errors in the handling or analysis of system infrastructure or information.
- Provisioned elements of the system, such as computers, storage disks, or software, that are modified to provide back-door access to an adversary.
- Provisioned components, either hardware or software, having vulnerabilities that may be exploited.
- Data that is stored unencrypted on disks and, therefore, available to anyone with physical access to the disk.
- Data may be encrypted, but underlying cryptographic keys may be managed in a way that allows them to be observed by insiders with varying degrees of difficulty.

Threats that apply to remote data centers:

- Networks may fail or experience unpredictable delays.
- Tenants must rely on the trustworthiness and competence of personnel to safeguard data—sometimes in unencrypted form—that is being stored or physically transported.

Threats that apply to computer systems with multiple tenants:

- Co-tenants and service users who pierce isolation boundaries to compromise confidentiality or integrity of the data, code, or communications of another tenant.
- Authorized service users who access the services from insecure clients.
- Co-tenants and service users who compromise availability in a cloud infrastructure, for example, by consuming too many resources.
- Access to shared infrastructure is denied.
- The failure of access control mechanisms to grant authorized access to shared resources.
- A service provider who underestimates the aggregate resource demands of all tenants and, as a consequence, under provisions resources.
- Malicious tenants and service users who damage the reputation of infrastructure components or other tenants, thus raising concerns about the integrity or reliability of data or operations.

Threats that apply to public clouds:

- Public clouds have many, many tenants, thereby increasing the threats experienced by systems with multiple tenants.
- Tenants in public clouds cannot generally control who their co-tenants are, even when co-tenancy requirements are specified.
- Tenants in a public cloud computing data center generally have little visibility into operations that might create potential or actual vulnerabilities.

- Tenants in today's public cloud computing data centers have little or no ability to review and assess the hardware and software that must be trusted for their applications to execute safely.
- Tenants have only minimal situational awareness and limited awareness of data leakage.

Each computing application will have different security requirements and must be defended against different classes of adversaries. Differences among adversaries stem from the resources, access, and knowledge they can employ in an attack, as well as the expected value from a successful attack. The decision to deploy a specific system in a cloud computing environment becomes a question of balancing risk against opportunity and cost for that specific system and anticipated adversary attacks.

Figure 11 presents a notional taxonomy of cyber adversaries. The pyramid shape is meant to convey the number of adversaries for each class in the taxonomy, showing that adversaries today mostly operate at Tiers I and II—so-called "script kiddies"—using malicious code developed by others. The primary defense against these types of attacks is improved computer hygiene, meaning user best practices for passwords, firewalls, and links. Tier II actors have some ability to develop their own malicious code. Their actions may be directed at achieving specific business or political objectives, such as the theft of information or alteration of financial data. These lower-tier actors can be effective because sophisticated tools and techniques developed and exposed by others are widely available.

Tier III and IV actors are characterized by their abilities to employ a broad range of technical capabilities to penetrate cyber systems. The distinction between Tiers III

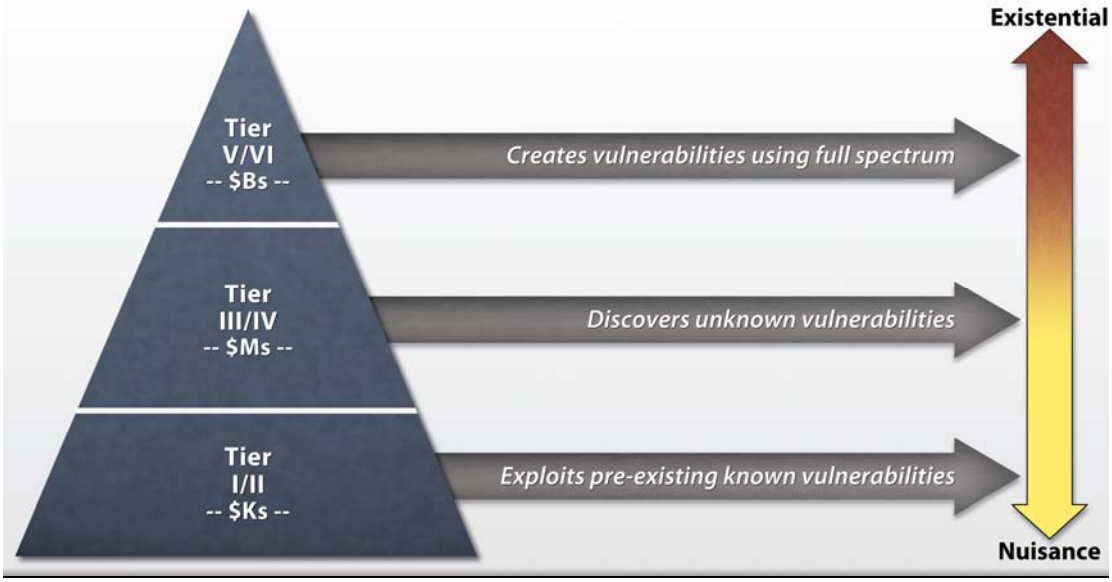


Figure 11. A notional cyber threat taxonomy

and IV is scale—Tier IV is characterized by larger, well-organized teams, either state or criminal.

Tiers V and VI comprise actors who are able to go beyond introducing malicious software via internet access. They are able to create new vulnerabilities in otherwise well-protected systems. Tier V actors can insert corrupt software or hardware at various points during a system's lifecycle for later exploit, including supply chain attacks. Tier VI organizations can employ full spectrum techniques, such as persons engaged in bribery and blackmail, as well as proximate physical or electronic means to gain system penetration. Tier VI adversaries may conduct many operations concurrently and benefit from coordinated attacks.

5.1.4 Security decision parameters

The operational environment of a data center, including a cloud data center, can be described by the following considerations:

- **Physical siting.** Where is the cloud data center located, and what are its physical security characteristics? What provisions have been made for availability of reliable power; local supply and support; proximity to high capacity, reliable network fibers; and protective measures for the equipment and surrounding area against kinetic attacks and electromagnetic eavesdropping?
- **Operator affiliation.** Who runs the data center? DoD personnel? Coalition partners? Conventional DoD contractors? Commercial providers?
- **Tenant population.** Does the data center only serve activities associated with a single mission? Diverse missions within DoD and intelligence community? Only U.S. Government applications? Commercial applications (which might or might not involve affiliations with foreign powers)?
- **Hardware.** What is the physical plant and hardware architecture? Who manufactures hardware components? Who assembles and integrates hardware components? How are hardware components tested, qualified, and monitored for secure operation?
- **Software.** Who develops and tests data center infrastructure software (resource allocation, provisioning, hypervisor, operating system)? Who supplies application components? Who provides software components? How are software components tested, qualified, and monitored for secure operation?
- **Data.** What is the value of the data? What are the consequences if there is a loss of data integrity?
- **Architecture.** Is forensic information that that can provide insight into safety and performance of the systems collected regularly? Can redundant data centers provide resiliency and consistent performance? Can operational data that is critical to carry out mission critical activities be staged or cached for access under network-

disadvantaged circumstances? Is there enough information to make a well-informed forward risk assessment?

- **Clients.** Who can access cloud resources? What are the client machine security requirements?

Making computing deployment decisions requires careful consideration of these factors. In each case, these decision parameters must not only be verified at design and installation, but must be meaningfully and effectively verified during ongoing operation.

5.2 Data Center Security

Security goals, as noted above, will depend on the application, and the means of enforcement will depend on the threat, so there can be no single answer for the characteristics of a data center used by DoD. The following lists some considerations for DoD cloud computing data center security.

5.2.1 Physical security

Large commercial cloud computing data centers observe excellent physical security measures, and the amortized cost of good physical security in a large cloud computing data center is relatively modest. Data centers live and breathe on connectivity, so a cloud data center's management of network activity is usually quite good. They are often located far away from populated areas, and few, well-authenticated people are granted admittance. Good perimeter security is in place, including intrusion detection, electronic physical access control, and video surveillance.

To host very high security applications at a cloud computing data center, the DoD would undoubtedly take direct military control of a physical plant and add physical security measures. Fortunately, protection and risk assessment of kinetic attacks as well as physical security is comfortably within DoD competencies. Most commercial data center facilities do not protect as rigorously against snooping via electronic emanations as is customary in high-value DoD location, nor do they protect against sophisticated kinetic attacks.

The DoD could also employ multiple data centers to provide resilience against attacks. Balancing the economics of much stronger physical protection for a small number of sites versus the protection derived from having many data center sites requires careful, mission-specific analysis.

5.2.2 Personnel security

The most important element in ensuring information security is having a highly skilled, adequately resourced team, backed by management that takes security seriously. Personnel security practice is usually good at commercial data centers, and

they are staffed by well-trained professionals. The DoD will almost certainly employ even more stringent personnel security practices in their data centers. The DoD would likely also develop and deploy additional technology to reduce access to confidential data by data center personnel, as well as to remotely verify operations. As with physical security, requirements for personnel security will vary based on classification of what is being protected.

In addition to a core security team, red teams are needed, made up of experts who try to attack systems in an effort to determine risks. Modern computing systems are too complex for direct analysis alone to determine all real-world risks. Finally, incident teams are needed that are skilled in mitigating ongoing attacks quickly and identifying root causes.

For application software, DoD must employ individuals skilled in developing scaled and secure cloud software so that applications will be written to the same rigorous security standards as systems software. Many cloud applications will be developed using a small number of common frameworks and libraries, so investments in secure coding of these building blocks can be leveraged.

5.3 Secure Cloud Computing Software

Software is possibly the most critical security element for cloud computing deployments. The critical components include image management and resource allocation software, hypervisor and management partition software at node instances, and audit and forensic software.

Conventional DoD software, acquired over many years and written without specific concern for resisting cyberattacks, is often vulnerable to attack by relatively unsophisticated adversaries. Even if properly updated, such software is likely to have easily exploited vulnerabilities. Further, its complexity makes proper configuration of most software difficult and time consuming—even for experts—providing yet another avenue of attack for an adversary. Software that is employed in a DoD cloud computing data center is likely to come from the same sources as software used by commercial data centers, and the high costs of software development suggest this is unlikely to change.

Today, these may be a combination of commercial-off-the-shelf (COTS), open source, and custom software.

- **Commercial software** such as VMware and Windows are commonly used in cloud computing data centers. It is important to realize that all such commercial programs have proprietary source code and are opaque to security specialists, both in terms of source code and in their manufacturing provenance.
- **Open source software** is also likely to be used. These programs are maintained by an international community and under the condition that all source code is published. Open source software offers the advantage of transparency to facilitate

security assessments, but currently available offerings might not have all of the features required by DoD.

- **Custom operating software** may also be developed specifically for DoD cloud computing data centers. However, custom software has proven to be more expensive than either commercial or open-source and can lead to very long deployment delays.

Whether commercial software or open source software is being run, DoD must carefully examine these software components to ensure that they can provide sufficient operational security assurance. It is critical that DoD develop a well-conceived process for analyzing and gaining assurance when using open source and commercial proprietary software. The DoD must have some basis to believe that a chosen operating system is trustworthy.

The task force found that it is unlikely to be economically feasible or even advisable for the DoD to write or buy all of the software it runs in cloud computing data centers. The task force found that the best strategy will likely leverage existing commercial and open source development. If such a hybrid case were pursued, DoD would only need to develop software when commercial and open source communities are unlikely to offer the needed functionality in a timely manner.

If DoD-authored improvements are contributed to the open source community, the normal community processes will mean that the improvements remain usable as the software evolves, and the DoD-authored improvements could be adopted by commercial cloud providers, improving their security. The DoD, in fact, already has taken important steps by making contributions, albeit somewhat inconsistent and uncoordinated, to review and assess the security of open source systems. Of course, certain sensitive DoD software might have to be developed by the DoD itself, without disclosure. This route should not be taken casually, because the DoD then incurs all subsequent costs of maintaining this software.

Reasons that the DoD might consider partnering with public cloud service partners to develop cloud computing software include:

- The DoD can leverage software capability and security capabilities developed by the open source community, including existing schemes for isolation of tenants, key management, and strong authentication and authorization.
- Even in a DoD cloud computing data center, the best commercial security practices developed by large scale commercial cloud service providers with deep experience will contribute to running a more secure system. That is, as compared to less experienced contractors with experience limited to building small-scale, private clouds. In fact, large commercial cloud providers face a threat profile that has many of the same elements as the threat the DoD faces. For example, insider threats are critical in public cloud data center operations just as they are for the DoD.

5.4 Secure Cloud Computing Hardware

5.4.1 Hardware supply chain security

Commercial cloud providers and the DoD generally buy COTS hardware from commercial vendors, although some large commercial cloud providers increasingly build custom hardware made from COTS components. Thus, commercial providers and the DoD are today at risk to supply chain attacks by well-funded organizations (*i.e.*, tier V and VI adversaries). A supply chain attack can be perpetrated at any point in a production—design, manufacturing, testing, distribution, or installation—and this life-cycle exists concurrently for components, boards, subsystems, and entire systems.

Some hardware components, like central processing units, can be economically designed and manufactured by only a very few vendors, who must manufacture at scale to recover development costs. While DoD has relied on trusted foundries for the production of certain sensitive parts with minimal exposure to supply chain risks, this strategy is not feasible for all of the parts that comprise a cloud data center.[11] Today, many critical components in computers are manufactured only outside the United States and some are especially vulnerable to supply chain attacks.

Testing and procurement strategies will be needed to mitigate supply-chain risks.[12,13] Among other strategies, a DoD cloud data center design could limit the number of components in each design required for secure operation, could vary the sources of components it procures, or could attempt to be secretive about procurement sources. By far, the best solutions—in analogy with software—are designs that focus on reducing the trusted computing base for hardware.

5.4.2 Client hardware security

Client hardware security is just as essential for cloud computing as is the security of cloud servers. In fact, many famous cloud computing security failures were caused by compromised client machines. Client machines operated by authorized persons can initiate critical operations in cloud computing data centers and transmit sensitive data to other clients. For these reasons, client hardware and client software must incorporate the same level of protection as other cloud components.

For example, a client must be able to quickly verify the identity and security posture of a server running in a cloud computing data center, just as that server must be able to

11. Trusted Foundry Program. Available at press time at http://www.trustedfoundryprogram.org/
12. Defense Science Board, "High Performance Microchip Supply" (2005). Available at press time at http://www.acq.osd.mil/dsb/reports/ADA435563.pdf
13. W.J. Lynn, III, "Defending a new domain," *Foreign Affairs* (September/October, 2010). See also, "The Pentagon's Cyberstrategy, one year later," *Foreign Affairs* (September, 2011).

authenticate the client system. Even thin client components must still be capable of authenticating to the cloud and being authenticated by the cloud.

5.4.3 Building secure cloud computing environments

In addition to physical security, a secure cloud computing data center must run reliable components to ensure isolation and facilitate remote authentication of software components. It must also support authentication of clients and servers, authorization for access to software and data, verified audit of data access and resource usage, secure and flexible key management, safe and effective resource management, and provisioning—of images, network configuration and partitioning and storage management—that can be verified under circumstances that support economical and agile data center operations.

Two interconnected capabilities can be deployed to provide confidentiality and integrity of both data and software. The first is an ability to authenticate that a critical software component is one that appears on a list of well-understood, trusted implementations that have been isolated from other software. The second is a basis to have assurance that an application was loaded and executed without tampering. Because cloud computing resources will be used freely by developers, all software executed by the cloud cannot be expected to come from a short "white list" denoting approved components. Rather, system software must automatically ensure that the developer's code, even though it might exhibit bugs, does not have the power to access or corrupt production software or data in the system.

Assurance that an application was loaded and executed without tampering can be achieved if the system can store cryptographic keys in such a way that these keys can be used only by some pre-specified uncorrupted software element executing in an environment that ensures it is isolated from other tenants. The isolation environment and key management scheme must be resilient to insider attack and must be remotely verifiable.

Modern processors often have hardware that can be used to enable these two interconnected capabilities. These processors are typically not deployed in cloud data centers despite the low incremental hardware cost of technologies. These hardware attestation technologies are known as hardware security modules, or Trusted Platform Modules. The DoD has mandated the use of this technology in most client computers.

Where hardware attestation is available on cloud computing processors, the basic approach for secure isolation and key management is to employ encryption and cryptographic integrity verification for all data in transmission and storage, coupled with secure access control that enables decryption only within those isolated software components that are run by authorized users under policy control enforced by cryptographically protected credentials issued by data owners. These same isolation

and key management facilities then also provide an environment for implementing secure forensics and audit.

5.5 Secure Data Center Operations

Many security considerations apply whether the DoD is using dedicated data centers for specific mission, shared defense clouds, vendor or coalition operated clouds or, indeed, public clouds. Commercial cloud data centers must have efficient means for deploying applications, which lead to timely installation of updates—something that is particularly valuable for patches that mitigate against newly discovered vulnerabilities. Centralization in cloud data centers enables better and more comprehensive forensics, especially for detecting subtle attacks that are less visible in a small environment. Moreover, the naturally dynamic nature of resource assignment in a cloud data center makes persistent attacks on critical functions much more expensive.

Security requirements and gaps in current technology, however, will likely force the DoD in some cases to operate their own private clouds under operational conditions similar to those of existing, non-cloud systems. Regardless, the following operating guidance is offered to any cloud computing data center implementation.

5.5.1 Data sharing, confidentiality, and integrity

Data centers employ cryptographic mechanisms and operational procedures to ensure confidentiality and integrity of data and program code. DoD, in conjunction with NIST, has done an admirable job in contributing to interoperable algorithms, protocols, and format standards to support cryptography. Yet data is still often stored in unencrypted form on data center disks, or it is transmitted within or between data centers with no encryption.

Current data centers also often employ poor key management. For example, cryptographic keys that form the basis for protection may be stored in a way that makes them visible to insiders. Even when it has been encrypted, data at rest and data in transit can remains vulnerable to attacks by operators, technicians, and even other tenants. Inadequate key management is not unique to cloud computing, and these vulnerabilities are not always easy to remedy without impacting performance. Ultimately, the use of encryption creates trade-offs between confidentiality and legitimate monitoring for abuse or other kinds of situational awareness.

Robust, flexible, and verifiable authentication and authorization frameworks to enable secure data sharing among authorized applications have been developed, but are not widely employed in today's cloud computing systems. A good example is provided by the Accumulo system and the associated authentication, authorization, and access control framework. Strong authentication of software components in a data center and end-clients is crucial for ensuring that the principals requesting access to data are legitimate.

To improve verifiability and gain protection from insider attacks, operators should require appropriate processes for making hardware changes and to deploy system initialization routines that do not depend on the competency or altruism of insiders. The goal is to allow safe key management and remote verification of operation. Keys used by programs should be available only to user programs that have been authenticated and whose execution is isolated from potentially malicious programs. In most traditional and cloud computing systems, this isolation is implemented by the operating system software and database-level file and process access lists, which are difficult or impossible to protect from insider tampering. However, such complete trust in software is not necessary—hardware such as TPMs exists to strengthen access controls. Without it, confidentiality and integrity of programs and data is difficult to assure. The DoD could lead the way in fostering the widespread availability of cloud data centers that incorporate these features.

5.5.2 Forensics, monitoring, and intrusion detection

Many cloud architectures provide data collection and analysis functionality to support detection, analysis, and rapid remediation of attacks, although such capabilities vary widely among providers. Indeed, one possible security benefit of cloud architecture is the ability to gather such comprehensive situational data.

Cloud computing data center providers generally are not, however, economically motivated to provide the data to their tenants, nor to release the information, because it could adversely affect their reputation or reveal information to attackers attempting to avoid detection. Discretion is sometimes required during assessment and investigation to avoid alerting adversaries; other times, complex equities must be balanced. Some providers and vendors have very good incident response teams; others share information under *ad hoc*, personal arrangements and publish vulnerabilities only after they are fixed. Sharing this information more rapidly allows others to learn about threat signatures and potential vulnerabilities and make repairs. Silence about attacks can mean that risk is poorly understood by those responsible for overall mission results.

5.5.3 Red teaming and incident response

Vulnerabilities often are the result of subtle design and implementation problems, and historically the most effective mechanism to assess the security of a system is red teaming. Unless a system is regularly subjected to red teaming, there is no way to understand the security risks. Because systems tend to evolve over time (as software components are updated, for example), a system's vulnerabilities will also change over time. A list of problems identified in prior red team attacks might say little about a current system. The need for recurring security assessment is a major issue in cloud data centers where operations are opaque and vulnerabilities and successful intrusions are sometimes either undetected or unreported.

Figure 12 depicts a block diagram for one red and blue teaming model. Effective red teaming includes reviewing the current threats, assessing the current system performance, and extrapolating to estimate future threats and system vulnerabilities. Blue teaming investigates techniques for reducing the impact of current and future threats and vulnerabilities. The results of blue teaming can then set the direction for future cyber system developments, new tactics, techniques and procedures, and improved technology investments. The benefits of red and blue teaming are improved significantly when soldiers, sailors, airmen, and marines with recent operational experience participate.

Commercial, government, and DoD cloud providers currently have widely varying expertise in red teaming, incident response, and analysis. There are pockets of deep knowledge in industry, the intelligence community, and in the DoD. Improved and coordinated efforts to systematically build a basic understanding, principles, tools, and procedures for evaluating security of cloud systems would benefit both DoD and commercial providers.

5.5.4 Operating under degraded conditions

Cyber attacks, system failures, and human error all can cause operational cyber systems to exhibit degraded performance. Cloud computing systems are not immune to most of these disruptions and, in fact, in some cases may be more affected. While commercial cloud computing providers have developed effective detection and mitigation techniques for degradations in communications, network outages can still result in complete service disruption for many commercial cloud computing services. Such disruption could be catastrophic for some DoD applications.

Before any cyber system becomes central to DoD military operations, it is essential that the operational community understand the implications of degradation of that

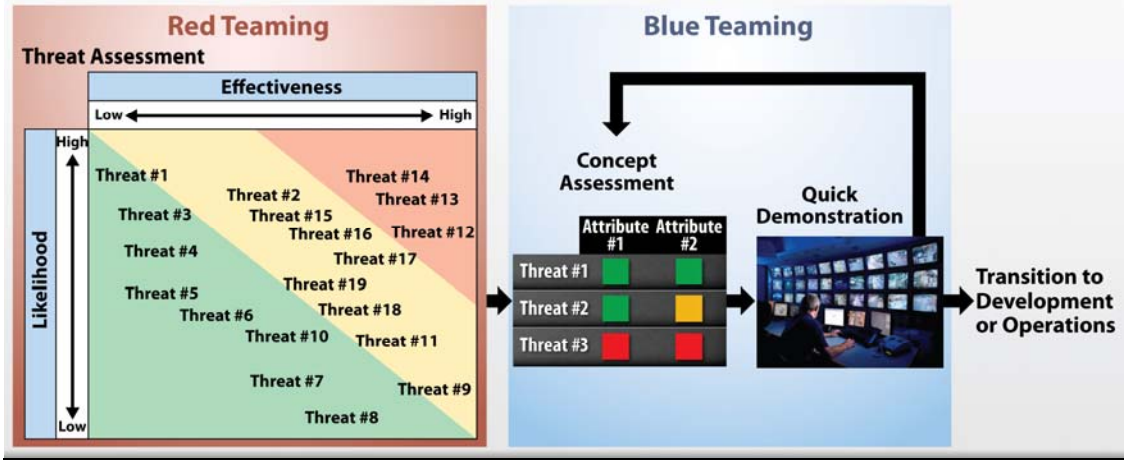


Figure 12. A red-blue teaming model for enhanced assessment of advanced threats and possible mitigations

system. The goal is to use realistic training and exercising to develop tactics, techniques, and procedures (TTPs) to ensure adequate mission assurance under the degraded conditions. These TTPs may include such changes as more distributed cloud computing resources or more redundant cyber and communication systems.

The 2010 Defense Science Board Summer Study on Enhancing Adaptability of our Military Forces investigated the prevalence of military training and operational exercises with degraded cyber and communication system conditions.[14] The study found that most military operational exercises did not include scenarios with realistic examples of degraded cyber system conditions. At the time of the study, the U.S. Pacific Command's *Terminal Fury* series of exercises was a notable exception because of its incorporation of an increasingly complex set of cyber threats.

The DoD conducted a number of support exercises during fiscal year 2012 that included degraded operations. Several commands participated, including the U.S. Strategic Command in *Global Thunder and Global Lightning*, the U.S. Cyber Command in *Cyber Flag*, the U.S. Pacific Command in a follow-on to *Terminal Fury*, the U.S. European Command in *Austere Challenge*, and the U.S. Transportation Command in *Turbo Challenge*. Although incorporating degraded cyber operational conditions improved these exercises, stronger red and blue teaming will be needed to improve the realism of degraded cyber systems in training and operational exercises.

To effectively design and deploy cloud computing services, DoD will need to train and exercise under degraded conditions and to incorporate mitigating actions in those activities. Effective mitigations for communications outages might include the use of local data centers that have cached mission-critical information, the use of alternative network provisioning, and the fielding of thick clients that can provide local, albeit degraded, support for mission critical capabilities, or restricting some applications from deployment on cloud computing systems.

5.5.5 Operating in partnership with public cloud computing providers

Many of the risks discussed in this chapter can be reduced or eliminated by using in-sourced or out-sourced private cloud computing facilities, or non-cloud local computing resources.

The vast information resources available in public clouds, however, is also important to the DoD. In addition, some DoD missions will appropriately operate in the public cloud. Working in partnership with public cloud computing providers also offers some operational advantages. As the U.S. government encourages more secure public cloud computing infrastructures, these facilities may become an important option for the DoD in emergencies. As well, large public clouds may have more secure software stacks, and

14. Defense Science Board, “Enhancing Adaptability of our Military Forces” (2011). Available at time of press at http://www.dtic.mil/docs/citations/ADA536755

larger and better-trained security teams than newly constructed private clouds. Further, careful development of critical software that takes economic advantage of shared benefit in development may make critical cloud components more naturally resilient to hardware and software vulnerabilities.

If a public cloud data center implemented the best practices recommended here for DoD private facilities, then public cloud computing would not generally be any less safe than in-house operations. These practices may include externally verifiable key management, verifiable reliable access control over tenant-owned-but-shared resources, adequate isolation of execution, verification of critical data center operations, and reliable authentication of client systems to ensure their safety and identity. In addition, the naturally dynamic nature of resource assignment in a large public cloud computing data center might make persistent attacks on critical functions much more expensive.

Today, commercial data centers do not generally disclose hardware infrastructure in sufficient detail for a potential tenant to reasonably assess risk. To work in partnership, therefore, the DoD must be prepared to negotiate contractual provisions with commercial data centers to ensure the integrity of the hardware infrastructure.

Findings

Finding 4: Cloud computing is not intrinsically more secure than other distributed computing approaches, but its scale and uniformity facilitate and enable the wholesale and consistent application of security practices. Secure aspects include large scale monitoring and analysis of data to detect attacks, and automated and persistent provisioning and re-provisioning to foil intrusions. For these reasons, well-operated cloud computing facilities can exhibit better security hygiene than conventional data centers. However, the centralization of resources in a huge data center also encourages more determined attacks, especially on critical components broadly affecting security, just as in conventional systems, attacks are observed to focus on central directories.

Finding 5: The scale of cloud computing enables the analysis of packet and log data that provides new capabilities for event forensics and real-time detection of malicious behavior. The ability to manage very large, diverse datasets facilitates a data-centric security model, in which users are authorized to work with data based upon their security credentials and the security markings on the data, rather than the conventional enclave-centric security model in which users are provided access to an enclave and can access all the data in the enclave.

Finding 6: No cloud computing deployment model is uniformly suitable for hosting all DoD applications. In general, sensitive, classified, and time-critical DoD applications should be deployed only in private clouds or conventional non-cloud approaches.

Finding 7: The case for transitioning a DoD application to a cloud computing data center must include a security assessment detailing the impact of the transition.

Whether security will be improved by transitioning an application to a cloud computing data center will depend on factors specific to the application, to the cloud computing data center, and to the transition process.

Finding 8: The DoD has not established effective plans for cloud computing facility backup, or for dealing with any anticipated degradation of communications between the cloud computing facilities and the end user.

6. The Economics of Cloud Computing

The movement to cloud computing is an oft-cited contemporary strategy for achieving efficiencies within information system enterprises. In the past three years, dramatic price reductions have been offered by commercial cloud computing service providers. As depicted in Figure 13, Amazon Web Services (AWS) lowered prices consistently over three years. Moreover, the commoditization of infrastructure resources by commercial cloud service providers has created a new marketplace with ever decreasing price floors and a trend for increasing service fungibility.[15,16]

Significant differences in management models (*i.e.*, private, in-sourced private, out-sourced private, and public); service models (*i.e.*, software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS)); and support for security and compliance introduce architecture and implementation differences that impact investment and operating costs of cloud data centers. Hence, when examining the economics of cloud computing, decision-makers must be wary of unsubstantiated estimates for return on investment. Consequently, a careful analysis that incorporates expected cost and expected security is essential in order to ensure savings and function for DoD mission.

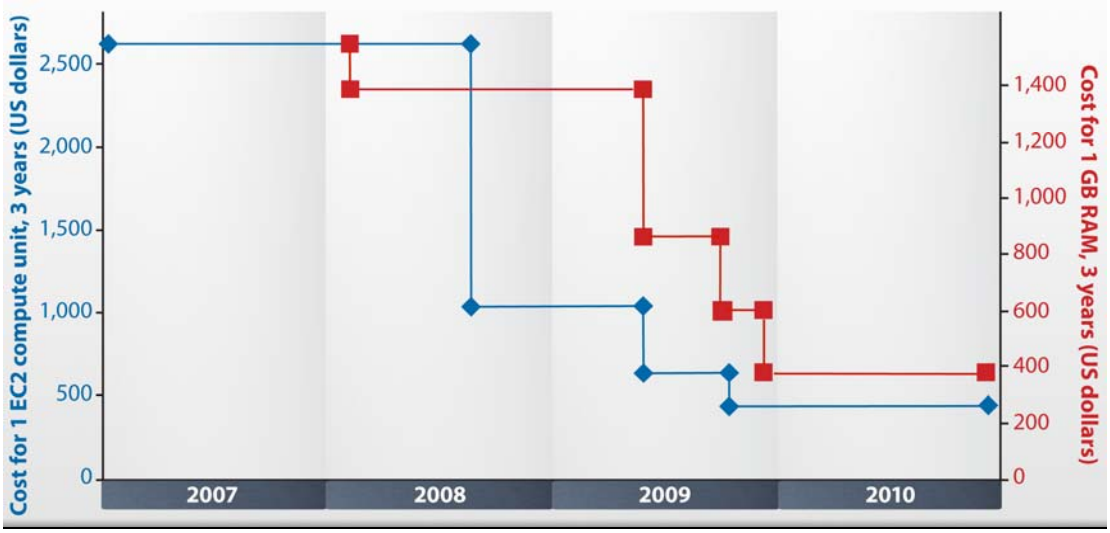


Figure 13. Price in U.S. dollars for a 3-year rental of Amazon Web Services resources, 2007–2010

15. Amazon Web Services blog, "Dropping Prices Again—EC2, RDS, EMR and ElastiCache" (March 5, 2012). Available at time of press at http://goo.gl/lGZHs
16. B. Howe, "Cloud Economics: Visualizing AWS Prices over Time" (eScience Institute, November 28, 2010). Available at time of press at http://goo.gl/ASGbS

6.1 Cloud Service Economic Drivers

Online pricing of commodity cloud providers suggests lower prices for consumers. How service providers are achieving these cost reductions is discussed here, as well as the changes to traditional data center practices, technology, and information technology culture that have resulted in these savings.

6.1.1 Improving the administrator to server ratio

Changing the number of servers that a system-administrator can effectively support has significantly affected costs. In traditional data centers, this ratio has averaged around 20 to 30 machines per administrator, depending on the types and complexity of machines, the commonality of underlying system software, and the similarity of machine configurations.

This machine-to-operator ratio improves for cloud computing architectures. The use of a small number of approved baseline machine images simplifies configuration management and the operation of the machines.

The ratio in the average enterprise of administrators to VM is about one to 77.[17] For most organizations this effectively halves the number of operations staff required. Trade press reports routinely suggest even higher numbers for public commodity cloud computing providers.

6.1.2 Increased automation

Many commodity cloud service providers are taking advantage of substantially increased automation to lower labor costs for operations staff. Automation in a cloud data center can supplant many labor-based data center processes. For example, provisioning activities for cloud computing resources is typically fully automated and made directly available to the consumer. This saves on center operations staff time.

Some cloud data centers include automatic support for scaling the numbers of processors in response to system demand, recovery of failed application components, automated metering and billing, and automated backup of server states and images.[18] All of this frees the operations staff to focus on other concerns.

17. Enterprise Management Associates, “Best Practices in Virtual Systems Management: Virtualization Metrics and Recommendations for Enterprises,” (2009, page 2). Available at time of press at http://goo.gl/N8xEV
18. As an example, see a commercial offering for cloud automation solutions from enstratus. Available at time of press at http://goo.gl/biAx0

6.1.3 Leveraging virtualization

With server virtualization, multiple VMs, each running their own, perhaps different, operating system, can share the same underlying hardware within separate, isolated partitions. The cost savings when applications use virtualized resources may include:

- **Reduction in data center rack space**. If the applications are lightly or occasionally used but need to be available to users continuously, then sharing the resources of the underlying hardware can reduce the number of servers needed to host a given set of applications. A recent survey of 346 CEOs found cloud computing adoption is widespread. The primary reason given for adoption was reducing total cost of ownership.[19] In one example cited, the equivalent of thirty traditional servers could be placed in one rack-mounted computer because they have low processor utilization.
- **Reduction in total cooling and power requirements**. A reduction in the underlying number of servers also reduces the air conditioning (HVAC) required for cooling, the backup batteries required, and the kilowatts (kWs) required for powering the infrastructure. This savings includes one-time costs for building the HVAC and power support structures, and the continuing operational costs.
- **Supports straightforward continuity of operations**. VMs can be saved off-site and quickly re-started if needed for continuity of operations purposes. Because the entire VM, containing a specific instance of an operating system at any patch level with any combination of applications, can be saved as an image, the installation can be quick and low-risk. This can be a cost-effective approach to ensuring availability versus active-active failover or multi-site clustering for some applications.

6.1.4 More effective power

Electric power impacts cloud computing basic economics in at least two significant ways. First, the more efficient use of processing resources and associated HVAC reductions, due to virtualization and better utilization of hardware, changes the total amount of power required for the same computing. Second, many service providers move their data centers to take advantage of lower power costs prices at certain locations.[20] Figure 14 shows how the cost of power varies across the United States—with numerous locations where the cost of power is relatively low.

Moving data centers to locations where electrical power is less expensive is a very effective way to keep the costs of cloud computing low. Further, moving to locations where colder air is abundant and can be used to augment cooling and reduce HVAC costs can also have a considerable impact on the bottom line.

19. J. McKendrick, "Cloud Providers Pitch Cost Savings, But Enterprises Want More: Survey," *Forbes* (August 16, 2012). Available at time of press at http://goo.gl/QVkxV
20. Energy Information Agency, "Electricity: Wholesale Market Data." Available at time of press at http://www.eia.gov/electricity/wholesale/index.cfm

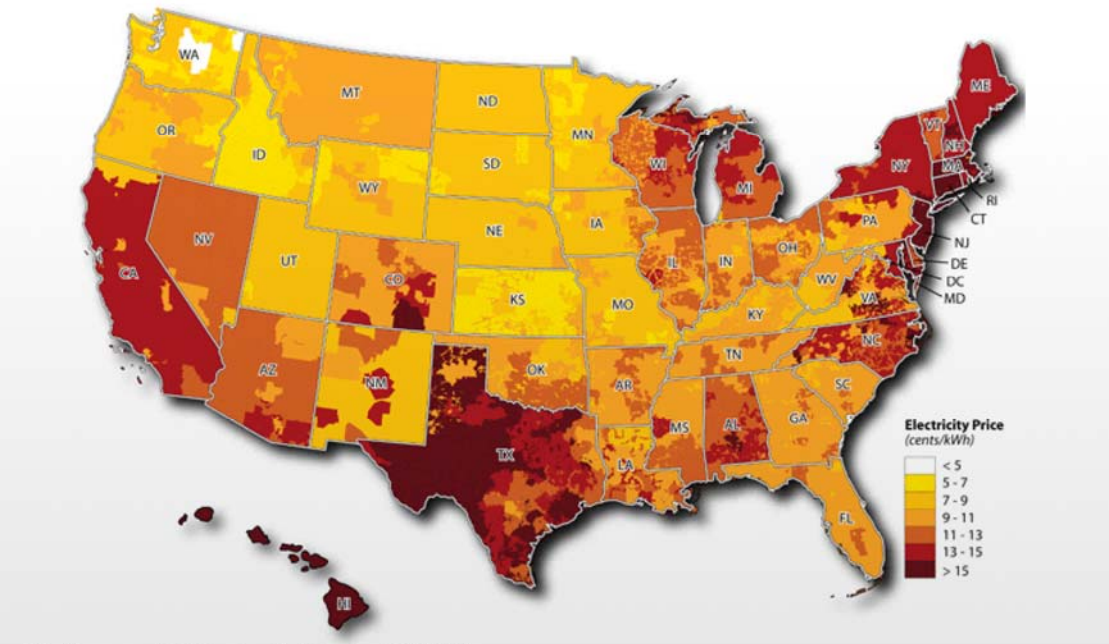


Figure 14. Map of average U.S. residential electricity price by utility service territory

6.2 Business Case Considerations for Cloud Service Use

In deciding to deploy applications into cloud computing data centers, the size of the investment, performance, risks, and investment characteristics (*i.e.*, capital investment versus operating expense, fixed versus variable costs) must all be factored into the business decision model.

Transition to cloud-based services must consider initial investment costs as well as recurring costs. Moving applications to clouds may require licensing and maintaining virtualization software. The use of virtualization may require new security software, identity management software, and management software for provisioning and backups. Data migration, integration, and testing costs associated with moving applications to a virtualized environment must also be incorporated in the cost model. Depending on the architecture of legacy applications currently deployed, there may also be porting costs.

Figure 15 shows a notional graph of the investment over time required to move an application to cloud computing. Some initial investment may be required initially, but over time the overall costs should be expected to decline.

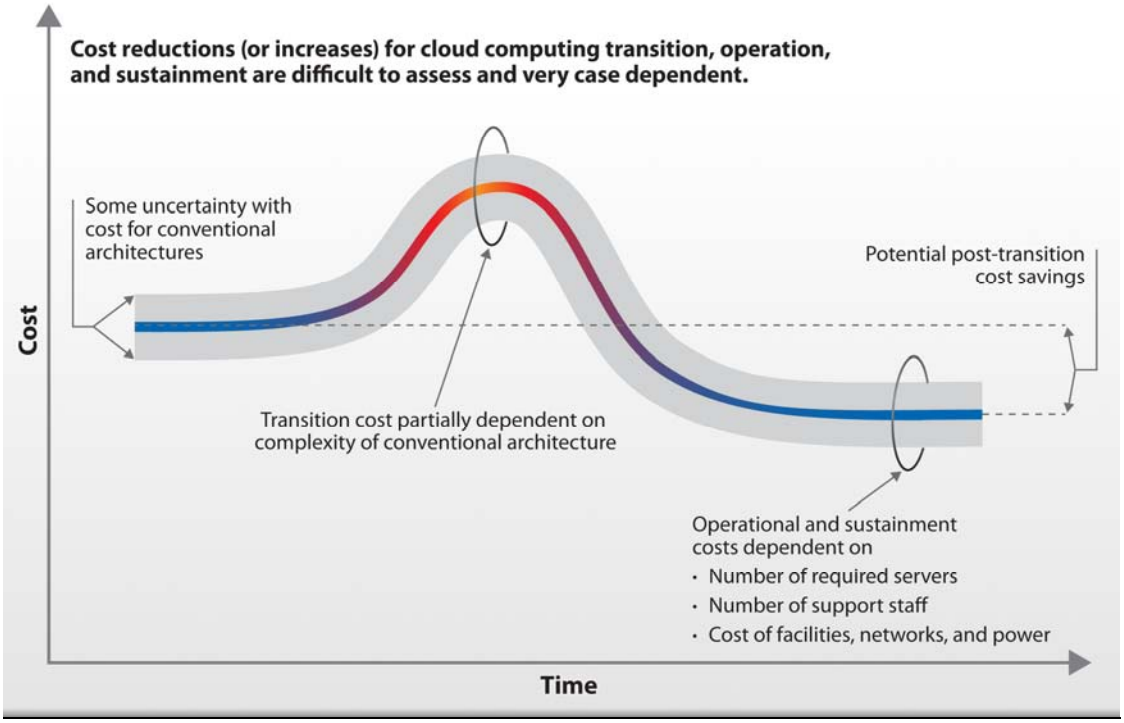


Figure 15. Notional cost over time to implement cloud computing

As mentioned, users generally access cloud computing applications over browsers. This may impact network loads and, in turn, increase costs if upgrades to networks are needed to meet agreed-upon service level agreements (SLAs).

6.3 Service Level Agreements

SLAs affect in-sourced private, out-sourced private, and public cloud computing. Well-understood SLAs are essential for mission success, and they can directly drive costs for the service provider. When SLAs are negotiable, contractual terms for SLAs can include characteristics such as:

- response time
- hours of operation
- service availability
- expected throughput and utilization ranges
- maximum permitted down-time
- performance measurement and reporting requirements
- performance-based pricing
- problem resolution thresholds
- problem escalation and priorities

When possible, an SLA should specify steps the consumer can take when service is not meeting terms specified in the SLA. These remediation steps should include points-of-contact, contact information, and escalation procedures. The time-to-resolve performance should be specified in the contract based upon the severity of the problem.

It is useful to be precise in the definition of metrics, and specify when and where they will be collected. For example, network performance metrics could have different values when measured from the consumer or provider, due to the propagation delay of the network. Performance metrics should measure characteristics under the control of the vendor or they will be unenforceable. Finally, the SLA should describe a mutual management process for the service levels, including periodic reporting requirements and meetings for management assessments.

6.4 Cloud Computing Case Studies

Case studies can be helpful for understanding the business case and potential cost savings associated with deploying or re-deploying an application to a cloud computing data center. The case studies cited here indicate significant cost savings.[21]

6.4.1 Cloud Providers

- The U.S. Air Force's 45th Space Wing at Patrick Air Force Base estimates that they save $180,000 annually through their use of virtualization. They found that even at peak load times, very few of their servers were running at more than 5 to 6 percent of load according to Glenn Exline, manager of enterprise networks at Computer Sciences Raytheon, which supports the 45th Space Wing.[22] To improve utilization, and lower hardware and energy costs, the Wing reduced 60 physical servers to four running a VMware virtualization solution. Here, the elements of cost savings were: $104,000 in hardware costs, $30,000 in power to cool what used to be 60 file servers, $28,000 in maintenance costs, and $18,000 in other expenses.[23]
- Over a four-year period, the Department of Energy Los Alamos National Laboratory has removed 100 physical servers and replaced them with 13 servers running hundreds of virtual machines, resulting in a cost savings of $1.4 million.[24] Recently they launched private cloud capabilities that allow employees to request and provision Windows, Linux, or Sun virtual server environments through a self-

21. Federal Chief Information Officer, "The State of Public Sector Cloud Computing" (May 20, 2010). Available at time of press at http://goo.gl/alOzh
22. R. Yasin, "Virtualization Takes Wing at Patrick Air Force Base," *GCN* (April 1, 2010). Available at time of press at http://goo.gl/NAm94
23. D.M. West, "Saving Money Through Cloud Computing" (Brookings Institute, April 7, 2010, page 9). Available at time of press at http://goo.gl/ZJH3y
24. R. Yasin, "Los Alamos Lab Launches Private Cloud," *Federal Computer Week* (September 8, 2010). Available at time of press at http://goo.gl/4zkMc

service portal. They've implemented management and chargeback capabilities as well. Chargeback is also important because users have an idea that virtual servers are free, comments Anil Karmel, solutions architect at Los Alamos' Engineering Division in an interview with *Federal Computer Week*.

6.4.2 Cloud Consumers

- The Recovery, Accountability, and Transparency Board moved Recovery.gov to Amazon's Elastic Compute Cloud (EC2) for a projected savings of $334,000 in 2010 and $420,000 in 2011.[25] Recovery.gov is the official website for Recovery Act data and EC2 is a commercial, publicly available IaaS cloud offering. Additionally, there is additional value to the Board in the capabilities EC2 provides, including uptime and backup capabilities.
- The General Services Administration (GSA) moved 17,000 users to cloud-based Google Apps for Government in 2011 to provide email and collaboration capabilities.[26] Martha Johnson, GSA Administrator, said, "We expect that using a cloud-based system will reduce email operation costs by 50 percent over the next five years and save more than $15.2 million for the agency in that time. A large part of these savings will come from a decrease in the number of costly data centers requiring hardware, software licenses, maintenance, and contractor support."[27] Google Apps for Government has received an authority to operate at the FISMA-Moderate level and includes more secure versions of cloud tools like Google Docs, Google Sites and Google Reader.[28]

Finding

Finding 9: Potential cost reductions, or increases incurred during the transition to and sustainment of cloud computing infrastructure, depend on the specifics of the implementation. Potential cost-reduction factors include a higher utilization of servers, lower professional support staff needs, economies of scale for the physical facility, and the flexibility to locate data centers in areas with lower-cost power.

25. J.N. Hoover, "Recovery.gov Moved To Amazon Cloud," *InformationWeek Government* (May 13, 2010). Available at time of press at http://goo.gl/cE7LL
26. R. Yasin, "GSA wins race to the e-mail cloud," *GCN* (October 18, 2011). Available at time of press at http://goo.gl/XehEj
27. M. Johnson, "GSA Is In the Cloud," U.S. General Services Administration blog (July 26, 2011). Available at time of press at http://goo.gl/N217t
28 Google, "Google Apps For Government." Available at time of press at http://goo.gl/iRM3P

7. Technology Investment and Research Opportunities

Cloud computing technologies developed in the private sector provide significant capability. In particular, these capabilities include utility computing for business services on public clouds, and large scale analytics to process increasing amounts of web traffic, and search and user data on private clouds. In this context, security is not the main concern, while business agility is. Additionally, many applications supported by today's clouds are embarrassingly parallel (employing simple parallelism). While it is certainly possible to encode parallel sparse linear algebra in MapReduce or over a collection of EC2 instances, it is inefficient.

In this chapter, research investments are discussed in key areas, such scalability, security, and usability, that could take cloud computing from predominantly embarrassingly parallel utility computing to computing that is secure, reliable, and more tightly coupled with adequate support at the tactical edge. Many technology challenges highlighted in this chapter are active research areas with often decades of relevant results (such as scheduling, performance optimization, fault tolerance, databases, and statistical learning techniques). However, the scales of current and emerging computing systems and the DoD and intelligence community application requirements dictate the need for new technical approaches to these, at times, classical problems.

It is important to note that DoD's computing requirements are quite broad, ranging from applications that are well-served by today's utility, data, and storage clouds, to applications that use specialized high performance computers, computing clusters with high performance specialized interconnects, and computing clusters with graphic processing units. In addition, today's clouds were never designed to support war fighters at the tactical edge, in which communications may be interrupted or seriously degraded.

An interesting development is the emergence of commercial high-performance computing cloud offerings, in which midrange symmetric multiprocessor clusters are being offered as infrastructure as a service, as well as computers with large memory and high performance storage. In addition, clouds at universities and research laboratories are being prototyped that incorporate graphic processors, solid state memory, and other specialized hardware. This is changing the type of applications that can be supported by infrastructure as a service.

Also in this chapter, a survey of research is discussed that may lead to a broader range of applications that can be supported by today's cloud, as well as clouds built from some of the more specialized types of hardware just mentioned.

As an example, today there is a significant amount of work being done so that clouds can support mobile clients, which may experience degraded communications or, at times, be disconnected. Some of this work may prove useful in the future to supporting

clients at the tactical edge. From this point of view, supporting devices on the tactical edge using clouds for DoD has some similarities to supporting mobile clients using commercial clouds in the private sector. With the appropriate research and development, thick clients with intelligent caching policies may be used in order to mitigate bandwidth constraints so that remote clouds can more effectively support devices at the tactical edge.

For many potential users of public cloud computing data centers, security is still one of their most important concerns. There are significant differences between how security is implemented and managed between different commercial clouds. Some of the large commercial clouds have security groups that are larger than the security groups at many DoD facilities, while other commercial clouds pay much less attention to security. As mentioned in Chapter 5, users of commercial cloud computing services are usually not given sufficient information about the computing infrastructure, monitoring of the infrastructure, forensic investigations, and so on, to satisfy the security requirements required for DoD applications.

Within the DoD, DARPA is currently making investments in several relevant technical areas. For example, its Mission-Oriented Resilient Cloud program is developing technologies to detect, diagnose, and respond to attacks in the cloud, effectively building a "community health system" for cloud computing. The DARPA Programming Computation on Encrypted Data (PROCEED) program is developing methods that allow computation on encrypted data without first decrypting it. One approach, homomorphic encryption, would have a client encrypt the data before sending it to the cloud. The client would also provide the cloud with executable code to allow it to work on that data without decrypting it. Current homomorphic encryption approaches are computationally infeasible, and researchers at DARPA are seeking to make them practical. DARPA is also funding other related approaches that may incur less overhead.

Table 2 lists recommended technology investments, categorized into three areas For each column, the list is presented approximately in order of increasing difficulty. In the section that follows a few particularly high-impact technology areas are described in more detail (with emphasis on security). While the technology areas are broken down into the three categories, technologies often require advancements across categories. Furthermore, progress within a category has impact in other categories. For example, while scalable runtime code optimization will benefit from improvements in program analysis, such improvement will also benefit research and development of malicious code detection.

7.1 Scalability

7.1.1 Quality of service and fault tolerance

While existing technologies provide fault tolerance through elasticity (replication) of computing resources, this is not sufficient to run competitive DoD applications that use high performance computers. To address the needs of these applications, fault tolerance must be achieved in the context of complex parallelism or distribution, as opposed to simple parallelism models and parallel computations. To continue scaling with both data sizes and computation complexity, one must address computational efficiency—achieved operations per second as compared to peak operations per second—and scalable application-to-computing-architecture mapping.

Table 2. Recommended DoD Research and Development for Cloud Computing Technology

Scalability	Security	Usability
• Quality of service metrics and guarantees • Dynamic scheduling and resource allocation • Program analysis • Dynamic program analysis • Automated diversity • Software, middleware for heterogeneous hardware • Automated performance optimization • Compute optimization • Memory, communication optimization • Power optimization • Automated diversity, fault tolerant runtime environments • Resilient, scalable storage systems	• Security instrumentation and metrics • Authentication and access control • Efficient, secure hypervisors • Traffic and data flow analysis • Hardware provenance and remote program authentication • Key management • Data/information flow labeling, isolation, and tracking • Cyber offense • Streaming analytics • Statistical analytics • Formal methods for correctness • Secure multiparty protocols • Tractable, practical homomorphic encryption • Specialized hardware architectures, co-processors for homomorphic encryption	• Fused data representation • Data and resource visualization • High-performance, massively parallel databases • Optimized indexing, search, and retrieval • Heterogeneous client—thin, thick—programming models • High-level composable application programmer interfaces (APIs) • Cloud-client application partitioning • Scalable parallelization-distribution framework • Specialized, general code generation

*The ordering in each column is in approximate order of increasing difficulty.

Technology initiatives to address this include development of instrumentation tools that could measure performance in context of the SaaS cloud computing paradigm, distributed and heterogeneous programming models, parallelization techniques that are agnostic and separate from the nature of the computation, and check-pointing tools (similar to traditional high performance computing).

7.1.2 Program analysis

Advances in program analysis tools and techniques would benefit both scalability and security. Efficient program analysis would allow performance prediction and optimization of data-intensive codes, analysis of security flaws in the code, and detection of injected malicious software (given signatures or family of signatures). As many large software systems within the DoD leverage open source components, program analysis tools could significantly increase safety and reliability of those components.

Program analysis challenges that are specific to DoD cloud computing include highly heterogeneous software and hardware environments, virtualization (analyzing code within a virtual machine), and efficiency at scale (developing program analysis techniques with minimal overheads).

7.1.3 Automated diversity

Even with the most diligent attention to preventing software flaws, some exist. Often exploiting these flaws depends critically on specific implementation-artifacts that must be present in every copy of the program. Cloud computing applications, with high development costs and wide distribution, are especially vulnerable. Artificial diversity techniques, such as address space layout randomization, have been developed to provide economical resistance to some attacks.

Additional techniques, like control flow integrity, offer further protection with limited human investment in individual program copies. Attackers respond to defenses, so the defense remains useful only if development continues. And, given the vulnerabilities, this research is critical for cloud computing.

7.1.4 Automated performance optimization

Cloud data-intensive computing can be highly inefficient by classic performance measures. Reasons for inefficiency include the common use of languages that trade off computational efficiency for programmability (Java); coarse parallelism programming models (MapReduce); data characteristics (sparse, massive data); and hardware inefficiencies (hardware platforms that are designed for high locality of data accesses).

One way to address some of these challenges is runtime performance optimization of existing codes. Research efforts in this area could include a runtime optimization environment for MapReduce, leverage of flexible communication models for parallel

codes such as publish-subscribe, and leverage of flexible programming models such as message passing interface and array-based programming. Additionally, for utility computing, tools can be developed to optimize power efficiency and provide guarantees for quality of service.

7.2 Security

7.2.1 Instrumentation and metrics

Often, there simply is a lack of available data regarding a wide range of security incidents. Instrumentation that minimally interferes with performance could significantly improve the development of meaningful metrics, provide data for development of analytics, and enable effective remediation of security incidents. Research efforts could focus on instrumentation at different levels in the software stack—from virtual machine to specific services or computation modules.

7.2.2 Authentication and access control

Cloud users and providers are rarely co-located and, yet, establishing their identity is foundational to all access control. While there are good cryptographic techniques for doing this authentication, wide-scale use of these techniques incurs all sorts of problems, from the manipulation of physical tokens, to corruption of central databases like Active Directory (whose compromise can defeat an entire enterprise's security in one fell swoop), to sensor-based confirmation or audit of identity (biometrics, gaze tracking), and so on. Further, access control in cloud systems and mobile devices depends not only on authenticating the human user, but also on authenticating some program performing the task. In fact, most mobile applications and cloud applications depend solely on program identity to determine access rights. Passwords are problematic, and usability of authentication mechanisms represents a major issue. Much work remains to be done.

The ability to read or write data, run programs, enter physical facilities, and receive keys to decrypt information must be controlled so that only authorized parties (people, programs, organizations), under authorized conditions (location, time) have the access. As with authentication, access control with clouds is an inherently distributed systems problem, requiring high assurance and flexibility (to describe who has what rights to what objects under what conditions). Cryptographic techniques using supporting infrastructure (physical tokens, isolated, measured software components, and perhaps a public key infrastructure) have been employed but, as with authentication, current deployments are fragile in the face of use by real people trying to get their jobs done. This is a large research area.

7.2.3 Labeling, isolation, and tracking

As stakeholders become physically remote from the computers that run programs on their behalf, technology must play a larger part in ensuring the isolation, confidentiality, and integrity of those computations from other cloud users, as well as the data center operators. Hardware has already been developed to enable programs to "prove" isolation and integrity properties to their providers and remote users.

This technology also allows cryptographic key provisioning in a manner that ensures even data center operators cannot access keys. Programs can now encrypt all stored and transmitted data as a means to ensure confidentiality and integrity. This technology also can be employed by data center framework components—that can be remotely verified by users—to ensure fair resource allocation, inter-job sanitization, and reliable auditing.

However, there is much work to do to make this existing technology useful to DoD, as well as to ensure usability by clients and by datacenter managers. Further research and software development is required to enable cloud-client systems to leverage these mechanisms for enforcing isolation boundaries that are crucial for ensuring confidentiality and integrity of data, results, and provenance of information. Further investigation related to the resilience of these systems is also needed, to address implementation flaws and environmental factors; for example, those that enable side channel attacks. A further area of interest is compositional vulnerabilities introduced when two otherwise safe components interact in an unexpected way.

7.2.4 Cyber adversaries

Cyber security is often described as computing in the presence of an adversary. Economics often involves the study of parties competing for resources. In both cases, understanding adversarial capabilities is critical to defense. Security experts often place confidence in assessments of system vulnerabilities as a result of diligent attacks by well-informed and well-trained red team members. This sort of analysis is much more representative of the actual threat environment than contrived, piecemeal analyses. Red team attacks have proven to be the best source of information about improvements in the design and operations of a security system.

For security reasons, this knowledge is often not widely held or employed by system researchers. Further, other parties, such as antivirus vendors, are occasionally better positioned to discover attacks and attack techniques. While some techniques should properly be closely held, developing capable attack techniques is surely critical to developing safer systems and this research must proceed.

7.2.5 Streaming statistical analytics

Existing streaming analytics, whether for detection of cyber security events or analysis of large-scale datasets and databases, usually either perform relatively simple

computations on large datasets or complex computations on small portions. However, the application of complex analytics to large datasets is needed to address the challenges presented by big data, namely detection of weak signature events in multi-intelligence or multisource datasets in an efficient and timely manner, in absence of a cue. Technology efforts could focus on developing novel algorithmic techniques along with data representations. These algorithms would detect statistical anomalies at global scale and alert either human analysts or another algorithm layer. This approach would fundamentally change the current analytic paradigm where the initial processing step is either omitted, based on reliance on an external cue, or performed manually. New algorithmic techniques would also benefit from new programming models and performance optimization tools.

7.2.6 Hardware provenance

A disturbing trend is the increased susceptibility of hardware components to attack by unknowingly using attacker-supplied circuits in fabrication. Most DoD components—and almost all national infrastructure components—are manufactured by commercial suppliers, operating on thin economic margins, with the associated motivation to economize with respect to security and assurance. These commercial suppliers operate largely in foreign countries that may themselves have good reasons to interfere with equipment used by DoD. Limiting the number of critical components required to "fight through" a corrupt hardware chain is a new design imperative, and the area is one that requires new research. Discovering hardware modifications or assuring their absence is also an important area for further research.

7.2.7 Methods for assurance

The customary DoD processes for testing, verifying, and certifying software systems are inadequate. Certified systems often delay deployment until long after vulnerabilities are discovered and much safer (but not-yet-certified) versions are available. Certification cost is prohibitive, driving innovation out and, worse yet, delaying availability of key new capabilities (like big data analysis tools) that are rapidly developing. Every single element of this process must change. Better and more automated testing; automated verification and risk assessment; and an economical, streamlined certification process are absolutely critical research areas.

7.2.8 Homomorphic encryption

Homomorphic encryption enables operations on the data in encrypted form (performing multiple functions without decrypting the data). A key challenge in homomorphic encryption is computational tractability. Existing techniques have demonstrated the capability for simple operations such as addition and multiplication; however, the computational complexity is currently prohibitive. Advances in tractable

homomorphic encryption could enable both highly secure utility computing and data intensive computing across multiple classification levels. These techniques could also provide situational awareness and aggregate statistics without sacrificing privacy or sensitivity of the original data sources. The research challenges need to be tackled algorithmically through design of novel hardware. Near-term research targets could include advances in searchable encryption, where recent results already allow practical database-like operations on encrypted data in a cloud, preserving the confidentiality of the data.

7.3 Usability

7.3.1 Data and resource visualization

A key challenge in processing massive datasets is visualizing both the raw data and the results of computations on the data. Similarly, when distributing applications and computations across multiple resources, it is often desirable to visualize those resources and data distributions. Research efforts into visualization (especially combined with statistical analytics of data) could significantly improve analysts' ability to understand massive datasets and detect events of interest (for example, malicious activity or exfiltration).

7.3.2 High performance massively parallel databases

While existing distributed database technology enables ingest of large datasets, complex queries still present a significant challenge. As algorithms for processing large data continue to advance, support for increasingly complex queries will become necessary. Research efforts could focus on database architecture, query language, and performance optimization.

7.3.3 Composable, high-level application programmer interfaces

Ability to interact with large datasets in context of data-intensive cloud computing requires efficient, scalable, and intuitive APIs. While existing APIs (such as MapReduce) provide reasonable capabilities, they do not naturally allow for implementation of complex parallelism and support for array-based algebra operations. Array-based algebra allows for both implementation of complex analytics (principal component analysis, feature extraction, large graph analysis, complex queries) while providing an intuitive parallelization interface (mapping the array). Research efforts could focus on developing novel languages, creating libraries using existing languages, developing parallelization frameworks, and performance optimization techniques.

7.3.4 Cloud–client application partitioning

Cloud computing does not exist in a vacuum—it must be provisioned, and often it provides value by supplying information to remote clients. Service is often delivered to users by browser-based clients or moderated through productivity applications, such as Adobe Reader or Microsoft Office, that can hide attacks, promulgate them, and deliver exfiltrated information to attackers or corrupt information to users. Browsers are particularly worrisome, given the large and unending stream of catalogued exploitable vulnerabilities. A cloud application is really the combination of a cloud component and a client component, and the security goals should be to protect this combined system—not to protect just the cloud. Increasing the proportion of processing done on a safe cloud can improve updates and maintenance; however, the remaining components on a client must also be made safe. Research into safer browsers and browser-extension safety is required. Access policies, configuration safety, and verification of client components are needed, as well as the ability of both client- and cloud-computing components to mutually verify partner code. This is clearly a critical research area prior to deploying a broad set of applications between client devices and a cloud computing center.

7.4 Combining Technologies

As described here, these technology areas can provide significant improvement in scalability, security, and usability in their own right. However, combining these technologies could lead to revolutionary changes in computing. For example, combining homomorphic encryption with a high-level composable, parallel API (such as an associative array) could enable processing of massive, multi-intelligence datasets, thereby providing actionable intelligence to the user at the tactical edge without concern for mixing classification levels. Being able to provide big data analytics at the tactical edge (without a cue, which is consistent with the changing nature of the conflict) is an example of significant capability that could be enabled by investment in cloud computing technologies.

Finding

Finding 10: The DoD has active research and development efforts in technology areas applicable to cloud computing performance and security. Sustained DoD investment in cloud computing security technology is critically important to allow DoD data centers to continue improving their defenses against evolving threats. Research and development in software stack protection, monitoring, and forensics of very large datasets, secure hypervisors, and advanced encryption offer significant possible security benefits.

8. Findings Summary and Recommendations

8.1 Findings Summary

The Significance and Impact of Cloud Computing

Finding 1: Although cloud computing is an overloaded term, cloud computing providers are offering services that are fundamentally new and useful, typically delivering the:

- ability for massive scale-up of storage and computing
- rapid, agile, elasticity with the ability to increase and decrease storage and computing capacity on-demand, when the community of tenants don't all require that capacity at the same time
- metered services where the user pays only for what is used
- self-service start-up and control

Finding 2: Modular data centers offer an approach to quickly set up cloud computing capacity, to add additional capability to existing cloud computing data centers, and to easily refresh or update existing capability. This concept is illustrated in Figure F-1.

Finding 3: Cloud computing services can scale to data centers or "warehouse-scale" computing. Elastic, warehouse-scale cloud computing is fundamentally new and can provide DoD with important new capabilities.

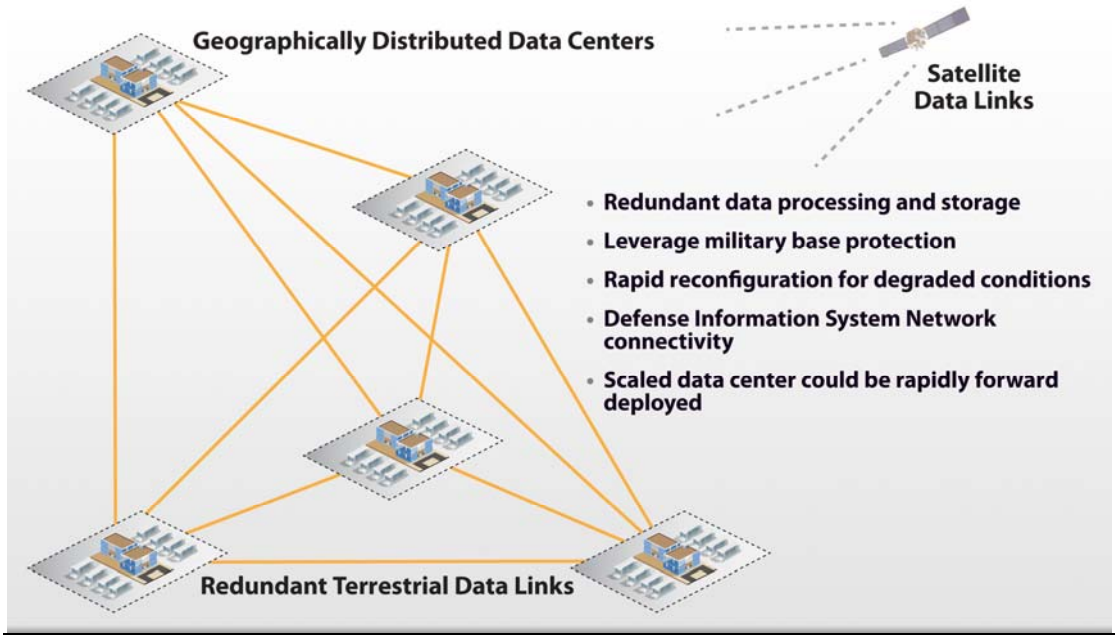


Figure F-1: Concept for a geographic distribution of DoD data centers

The Security of Cloud Computing

Finding 4: Cloud computing is not intrinsically more secure than other distributed computing approaches, but its scale and uniformity facilitate and enable the wholesale and consistent application of security practices. Secure aspects include large scale monitoring and analysis of data to detect attacks, and automated and persistent provisioning and re-provisioning to foil intrusions. For these reasons, well-operated cloud computing facilities can exhibit better security hygiene than conventional data centers. However, the centralization of resources in a huge data center also encourages more determined attacks, especially on critical components broadly affecting security. This is similar to conventional systems where attacks are observed to focus on central directories.

Finding 5: The scale of cloud computing enables the analysis of packet and log data that provides new capabilities for event forensics and real-time detection of malicious behavior. The ability to manage very large, diverse datasets facilitates a data-centric security model in which users are authorized to work with data based upon their security credentials and the security markings on the data rather than the conventional enclave-centric security model in which users are provided access to an enclave and can access all the data in the enclave.

Finding 6: No cloud computing deployment model is uniformly suitable for hosting all DoD applications. In general, sensitive, classified, and time-critical DoD applications should be deployed only in private clouds or conventional non-cloud approaches.

Finding 7: The case for transitioning a DoD application to a cloud computing data center must include a security assessment detailing the impact of the transition. Whether security will be improved by transitioning an application to a cloud computing data center will depend on factors specific to the application, to the cloud computing data center, and to the transition process.

Finding 8: The DoD has not established effective plans for cloud computing facility backup or for dealing with any anticipated degradation of communications between the cloud computing facilities and the end user.

The Costs Associated with Cloud Computing

Finding 9: Potential cost reductions or increases incurred during the transition to and sustainment of cloud computing infrastructure depend on the specifics of the implementation. Potential cost-reduction factors include a higher utilization of servers, lower professional support staff needs, economies of scale for the physical facility, and the flexibility to locate data centers in areas with lower-cost power.

Research and Development for Cloud Computing Technologies

Finding 10: The DoD has active research and development efforts in technology areas applicable to cloud computing performance and security. Sustained DoD investment in cloud computing security technology is critically important to allow DoD data centers to continue improving their defenses against evolving threats. Research and development in software stack protection, monitoring, and forensics of very large datasets, secure hypervisors, and advanced encryption offer significant possible security benefits.

8.2 Recommendations

Overarching Recommendations

Recommendation 1: For some sensitive, classified, and time-critical applications, the DoD should pursue private cloud computing, provided that strong security measures are in place.

In particular, cloud computing-based solutions should be considered for applications that require the agility, scale-out, and ability to integrate and analyze massive data that cloud computing can provide. Examples of such applications include: big data analysis and all-source intelligence integration; processing, exploitation, and dissemination of data gathered through intelligence, surveillance, and reconnaissance (ISR); large-scale modeling and simulation; open source data collection, storage, and assessment; and advanced decision support systems.

Recommendation 2: The DoD CIO in partnership with the military Services should deploy interconnected, modular cloud computing data centers located at secure locations, such as military bases.

The development of large, private community clouds in DoD will enable greater computing and storage elasticity and the improved ability to operate under degraded conditions. The DoD CIO should guide this development with an eye on both current and future DoD computing needs.

A DoD private community cloud may include in-house, in-sourced, or out-sourced private clouds. Implemented through interconnected, modular cloud computer data centers, this can be operated as an integrated unit to improve the potential reducing costs.

Because large data centers can also be attractive targets, geographically distributed modular data centers are recommended that are operated as a single, large-scale, distributed cloud. The design should include a distributed data center architecture that allows access by multiple Services and Agencies. Cost savings would be achieved through shared development, operations, and maintenance support.

These modular data centers could be located on military bases in order to provide good physical security. The location should also be influenced by the cost and availability of reliable

electric power. It is anticipated this will be similar to the National Security Agency private cloud models. Shared cyber security event response and rapid forensics would be an enhanced capability.

By designing and acquiring these data centers as a system, the DoD can achieve the economies of scale typically associated with large data centers.

Recommendation 3: The DoD CIO and DISA should establish clear security mandates for DoD cloud computing.

Security mandates should be aimed at reducing the number of cloud compromises and to mitigate those that occur. Some examples of potential mandates include:

- Hypervisors hosting DoD operating systems should have effective cryptographic sealing, attestation, and strong virtual machine isolation.
- Data at rest should be stored in encrypted form with keys protected using hardware attestation, such as a trusted platform module (TPM).
- Data in transit on communication lines should be encrypted with keys protected using hardware attestation, such as a TPM.
- Access to cloud computing systems should require multifactor authentication.

Recommendation 4: The DoD CIO should establish a central repository to fully document cloud computing transition and sustainment costs and best practices for programs underway or completed.

Because the cost savings to be gained through cloud computing are case-dependent, a central repository documenting DoD cloud computing programs is needed. The goal of this repository is to improve the understanding of the following:

- system costs before the switch to cloud computing, costs during transition, and sustainment costs
- enhanced functionality attributable to cloud computing architectures
- best practices for cloud computing security
- issues surrounding service license agreements
- metrics for availability and reliability

This repository will enable leveraging the lessons learned from several DoD cloud computing initiatives underway, including:

- NSA development and use of private clouds
- DISA Rapid Access Computing Environment (RACE)
- Army Enterprise Email

Recommendations to Improve DoD's Implementation of Cloud Computing

Recommendation 5: The DoD USD AT&L and the DoD CIO should establish a lean, rapid acquisition approach for information technology infrastructure, including cloud computing hardware and software.

Acquisition guidelines for all information technology—not only cloud computing hardware and software—should strive to create a lean, capabilities-based approach with strong, clear security mandates. Rapid certification and accreditation (C&A) and other characteristics to streamline acquisition of cloud computing hardware and software should be developed and implemented quickly.

Recommendation 6: The DoD CIO and DISA should establish standard service level agreements for private and public cloud computing.

Key attributes that should be included in service level agreements include availability, authentication and authorization approaches, data processing and storage locations, software and data back-up approaches, cyber attack event notification, required staff clearances and background checks, software and data disposition, risk disclosure requirements, and contingency plan. Transparency in all of these aspects for DoD service providers will help set standards for secure cloud computing across the economy.

Recommendation 7: The DoD CIO and DISA should participate in the public development of national and global standards and best practices for cloud computing.

A key outcome of this activity will be to inform the private sector and open source developers about the agility and auditability requirements for DoD cloud computing.

Recommendations to Improve Cloud Computing for Degraded Operations

Recommendation 8: The DoD and the intelligence community leadership should develop a unified approach for training and exercising with degraded information infrastructure, including cloud computing hardware and software.

Degraded operations in a realistic operational exercise must be implemented organically, *i.e.*, beyond simply holding up a white card to introduce a cyber event to an exercise. Advanced cyber security threats should be exercised, including a gradual ramp-up of threat and loss of disadvantaged communication and data links as well as primary capabilities. Enhanced red and blue teaming should be established along with operational exercises incorporating degraded cloud computing infrastructure. Participants should demonstrate a rapid forensics response and effective backup plans.

Recommendation 9: The Joint Chiefs of Staff and Combatant Commands should establish effective back-up plans for operations with degraded information infrastructure, including cloud computing hardware and software.

Candidate plan attributes include implementing thicker clients and forward caching of data as well as backup data networks, processors, and storage. Each organization should also develop operational contingencies for degraded networks. Potential strategies also include using local network connectivity for forward clients and narrowband, analog communication links for situational awareness and warning.

Recommendations for Investment

Recommendation 10: The DoD should continue investing significantly in information security research and development, including research and development for secure cloud computing technology.

To best leverage state-of-the-art cloud computing technologies for DoD, significant investment should continue for technology research and development activities in areas such as: efficient operations of cloud computing data centers; cloud security; secure, lean hypervisors; micro-virtualization; advanced TPMs; homomorphic computing; and cloud situational awareness software.

8.3 Concluding Remarks

The DoD should pursue cloud computing to enhance mission capabilities, provided that strong cyber security measures are in place. Stronger red teaming and realistic exercises are needed to provide critically important improvements to DoD cyber security.

Missions that may enjoy new capabilities as a result of cloud computing include: big data analysis and all-source intelligence integration; ISR processing, exploitation, and dissemination; and large-scale modeling and simulation. New modular cloud data center hardware offers the DoD an opportunity for rapid enhancement of the overall computing enterprise in a secure and resilient manner.

The potential to reduce costs or to increase them through implementation of cloud computing is very case-dependent. Some factors that can lead to cost reductions include higher usage of data center, lower support staff-to-server ratios, better management of peak loads, and economies of scale.

Terms of Reference

DEPUTY SECRETARY OF DEFENSE
1010 DEFENSE PENTAGON
WASHINGTON, DC 20301-1010

MAY 19 2011

MEMORANDUM FOR CHAIRMAN, DEFENSE SCIENCE BOARD

SUBJECT: Terms of Reference – Defense Science Board (DSB) Task Force on Cybersecurity and Reliability in a Digital Cloud

You are hereby directed to establish a Task Force to evaluate all aspects of providing reliable, secure, and responsive services for military and intelligence applications using these technologies. Specific purposes of this Task Force include:

- Characterize the operational properties of clouds and virtualized infrastructure, and the quality of service that can be delivered to connected users, paying particular attention to attacks on communication that would destroy or delay delivery of services and information for time-critical uses;
- Consider alternative designs and implementations of these technologies and evaluate their use for varied military and intelligence applications;
  - Discuss options for inter-cloud services and other communications between clouds, especially between clouds that operate at different classification levels, have different authorities, or are operated on behalf of different agencies and organizations;
- Evaluate the vulnerability and risk mitigations of a cloud infrastructure to various attacks, compared to alternative infrastructures;
- Evaluate the vulnerability and risk mitigations of virtualized infrastructure to various attacks, compared to alternative infrastructures;
- Determine how to avoid the danger of concentrating data and computation; for example, suggest how diverse (non-homogeneous) software and hardware can be deployed in the cloud to enhance reliability and security;
- Review and project the consequence of current trends in digital technology on cloud deployments, including social computing;
- Comment on customer practices and modes of interaction with the cloud that might aid in increasing security;
- Make recommendations on what dimensions (pros and cons) of these technologies should be considered to increase reliability and to ensure security as the military and intelligence communities evolve their digital infrastructure;
- Comment on workforce implications (skills, qualifications required, etc.) the Department might expect in transitioning from current environments to cloud implementations;
- Identify research opportunities and estimate the level of investment to achieve results consistent with DoD needs;

- Assess cost/benefit and effectiveness/suitability issues associated with Software as a Service, Platform as a Service, and Infrastructure as a Service, as those technologies apply to both garrison and deployed computing requirements; and
- Identify methods to leverage the rapid innovation and operational maturity being delivered by public cloud providers while maintaining the Department's ability to service its own IT needs effectively and operate at multiple classification levels.

The military and intelligence communities are increasingly articulating a need for information systems that rely upon cloud computing, virtualization, and related technologies. A digital cloud has a richly networked, distributed core of data storage and computational resources that provides shared information and services on demand to individual users located outside the cloud, but connected via a network. Cloud advocates assert that infrastructure incorporating cloud-based technologies and virtualization can deliver both higher reliability and more assured cybersecurity.

Administration support and funding will be provided by the Under Secretary of Defense for Acquisition, Technology, and Logistics. Additional support will be provided by the Assistant Secretary of Defense for Networks and Information Integration/DoD Chief Information Officer (ASD(NII)/DoD CIO), the Director, Operational Test and Evaluation (DOT&E), the Vice Chairman of the Joint Chiefs of Staff, and the Commander, U.S. Cyber Command. All Task Force members, consultants, and supporting personnel will be appointed or designated in accordance with DoD Instruction (DoDI) 5105.04, "Department of Defense Federal Advisory Committee Management Program."

The Task Force will be established and operated in accordance with the provisions of the "Federal Advisory Committee Act" (5 U.S. Code Appendix, as amended), DoDI 5105.04, the DSB Charter, and all applicable laws, policies, and regulations. It is not anticipated that this Task Force will need to go into any "particular matters" within the meaning of section 208 of title 18, U.S. Code, nor will it cause any member to be placed in the position of acting as a procurement official.

Task Force Membership

Co-Chairs
Dr. Eric Evans, MIT Lincoln Laboratory
Dr. Robert Grossman, University of Chicago and Open Data Group

Executive Secretary
Mr. John Mills, Department of Defense CIO

Members
Mr. Gregory Neal Akers, Cisco
Dr. Wanda M. Austin, Aerospace Corporation
Dr. Ron Brachman, Yahoo!
ADM William J. Fallon, USN (RET), Private Consultant
Mr. Al Grasso, MITRE Corporation
Dr. Eric Grosse, Google
Hon. (Dr.) Anita K. Jones, University of Virginia
Mr. Larry Lynn, Private Consultant
Dr. John L. Manferdelli, Intel
Mr. Larry Meador, MGI Strategic Solutions
Mr. Alden V. Munson, Jr, Potomac Institute for Policy Studies
Dr. Fred B. Schneider, Cornell University
Mr. James D. Shields, Draper Laboratory
Mr. Steve Schmidt, Amazon
Dr. Marc Zissman, MIT Lincoln Laboratory

Senior Advisors
Dr. Craig Fields, Private Consultant
Dr. Theodore S. Gold, Private Consultant
Hon. R. Noel Longuemare, Private Consultant
Hon. (Dr.) William Schneider, Jr., International Planning Services

Government Advisors
Mr. Lee Badger, National Institute of Standards and Technology
Dr. Patrick W. Dowd, Department of Defense
Ms. Cecilia Phan, Department of Defense CIO

Defense Science Board
CDR Doug Reinbold, U.S. Navy

Staff
Dr. Toni Marechaux, Strategic Analysis, Inc.
Ms. Kelly Frere, Strategic Analysis, Inc.

Presentations to the Task Force

	Organization	Title of Brief
May 17–18, 2011		
Ms. Jennifer Bayuk	Stevens Institute of Technology	Systems Security Engineering
Mr. Brad Bleier	National Cyber Investigative Joint Task Force	National Cyber Investigative Joint Task Force
Ms. Jamie Dos Santos	Terremark, Inc.	National Capital Region NAP
Dr. Timothy Fraser	Defense Advanced Research Projects Agency	Dynamic Quarantine of Worms; Cyber Gnome
Ms. Margie Gilbert	Office of the National Counterintelligence Executive	National Counterintelligence
Dr. Keith Gremban	Defense Advanced Research Projects Agency	Cloud to the Edge
Mr. Barry Horowitz	University of Virginia	Research Roadmap
Mr. Dan Kaufman	Defense Advanced Research Projects Agency	Cyber Analytical Framework
Dr. Carl McCants	Defense Advanced Research Projects Agency	TRUST, ISIS Programs
Dr. Jenendra Ranka	Defense Advanced Research Projects Agency	National Cyber Range
Mr. Tony Sager	National Security Agency	Vulnerability Assessment: Virtual Machine
Mr. Steve Stone and CAPT Mike Murray	U.S. Transportation Command	TRANSCOM Security issues
June 15–16, 2011		
Mr. Lee Badger	National Institute of Standards and Technology	U.S. Government Cloud Computing Technology Roadmap
Mr. Patrick W. Dowd	National Security Agency	Cloud Computing Architectures
Mr. Steve Schleien	Office of the Secretary of Defense	DoD Strategy for Cyberspace
Mr. Rob Vietmeyer and Mr. John Shea	Department of Defense CIO	DoD Cloud Computing for the Defense Science Board
Mr. Greg Wilshusen	Government Accountability Office	Information Security Issues for Cloud Computing

	Organization	Title of Brief
Mr. Jim Young and Ms. Michelle WeslanderQuaid	Google	Overview
July 12-13, 2011		
Mr. Asher Aziz	FireEye, Inc.	Advanced Persistent Threat Landscape and the Subversion of Existing Defenses
Mr. Rod Beckstrom and Mr. Whitfield Diffie	Internet Corporation for Assigned Names and Numbers (ICANN)	Internet Corporation for Assigned Names and Numbers (ICANN)
Mr. Bill Burns	Netflix	Netflix Cloud Security
Mr. Edmundo Costa	Catbird	Security and Compliance for Virtual and Cloud Infrastructure
Mr. Michael Donovan and Ms. Kelly Collins	HP Enterprise Services	Cloud Security Issues; HP Fortify Federal: Protecting the Software that Runs the Mission
Mr. Bret Hartman	RSA	The Intelligent Security Operations Center and Advanced Persistent Threats
Mr. Chris C. Kemp	OpenStack Cloud Software	OpenStack Cloud Software
Mr. Pravin Kothari	CipherCloud	Cloud Data Protection
Dr. John C. Mitchell	Stanford University	Innovation in the Big Cloud
Mr. Eric Olden	Symplified	Small Business Innovation in the Big Cloud
Mr. Don Proctor	Cisco Systems	Cisco Systems
Mr. Justin Somaini	Yahoo!	Yahoo!
Mr. Rich Tener	Zynga	Public Cloud Security Topics
Dr. Pete Worden	NASA Ames Research Center	Welcome to NASA Ames Research Center
Mr. Ken Xie	Fortinet	Small Business Innovation in the Big Cloud
August 10–11, 2011		
Mr. Lonny Anderson	National Security Agency	View from the CIO

	Organization	Title of Brief
Mr. Phillip Bodkin	National Security Agency	NSA Threat Operations Center
Mr. Justin Christian	National Security Agency	OZONE Cloud Interface
Mr. Simon Crosby	Bromium	Three Tales of the Cloud
Dr. Patrick W. Dowd	National Security Agency	Cloud Strategy
Mr. Oren J. Falkowitz	National Security Agency	ABR
Mr. Sterling S. Foster	National Security Agency	Data Cloud; Metadata Repository Increment 2
Mr. Glenn Gaffney	Office of the Director of National Intelligence	Cloud Computing: Key Questions
Ms. Melissa Hathaway	Hathaway Global Strategies	Observations on Cyber Strategy
Dr. John D. Howard	Defense Information Systems Agency	DISA Common User Services: The "Cloud" and the Future of DoD IT
Mr. Brian Hughes	National Security Agency	IT Efficiencies at NSA
Mr. John Ingliss	National Security Agency	Overview of NSA
Mr. Jaime Jenandez-Cordero	National Security Agency	Human Language Technology Demonstration
Ms. Catherine Lotrionte	Georgetown University	Georgetown Cyber Law and Policy
Ms. Stacey Olexy	National Security Agency	Content (PRESSUREWAVE)
Ms. Teri Takai	Department of Defense CIO	DoD CIO Challenges and Priorities for the DoD Enterprise
Mr. Neal Ziring and Mr. Adam P. Fuchs	National Security Agency	Cloud Security
September 14–15, 2011		
Mr. Kevin Cooley	U.S. Navy	U.S. Navy, Fleet Cyber Command Information Officer
Mr. John Hale	Defense Information Systems Agency	DoD Enterprise E-Mail
Mr. Frank Konieczny	U.S. Air Force-Chief Technology Officer	Air Force Perspectives on Cloud Computing
Mr. John McLaughin	IBM Corporation	Mission Oriented Cloud Architecture Overview

	Organization	Title of Brief
Mr. Mike McCarthy	U.S. Army-Brigade Modernization Command	Connecting Soldiers to Digital Applications
Brig. Gen. Linda R. Medler	The Joint Staff	Cloud Services for the Warfighter
RDML Jan Tighe	U.S. Navy	Cyber security in Military Exercises
November 30–December 1, 2011		
LTC Arthur Sellers and Mr. Tony Gillespie	U.S. Special Operations Command	U.S. Joint Special Operations Command
Ms. Jamie Dos Santos	Terremark	Overview of Terremark's National Capital Region Network Access
Dr. Howard Shrobe	Defense Advanced Research Projects Agency (DARPA)	Cloud Investigations at DARPA
MG Mark Bowman and Col Gary Langston	U.S. Army, G6	Enterprise Email
January 18–19, 2012		
Mr. Kevin Dulany	DoD Office of the Chief Information Officer	Federal Risk and Authorization Management Program
Mr. Kevin Gates	House Armed Services Committee	Data Servers and Centers
Mr. John Howard	Defense Information Systems Agency	Technical Strategy, Target Architecture, and RACE
Mr. Gus Hunt	Central Intelligence Agency	Cloud Strategy
Gen. Stanley McChrystal, USMC (RET)		Observations on Operations
Mr. Mark Morrison	Defense Intelligence Agency	DNI IT Efficiencies
Mr. Albert Reuther	MIT Lincoln Laboratories	Some Cloud Computing Economics
March 14–15, 2012		
Mr. George Slessman	IO Data Centers	IO - Intelligent Control
Mr. Bill Newhouse	National Institute of Standards and Technology	An Overview of the Cloud Assumption Buster Workshop
Mr. Orv Stockland	National Reconnaissance Office	Cloud Computing

	Organization	Title of Brief
Mr. Rob Vietmeyer	Office of the Secretary of Defense	Cloud Update
Mr. Neal Ziring	National Security Agency	Cloud Security

Abbreviations and Acronyms

ALIRT	airborne ladar imaging research testbed
API	application programming interface
ARGUS-IS	Autonomous Real-Time Ground Ubiquitous Surveillance Imaging System
AWS	Amazon Web Services
C&A	certification and accreditation
CIO	Chief Information Officer
COTS	commercial off the shelf
DARPA	Defense Advanced Research Projects Agency
DISA	Defense Information Sysytems Agency
DISN	Defense Information Systems Network
DoD	Department of Defense
EC2	elastic compute cloud
EIA	Energy Information Administration
GIG	Global Information Grid
GSA	General Services Administration
HPC	high performance computing
HVAC	heating, ventilation and air conditioning
HYCAS	hyperspectral collection and analysis system
IaaS	infrastructure as a service
IEEE	Institute of Electrical and Electronics Engineers
ISR	intelligence, surveillance, and reconnaissance
kWs	kilowatts
NIST	National Institute of Standards and Technology
NoSQL	not only structured query language
NSA	National Security Agency
PaaS	platform as a service
PROCEED	Programming Computation on Encrypted Data
RACE	Rapid Access Computing Environment
RAM	random access memory
S3	Amazon Simple Storage Service
SaaS	software as a service
SLA	service level agreement
TPM	trusted platform module
TTP	tactics, techniques, and procedures
USD AT&L	Under Secretary of Defense for Acquisition, Technology, and Logistics
VM	virtual machine
WISP	wideband infrared scene projector

DEPARTMENT OF DEFENSE

DEFENSE SCIENCE BOARD