A George Mason University Reader on

Cyber Operations

Prof. Paul A. Strassmann

None the text appearing on these pages can be considered as official DoD material. The text consists either of original article published by AFCEA Signal, or has previously appeared on http://pstrassmann.blogspot.com/ as original text. Source of information are shown in footnotes.

The text cannot be considered as offering consulting services. It represents professional views of prof. Paul A. Strassmann.

Graphics and Composition: Paul A. Strassmann

ISBN Number and Library of Congress Registration is pending

Publisher: THE INFORMATION ECONOMICS PRESS

Strassmann, Paul A.

1. Strategic Planning; 2.Information Technology; 3.Business Management; Title 658.4

Version 1.1 – September 2011

Table of Contents

Introduction

This book is a collection of text that has been published either as a blog (http://pstrassmann.blogspot.com/), or as miscellaneous 2009 through 2011 AFCEA publications. It is an attempt to offer an executive level exposure to cyber operations issues that would be of interest to cyber commanders who need to acquire competency in dealing with cyber systems matters.

This collection has been conceived to meet the needs of military and civilian personnel who will be increasingly engaged informal briefings or in formal academic training at the graduate or post-graduate levels.

The topics assembled here are not arrayed in the sequence of lectures to students in the graduate level AIT 690 course. The printed version of this collection will keep track of subjects as new developments occur. Since this book is published on demand each new print version will include new topics but also delete subject matter that is of declining interest.

Paul A. Strassmann

New Canaan, Connecticut, September, 2011

PART 1: THE CYBER CONTEXT

The Need for a Federal Enterprise Architecture

One mandate of the Clinger-Cohen Act of 1996 was creation of the Information Technology Architecture. In subsequent 1999 guidance, the Federal Chief Information Officers Council defined the Federal Enterprise Architecture as the process for developing, maintaining and facilitating the implementation of integrated systems.

Chief information architects then were appointed at the federal and Defense Department levels. However, as of June 2011, the Government Accountability Office (GAO) reports that the enterprise architecture methodology was not deployed. Between 2001 and 2005, the GAO reported, the Defense Department spent hundreds of millions of dollars on an enterprise architecture development that was of limited value. None of the three military departments so far have demonstrated that they have made commitments to deploy an architecture needed to manage the development, maintenance and implementation of systems.

Senior defense information technology executives have stated that the development of an architecture methodology has been pushed back due to budget limitations. As yet, no time frames have been established for producing a Federal Enterprise Architecture (FEA). There is no specific date when the enterprise architecture would be delivered. The current focus is on other resource-intensive commitments.

Nevertheless, the use of well-defined enterprise architectures remains nowadays an attribute of managing information technology in successful commercial firms. A centrally directed architecture continues as the basis for system integration and for delivering lower information technology costs. In spite of the significant potential benefits, the enterprise architecture has not guided Defense Department systems over the past 15 years.

While the Defense Department was promoting departmental and agency wide enterprise concepts, actual implementation of integration was missing except in isolated cases. Consequently the department ended up lacking a coherent blueprint for creating a technology environment that would be interoperable, economically efficient and easily accessible for innovation. The absence of a working architecture has prevented the Defense Department from making progress at the same rate as leading commercial firms.

The absence of a guiding Defense Department architecture also opened the floodgates to excessive project fragmentation, to technology incompatibilities and to operations that were contract-specific rather than enterprise integrating. That increased not only the costs of maintenance and of modernization upgrades, but also put the brakes on the ability to innovate, which would meet the rapidly rising demands to achieve information superiority. If there was innovation, it had to take place as stand-alone new projects and not as an enhancement to systems that could be improved at a lesser cost.

The Clinger-Cohen Act also passed to the chief information officers (CIOs) the responsibility for managing all information technology spending. That did not take place as originally legislated.

Currently, CIOs cannot be held accountable for the contracting decisions that are made by acquisition officers who use regulations found on 2,017 pages printed in small type. The management of information

technology, as of March 2011, is split into 2,258 business systems with only spotty direct-component CIO oversight. Meanwhile, the ability of the Defense Department CIO to control rising costs continues to be limited.

There are also more than 1,000 national security systems that are not listed in the Defense Department Information Technology Portfolio Repository (DITPR). Such projects include intelligence systems, military command and control networks, and equipment included as an integral part of weapons. Whereas in the Cold War years, when national security systems could be set up as stand-alone applications, modern warfare conditions now require a real-time interoperability between national security systems and selected business applications such as logistics.

As the consequence of having no architecture as well as weak CIO control over information technology costs, the Defense Department ended up with largely isolated enclaves, that is silos, of projects.

Information technology projects now are managed as separate contracts. Applications are not interoperable across the department because they cannot share enterprise wide data. Information technology projects are run in more than 700 data centers as well as in an untold number of special-purpose servers that often have data center capabilities.

Such outcomes originally were not envisioned in 1996. The flawed outcomes, predicted in 1995 Senate hearings, pointed to the inadequacies of Clinger-Cohen to meet post-Cold War needs. The 1996 legislation did not offer changes in the organizational structure for managing information technology. It did not alter the Defense Department component-entrenched budgeting practices. It also did not set out to deliver a shared information technology infrastructure that would compare favorably with best commercial practices. Security was not as yet on the agenda. Rising Defense Department requirements then would be met with rapidly rising spending for information technology.

Given the current status where both policy and implementation are lacking, two fundamental questions must be answered immediately. First, is it still possible to complete the originally planned FEA to guide Defense Department systems development? Second, is it possible to realign the positions of the Defense Department CIOs for acquiring the budgetary and organizational controls that would allow the management of information technology spending?

The answer to both questions is negative. The assumptions that have been applied for 15 years do not work anymore. Fortunately, the rapid development in the delivery of information technologies as a shared service—cloud computing—now has opened new options for the Defense Department to start redirecting its information technology policies. What is needed now is a FEA2 architecture for the placement of platform-as-a-service (PaaS) as the core around which to reorganize the conduct of information technology.

Instead of FEA1 trying to direct how thousands of separate projects would be designed and operated, FEA2 would define how to set up only a handful of PaaS cloud services. Staff located in remote locations away from where PaaS services are delivered would not develop FEA2. Instead, FEA2 staffs would be embedded where PaaS is executed.

FEA2 would be realized as an evolutionary process, not as a preset blueprint. The management of projects would not be split into stand-alone engineering, development, programming and operation phases assigned sequentially to separate organizations. Instead, in the spirit of the late Adm. Hyman Rickover, USN, programs would be managed as unified and tightly coupled ventures. FEA2 would concentrate on quarter-by-quarter migration to guide a rapid progression from the current legacy code to eventual arrival in the PaaS environment.

The planning horizon for FEA2 should be at least 10 years. Management of programs should be in the hands of senior, long-term officers, not rapid-rotation appointees. With information technology now

classified as a weapon rather than as an auxiliary function, all senior appointments should qualify as information systems specialists.

FEA2 would provide guidance on how applications would be stripped gradually from their underlying infrastructures. The objective would be to convert each application into a form that can be delivered as a service on demand on any of the Defense Department's standard PaaS offerings. New applications would be placed on top of a standard PaaS instead of the existing proliferation of infrastructures.

FEA2 would define PaaS as addressing the following standard services to be used as shared capabilities: networking, storage management, management of servers, virtualization of the hardware, maintenance of operating systems, implementation of security assurance, middleware for managing the transition of legacy code and the design of run-time interfaces.

Security services, which will become the most critical part of all Defense Department systems, will be embedded primarily inside each PaaS and then updated in a relatively small number of software assets. Leaving security features—such as access authentication—within applications will make it possible to consolidate within a small number of PaaSs what presently are the most costly information assurance functions.

Excluded from PaaS would be the application codes and the applications-related data. Metadata directories would be managed as an enterprise asset to assure interoperability. The separation of PaaS services from the applications would be controlled by tightly defined standard interface protocols.

FEA2 largely would simplify how the Defense Department manages its systems and how it can reduce costs. Instead of thousands of contractor-designed and -maintained infrastructures that now account for an excessive 57 percent of all costs, the department will concentrate on maintaining only a handful of PaaS environments, probably at least one PaaS for each military service and for each agency. PaaS services would be available in the form of private or public clouds, though infrastructure-as-a-service (IaaS) always would be present as a transition offering on the path toward PaaS.

FEA2 will depend on a strict adherence to open standards to assure interoperability across PaaS clouds. A Defense Department memorandum of October 16, 2009, recommends a preferred use of open-source standards to allow developers to inspect, evaluate and modify the software based on their own needs, as well as to avoid the risk of contractor lock-in. The department cannot allow each contractor to define its own PaaS. Nor can it allow the operation of a hotel where guests can check in but never check out. Existing PaaS solutions offered by vendors such as Amazon and Microsoft Azure provide services that operate in this manner. To ensure competitiveness, only operations based on open-source standards can be allowed. An independent test of open-source interoperability would verify that applications could be relocated from one PaaS to another.

The roles of the CIOs and the acquisition personnel would be enlarged to oversee the rates charged by PaaS services. To assure competitiveness and for comparability of charges, the transaction pricing structure for each service would have to be uniform for every PaaS provider. The ultimate test of FEA2 will be cross-platform portability. Defense Department customers should be able to relocate any application from any one PaaS to another with only minimal transfer expenses.

The technical aspects of PaaS can be described as a method that allows a customer's unique data and applications to be placed on top of a defined computing "stack." The PaaS stack takes care of every infrastructure service. The customer can be left to worry only about applications and data.

The Defense Department will have some diversity in PaaS offerings because various components will make progress at different speeds. This will require the placement of a software overlay on top of the existing information technology coding stacks. Such overlays will act as intermediaries between what belongs to the

customers—the services and agencies—and what belongs to the PaaS during the progression from legacy systems to PaaS. The placement of an intermediation software layer between the PaaS and the applications will allow for the diversity of applications during the long migration it will take to reach the ultimate goal. It may take a long time for such migration to take place as legacy systems eventually are replaced with PaaS-compatible solutions.

The PaaS arrangement makes it necessary for applications and data to be constructed using standard development frameworks. Such standardization is necessary to enable applications to relocate easily from one PaaS cloud to another, whether these clouds are private or public. With such standardization applications and data, users can relocate to take advantage of competitive offerings from different PaaS vendors.

To prevent PaaS contractors from offering cloud solutions that capture customers by means of proprietary run time and middleware solutions, it is necessary to control the interoperability across several PaaS services as well as across any interim IaaS solution. To accomplish this goal, the Defense Department must establish a policy to ensure that interfaces depend on open-source solutions that can be verified for interoperability.

Achieving PaaS standards only through policy is insufficient. The PaaS technologies are global, whereas the reach of the Defense Department is limited by spending less than 1.5 percent of the global information technology costs. The ability of customers to migrate from one PaaS vendor to another PaaS vendor must be preserved. This can be assured if the department works with commercial firms to adopt standards that prevent lock-ins by large prime contractors—the type of lock-ins that could prevent smaller firms from offering PaaS services.

The insertion of a limited number of PaaS services into the Defense Department will result in large cost reductions. Contractors will be shifting to proprietary PaaS services to gain larger profit margins unless the department sees to it that competition can prevail.

Transferring applications to a cloud offers enormous benefits. It also can be a trap. After placing an application on IaaS to take advantage of virtualization of servers, such a move can be wedged into a unique software environment. For all practical purposes applications cease to be transportable from any one IaaS to another IaaS and certainly not to PaaS.

Hundreds of cloud services already operate in a proprietary manner, and the Defense Information Systems Agency is now considering such moves. Defense Department policy must require that all migration moves fit the ultimate objective of operating as a PaaS. IaaS solutions are useful in offering raw computing power but may not be sufficiently flexible to enable redeployment when conditions change.

PaaS services therefore must offer the following features. First, the interface between customer applications and the PaaS must be in the form of open-source middleware, which complies with approved IEEE standards or prevailing industry best practices. Standard open-source middleware must allow any application to run on any vendors' PaaS cloud. Regardless of how an application was coded, it should remain transportable to any Defense Department-approved PaaS cloud, anywhere.

Second, the isolation of the customer's applications from the PaaS software and hardware is necessary to permit the retention of the Defense Department's intellectual property rights. This must apply even if the department chooses to host some of its applications on a public cloud.

Third, the cloud provider must certify that applications will remain portable regardless of configuration changes made within its PaaS. This includes assurances that applications will retain the capacity for fail-over hosting provided by another PaaS vendor.

Finally, there must be assurances that the customers' application code will not be altered when hosted in the PaaS cloud, regardless of the software framework used to build it.

Any Defense Department plans to migrate systems into a PaaS environment henceforth will have to consider the ready availability of off-the-shelf software that will make the migration to PaaS feasible at an accelerated pace.

Commercial software that is already available aims to allow developers to remove the cost and complexity of configuring an infrastructure and for a run time environment that allows developers to focus on the application logic. This streamlines the development, delivery and operations of applications. It also enhances the ability of developers to deploy, run and scale applications into the PaaS environment.

The objective of FEA2 is to deploy an application without becoming engaged in set-ups, such as server provisioning, specifying database parameters, inserting middleware and then testing that it is all ready for operations after coordinating with the data center operating personnel.

With a redirection of Defense Department information technology to a new architecture and with focusing on PaaS services, the department can overcome its budget limitations while freeing funds for attaining information superiority.

FY11 DoD IT Spending

The IT dashboard, published by the Office of the Federal Chief Information Officer, provides comprehensive information about US Government IT spending for FY11.[1]

Since the Department of Defense uses 46.2% of total Federal IT spending, the structure of its spending offers useful insights.

Half of systems projects have budgets of less than $1 million. These projects, usually executed as separate local contracts, depend on unique equipment and software configurations. Although these projects account for only 1% of total spending, there is a large population of users that depend on these systems for everyday support. How to migrate such projects into a shared cloud environment is a problem that is waiting for resolution,

The 78 systems with FY11 budgets of more than $100 million are multiyear programs. How to insure interoperability of multiple applications in near real-time, each with diverse databases is a problem that calls for a better resolution. Since these systems are largely under the control of contactors, the absence of an overall DoD shared architecture is likely to inhibit a transfer into shared cloud operations.

[1] http://it.usaspending.gov/data_feeds

Range of FY11 Systems Budgets ($Millions)	Number of Systems	Budgets of Systems ($Millions)
$500 - $3,000	8	$7,725
$200 - $500	22	$6,803
$100 - $200	48	$7,227
$50 - $100	78	$5,546
$20 - $50	137	$4,326
$10 - $20	141	$2,087
$1 to $10	604	$2,276
Less than $1 million	1,065	$301
Total for FY11 Systems	2,103	$36,292

A more interesting perspective on DoD spending is an examination from the standpoint of the functions IT supports.

Systems Functions	Number of Systems	Budget ($Millions)	% of Budget
Information & Technology Management	435	$19,510	54%
Defense and National Security	248	$9,004	25%
Supply Chain Management	571	$2,825	8%
Human Resource Management	304	$1,737	5%
Health	75	$1,128	3%
All Other Systems	86	$838	2%
Financial Management	225	$782	2%
Administrative Management, Planing	159	$467	1%
FY11 Systems Totals	2,103	$36,292	100%

The largest share of the budget – 54% - is devoted to "Information & Technology Management", which largely represents investments and operations in DoD IT infrastructure. The "warfighter's" domain consumes 25% of the IT budget, with logistics using 8% of the total. If one considers that the primary objective of DoD is to support warfighter missions, then a quarter of all IT spending appears to be minimal. However, one must consider that fact that the current IT statistics does not include the costs of military and civilian personnel. It also excludes all IT that has been integrated into weapons. Consequently the relatively small reported share of IT is likely to be misleading.

The remaining 25% is applied to business operations where Supply Chain Management and Human Resources Management are taking the largest shares. This domain is the responsibility of the Chief Management Officer and the Deputy Chief Management Officer.

Systems included in "Information & Technology Management" are critical to the effectiveness of IT. They include the costs of the entire infrastructure, which supports all other functional domains. The top 25 systems, accounting for 62% of this spending are listed as follows:

Top 25 Infrastructure Systems	FY11 Budget ($Millions)
DEFENSE INFORMATION SYSTEM NETWORK	$2,080.2
NEXT GENERATION ENTERPRISE NETWORK	$1,853.0
DEFENSE ENTERPRISE COMPUTING CENTERS	$800.1
DCSIM/DOIM STAFF OPERATIONS COSTS	$672.4
NON DISN TELECOMM	$593.7
COMPUTER NETWORK DEFENSE	$589.1
CRYPTOGRAPHIC MODERNIZATION	$566.3
NETWORK ENTERPRISE TECHNOLOGY COMMAND	$482.9
BASE LEVEL COMM. INFRASTRUCTURE	$411.3
LONG HAUL COMMUNICATIONS PAYMENT TO DISA	$404.4
DLA COMPUTING INFRASTRUCTURE	$388.3
COMMUNICATIONS AND COMPUTING INFRASTRUCTURE	$372.0
BASE LEVEL COMMUNICATION INFRASTRUCTURE	$305.7
TRI-SERVICE INFRASTRUCTURE MANAGEMENT	$294.0
LEASED TELECOMMUNICATIONS	$277.4
DISN GLOBAL INFORMATION GRID (GIG)	$234.1
BASE INFORMATION INFRASTRUCTURE	$227.2
COMPUTER NETWORK DEFENSE	$226.6
ARMY INFORMATION MANAGEMENT INSTALLATION	$199.1
BASE LEVEL COMMUNICATIONS INFRASTRUCTURE	$199.0
CRYPTOGRAPHIC KEY MANAGEMENT	$194.4
BASE LEVEL COMMUNICATIONS INFRASTRUCTURE	$191.2
INFORMATION TECHNOLOGY AGENCY	$187.8
IA GENERAL SUPPORT	$181.1
INSTALLATION INFRASTRUCTURE MODERNIZATION	$180.4
FY11 Budget for Top 25 Infrastructure Systems	$12,111.6

Most of the above systems can be classified as supporting an infrastructure that has been put in place to support all other systems. The remaining 410 Information & Technology Management systems include a number of systems that can be classified as "infrastructure" (which includes security) rather than supporting applications.

SUMMARY

A review of DoD FY11 budgets shows that a disproportionately large share of resources is committed to paying for numerous infrastructures. When this is compared with commercial practice, DoD spending on Information & Technology Management can be viewed as disproportionate.

One of the explanations for such proliferation is historical. Most of the large multi-year and multi-billion programs have involved contracts that included the creation of unique infrastructures.

Perhaps the most critical reason for such diversity is the separation of accountability for spending. Warfighter and business application budgets are included in Army, Navy, Air Force and Agencies development costs. After that the diverse Acquisition organizations handed over operations either back to the Army, Navy or Air Force, or to Agencies that starting taking over the management of infrastructures.

Why DoD Must Reduce Costs

On July 22, 2010 a Task Group of the Defense Business Board, appointed by the Secretary of Defense, delivered initial observations on "Reducing Overhead and Improving Business Operations". The observations include a number of facts that should be relevant to cyber commanders:

1. Increased spending from 1988 through 2010 supported less active duty personnel, fewer reserves, fewer ships, less Army divisions and a smaller number of fighters (SOURCE: National Defense Budget Estimate for FY2011 as of April 2010):

Category	FY2011 Budget Estimate	Change 1988-2010
Total Budget Authority ($B -Constant $)	$553	7%
Total Budget Authority ($B -Current $)	$553	92%
Active Duty Personnel (K)	1,484	-33%
Civilian Personnel (K)	785	-28%
Reserve and Guard Personnel (K)	845	-27%
Active in Commission Ships	284	-50%
Army Divisions (active)	10	-50%
AF Fighter/Attack (Total Active Inventory)	1,280	-58%

Conclusion: Spending more bought substantially less.

2. 40% of active duty personnel were never deployed. 11.4% was deployed three or more times. 339,142 active duty personnel, costing on the average $160,000/year were performing commercial (not war fighting-related) activities.

Conclusion: A small number of active duty personnel do most of the fighting.

3. There has been a steady rise in non-payroll costs for DoD personnel. In FY10 this accounted for $22.5 Billion such as for TRICARE, family separation allowances and survivor benefits. Total retiree outlays for military personnel were $46.7 Billion in FY10.

Conclusion: The trend of rising non-payroll costs will reduce the amount of money available for active duty and civilian personnel.

4. The Federal Deficits, as a % of GDP, can be expected to increase from 9.9% for FY2009 to 24% in FY2040 if continues without change. (SOURCE: Peterson Foundation – A Citizen Guide April 2010).

Conclusion: A 24% share of the GDP deficit is unlikely to be sustainable.

5. The FY 2010 $3.5 Trillion Federal Budget, in constant 2009 $s, will allocate 40% of total to Social Security & Medicare, 6% to net interest, 34% to all other and 20% to Defense. The FY 2040 $12.3 Trillion Federal Budget, in constant 2009 $s, will allocate 52% of total to Social Security & Medicare, 30% to net interest and 7% to all other. 11% of the Federal Budget would be available for Defense (SOURCE: Peterson Foundation – A Citizen Guide April 2010).

Conclusion: If the Federal Budget, in 2009 $s of $112.3 Trillion is available, it would support $1.4 Trillion for Defense, or double the FY 2010 budget. However, the rise in non-payroll costs could absorb most of Defense budget increases over a thirty-year period.

6. DoD overhead costs, which are concentrated in Administration, Logistics, Finance and HR systems are estimated to cost about $212 Billion, which is approximately 40% of the total DoD budget.

Conclusion: Information technology should be considered as a premier tool for reducing overhead costs by means of systems simplification and business process improvement.

SUMMARY

The costs of DoD information technologies for FY10 is $33.7 Billion for O&M costs plus compensation of an estimated number >200,000 of military and civilian personnel at an average cost of $130,000/year of approximately $26 Billion. This represents about 10% of the total DoD budget.

Since information technology is deflating at a rate that is greater that 15% it offers an attractive opportunity for cost reduction. If information technology is simultaneously applied to business process redesign the resultant manpower savings can be even greater than what can be accomplished though cost cuts in technology spending.

PART 2: INFORMATION TECHNOLOGY

Reducing the Costs of DoD Software

According to Capers Jones — one of the foremost authorities on the productivity of computer programming — the following insights are relevant when evaluating the potential cost reductions in DoD Software (for details see: Productivity Comparisons of Selected Software Methods, Version 10, Copyright 2010, Capers Jones & Associates):

1. The difference between acceptable and excellent development costs is about $1,748,043 - $1,034,300 = $713,746 (CMM3 - CMM5) per 1,000 function points. (NOTE: CMM is the Capability Maturity Model of the Software Engineering Institute).

With thousands of DoD contractors and subcontractors writing and maintaining computer code on >5,000 applications with an estimated >500 function points/application the potential savings are very high, especially since much of the contractor's code is at the CMM1 level. However, such savings cannot be realized in the existing contract acquisition environment.

2. It is unlikely that DoD contractors can ever deliver code plus maintenance at CMM5 levels. Therefore, the only remedy is for DoD to take applications and break them into code for the shared infrastructure (such as data management and security) and code that manages application procedures . The shared infrastructure parts of an application, perhaps as much as 50% of the code, could be then constructed by contractors who deliver at least 50% re-use of certified components.

3. If DoD succeeds in imposing the separation of infrastructure code from procedural code, half of the code would then benefit from the difference between CMM3 costs and 50% Reuse costs ($1,034,300 - $752,773 = $281,527, with incremental savings of $351,908,750.

4. Additional savings are available by adopting "85% certified re-use methods". This reduces the costs per 1,000 functional points to a Total Cost of Ownership of only $287,518. By far the most promising development that would support such approach is SourceForge with its >250,000 library of available code components that has been refereed with >99% reliability and downloaded by ten thousands of users. SourceForge is an open source library, readily accessible at no cost. For instance, it includes code listing for 62,949 development projects, 13,453 projects for database management and 7,037 projects for security.

5. I believe such savings would be realistic because DoD's increasing emphasis on cyber security will dictate that infrastructure will have to be taken out of the hands of thousands of contractors plus subcontractors and concentrate in the hands of a few software firms that can deliver CMM5 secure code.

6. In addition to the TCO of the development of code there are additional costs incurred by DoD military and civilian personnel as well as data center charges. CMM1 software creates additional cost penalties incurred by users of poorly conceived applications.

SUMMARY:

There is a large cost reduction potential in DoD's software development processes.

Computers for Shooters

Two weeks ago I listened to a Marine Corps Brig Gen making a plea for a lightweight personal computer for use by shooters at the squad level. All of the talk he heard about net-centric networks was meaningless because it did not reach where it was needed.

The planner's slides that promised connectivity for everyone were fiction. The existing radios were just too heavy and the antennas gave snipers targets. If the civilians could walk around with Black Berries why could not DoD provide comparable services?

There is no reason why we should not provide our fighters with a shirt pocket five-ounce device with a 3.7" color touch screen, GPS, camera and at least a seven-hour power supply for less than $300.

There are several programmable commercial products that can do that as illustrated below:

A Programmable Android Cell Phone

There are several issues that must be solved before we can proceed:

1. Training

The key to adapting computers in the combat environment is simplicity and persistence. Soldiers should be able to use a variety of computing devices regardless how the technology changes. Recruits ought to receive their shirt pocket appliance at the same time when they get their rifle. The graphic buttons on the appliance would be standard icons, with added variations for the Marine Corps, Army, Navy and Air Force. Unique buttons could be designed for specific purposes or for designated individuals. This approach guarantees training continuity over decades. Such proprietary buttons can be programmed using device specific Application Programming Interfaces.

2. Communications

3G cell towers or Wi-Max transmitters can be erected in the battlefield or on military bases for encrypted transmission. Protected commercial circuits can be also used if additional safeguards are installed. The visual persistence of the shirt pocket devices can apply also to desktops, laptops or note pads. Regardless of technology all accesses to the DoD private networks can be identical.

3. Security

The shooter's computer is stripped of every application that is not accessible by means of a standard graphic "button". Standard code reduces the attack profile to intrusions. Consequently the code for every function will represent mature software that can be modified only by the designers. Each "button" then offers access privileges based on the roles that are assigned to an individual, regardless of location. Central network control monitors all traffic including awareness as to the uses of the phone.

4. Social Computing

One graphic "button" can be reserved for access to the public Internet. It offers access to a virtual server that is completely isolated from military networks provided that bandwidth capacity is available.

5. Performance

Access to a screen should take less than a second. Combat requires response times of less than 250 milliseconds. Redundancy in communications must guarantee scheduled availability at all times. To meet these requirements will require a complete overhaul in the ways in which DoD manages its data centers and its networks.

SUMMARY:

Creating a uniform communications environment for our war fighters is not only feasible, but is also reduces costs. It scales down the time needed for learning how extract data from diverse sources. It improves security by relying on "thin" computing for access to intelligence regardless of location. Simplification of the user interface creates reusable software components, which increase the reliability of all communications.

The shooter's computer is feasible because the technology risks are manageable. There is no reason to wait any longer.

Enterprise E-mail for DoD

The Army reports that it is spending over $400 million annually in operating costs to sustain organization-specific e-mail systems.[2] That supports 1.6 million mailboxes as a cost of $250 per mailbox. This does not include the costs of communications (reported by DISA) or any development costs.

An examination of available Software-as-a-Service (Saas) e-mail services shows that richly featured e-mail services are now available for prices as low as $8/seat, which may include several mailboxes. There other hosted e-mail services available, with higher prices, but with a wide range of available features. All of

[2] *DefenseSystems.com, July 2011, page 22*

the available SaaS e-mail services report a lower than 0.01% downtime. Nevertheless, an estimate of possible savings of a standard DoD enterprise-wide e-mail offering could offer more than $1 billion of readily available operating costs from e-mail alone. What steps can be taken to deliver such savings?

The initial migration steps toward a cloud hosted SaaS service, to be offered by DISA is now taking place. The Army is replacing Army Knowledge On-Line (AKO) with an offering that essentially replicates a Microsoft's proprietary web-based offering. With only 4% of mailboxes moved by end of June 2011, the Army has experienced outages of e-mail service of over five hours. What is the cause of such outages during migration is not clear, though the capacity of every one of the nine DISA DECCs to maintain better than 99.999% availability needs to be demonstrated.

It appears that the Army is pursuing an approach that the migration of individual mailboxes will take place without a noticeable effect on a user. This means that even small variations on the Active Directory must be "cleaned up" for conversion. That is hard to do since the Active Directories are maintained at about 300 separate service sites, each with a slightly different variation in software. If any migration also requires the transfer of existing records of past e-mails, documents and attachments, the challenges to making a smooth transition are formidable. There are also added complexities, such as variations in local licensing agreements and security processes that make any migration a demanding task.

The Army's migration to the DISA cloud for the delivery enterprise e-mail services is seen a prototype for the rest of DoD to follow. it is receiving ample attention. For instance, a Committee of the Congressional Armed Services Committee has slashed the Army's e-mail services plans by 98% of the FY12 request for $85.4 million until better justification of spending plans is received.

SUMMARY

I the Army's approach to an enterprise e-mail system is to serve as a prototype for everyone in DoD, the approach to migration to a customized a cloud enterprise standard must be simplified. It is inconceivable how the enormous variety of existing e-mail implementations in the Air Force, Navy, Marine Corp and a multiplicity of Agencies can be wedged effectively into a standard SaaS e-mail offering.

Consideration should be given to choosing a single low cost, open source, highly secure, DoD interoperable and upgradeable SaaS system that is extensible to additional commodity services such as cooperation, information sharing and document management. E-mails have a limited shelf life of only few days. DoD components could operate in a private secure SaaS cloud instantly, with short switch over time. For archival purposes, DoD components could then operate dual e-mail systems until such time when the DoD standard processes takes over all e-mail functions. If selected archival records require retention, conversion utilities would be available to do that at a fraction of the enormous cost it would take to impose on the entire migration upward compatibility without any change.

Can DISA Operate the DoD Cloud?

2011 global IT spending is expected to reach $3.4 trillion. DoD has a $36.3 billion IT budget, plus an estimated spending for 150,000 military and civilian personnel of $15 billion. Consequently, the DoD has by far the single largest IT budget in the world. It consumes 1.5% of total global costs.

The 2010 DoD Operations & Maintenance IT expense was $25.1 billion. Over a half of that was spent in data centers.

The potential size of all DoD cloud operations would be approximately $12.5 billion. Such a huge computer operation does not exist anywhere. How such as complex could be managed is not known. The two largest public commercial cloud firms, Amazon EC and Rackspace, have revenues of only about $1 billion. To take over cloud computing for DoD the DISA operation would have to grow at an unprecedented scale. Its organization would have to be restructured.

The most detailed information about DISA computing operations come from a presentation made in February 2009. DISA operated thirteen Enterprise computing (DECC) data centers, averaging 34,000 square feet of raised floor. That is comparable to the US data center I built in 1972 for Xerox but which is now economically not viable because it is too small. From the standpoint of economies of scale, newly built data centers to house cloud operations is in the 300,000 to 500,000 square foot range, which makes DISA data centers too small. There are now hundreds of such data centers.

The current DECCs are the result of the consolidation of marginal data centers authorized by DMRD 918, which was approved in 1992. This was based on work that started in 1990 when I was the Director of Defense Information. The DECCs were supposed to be completed by 1996 so that a next phase of consolidation of data centers that would consolidate Components' operations could take place. That never happened. While the DISA data center count held steady at thirteen, the total number of DoD data centers grew to a total of 772.

New commercial cloud data centers are currently being constructed to house over 100,000 servers. At present, DISA has in all data centers only a total of 6,100 servers plus 34 mainframes, which are used mostly for legacy applications and would not be suitable for cloud operations. Again, the scale of DISA data centers is too small.

In 2009 the total number of files in all DISA data centers was 3.8 petabytes. The only benchmark comparison we have is the 29 petabytes that can be housed in 320 square feet. DISA storage capacity does not have anywhere such density.

SUMMARY

The current claim that DISA could become the processing cloud for DoD is not as yet credible. How DoD could migrate into cloud computing is a question that still needs more work.

DoD Server Computing

No data is available about the amounts of data processed through DoD's $36.3 billion IT. However, there is information available of the amount of information now processed by servers. With the DoD IT budget ten times larger than the IT budget of any one US corporation, it is possible to infer what are some of the information processing issues. In 2008 the US has installed 38% of the global server population.[3]

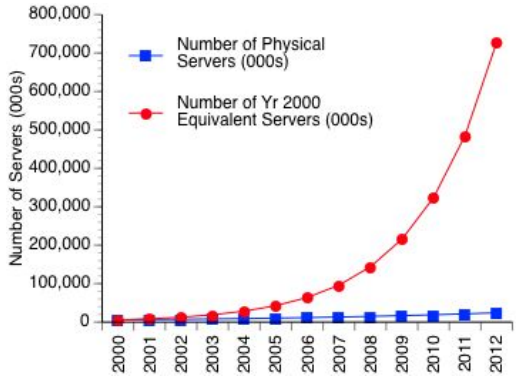
[3] http://hmi.ucsd.edu/howmuchinfo.php

According to studies by the University of California, San Diego, in 2008, the world's servers processed 9.57 zettabytes of information, which is 10 to the 22nd power. That is ten million million gigabytes or 12 gigabytes of information daily for each of the 3.2 billion workers in the world's labor force. The world's 151 million world businesses process 63 terabytes per company of information annually.

There are about 2.5 megabytes in a long book. It would be necessary to stack such books from the Earth to Neptune and back about 20 times to equal one year's capacity of existing servers. Each of the world's 3.2 billion workers would have to read through a stack of books 36 miles long each year to read 9.57 zettabytes of text.

The total number of servers in the world in 2008 was 27.3 million, with 10.3 million in the USA. 95% of the world's total zettabytes in 2008 were processed by low-end, servers costing $25,000 or less. The remaining 5% was processed by more expensive servers. 87.8% of processing on US servers in 2000 was performed by means of low end servers. By 2008 that number has risen to 96.0%, which indicates that the low-end servers will continue to carry a greater share of the processing workload.

The follow graph shows the total number of USA servers, including an estimated equivalent number of low-end servers based on 50% performance/price/year improvements:[4]



The total annual world server sales in 2008 were $53.30 billion, with entry level servers at $30.8 billion. It is interesting to note that large computer complexes, such as operated by Google, Yahoo or Facebook depend on small scale servers for information processing. High end servers show slower gains in performance/price/year than low end servers and are not preferred in billion dollar computer complexes.

It follows that most information processing in the world is performed by low end servers (392 billion million transactions in 2008), with only 10 billion million transactons executed by high end servers. This pattern is expected to persist. The purchase of computers for cloud computing is not likely to shift in favor of mainframe equipment.

Transaction processing workloads amounts to approximately 44% of all the total bytes processed. Such transactions are "commodity" applications, such as e-mail, that are not handled efficiently in low-end servers. The overhead costs for administration of a large number of low-end servers are excessive unless they are aggregated into huge complexes that use the identical harware management and software control architecture.

[4] Figure 5 in http://hmi.ucsd.edu/howmuchinfo.php

SUMMARY

More than 2,000 of DoD system projects, with budgets under one million per year, are likely to be processed in low-end servers. The the performance/price of servers has been increasing since 2000, the share of the work performed by low-end servers has been increasing. The current OMB guidelines that count as data center only operations with more than 500 sq. ft. are becoming irrelevant. A rack mounted server costing less than $25,000 occupies only 10 sq. ft. of space. It has greater processing power than a 1990 mainframe. Consequently the current DoD approach to reducing the number of data centers will miss the attempt by contractors to continue installing low-end servers in increasingly stand-alone configurations.

Most of the workload on low-end servers consists of Web services and of commodity computing. There are large economies of scale availabale through virtualization of workloads and consolidation for processing by large servers. This will deliver economies of scale and reduce that number of operating personnel needed to support a large number of low-end stand-alone servers.

What is now emerging from the proliferation of servers is the "big data" challenge. Server capacities are almost doubling every other year, driving similar growth rates in stored data and network capacity. Computing is now driven by increasing data volumes, the need for integration of ever increasing sources of heterogeneous data. There is a need for rapid processing of data to support data-intensive decision-making.

It is necessary to re-examine the current approaches to the realization of IT for economies of scale in DoD. This places pressure in the direction of more centralized management of IT resources.

Continuity of Operations

One of the primary benefits from cloud data operations is the capacity to perform complete backups and to support Continuity of Operations Plans (COOP). COOP recovers operations whenever there is a failure.

In the past data centers were small. Files that needed backup were relatively small, hardly ever exceeding a terabyte. Real-time transactions were rare. High priority processing could be handled by acquiring redundant computing assets that matched the unique configurations of each data center. COOP was managed as a bilateral arrangement. Since hardware and software interoperability across data centers was rare, to restart processing at another side was time consuming. Luckily, data centers were operating at a low level of capacity utilization, which made the insertion of unexpected workloads manageable.

The current DoD environment for COOP cannot use the plans that have been set in place during an era when batch processing was the dominant data center workload. Files are now getting consolidated into cross-functional repositories, often approaching petabytes of data. The urgency of restoring operations is much greater as the processing of information supports a workflow that combines diverse activities. Desktops and laptops, that used to be self-sustaining, are now completely dependent on the restoration of their screens with sub-second response time. There has been a rapid growth in the number of real-time applications that cannot tolerate delays. What used to be bilateral COOP arrangement is not acceptable any more as DoD is pursuing data server consolidations. The merger of data is based on virtualization of all assets that eliminates much of the spare computer processing capacity.

The current conditions dictate that for rapid fail-over backup data centers must be interoperable. Hardware and software configurations must be able to handle interchangeable files and programs. A failure at any one location must allow for the processing the workloads without interruption at another site. To achieve

an assured failover the original and the backup sites must be geographically separate to minimize the effects of natural disasters or of man-caused disruptions. What used to be the favorite COOP plan of loading a station wagon and driving relatively short distances with removable disk packs is not feasible any more. Petabyte data files cannot be moved because they do not exist in isolation. Data center operations are tightly wrapped into a complex of hypervisors, security appliances, emulators, translators and communications devices,

Under fail-over conditions the transfers of data between data centers must be manageable. The affected data centers must be able to exchange large amounts of data over high capacity circuits. Is it possible to start thinking about a COOP arrangement that will operate with fail-overs that are executed instantly over high capacity circuits? How much circuit capacity is required for claiming that two (or more) data centers could be interchangeable?

The following table shows the capacities of different circuits as well as the size of the files that would be have to be transferred for a workable COOP plan:

Transmission Time	10 Terabytes (in hrs)	100 Terabytes (in days)	1 Petabyte (in weeks)
1 MB/Sec	2,778	1,157	1,653
10 MB/Sec	278	116	165
100 MB/Sec	28	12	17

Transfers of files between small data centers (each with 10 Terabytes of files) would take up to 28 hours using an extremely high capacity circuit (100 MB/Sec). That is not viable.

DoD transaction processing, including sensor data, involves processing of at least one petabyte per date at present, and probably much more in the future. It would take 17 weeks to back up a single petabyte of data from one data center to another even if the highest available circuit capacity of 100 MB/sec is used. That is clearly not viable.

We must therefore conclude that the idea of achieving fail-over backup capabilities by electronic means cannot be included in any COOP plans.

SUMMARY

COOP for the cloud environment must be based on multiple data centers that operate in synchronization with identical software, with matching technologies of computing assets and with comparable data center management practices. This does not call for a strict comparability of every application. What matters will be an identity of the DoD private cloud to act as a Platform-as-a-Service utility, with standard Application Processing Interfaces (APIs).

OMB has mandated that for FY12 all new DoD applications will have to cloud implementation options. Rethinking how to organize the DoD cloud environment for COOP will dictate how that can be accomplished.

Data Center Consolidation

There were 2,094 Federal Data Centers. DoD operated 772 data centers. In each case, a "data center" was defined primarily as any room that is greater than 500 square feet and devoted to data processing.

The 500 square definitions do not hold up any more as the technologies are shrinking. It is possible to fit into 20x8 ft. shipping containers with the formidable the capacity of up to 29.5 petabytes of storage and up to 46,080 CPU cores of processing power.

The economies of scale of data centers of 300,000 to 500,000 sq. ft. show a dramatic lowering of costs of information processing, huge decreases in operating expenses, reductions in staff, while also increasing the reliability and latency of what would then become a "server farm." Examples of the construction of such huge data centers can be seen from firms such as Apple:

Apple data center in Maiden, North Carolina

Facebook data center in North Carolina

There are many new firms that build and equip data centers. These are investment ventures. They build highly efficient large facilities and then lease them either as totally dedicated facilities for a particular organizations, or as "cages" available for partial occupancy. In the case of companies, such as Apple or Facebook, the computer configuration is standardized to meet the company's proprietary system architecture. Partial occupancy pages offer greater flexibility to a customer to install specific software.

As an illustration of only a small sample of firms, the data center venture firm of Sabey is currently building a 350,000 sq. ft. data center for Dell.[5] The firm of CoreSite offers several locations with finished "wholesale" data center space for users who seek turnkey space that can be deployed quickly.[6] The Equinix operates 22 data centers in the USA, seven in Europe and five in Asia. Some of the large Equinix data centers are larger than 200,000.[7] Rackspace operates nine data centers, which includes managed hosting.[8] The firm of Interxion operates 28 sites in Europe.[9]

SUMMARY

US Government and particularly DoD data centers have been constructed over the past thirty years. The do not reflect the economies of scale that have become available primarily on account of ample optic transmission capacity. Meanwhile the costs of hardware have shrunk, while the cost of electricity and operating manpower has been steadily rising to meet an exponential growth in demand for services.

The existing data centers have their origins in separate contracts, which dictated how computing facilities would have to be organized. It is clear that the currently proliferation of data centers that cannot support any more the increasing demands for security and for reliability. With capital costs for the construction of economically viable data centers now approaching $500 million, it is unlikely that the needed capital could be available with the looming budget cuts in DoD.

The current FY10 O&M (operating and maintenance costs) of I.T. are $25.1 billion. This makes the leasing of DoD-dedicated data centers to be constructed by any one of many commercial firms affordable.

The total DoD user population is about eight million. Its workload would be only a fraction of the estimated high frequency users of firms such as Facebook that process transactions for over 600 million customers in less than 200 milliseconds. Even with provisions for redundancy, the consolidation of computing services into half a dozen Facebook-like data centers is an option.

---

[5] http://www.sabey.com/real_estate/data_centers_main.html

[6] http://www.datacenterknowledge.com/archives/category/crg-west/

[7] http://www.equinix.com/data-center-expertise/platform-equinix/

[8] http://www.rackspace.com/managed_hosting/private_cloud/index.php

[9] http://www.interxion.com/About-Interxion/

PART 3: ATTACKS ON OPERATIONS

Attack on RSA

RSA (names after the inventors of public key cryptography the inventors of public key cryptography: Ron Rivest, Adi Shamir and Leonard Adleman) is one of the foremost providers of security, risk and compliance solutions. When RSA SecureID token was recently attacked and compromised, this raised the question how good are the safeguards of the keepers of everybody's security safeguards. [10]

The RSA attack was waged in the form of an Advanced Persistent Threat (APT). Information was getting extracted from RSA's protectors of the RSA's SecurID two-factor authentication products.

"The attacker in this case sent two different phishing emails over a two-day period. The two emails were sent to two small groups of employees; who were not high profile or high value targets. The email subject line read '2011 Recruitment Plan. The emails was crafted well enough to trick one of the employees to retrieve it from their Junk mail folder, and open the attached excel file. It was a spreadsheet titled '2011 Recruitment plan.xls. The spreadsheet contained a zero-day exploit that installs a backdoor through Adobe Flash vulnerability (CVE-2011-0609)." [11]

The attack on RSA can be considered to be a textbook example of a targeted phishing attack, or a "spear fishing attack". What the attacker goes after and obtains once inside the compromised network largely depends on which user he was able to fool and what that victim's access rights and position in the organization are.

The malware that the attacker installed was a variant of the well-known Poison Ivy remote administration tool, which then connected to a remote machine. The emails were circulated to a small group of RSA employees. At least one must have pulled the message out of a spam folder, opened it and then opened the malicious attachment.

In studying the attack form RSA concluded that the attacker first harvested access credentials from the compromised users (user, domain admin, and service accounts). Then proceeded with an escalation on non-administrative users that had access to servers that contained the critically protected "seed" number that is used to generate SecureID numbers ever 60 seconds.

The process used by the attacker was not only sophisticated but also complex, involving several methods: "The attacker in the RSA case established access to staging servers at key aggregation points; this was done to get ready for extraction. Then they went into the servers of interest, removed data and moved it to internal staging servers where the data was aggregated, compressed and encrypted for extraction. The attacker then used FTP to transfer many password protected RAR files from the RSA file server to an outside

[10] http://www.rsa.com/node.aspx?id=3872

[11] *Adobe Flash vulnerability (CVE-2011-0609)*

staging server at an external, compromised machine at a hosting provider. The files were subsequently pulled by the attacker and removed from the external compromised host to remove traces of the attack."[12]

SUMMARY

The successful penetration of a highly guarded and well protected source of an RSA security offering should be seen as a warning that a persistent and highly skilled attacker can break down even the strongest defenses.

In this case we have a "spear fishing" exploit, which shows that the attacker must have possessed a great deal of inside information in order to direct the placement of the Poison Ivy tools. Using a known vulnerability (in Adobe Flash) as a vehicle only shows that multiple exploit vehicles can be exploited simultaneously to achieve the desired results.

As is always the case, it was a human lapse that allowed the attack on RSA to proceed. Opening a plausibly labeled attachment to e-mail is something that can happen easily, even by people who have special security training.

The only known remedy in a situation like the RSA attack, assuming that somebody, somewhere would be easily fooled to open an attachment, is to enforce the discipline of permitting the opening of e-mails only from persons whose identify is independently certified. Even then there is always a possibility that an invalid certification of identity may somehow creep into DoD. Consequently, a high priority must be placed on instant revocation of any PKI certification of identity.

According to IEEE Security & Privacy, July 2011, the RSA hacker exploit was based on a bug in the Adobe Flash Player. Attackers broke into the RSA network by sending e-mail messages to a number of RSA employees. Attached was an Excel spreadsheet. The Excel spread sheet contained an embedded Flash file with a vulnerability that was previously unknown to Adobe. This vulnerability allowed the attackers to take over an RSA employee's personal computer and install a version of the Poison Ivy remote administration tool (Poison Ivy is a remote administration utility, a backdoor Trojan, which bypasses normal security mechanisms to secretly control a program, computer or network). This enabled the attackers to steal user credentials, access other RSA computers and then transfer to themselves sensitive information.

This situation could have been averted. RSA employees should have strong access authorization that would identify the Poison Ivy source as illegal. The RSA network administrators should have been able to detect a communication anomaly and to immediately intercept it.

[12] https://threatpost.com/en_us/blogs/rsa-securid-attack-was-phishing-excel-spreadsheet-040111

Detecting Web Based Malware[13]

Google engineers analyzed four years worth of data comprising 8 million websites and 160 million web pages from its Safe Browsing service, which warns users when they hit a website loaded with malware. Google said it displays 3 million warnings of unsafe websites to 400 million users a day.

The detection process is becoming more difficult due to evasion techniques employed by attackers that are designed to stop their websites from being flagged as bad.

The company uses a variety of methods to detect dangerous sites. It can test a site against a "virtual machine honeypot" where it can examine malware. It can make a record of an attack sequence. Other methods include ranking a website by reputation based on its hosting infrastructure, and another line of defense is antivirus software.

One of the ways hackers get around detection is to require the victim to perform a mouse click. This is a kind of social engineering attack, since the malicious payload appears only after a person interacts with the browser.

Browser emulators can be confused by attacks when the malicious code is scrambled, a method known as obfuscation. Google is also encountering "IP cloaking," where a malicious website will refuse to serve harmful content to certain IP ranges, such as those known to be used by security researchers. Google found that some 200,000 sites were using IP cloaking.

Antivirus software programs rely on signatures as one method to detect attacks. That software often misses code that has been "packed," or compressed in a way that it is unrecognizable but will still execute. Since it can take time for anti-virus vendors to refine their signatures and remove ones that cause false positives, the delay allows the malicious content to stay undetected.

While anti-virus vendors strive to improve detection rates, in real time they cannot adequately detect malicious content. Attackers use anti-virus products as test-beds before deploying malicious code.

SUMMARY

Malware detection software is progressing, but attackers are learning also. Interception of suspicious web pages is available, but is still insufficient. The best defense remains in extreme personal caution in opening any messages.

[13] *Based on http://tech.slashdot.org/story/11/08/19/1328237/Google-Highlights-Trouble-In-Detecting-Malware?utm_source=headlines&utm_medium=email*

TDL-4 Botnet Threats

The TDL-4 botnet is a collection of Trojans with the capacity to inflict damage through increased technical sophistication as well as improved commercial exploitation.[14] A botnet contains compromised computers connected to the Internet used mostly for malicious purposes. When a computer becomes compromised, it becomes a part of a botnet. Botnets are usually controlled via standards based network protocols such as Internet Relay Chat (IRC). TDL-4 uses the KAD peer-to-peer network for managing its control communications.

Millions of personal computers have been infected. The TDL-4 botnet is sneaky, evasive, hard to detect and difficult to disinfect. TDL-4 is the fourth generation of the TDL malware. TDL-4 packs all kinds of tricks to conceal deep within hard drives, evading most virus scanning software as well as more proactive detection methods. It communicates in encrypted code, and contains a rootkit program that allows an operator access to a computer even while hiding itself from the user, network administrators and automated security measures.

TDL-4 is malicious because it facilitates the creation of a botnet--a network of infected computers that can be used in concert to carry out tasks like distributed denial-of-service attacks, the installation of adware and spyware, or spamming. It currently has 4.5 million machines under its control and counting. The infecting file is usually found lurking around adult sites, pirated media hubs, and video and media storage sites.

The TDL-4 malware originators have extended the program functionality to encrypt communications between bots and the botnet command and control servers. The controllers of TDL have created a botnet that is protected against countermeasures and antivirus companies. Antivirus vendor, Kaspersky, has suggested that TDL-4 has installed nearly 30 different malicious programs onto the PCs it controls.

TDL-4 installs itself into the master boot record (MBR), which makes it difficult for the Operating System or any antivirus or security software to detect its code. Once inside a personal computer, TDL-4 takes up residence in the MBR, which means it can run before the computer is actually booted up. This MBR is rarely combed over by anti-virus software giving TDL added invisibility. Then, TDL-4 runs its own anti-virus program. It contains code to remove around 30 of the most common malicious programs, wiping an infected machine clean of everyday malware that might draw a user's attention or cause an administrator to take a closer look. It can then download whatever malicious software it wants to in the place of the deleted programs. This version of TDL-4 also has added modules, which can be used to hide other malicious cyber actions.

An advanced encryption algorithm ensures that security and anti-virus products are unable to 'sniff' packets that it sends out onto the network. This helps to cloak information that is being sent from Command and Control (C&C) servers, and the information being returned by the TDL-4 Trojan. Any attempt to take down the regular C&Cs can be circumvented by updating the list of C&Cs. Any C&C has a means to directly communicate over the encrypted channel to any host, so that it is virtually indestructible.

TDL-4's controllers use the botnet to plant additional malware on PCs, rent it out to others to conduct spam and phishing campaigns or for distributed denial-of-service attacks. Duckling points out rightly that you cannot buy the source code per se and that you can only rent time on a botnet that is built using the TDL4 toolkit, in essence replicating the business model of Software-as-a-Service.

[14] *IEEE Computer, August 2011, p.16*

The owners of the rootkit go to great lengths to make sure that its turf, which are literally the millions of computers that are part of its army, are protected from other rogue malware. The defense mechanism includes its own antivirus to take out other competing malware and eliminate the risk of potential conflicts as well as the use of public P2P networks to link the slave computers to Command and Control servers.

The TDL-4 network is rented out at a high price to criminal organizations. With 4.5 million zombie PCs working for them, the owners of TDL-4 can launch impressive spamming and phishing campaigns, which can rake in fees. TDL-4 can be also used to plant other malicious pieces of malware, including "spybots", hijacking toolbars, and even fake antivirus software. When a contract runs out, TDL-4 can remove these programs easily. TDL-4 is also removing the competition (malware it doesn't sanction) while opening captured computers to software it prefers. It's definitely the cyber version of organized crime.

The continual development of the TDL-4 network, its advanced tactics, and its wide dispersal is the work of a concentrated criminal network with thousands of dollars devoted to development of its cyber operations. "Partner-programs", most often operating through websites offering adult content, bootleg videos, or file storage, are paid $20-$200 for every 1000 computers they infect with TDL. Kaspersky estimates that TDL-4 has cost its controllers about $250,000 to set up their network. Daily revenue from a botnet the size of TDL-4 can be in the many tens of thousands of dollars.

SUMMARY

At one point the "Conficker" Trojan was going to destroy the entire Internet as we knew it, but it is now contained. TDL-4 will continue to confound and frustrate security experts for years but this too shall pass, causing damage meanwhile. The problem is that the TDL-4 continues to evolve as defenses become more capable. It is multigenerational persistent malware, with new attack forms getting launched as profits from botnets keep rising.

Sandboxing for Security

Sandboxing protects a system by limiting what an application can do, such as accessing files on disk or resources. Limiting the capabilities of an app to just those operations needed for viewing social computing messages will keep the rest of a system totally secure in the event that a message or an app are compromised.

The exploitation by one virus is what makes it possible for downloaded malware to corrupt an entire machine. Web browsers and their plug-ins can infected Web pages. Malicious PDF or Word document can become a conveyor of infection. Firewalls, anti-malware software and other products aren't much help in cases of "spear-fishing" or zero-day attacks. Social computing communications, such as messages received over Face Book or Twitter, are one of the principal sources of malware, since they usually originate from personal computers from members of the families of DoD personnel.

If a DoD person, using a secure desktop, laptop or smart phone, receives a social computing message, one cannot be ever sure that the message is not also acting as a conveyor of malware. The right solution is to place all incoming traffic that originates from addresses other than .mil (from any unauthorized source) directly into a sandbox where it can be examined, but not transferred anywhere on the DoD network.

A sandbox is an isolated zone designed to run applications in a confined execution area where all functions can be tightly controlled, if not prohibited. Any installation, modification, or deletion of files and/or system information is restricted. From a software security standpoint, sandboxes provide an extremely

limited code base. It prevents any decision-making on the user's behalf except to examine the incoming message. This protection is invisible and cannot be changed by the recipient.

Sandboxes should be also used to prevent the downloading of "Applets" from diverse libraries such as Apple, Google and Amazon. Any such download would be automatically routed to a user's sandbox until such time that network control would can test, verify and legitimize a new application.

SUMMARY

All sandboxes must run as isolated virtual computers on separate servers that are controlled within an IaaS or PaaS cloud environment, on a private DoD cloud. Under no circumstance should DoD allow the creation of sandboxes on client desktop or laptop machines. The virtual desktop will then display the contents of a virtual desktop as a separate and isolated window, which will prohibit pasting or cutting sandbox text or data unless authorized to do so by the network control center.

Trojans Inside Semiconductors[15]

The globalization of the semiconductor industry and associated supply chains has made integrated circuits increasingly vulnerable to Trojan programs inside a microprocessor that executes designed microcode**.** A Trojan is a destructive program that masquerades as an application. The software initially appears to perform a desirable function for the user prior to installation, but steals information or performs illegal the system functions. Vulnerabilities in the current integrated circuit (lC) development process have raised serious concerns about possible threats from hardware Trojans to military, financial, transportation, and

An adversary can introduce a Trojan through an IC that will disable or destroy a system at some specific future time. Alternatively, an attacker can design a wire or some IC components to survive the testing phase but fail before the expected lifetime. A hardware Trojan can also covertly cause a system to leak confidential information.

Trojans can be implemented as hardware modifications to application-specific integrated circuits (ASICs), commercial off-the-shelf (COTS) parts, microprocessors, microcontrollers, network processors, or digital signal processors (DSPs), or as firmware modifications-for ex- ample, to field-programmable gate array (FPGA) bit streams.

To ensure that an IC used by a client is authentic, either the developer must make the IC design and fabrication processes trustworthy or the client must verify the IC for trustworthiness. Because the former approach requires a trusted design center and foundry, it is expensive and economically infeasible given current trends in the globalization of IC design and fabrication. On the other hand, verifying trustworthiness requires a post-manufacturing step to validate conformance of the fabricated IC to the original functional and performance specifications- nothing more and nothing less.

Most Trojan detection methodologies assume the existence of secure IC circuits, which are obtained by arbitrarily selecting chips from a large batch of fabricated ICs and thoroughly testing them. This procedure

[15] Extracted from Computer, IEEE Computer Society, July 2011

assumes that Trojans are inserted into random ICs, but to do so, an attacker must use a different set of masks for selected chips, making such an effort unattractive. It is more viable for an attacker to insert a stealthy Trojan into every fabricated IC that passes manufacturing tests and trust validations, obviating the need for additional expensive masks. This raises the challenge of detecting Trojans in ICs without relying on a proven secure IC.

Current design methodologies provide multiple opportunities to insert Trojans that can go undetected. It is important to incorporate new design-for-trust strategies that prevent attackers from inserting Trojans into a design as well as effectively detect Trojans in fabricated circuits. ICs must be designed such that undetected changes are nearly impossible.

COTS components are commonly used in today's systems. These components are usually designed and fabricated offshore and thus cannot be trusted. The challenge is to develop testing methodologies that consider COTS components' specifications and functionality without having access to their internal structure. The internal details of components are no longer supplied by the original equipment manufacturer.

SUMMARY

Hardware has become a vulnerable link in the chain of trust in computing systems and must be overcome. The problem of hardware security has gained significant attention during the past several years. The assumption that hardware is trustworthy and that security efforts need only focus on networks and software is no longer valid given globalization of ICs and systems design and fabrication. Until DoD develops novel techniques to secure hardware any computer application potentially can be considered untrusted while in the field.

Advanced Persistent Threats

McAfee's VP of threat research in a recent blog post noted "The targeted compromises--known as 'Advanced Persistent Threats (APTs) … we are focused on are much more insidious and occur largely without public disclosures. They present a far greater threat to companies and governments, as the adversary is tenaciously persistent in achieving their objectives. The key to these intrusions is that the adversary is motivated by a massive hunger for secrets and intellectual property; this is different from the immediate financial gratification that drives much of cybercrime, another serious but more manageable threat."

The actual attack method is familiar. The compromises follow standard procedures of targeted intrusions: a "spear-phishing" e-mail containing an exploit is sent to an individual with the right level of access at the company. The exploit when opened on an unpatched system will trigger a download of the implant malware.

That malware will then execute and initiate a backdoor communication channel to the Command & Control web server and interpret the instructions encoded in the hidden comments embedded in the webpage code. This will be quickly followed by live intruders jumping on to the infected machine and proceeding to quickly escalate privileges and move laterally within the organization to establish new persistent footholds via

additional compromised machines running implant malware, as well as targeting for quick exfiltration the key data they came for.

In a recent study by the Intrepidus Group, which is behind the PhishMe.com awareness service allowed companies to attempt to phish their employees. Findings based on 32 phishing scenarios tested against a total of 69,000 employees around the world. Here they are:

- 23% of people worldwide are vulnerable to targeted/spear phishing attacks;
- Phishing attacks that use an authoritative tone are 40% more successful than those that attempt to lure people through reward-giving;
- On an average 60% of corporate employees that were found susceptible to targeted spear phishing responded to the phishing emails within three hours of receiving them;
- People are less cautious when clicking on active links in emails than when they are requested for sensitive data.

To distinguish cyber attacks that are "*highly targeted, thoroughly researched, amply funded, and tailored to a particular organization -- employing multiple vectors and using 'low and slow' techniques to evade detection*" from hacker exploits, the US Air Force has coined the term APT.

APT infiltrations can originate from nation-states and their hired attackers, from industrial competitors, or from organized crime. The standard approach to fortifying the perimeter of an organization, such as network encryption, is a losing battle. Attackers are not trying to insert malware through existing encrypted channels. A successful defense has to change from "keeping attacks out" to accepting that "sometimes attackers are going to get in" regardless of protective measures.

The first line of defense is therefore the ability to detect attacks and then to minimize the damage instantly. Zero-day attacks are used with increased frequency. No pre-planned defense will counter that. One must assume that every organization has been already be compromised and then immediately proceed with countermeasures.

An approach to cyber defense must therefore rely on the presence of highly automated network control centers that have installed triggers, often using artificial intelligence or neural networks, to detect intrusions. If an organization has more than a thousand networks and several hundred data centers (such as is the case in DoD), it has neither the personnel, nor the resources or organization to stand up a rapid response line of defense. The only way to address the organization of secure network control centers is to limit their numbers through a consolidated management of networks that operated with only a limited number of Platform-as-a-Service (PaaS) clouds.

The second line of defense is to control tightly the access desktops, laptops or smart phones. With millions of such devices in DoD it is neither practical nor affordable to install into every device firewalls, virus protection and malware detection means. Access to desktops is always based on personal authentication privileges, regardless of location or computer technology used. Configuration updating of virus, firewalls or malware therefore becomes an unmanageable task for controlling access to a very large number of points of access. Security enforcement should be done at the server farm level where up to hundred thousand virtual desktops can be controlled centrally.

Rapid migration to cloud computing, in the form of a private PaaS, is the only affordable and feasible way for protecting DoD against cyber attacks.

SUMMARY

Given the tendency of users to be open to targeted attacks, the only solution is to isolate all traffic originating from un-authorized locations – that is sources not on a "white" security list - into an isolated "sand boxes".

Sandboxing protects the system by limiting what an application can do, such as accessing files on an internal disk or any other desktop over the network. Limiting an app inside the sand box to just operations that it needs to perform keeps the rest of the system secure in case a downloaded app is corrupt or compromised.

Since all social computing in DoD, which now constitutes a large share of total transactions, is the primary sources of targeted spear fishing, DoD should set up its desktops on PaaS based virtual computers at central servers where all transaction are subject to automated surveillance. As a first priority, DoD should proceed with providing completely isolated "sand boxes" on all desktops, laptops and smart phones.

Fictitious Identities

The attractive person you encounter on Facebook, MySpace, LinkedIn, Nexopia, Bebo, Friendster, Orkut or many other social web sites outside of the U.S.A. could be actually a fake. There are many ways of constructing such fictitious individuals, including persons invented by a government agency.[16]

An Internet IP address can be registered for as little as 99 cents/year for an <.info> domain or for $4.99/year for a <.com> domain.[17] An operator can create a large number of personas, replete with background, history, supporting details, resumes, pictures and cyber presence that are technically, culturally and geographically credible.

Such fakes enable an operator to display a number of different online personalities from the same workstation. This can be done without fear of being discovered. E-mail, blog and collaboration applications can appear to originate from any part of the world for interactions through conventional online services or social media platforms. The fake includes user-friendly indications that maximize situational awareness, such as displaying real-time local information or weather.

Communications from fake personalities can have a wide range of motivation. This includes sexual enticement, accusation of misconduct, fictitious reports, bullying, slander or libel. The possibilities of abuse are limitless, especially if the allegations originate from different sources that appear to be credible. Fake sources are also ideal for spreading propaganda and can be used to spread misinformation about political matters. If a faker's bona fides are questioned, a variety of references can be provided from multiple fake addresses.

[16] http://www.bnet.com/blog/technology-business/so-why-does-the-air-force-want-hundreds-of-fake-online-identities-on-social-media-update/8728

[17] http://order.1and1.com/xml/order/Home;jsessionid=D5C95BB2FA9082F5EABCA0776C314EE2.TCpfix142b

SUMMARY

Except in cases where certification is authenticated by a government issued identity document, such as a CAC card in the case of DoD, other origins of Internet communications will remain untraceable.

With proliferation of Internet fake personalities, protective measures will have to be taken. For instance, in the case of DoD social computing, a government issued identity certification may have to be issued to safeguard communications between the military and private addresses.

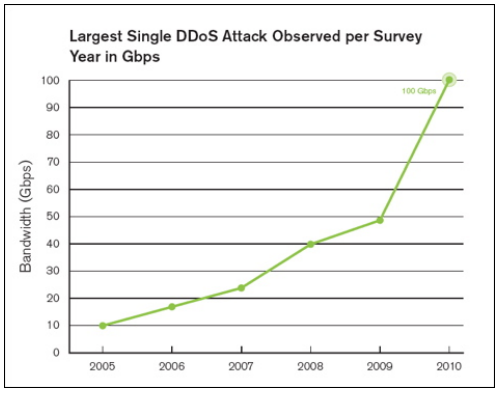
In the case of commercial communications the existing certification authorities, such as obtained from Verisign, would require additional authentication of an individual by confirming the validity of a government issued driver's license or passport. This would create unprecedented traffic on the Criminal Justice Information Network (CJIN) currently used by law enforcement agencies.

Fake personalities on the Internet are emerging as a new threat to communication. Right now there are too many easy ways how to establish Internet personalities. In due course this risk will have to be contained.

Denial of Service Attacks

Arbor Networks, in a just published Infrastructure Security Report, states that in 2010 there has been an increased severity in Distributed Denial of Service (DDoS) attacks. For the first time a 100 Gbps attack was reported. [18]

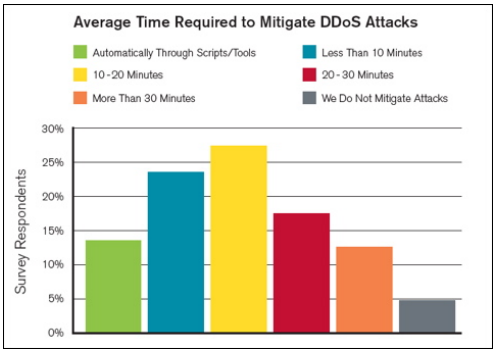
That represents a dramatic escalation in the amount of information that is piled up on a network in order to shut it down:



[18] http://www.arbornetworks.com/report

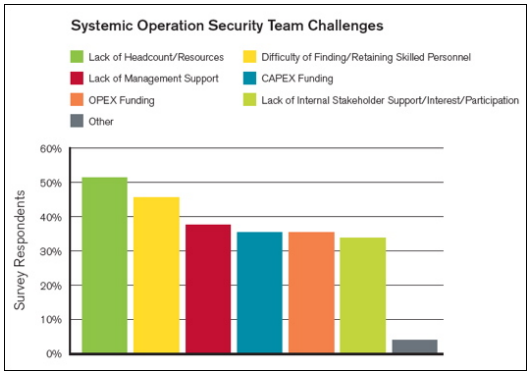
Since the most frequently deployed defense against DDoS is to shut down the computer links that have been jammed, a 100 Gbps attack can possibly unleash large amount of damaging transactions before all connections are finally severed.

The delays between DDoS detection and when the shut down happens can be seen from survey results of 111 technical network managers of Information Services Providers (ISPs):



Shutting down and then restarting a network hit by DDoS is not automatic (13% of responses). It can be a time consuming affair.

The network defenders also suffer from a scarcity of qualified personnel. To stand sentry-duty in a data center could be a position that is hard to fill, as illustrated by the following:



DDoS attacks are launched from "bot" computers that have implanted programs capable of launching attacks against designated IP addresses. Attacks occur when the controller (known as the "herder) of a "botnet" triggers the release of a rapid sequence of messages. It is interesting to speculate how many "bots" would be necessary to generate a simultaneous stream of 100 Gbps traffic.

Over 50% of the observed Internet attack traffic in the last quarter of 2010 originated from 10 countries, with USA, Russia and China accounting for 30%.[19] The global average Internet connection speed is now about 2 Mbps, though it ranges from average speeds as high as 14 Mbps (South Korea) or 7 Mbps (Delaware). Therefore, to deliver a 100 Gbps attack would take anywhere from 7,000 to 50,000 bots.

[19] Akamai State of the Internet, 2010

Botnets have been known grow into large collections. The Dutch police found a 1.5 million-node botnet. The Norwegian ISP Telenor disbanded a 10,000-node botnet. In July 2010, the FBI arrested a "herder" responsible for an estimated 12 million computers in a botnet. One can therefore conclude that assembling DDoS capable botnets is well within the scope of malware operators. The chances of future attacks that would exceed 100 Gbps is high.

SUMMARY

With an estimated 15,000 networks in place, according to DEPSECDEF Lynn, DoD is now vulnerable to more powerful and most likely more frequent denial of service attacks. How to defend against that is a matter of tradeoffs between the availability of highly trained people, or investments into an installation of automating shutoffs or in ways how to acquire fail-over capabilities.

The defense of 15,000 individual networks against DDoS by human operators is neither affordable nor executable. A defense that depends on automatic shut-offs would require retrofitting existing software with such features. It is unlikely that there is either the time or the money to do that.

The best option is to set up DoD data centers with virtual servers that can fail-over to one or more back-up servers whenever a DDoS hits. That would require migration into a virtualized environment, which is likely to show relatively fast paybacks and which can be executed by means of hypervisor software.

Password Cracking

Password cracking is the process of discovering passwords from data that has been stored in or transmitted by a computer system. A common approach is to repeatedly try guesses for the password. In cyber operations the purpose of password cracking is to gain unauthorized access to a system, or as a preventive measure to check for easily crackable passwords.

The top ranking password cracking software packages, out of a large collection, are as follows:

Cain & Abel is a password recovery tool for Microsoft OSs. It allows easy recovery of various kind of passwords by sniffing the network, cracking encrypted passwords using Dictionary, Brute-Force and Cryptanalysis attacks, recording VoIP conversations, decoding scrambled passwords, recovering wireless network keys, revealing password boxes, uncovering cached passwords and analyzing routing protocols.[20]

John the Ripper is a fast password cracker, currently available for Unix, Windows, DOS, BeOS, and OpenVMS. Its primary purpose is to detect weak Unix passwords.[21]

Hydra is a software project developed by the "The Hacker's Choice" (THC) organization that uses a dictionary attack to test for weak or simple passwords on one or many remote hosts running a variety of different services. THC-Hydra offers most developed password brute forcing.[22]

[20] http://www.oxid.it/cain.html

[21] http://www.openwall.com/john/

L0phtCrack offers hash extraction from 64 bit Windows, multiprocessor algorithms and password recovery.[23]

Password strength is a measure of the effectiveness of a password in resisting guessing and brute-force attacks. It estimates how many trials an attacker who does not have direct access to the password would need, on average, to guess it correctly. The strength of a password is a function of length, complexity, and randomness.

It is usual to estimate password strength in terms of information entropy, measured in bits, a concept from information theory. A password with, say, 42 bits of strength calculated in this way would be as strong as a string of 42 bits chosen randomly, say by a fair coin toss. Put another way, a password with 42 bits of strength would require 242 attempts to exhaust all possibilities during a brute force search.

SUMMARY

In cyber operations it is mandatory for the monitoring software at the Network Control Centers (NOCs) to run periodic verifications of every user security classification of how easy is to crack their passwords. DoD must include in every application a Password Assistant window that will reflect the implementation of security assurance policies. As a general rule in the case of cyber operations this will require at least twelve characters made of numbers and letters.

[22] http://www.darknet.org.uk/2007/02/thc-hydra-the-fast-and-flexible-network-login-hacking-tool/

[23] http://www.l0phtcrack.com/

PART 4: INTERNET NETWORKS

Internet Vulnerability is in the Infrastructure

Protection of the Global Information Grid (GIG) has now evolved into global asymmetric warfare. Engaging in this warfare is the principal mission of USCYBERCOM because the infrastructure of the Internet is fundamentally insecure and the DoD continues to depends on the Internet to function.

There are ten thousands of defenders of the Internet infrastructure who must be vigilant around the clock, everywhere. Meanwhile small teams of attackers can strike undetected whenever they choose, from wherever they maybe in the world. This is why the contests between the defenders and the aggressors meet the definition of asymmetric warfare in its extreme form.

The reasons for the intrinsic vulnerability of the Internet can be found in the engineering of its switches, routers and network connections that are owned by the Information Service Providers (ISPs) and by the communication carriers. These flaws are pervasive. They were embedded forty years ago when Internet protocols were conceived. In this paper such risks are discussed from the standpoint of a strategically motivated and hostile attacker whose purpose is to disable the military and electrical power communications that operate over the Internet.

There are corruptions from software bugs, computer viruses, which apply to computer devices such as servers, firewalls, desktops, laptops and smart phones. The government owns such devices. Attacks include denial of service, malicious code insertion or password cracking. "Hackers" and cyber criminals employ the Internet as a delivery means. Such attacks have a limited scope and therefore are seen as carrying geographically containable security risks except in cases of "zero day" attacks that can be devastating. How to protect against such incursions is a topic that warrants separate attention.

This article focuses on the failures that can results from the malfunctions of the Internet infrastructure. This includes connections to Internet Service Providers (ISPs) via Points of Presence (POPs). It includes Local Area Networks (LANs), Wide Area Networks (WANs) along with the switches that aggregate traffic. It includes the high bandwidth infrastructure for traffic between the ISPs and Network Access Point (NAPs). It includes the backbone interconnections among the NAPs. Therefore, the attack scenarios on the Internet infrastructure concentrate on its switches and routers.

Internet switches are intelligent network components with a wide-ranging set of software-defined services. These services are remotely maintained and upgraded in real-time. They also generate remote diagnostics for control locations, which is one of the weak links.

SWITCH FLAWS

The data link layer is next to the bottom of the Open Systems Interconnections (OSI) model. It suffers from gaping holes that can be exploited. The usual attack consists of altering the manufacturer's code in the switches. Major attack categories are:

Flooding Attacks on a Switch: There are attack tools that can generate over 100,000 bogus entries per minute, which then overloads the switch so that it malfunctions.

Address Resolution Spoofing: Allows an attacker to sniff the data flowing to a local area network. The traffic is either modified, or a denial of service condition is created.

"Man-in-the-middle" Attack: Adds a third party destination without the legitimate recipients being aware. The third party can extract passwords and confidential data.

Denial of Service Attack: The switch will not deliver packets and will time out, stopping all traffic.

Switch Hijacking Attack: The switch will inject illegitimate connections that will pretend to be authentic. The added connections will take over control without the recipients being aware.

Spanning Tree Attack: Allows the inclusion of spare links as backup paths. Communications are then rerouted.

The Root Claim Attack: Bogus bridge protocols are used to designate the attacker's station as the new root bridge. Once in control a variety of malicious attacks can be launched.

Forcing Eternal Root Election Attack: Makes the network unstable by tampering with the routing algorithm to keep searching for the root switch, without ever finding it.

VLAN Hopping Attack: Subdivision into different local area networks will be compromised if an attacker manages to send messages to the wrong links. When LANs support separately the NIPRNET and the SIPRNET one of them can be used to initiate a denial of service attack on the other.

ROUTERS

ISPs and NAPs are connected through intermediate network devices known as routers. A router is a special-purpose dedicated computer that makes connections when it receives a transmission from one of its incoming links, takes a routing decision, and then forwards the packet to one of its outgoing links. The routing decision is made based on the current state of the connecting links as well as on the priorities that have been attributed to the various links in order to make the selection of the next connection efficient. Each router uses a routing table to keep track of the path taken to the next network destination. Consequently, routing tables will never remain static but will change dynamically as conditions change in real-time.

The management of routing tables must be automated for instant adaptation and for assuming additional functions, such as performing security operations in which reverse path verification is feasible. In this technique the router looks up the source address of a message. If there exists no route back to the source address, the packet is assumed to be malformed or involved in a network attack, and is dropped.

When a router receives an incoming packet, it passes it to the next router, defined as a "hop" to which the packet should be forwarded. The next router then repeats this process, and so on until the packet reaches its final destination. While the packets travel they are vulnerable if an attacker is able to tamper with the router's software.

ROUTER FLAWS

To attack routers requires information how the network is configured and where the routers are located. One approach is to find the default IP values, which indicate where are the destination addresses on a network path. Another way is to use one of the numerous commercial trace route software programs. The trace route tracks a packet from all computers on a delivery path and reports all the router "hops" along the way. In this way the network topology is discovered.

The following are the principal ways how to compromise routers:

Promiscuous Mode Corruption: A promiscuous router can monitor and redirect traffic to and from other routers. The router will pass all traffic it receives in a random sequence. This happens when an attacker can masquerade as a "super-user" with software control privileges. Many router operating systems make "super-user" privileges available for maintenance or for software updating reasons.

Router Table Attacks: An attacker can create messages that look legitimate and can be then inserted into the routing table.

Router Information Attacks: Route poisoning is a method used to prevent routing loops within networks. A "hop" count will indicate to other routers that a route is no longer reachable and should be removed from their respective routing tables. The desired destination for packets will cease to function.

Shortest Path Attacks: Each router passes the status of its links to its neighbors who in turn forward this information to other routers in the network. As result of such passing each router has the link information for all other routers and eventually has the picture of the entire network topology. In a compromised table the calculated shortest paths will be incorrect and the shortest path will be purged.

Border Gateway Attacks: The Border Gateway protocol does not assure data integrity and does not provide source authentication. This protocol is the core routing protocol of the Internet, but can be tampered with.

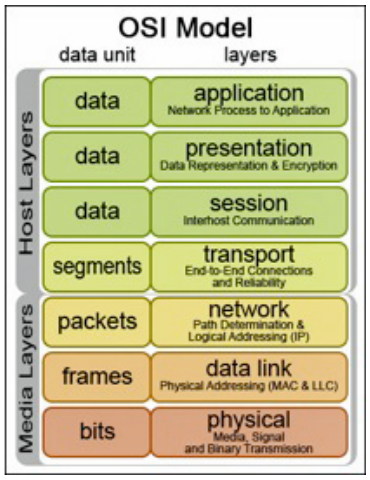
Border Gateway Poisoning: By making use of router vulnerabilities, various kinds of attacks can be launched to compromise the routing. A special case is the "Black Hole" attack where the router directs a packet to a network where packets enter but do not come out.

PRINCIPLES OF INTERNET OPERATIONS

The Internet infrastructure consists of a web of links that connect devices – switches and routers - that have the logical capability to keep redirecting traffic as it travels from origin to destination. The design of the Internet was to engineer this connectivity at the lowest cost possible to central organizations, such as telecomm carriers, while making tradeoffs that did not favor security. The original engineering of the Internet left it to other remedies, such as to virus protection software and firewall equipment, to remedy local security assurance.

With the emerging threats of cyber-attacks one can question whether retaining the existing tradeoffs between spending less on the Internet infrastructure and then boosting investments on local protection remains the best way for defending military networks

Internet communications can be seen as the passage of messages through layers of OSI protocols, as a transaction progresses from entry into an Internet switch until it arrives in its termination at a user's point of use. The OSI Model defines the entire path of an IP packet. OSI describes the standards that specify the electrical protocols to which all transactions must conform. This approach defines the processing of transactions into seven layers, which, from top to bottom, are the Application, Presentation, Session, Transport, Network, Data Link and Physical Layers, as shown below:



The Open Systems Interconnection Model

The OSI layered approach makes it possible for each subordinate layer to provide services to the next higher layer as a transaction is converted from lower to higher levels of abstraction. All of these abstractions travel from layer to layer as a series of binary "bits" because that is the only way in how microprocessors can handle the passage of a message as it traverses from layer to layer. Since all of the layers are interdependent if the Data Link or the Network layers are compromised any of the other layers will not be aware of this and the communications on the Global Information Grid will cease to function.

SUMMARY

Masquerading by the attacker, in many forms, is the root cause for Internet infrastructure attacks. The attacker either spoofs or disguises information, which is then inserted into switches and routers. When that happens the network is compromised and can be fixed only through actions that mitigate the intrinsic Internet defects.

The remedy for all the masquerading is the authentication of transactions as well as the vigilance of the operators in the Network Operating Centers to counter the attacker's disguises. Though the fundamental protocols of the Internet remain insecure, preventive measures can be taken provided that the thousands of defenders are better organized than the handful of people who are waging the attacks.

Defending the Internet infrastructure is an unequal contest. The attackers thrive from millions of local failures because they can gain knowledge every time they learn about the defenders. The aggressors do not need much money because they use the free Internet and their software tools can be easily acquired. The tools can be reconfigured to adapt to changing conditions. The defenders are meanwhile tied down by the technologies that must cover the entire network. They are shackled by budgetary limitations that cannot flex for rapid responses because protective measures must cover millions of potential points of exposure. This is why the defenders must rely on superior organization and on human intelligence for rapid responses to unexpected threats after their technological means become insufficient.

The security of the Internet remains the most advanced form of asymmetric arms race where the improved countermeasures by thousands of defenders have to compete against the handful of unconventional attackers who keep devising new schemes how to corrupt the Internet. This contest takes place not only in the form of technological countermeasures, but also in the form of superior competence of the defenders to maintain operations without error, negligence or acts of omission.

We must accept that the Internet infrastructure is faulty and will remain that way in the foreseeable future. It will take an exceptional USCYBERCOM, staffed by exceptional people to safeguard our military interests against failure that could have devastating consequences.

How to Acquire Enterprise Systems

Much has been written about the problems with the acquisition of large DoD systems. An accepted view is that the existing acquisition processes are "broken".[24]

There are, however, useful lessons to be learned from cases where complex systems have been delivered on time and under budget. Here are a few rules:

SOUND ENGINEERING FOR SUCCESS OF ACQUISITIONS

Compliance with elaborate policies, guidelines and instructions that dictate how systems are built and operated are unlikely to give assurance that a system will be delivered on time, on budget and with all of the features that the users requested. There is a long list of GAO reports that attest to the consistent failures of programs, each of which followed the thousands of documents required to comply with practices dictated by DoD Directives 5000.1 and 5000.2.[25]

Acquisition programs do not fail because they do not comply with practices dictated by DoD policies. Non-compliance is a consequence, not a cause. Failures are attributable to insufficient engineering leadership, to a lack of engineering expertise and the absence of good engineering practices.[26]

The primary challenges of program failure are technical and the ways programs are organized. The primary skill of Program Executive Officers (PEOs) should not be the skill how to steer an acquisition through a maze maze of regulations and procedural restrictions. The PEOs job is to make technology work and delivered with acceptable risks.

Program cost accounting and progress schedules are only indicators of how well the engineering and the organization of work is working. It takes more to measure of a PEOs success than what is accomplished during the acquisition phase. What matters are the follow-on operations and maintenance costs that will as well as the delivery of capabilities that, with enhancements and upgrading, will make a system viable for many decades after the PEO has moved on. Life-cycle operations and maintenance costs will always exceed the acquisition costs by a large multiplier!

ACQUISITIONS SUCCEED BECAUSE MISTAKE ARE AVOIDED

Acquisitions do not fail because of a few bad decisions. Nobody is perfect and mistakes will happen. They fail because major errors will add up and form a devastating failure-multiplier. It is the continuous

[24] http://www.fiercegovernmentit.com/story/hasc-says-dod-acquisition-process-it-broken/2010-03-14

[25] http://www.dtic.mil/whs/directives/corres/pdf/500001p.pdf

[26] http://www.usni.org/magazines/proceedings/2010-12/acquistion-reform-meyer-way

aggregation of often well-intentioned measures that will ultimately add up to a mess. That can be corrected only sometimes with more money or degradation in performance but hardly ever through an improved schedule.

Programs will fail when multiple flaws accumulate. It is the creep in defective engineering that will ultimately cripple any program. Two to five fundamental flaws can be fixed with budget supplements, schedule slippage or by deleting features. More than ten flaws are hard to overcome. When there is an excessive accumulation of what I call "sins of commission" that will result in either redefining what needs to be done or letting the entire venture fade quietly into oblivion.

Here is a list of twenty-six things to avoid:

ELEVEN SINS OF PROGRAM MANAGERS:

- Never schedule the completion of a program investment past its financial break-even point. Break-evens should be less than two years for each development increment.

1. Never plan a program that takes longer than a half of its useful technology life. Technology life these years is less than five years.

2. Never assume that the completion of all of the acquisition cycle milestones will guarantee the delivery of originally planned operating results. The farther the milestones are spread the greater the probability of changes in original requirements.

3. Never fail in the delivery of projected metrics for operation and maintenance, such as cost, latency, reliability and availability. Performance metrics should be used to keep track of the expected results during the entire program cycle.

4. Never assume that multiple organizations that depend on a system will reach agreement about the proposed features or the definitions of data.

5. Never trust requests for major system changes to be paid by someone who has no accountability for results.

6. Never commit to project schedules that take longer to implement than your customer's average time to the next reorganization. In DoD the time between reorganizations is about two years.

7. Never hire consultants to deliver undefined requirements subject to unclear specifications on a time-and-materials basis.

8. Never consider a program complete unless you have acceptance from a paying customer who is also directly accountable for results.

9. Never add major new requirements to a project after the budget and the schedule are fixed.

10. Never test a pilot program by relying on excess support manpower to demonstrate that the system will function for a selected small number of senior executives who have limited requirements. Pilot programs should be always tested at locations where operating personnel can fall back on back-up solutions.

EIGHT SINS OF TECHNOLOGISTS:

1. Never adopt a new technology unless the top management in charge of operations understands intimately how it works.

2. Never design a program that will simultaneously deliver new communications networks, new data centers, new databases and new concepts of operation as an integrated package.

3. Never introduce a totally new and untested technology for immediate deployment.

4. Never deploy totally innovative software simultaneously with totally innovative hardware.

5. Never code a system in a programming language that is unknown to most of your staff.

6. Never give programmers access to the network control center consoles that manage computer services.

7. Never automate a critical process that does not have a human over-ride switch.

8. Never rely on 100% reliability of a single communication link to retrieve data from a critical database.

SEVEN SINS OF PROGRAM PLANNERS:

1. Never consider a system acquisition program to be complete without with a full-scale pilot test.

2. Never assume the recovery of any system from complete failure without frequent re-testing of fail-over readiness.

3. Never depend on critical delivery dates without proof that a vendor has done it before.

4. Never design a system that will not be able to operate under severely degraded operating conditions.

5. Never consolidate data centers that operate with obsolete technology, dissatisfied customers and a management that is technically incompetent.

6. Never convert an old application or database to a new one without being able to retrace your steps in case of failure.

7. Never engage services of a contractor who has a program manager who lives in a trailer hitched to a pick-up truck (I did that, to my regret).

BUILD A LITTLE, TEST A LITTLE

The above flaws are only a partial list. Systems managers will always discover new ways of how to stumble. In the DoD there is an unlimited opportunity for adding to an inventory of misfortunes.

The best way for preventing a persistent accumulation of flaws is not to engage in the development of applications that take much longer than two years to implement after a prototype pilot has succeeded. Therefore, the installation of applications, in progressive increments, must also fit into an overall design to assure interoperability. This calls for mandatory compliance with enforceable standards.

All applications must be sufficiently small and modular so that they can fail. When that happens most of the parts can be then re-purposed for restarting a system at a lower cost because the price of all technologies keeps dropping.

DoD systems programs should never pursue a monolithic approach for the construction of a new telecommunications infrastructure, for a new operating environment and for new databases. New applications should ne never be developed to fit into their own dedicated telecommunication network, a separate operating environment and certainly not with a self-contained database. Applications can be launched fast, but networks, operating environment and databases take a long time.

The telecommunications infrastructure should be an enterprise service. It should be designed incrementally so that that it will support applications when they demand added services. Building a global grid calls for long-term funding and for a completely different management structure than what is required for applications.

The operating environment, and particularly the data centers, should be an enterprise service that can be built incrementally, so that its capacity, security and redundancy can support applications when they come on stream. A mix of government controlled and commercially managed operations calls for financial arrangements that transfer most of the technology risks to equipment vendors.

The databases are the keys to achieving the interoperability and the security of DoD systems. These should be constructed based on shared standards as well as enterprise-wide metadata and not on application-derived definitions.

FINDING PAYOFFS

The acquisition of enterprise systems by DoD must recognize that the total life cycle costs of preset investments into the DoD infrastructure can extend for many decades. What counts for such long-term investments is the discounted present net worth of cash flows for both the infrastructure as well as the applications.

From a financial analysis standpoint, the discount factor for investments for an application should be low if implemented rapidly. The discount interest rate would be a small multiple of short-term Treasury bonds.

The discount factor for investments in the DoD infrastructure must recognize the long time it takes to place communications and an operating environment in place while information technology takes place at a rapid pace. Consequently, the discount interest for infrastructure investments will be a large multiple of long-term Treasury bonds. On balance, that would favor spending money on infrastructure investments that speed up the implementation of applications at the lowest possible cost.

When viewed from the standpoint of life cycle costs the investments in new applications must be also traded off against long-term expenses for operations and maintenance. That will always favor spending money on how to accelerate the installation of applications.

Therefore, DoD policies should make enable the making of life cycle tradeoffs between acquisition costs and how systems are acquired and operated. DoD policies must also guide how investments are made on the infrastructure in comparison with applications.

The principal flaws in the implementation of current acquisition policies are:

1. The PEOs are not accountable for total life-cycle operations and maintenance (O&M) costs of systems. Systems acquisition budgets are funded separately from O&M. There is no visibility of what ongoing costs of applications may be.

2. The costs of the DoD infrastructure (currently consuming over 56% of total IT costs), which includes networks and the operating environment, cannot be traded off against the life cycle costs of applications. Infrastructure costs are widely distributed into enclaves such as DISA and DLA while each Service keeps building additions of their own infrastructure.

3. DoD IT costs are broken up between multiple organizations, each attempting to garner funds to attain self-sufficiency.

4. Only a small fraction of DoD IT life cycle costs is subject to a formal budgetary process where tradeoffs in investment vs. O&M are made. Meanwhile, individual project managers are left to their own resources to concentrate on keeping their acquisition costs below budget.

SUMMARY

The DoD systems acquisition processes can improve by signing up to the primacy of engineering over management by procedure. The acquisition of computer applications will gain from avoiding any of the twenty-six "sins" of how to manage.

However, the current acquisition policies will still fall far short in delivering what DoD needs. In the absence of enterprise level tradeoffs how investments between acquisition and O&M costs IT is unmanageable. The acquisition of IT systems differs from the acquisition of weapons. O&M costs, inclusive of military, civilian and contractor labor (which do not show up as IT costs) overwhelm the acquisition costs, which are watched closely.

While the demands as well as the costs to support cyber operations are rising the budgets are shrinking. Cost cuts and additional outsourcing will be insufficient to shift of DoD from an emphasis on kinetic weapons to technologies that support information-based warfare.

The fundamental flaws are not Directives 5000.1 and 5000.2. It is the funding mechanism that has allowed DoD to accumulate over 15,000 disjointed networks, close to a thousand of data centers and over 5,000 information technology projects that are not interoperable. The current method of funding projects has resulted in thousands of databases that are incompatible and therefore cannot support information warfare.

The DoD IT infrastructure has excessive costs while its performance is inferior as compared with commercial practices. Applications are redundant while their performance is inadequate.

The fundamental flaw in the management of DoD systems is not how acquisition is conducted but how money is spent on the DoD information infrastructure. At present DoD operates an expensive and ineffective infrastructure. In FY10 it cost $19.5 billion, or 58% of total IT spending. This is expensive because it supports 1,900 projects that can be found in different self-contained budgets, which are managed by separate organizations. Such diversity does not permit the making of the making of tradeoffs between short-term low risk projects and long-term high- risk infrastructure investments. As result, all tradeoffs are made only locally on a sub-optimal scale. Project management is then burdened with elaborate restrictions that try to achieve economic trade-offs by administrative means that ignore sound engineering methods.

The Global Information Grid

DoD operations continue to be hampered by the lack interoperability. In order to run war operations in the last decade DOD had to patch together disparate systems and networks. DoD has been also retrofitting systems after they are fielded to keep field operations working. This approach has been very expensive. It has been insufficient in meeting the DOD's stated goal of achieving a networked force where soldiers, weapon systems, platforms, and sensors are linked and able to function jointly.

DOD has been looking to the Global Information Grid (GIG) to solve the interoperability problems since 2002.[27] But progress to date has fallen short of its objectives.

The GIG is a large and complex set of technology programs intended to provide an Internet-like connectivity to every device, including wireless and radio. It is supposed to allow users at any location to access data on demand from anywhere. Its purpose is to enable the sharing of information in real time. GIG should enable collaboration in decision-making regardless of which military service is the source of information. The GIG would link weapon systems for greater joint command of battle situations as the US dependency on information-based warfare is rising rapidly.

According to a 2006 GAO report the GIG infrastructure will cost approximately $34 billion through 2011 though the rising costs of information assurance will be increasing that amount.[28] How much of the current annual IT costs of $36.5 billion is allocated to communications is not clear. However, the duplication in over 150 DoD networks is increasingly shifting the costs of information management from applications that support the warfighter to the underlying infrastructure.

DOD's investment in the GIG extends beyond development of the core network circuits. The purpose of the GIG is to integrate the majority of its weapon systems, application systems into a comprehensive network. Accomplishing these objectives involves reaching agreement on common standards and in aligning systems with GIG-like services.

There are three decisions processes that have so far impeded the progress in advancing the GIG:

1. The Joint Systems, which the DOD uses to identify, assess, and prioritize military capability needs has not come up with an architecture and design that can be the basis on which to build a functioning GIG;
2. The Planning, Programming, Budgeting, and Execution process, which guides how DOD allocates resources, has not been able to develop an acceptable fiscal and governance mechanism for funding enterprise-level investments;
3. Defense Acquisition System, which governs how DOD acquires weapon and information technology systems, has not been reformed to support a GIG-like venture in which the technologies are subject to rapid changes.

DOD's decentralized management approach does not fit the GIG. It is not designed for the development of a large-scale Joint integration effort, which depends on a high degree of coordination and cooperation. Though the GIG calls for clear leadership and authority to control budgets across organizational

[27] DoD Directive 8100.1

[28] GAO-06-211

lines no one is in charge of the GIG. There is no requisite authority or accountability for delivering GIG results.

The Office of the Secretary of Defense assigned overall leadership responsibility for the GIG to the DOD CIO, to include responsibility for developing, maintaining, and enforcing compliance with the GIG architecture; advising DOD leadership on GIG requirements; and providing enterprise-wide oversight of the development, integration, and implementation of the GIG. However, the DoD CIO has practically no influence on investment and program decisions by the military services and defense agencies, which determine investment priorities and manage program development efforts. Consequently, the services and defense agencies are unable to align their spending plans with GIG objectives.

DOD's decision-making processes are not structured to support crosscutting, department wide integration efforts. The existing processes were established to support discrete service- and platform- oriented programs rather than joint, net-centric never-ending programs. This situation remains in place to this day. The Joint Capabilities Integration and Development System (JCIDS) process has been implemented for almost a decade and has produced a large collection of policy papers but not much else. In the absence of collateral budgetary, PPBE and Acquisition process changes JCIDS plans have a limited use.

For instance, the DOD's acquisition process continues to move programs forward only if there is sufficient advance knowledge that technologies can work as intended. At the current extremely rapid rate of technological change, information systems investments will become obsolete by the time the entire multi-phase (five years+) Acquisition process can be ever completed.

Joint, net-centric capabilities depend on the delivery of several related acquisition programs. This calls for rapid-turnaround integration in at least quarterly time frames while the acquisition process with a time clock based on years instead of weeks is not suited for managing interdependencies among diverse programs, especially if cooperation from several services and agencies is instantly necessary for the correction of software defects.

SUMMARY

The Global Information Grid has been seen as the cornerstone of information superiority, as a key enabler of net-centric warfare, and as a basis for defense transformation. The GIG's many systems were expected to make up a secure, reliable network to enable users to access and share information. Communications satellites, next-generation radios, and a military installations-based network with significantly expanded bandwidth were supposed to pave the way in which DOD expects to achieve information superiority over adversaries. The focus of the GIG was to ensure that all systems could connect to the network based on common standards and protocols. Some progress has been made but only at the price of rising costs and the increasing disconnection between the technologies of DoD and commercial IT.

Increased budgetary pressures are starting to modify DoD's use of the term "GIG". That is undergoing changes as new concepts are emerging such as Cyberspace Operations, GIG 2.0 or the Department of Defense Information Enterprise (DIE). Such ideas are in the process of revising what was original version of GIG, which delivered mostly circuit bandwidth but little else.

However, unless there are revisions in the way in which Joint Systems requirements are defined, how the Planning, Programming and Budgeting processes are revised and how Acquisition is restructured, the existing management processes are inadequate for delivering the desired integration and interoperability goals.

Apple and Cloud Computing

Steve Jobs, the Apple CEO, announced the addition of cloud services to the Apple list of product offerings. The new cloud will make it possible to move from a primary reliance on Apple devices (Mac computers, iPhone and iPads) to also receiving computing support from Apple data centers.

Access to the Apple Cloud heralds the extension control to centralized and proprietary management by Apple. Apple customers will receive the benefits of lower capital costs as well as reduced operating expenses. Customers will have the incentive to shift from merely purchasing Apple devices and software to also renting computing services and software. New applications will be available directly from data centers so that a faster rate of innovation can take place.

Cloud computing represents major change how Apple can be expected to operate in the future. To understand the context of a different concept of operations the following describes what directions Apple is likely to be taking in the next decade:

Apple will continue with its dedication to offer devices and services to consumers and not to enterprises. Apple will not compete for market share in enterprise systems with firms like IBM, HP and Dell. However, cloud computing will allow an expansion in the scope of what Apple will be able to offer.

With 250 million active users Apple should be able to expand to a population that could eventually number billions of customers. The availability of cloud computing services will make the current rapid growth in the number of customers sustainable. Apple will not compete with Google who at present offer free services to consumers.

Apple will continue enjoying top ranked global brand name recognition. Its products will continue to receive premium pricing. Added services through the cloud will continue to maintain exceptional profit margins.

Apple is now the largest integrated IT firm in the world. The projected 2011 revenue exceeds $100 billion. Current revenue growth rate exceeds 50%/year. Its market capitalization of over $300 billion. This exceeds HP and Microsoft, who have comparable revenues of $127 but lower growth rates and lower profits than Apple. Cloud computing products will contribute to Apple's increase in market share.

Apple is the only IT organization with presence in every part of the consumer computing market with the exception of not participating in network services. HP dominates hardware while Microsoft dominates software. However, these firms do not deliver an integrated solution as done by Apple. The addition of cloud computing as an easy extension of existing services will only widen the gap between Apple and all other computing vendors.

Apple is unique in that it can claim to be the provider of integrated products and services by delivering both the hardware as well as the software that only fits on Mac computers or Apple devices. With the addition of proprietary cloud services, Apple has broadened its ability to serve dedicated customers at a much lower total cost of ownership than competitors. So far Apple has been able to obtain a modification of Microsoft's dominant Office suite to run in the Mac environment. Apple can be expected to pursue such approach by integrating seamlessly Software-as-a-Service offerings, such as SalesForce, into its cloud.

The global consumer information technology sector is growing faster rate than enterprise computing, in which Apple does not as yet participate to a significant extent. Apple's consumer sector is driven by rising population and by more demanding customer preferences. Apple is serving this sector by improving the ease of use through improvements in customer-to-computer interactions. The total cost of ownership of an Apple is lower than comparable costs of Microsoft centered devices. With easy access through cloud

computing Apple will now start adding the penetration of a large small business sector to its scope of services.

The enterprise sector of computing is consolidating by reaping improvements from more efficient technologies. In the short run the enterprise computer business will slow down while corporations concentrate on reducing the labor overhead costs that currently surround enterprise IT operations. Meanwhile, the consumer sector will grow explosively with a shift of consumer preferences to wireless connectivity. The Apple advantage in wireless connectivity, especially in high growth regions such as Latin America, will extend its cloud-based services to small-scale enterprises as well.

Apple has a limited but highly innovative product line. It is subject to rapid upgrading by means of proprietary software developed by over 100,000 registered developers who are not the Apple payroll. Apple will continue to extend the role of developers in less developed countries, especially through local applications hosted on the Apple cloud. In this way Apple will be able to gain global market share at a low marketing expense.

Apple product innovations, especially in customer interfaces, have surpassed IT competitors. Apple consistently excels in the design of devices and in the production of graphic displays. Cloud computing will continue to extend this capability through solutions than will depend on local developers who will be able to serve linguistically unique requirements.

Apple controls its products with only two Operating Systems, OSX and IOS, which share the same codebase. Competitive offerings from Microsoft and Linux are more diverse and create interoperability problems. Apple operating systems are updated frequently and have been able to maintain backward compatibility whereas competitors require additional investments to make diverse software components compatible. With cloud computing available for downloading of applications there is no reason to change the current policy of maintaining only two operating systems.

No competitor has been able to match Apple's complete hold over every part of its systems software offering. Apple exercises control over software development methodologies and programming tools, which includes application programming interfaces. It couples software and hardware under unified supervision. This speeds up that rate at which innovation can take place. Cloud computing will make it possible to maintain control over developers and prevent unauthorized corruption of software development methods.

Apple offers to customers an open source browser, which is closely coupled with the Apple operating systems as well as with web applications whether developed by Apple or accessed from other sources. The management of all software offers to Apple a competitive advantage that is unlikely to be matched by IBM, Microsoft, HP or Dell who are encumbered by too many legacy offerings. That advantage is now getting extended with the introduction of cloud computing as an integrated part of the entire spectrum of product offerings. It may take a new "clean sheet" offering of hardware-software solutions to match what Apple is placing into the marketplace.

Any product or software release from Apple has the advantage that it can be tested together for bug free operations across the entire Apple product line. A tightly controlled approach to software management makes this possible. Consequently Apple can deliver new products at a faster rate than competitors. Cloud computing will make it feasible to continue keeping such a close hold over product testing.

Apple has pioneered the introduction of products not dependent on the "mouse" devices that is associated with Microsoft. With portable computing devices (iPhone, iPad) this becomes a competitive advantage. The absence of a keyboard will continue to be an advantage in global market. Cloud computing will also permit the application-specific customization of input formats that can be swapped to deal with local needs, such as in health care or in the military.

Apple has pioneered innovative user access methods with reliance on high quality graphics. Apple now leads the industry in offering a preferred consumer experience with computing devices, which are standard across the entire product line, but can be instantly modified from the cloud.

Apple's App Store contains 425,000 applications that range in price from 99 cents to $29.99. App Store management conducts reviews of programming methods used prior to publishing to guarantee that all applications can be integrated with the unified Apple approach to software management. This differs from the largely unrestricted acceptance of developer applications in the Google Chrome Web Store and the Android Application Store. Cloud computing will now allow an almost infinite expansion of items available in customized application catalogues.

Apple reaps profits from the marketing of downloadable music (15 billion iTunes sold), downloadable movies (for rent or purchase), downloadable TV shows and as well as lectures, on-line training, academic course and podcasts. It has recently opened an iBookstore (130 million books sold). The App Store is pioneering new approaches to marketing of on-line information products. So far customers have opened 225 million accounts. The App Store sales so far are $4.3 billion. Cloud-based App Stores will make downloadable products also available to devices other than Apple such as Android phones or MS Windows devices. All App Store products will be downloaded, thus discontinuing the present practice of distributing products on a disk.

Consumers typically buy software for a designated machine so that vendors, such as Microsoft or Oracle, can collect license revenues by point of use. Purchases from the Mac App Store are attached to the identity of an individual. This is a departure from how consumer software is licensed. It allows Apple users to reuse App Store products regardless of device or location. There is an incentive for customers to keep purchasing more hardware, software, music, books and teaching materials without a concern about paying for changes in devices.

The extension of the processing capacity of the Apple cloud will allow adding to the App Store the ability to sell information services such as offered by payroll processing bureaus, tax services and video conferencing offers. The economics of on-line selling favors purchasing of products where the embedded security for collecting funds makes that a convenient choice. Cloud-based information services from App Stores are potentially a major contributor to Apple future profits while providing the producers with a low cost distribution channel. In this regard the major competitor of Apple would be Amazon.

Apple offers to its developers development tools, ten thousands of Application Program Interfaces (API) and marketing aids. In this way App Store becomes a closed system that assures integration while minimizing security risks. The estimated number of Apple developers is presently greater than 100,000 but cloud computing will permit a large enlargement of this number. There are large populations of highly trained but unemployed students world-wide who are already finding software development for Apple as a rewarding opportunity to start a software business.

70% of the proceeds from App Store sales are passed to developers. So far Apple has paid out $2.5 billions of fees. The highly motivated Apple developers should be seen as a virtual extension of Apple's own development organization and a source of technological power.

The newly announced Apple Cloud has been staged for incremental evolution. The cloud now becomes the link that connects all Apple devices.

The Apple Cloud should be seen as an Apple owned private and proprietary Platform-as-a-Service (PaaS) arrangement. It offers not the supporting global Internet infrastructure (Infrastructure-as-a-Service) for Apple applications but also standard middleware that will accept only applications that have been programmed according to Apple specifications.

The Apple Cloud also provides 50GBs of data storage for Apple customers at no added costs. Large collections of music and pictures may find cloud storage helpful, especially for archival storage.

The Apple Cloud will also hosts every application purchased from the App store. It will offer fail-over backup to another site in addition to the local "time machine" backups that are located in Macintosh computers. Several large Apple applications, that have frequent use, are already available from four data centers. The Apple's 500,000-square-foot computer operations, costing $500 millions, has the capacity of 120,000 servers.

All documents, music or pictures uploaded from any Apple laptop, iPhone, iPad Touch to the Apple cloud can be synchronized with every other Apple device. This will allow a customer to carry the contents of all applications regardless of locations or whether the device is portable or at a fixed location.

SUMMARY

Apple is the world's most successful information technology corporation and therefore warrants careful study about the patterns of the future directions how computerized information will be generated and marketed. With a highly focused strategy Apple can be expected to continue sustaining its leadership in the consumer segment of the IT industry in the foreseeable future. The keys to success will be now cloud computing, an augmentation of App Stores, interoperable consumer devices and a development virtual staff. Such strategies suggest that the current patterns of success will continue.

Yet, Apple with its $100 billion revenue occupies only 6% of the global $1,690 billion IT industry. The remaining 94% is fractured into thousands of suppliers, diverse software development practices and uncoordinated operating methods. How this vast market can be covered with the aid from cloud computing will be discussed in another blog. It is clear, however, that in the absence of integration across the entire application "stack", other methods will have to emerge as a unifying force. There is a technological and marketing opportunity how to achieve an improved economy how information technologies can be organized.

Are IPv4 Addresses Exhausted?

On June 9, 2003 the DoD/OSD CIO issued a memorandum that the DoD goal is to complete the transition from IPv4 addresses to IPv6 addresses by FY08 for all inter and intra networking. This was necessary to enable the transition of all National Security Systems and the GIG, to be completed by FY07. DISA would act as the Central Registration Authority for all DoD systems. The directed transition to IPv6 by FY08 never happened except for minor installations.

On September 28, 2010 the Federal Chief Information Officer issued a memorandum that the "The Federal government is committed to the operational deployment of Internet Protocol IPv6".[29] Agencies and Departments of the Federal Government will have to upgrade externally facing servers (such as web services, email, DNS and ISP services) to use IPv6 by the end of FY12. For internal applications that communicate with the public Internet servers, the upgrade to IPv6 would be implemented the end of FY14.

[29] http://www.networkworld.com/newsletters/frame/2010/111510wan1.html

The major benefits to be derived from migration from IPv4 to IPv6 are the much larger address spaces. IPv6 offers improved routing and enhanced security, especially how transactions are handled within Internet routers and switches. For instance, IPv6 reduces complexity of Internet services by eliminating the reliance on Network Address Translation (NAT) technologies. IPv6 also enables added security services for end-to-end mobile communications.

With a continuous growth of new facilities in DoD the question is whether IPv4 addresses are easily converted so that DoD systems will remain completely interoperable and show improved communications performance.

DoD has so far used only about half of all of the IPv4 addresses that have been assigned to it. As of February 2008 there were over 200 million IP addresses still available for DoD, which should maintain communications for a time.[30] If DoD proceeds with its adoption of IPv6 it would acquire 42 million billion billion billion IP addresses. That means that DoD would have enough IP addresses to give each grain of sand on earth 90 billion IP addresses. Such a number is nice to have, but the question is whether there are funds available to make the conversions to IPv6 with the urgency that has been dictated.

DoD has now backed off IPv6 implementation even though upgrading from IPv4 to IPv6 would allow for better network mobility, mission expedition and the widespread adoption of Radio Frequency Identification (RFID)s.[31] Although some DoD components have already started migration to IPv6 the differences between applications staying on IPv4 and those communicating using IPv6 will increase the complexity of network software. A mixed environment will require DoD to launch efforts that add to all IPv4 locations added interoperability capabilities until such time when all IPv4 addresses will be retired.

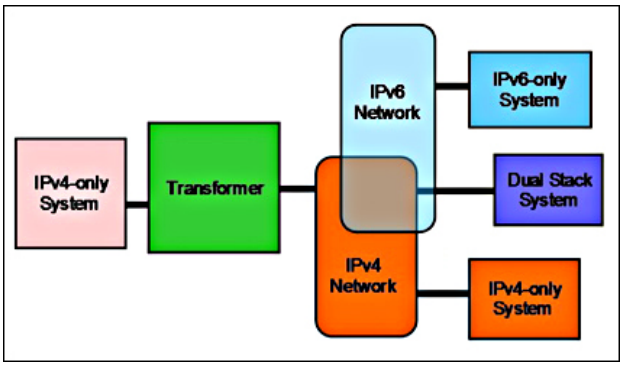
At this time there are no major funded programs for proceeding with IPv4 conversion on a tight schedule. OSD policy has now saddled all programs, whether they are legacy or new, with the need to acquire additional transformation software and hardware while in transition to IPv6. That will surely take longer than the policies have dictated.

Transition hardware and software is available from several vendors but it is questionable whether the current budgetary limits will permit spending money on projects with only a transitory life.[32] To maintain interoperability during the conversion from IPv4 to IPv6 thousands of DoD locations will need a capability to translate IPv4 addresses to more options, as illustrated below.

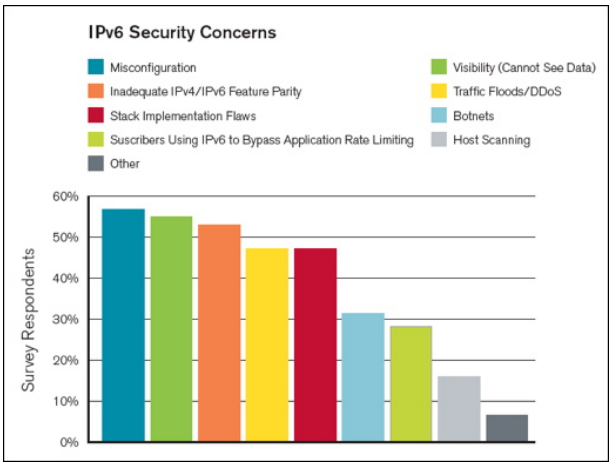
[30] http://royal.pingdom.com/2008/02/13/where-did-all-the-ip-numbers-go-the-us-department-of-defense-has-them/

[31] http://www.defense.gov/news/newsarticle.aspx?id=59780

[32] http://www.datatekcorp.com/index.php/ipv6-portal/ipv6-resources/117-sbir-work



Address transformation will add complexity to every site that communicates either within DoD or externally. Meanwhile the task of implementing IPv6 remains technically a demanding task. Due care must be taken to ensure that the existing communications are not impeded as more software is placed into the path of every transaction. GIG IPv6 network performance will also have to improve, especially for auto-configuration, prioritization, converged voice and video, multicast and mobility. A recent survey by Arbor Networks shows the following difficulties with IPv6 implementation:[33]



In more than half of the 111 reports from network technicians inadequate IPv4 vs. IPv6 parity had to be overcome with software fixes.

The global registry of IPv4 addresses, the Internet Assigned Numbers Authority (IANA), indicates further shrinkage of available IPv4 addresses, but by no means exhaustion. Only the organization that assigns Internet addresses to China and to India (APNIC) shows that they will be using up their address pool by the end of 2011. However, with a reallocation from existing poorly utilized address pools elsewhere in the world there are more than adequate IPv4 numbers available globally for an indefinite future.

Meanwhile Internet Service Providers (ISP's) are already upgrading network switches as well as routers to handle IPv6 addresses in addition to retaining the capacity to process every IPv4 addresses. Therefore, the availability of dual handling of addresses does not impose on DoD any short-term urgency to achieve IPv4 to IPv6 conversions.

[33] http://www.datatekcorp.com/index.php/ipv6-portal/ipv6-resources/117-sbir-work

SUMMARY

IPv4 allows 32 bits for the Internet Protocol and supports 4.3 billion addresses. IPv6 uses a 128-bit address and supports a practically infinite number of addresses. As of the end of 2010 only 533 million unique IP addresses have been assigned.[34] Though the USA currently has 26.4% of the global IP population, it has obtained more than 50% of the IP addresses, while the quickly growing China is exhausting its allocation. Clearly, there are enough IP addresses, on the average, except that they have been misallocated. An immediate rush into IPv6 cannot be therefore justified provided that IANA can take corrective actions.

Given poor progress in IPv6 implementation DoD contractors will have every incentive to continue enhancing IPv4 capabilities rather than working on the conversion to IPv6. IPv6 is not necessarily more secure than IPv4 provided that added security fixes are installed. Security features are now available for IPv4 from a number of sources. From the standpoint of DoD applications there will be few practical differences in security protection if the fixes are implemented. Therefore, keeping IPv4 in place makes sense unless DoD decides to proceed with a full implementation of RFIDS, which is not the case right now on account of enormous initial costs.

There is another option open: the IPv6 Native Dual Stack solution, now in testing. Can access services natively over both IPv6 and IPv4. Users do not need to use any IPv6 or IPv4 tunneling, translating, or NAT solutions. Access to both IPv6 and IPv4 can take place directly at high-speed. When the Dual Stack solution is ready, DoD may save money by avoiding costly software fixes.

Despite high-level policy mandates promulgated in 2003 and in 2010 the IPv4 to IPv6 conversions will not happen very soon. It will require a redirection how future DoD networks will be upgraded before DoD internal and external networks can start communicating using the identical address formats. The best choice for DoD is to proceed now with the adoption of IPv6 as a requirement for any upgrades rather than to confront every Component with fixed immediate deadlines.

[34] http://www.arbornetworks.com/index.php?option=com_content&task=view&id=1034&Itemid=525

PART 5: NETWORK OPERATIONS

A Brief History of Defense Networks

Many of today's original ideas about a global command and control system can be traced to Vice Adm. Jerry Tuttle, USN (Ret.), who served as director, Space and Electronic Warfare, from 1989 until his retirement in 1994. Faced with the need to restructure the Naval Telecommunications System to handle dramatically increased message traffic, Tuttle could have proposed buying bigger pipes. Instead, he created the Copernicus concept for evolving the Navy's networks. His immediate objective was to restructure the Naval Telecommunications System and then to extend it to other parts of the Navy as well as to other military departments. Copernicus concentrated on the Navy's immediate needs for increased bandwidth and for integrated communications.

Copernicus was not executed when the Navy decided to outsource the largest single share of its network operations to a contractor. Instead of open interoperability and easy information flows, the Navy received a closed and proprietary network with limited functionality. Meanwhile Navy networks, data centers and dependencies kept proliferating to the current count of over 100 networks and a large number of data centers.

In September 1992 Deputy Secretary of Defense Donald Atwood issued Defense Management Decision Memorandum (DMRD) 918. Its purpose was to achieve major cost reductions in information technology spending, which included the consolidation of networks into a centrally managed Defense Information Infrastructure (DII).

Almost all of the DMRD funding was applied to data center consolidation in order to create a small number of mega centers. In the period during which the Defense Information Systems Agency was engaged in consolidations there were dramatic cost reductions in server computing, which resulted in the creation of hundreds of new computing facilities throughout the Defense Department. Ultimately DMRD 918 made hardly any impact on enterprise-level communications. While mega centers were built, the military services responded by acquiring a large number of additional networks to support distributed computing.

The foremost exponent for migrating from a computing-centric to a network-centric approach was Vice Adm. Arthur Cebrowski, USN. His concepts of network-centric warfare were articulated in a widely circulated paper in 1998. This led to the formulation of the "power to the edge" systems architecture. In this concept, networks made it possible for dispersed military units to acquire information as needed instead of central agencies anticipating what could be useful. After October 2001, Cebrowski became the director of the Office of Force Transformation, reporting directly to the secretary of Defense. Further work on advancing network centricity had to take a back seat. Though attempts were made to implement applications based on the "edge" concepts, none of the major department investments followed this approach.

John Stenbit served as the chief computer executive in the Office of the Secretary of Defense from 2001 to 2004. He concentrated on net-centric warfare and operations that offered plenty of trusted bandwidth. The underlying concept was to replace the Defense Department's "smart push" regime with a new "smart pull" paradigm in which warfighters would be able, to retrieve whatever was needed to complete their missions. The vehicle to accomplish this goal was the Global Information Grid Bandwidth Expansion (GIG-BE) that would eliminate bandwidth constraints. The architecture for the net-centricity mission was then assigned to DISA as the Net-Centric Enterprise Services and Net-Enabled Command Capability programs, which mostly closed down in 2009. Although considerable funds were expended to

acquire long-haul circuits as well as to offer a wide range of collaboration applications, the military services never accepted these programs as a unifying design.

Network Virtualization

Army Gen. Keith Alexander, the head of the new cyber command stated that the Defense Department needs situational awareness across DOD's networks to protect its cyber defenses. "We do not have a common operating picture for our networks. We need to build that."

DOD is responsible for protecting more than seven million machines, linked in 15,000 networks, with 21 satellite gateways and 20,000 commercial circuits. Unauthorized users probe DoD networks 250,000 times an hour or over six million times per day, he added.

In the current situation the proliferation of networks, circuits and computers offer to attackers an enormous "attack surface" which is for all practical purposes indefensible.

Virtual Networks

Network virtualization combines hardware and software network resources into a software-based administrative environment, which can be managed centrally. Network virtualization enables the integration of numerous networks so that central services, such as consolidated security management, situation awareness and protective measures can be shared across every network.

The components of virtual networks are: Network hardware, such as routers, switches and network adapters; WANs and LANs; Network storage devices; Network media, such as Ethernet and Fibre Channels. Examples of virtual networks are switches that physically connect to external networks as well as services that allow system administrators to combine local area networks into a singly administered network entity for the purpose of intrusion prevention.

Network virtualization software allows systems managers to route traffic to diverse datacenter environments where support of business and warfare applications can take place.

In the past DoD components used to purchase multiple security protection measures and to set up failover and redundancy capabilities at each of thousands of data centers. The installation of network virtualization software makes it possible to migrate security services as a fully configured virtual service to each data center, regardless of geographic locations. This allows for migration from legacy environments to a virtual environment across datacenters across the world.

As data center resources become consolidated the network virtualization software allows for reduction in space requirements, in optimal server utilization and in the consolidation of controls into DoD-wide network control centers so that highly trained personnel can be utilized much better.

Implications

Establishing situational awareness and the much needed real time responses to attacks that emanate from 15,000 networks and 20,000 commercial circuits is not feasible using the existing network configurations in place in DoD.

The installation of network virtualization as an architectural direction for DoD will make it possible to consolidate points of control to a limited number of network control centers. Such a move will not only deliver large reductions in cost but also safeguard the security of millions of computer devices.

Time has come to start migrating to designs that will use network virtualization as the basis for cyber defense operations.

Cyber Defense and the DoD Culture[35]

According to Air Force LTG William Lord, 85 percent of cyber operations are in defense. That being the case, how should the Defense Department protect its network and computer assets? A 2009 RAND Corporation report on cyber deterrence asserts "…most of the effort to defend systems is inevitably the ambit of everyday system administrators and with the reinforcement of user vigilance." The report also states "…the nuts and bolts of cyber defense are reasonably well understood."

Such views encapsulate the current thinking about cyber defense, that such activity is primarily a back office service or a compliance matter. But these views are pernicious. They accept existing systems as they are, other than advocating for improved implementation methods. RAND does not admit that the current hardware, software and networks within the Defense Department are obsolete and dysfunctional. The department continues to operate within a culture that does not acknowledge that its computer systems are not suited for the age of cyber warfare.

Defense Department leadership appears to be viewing cyber defense issues primarily as a matter of policy and strategy that can be fixed incrementally. That is not possible. Cyber defense deficiencies have became deeply rooted as result of the defective ways in which the Defense Department acquired IT over the past decades. Cyber defense flaws are inherently enterprise-wide and are mostly not application specific.

The Defense Department has not as yet confronted what it will take to make systems and networks sufficiently secure. According to DEPSECDEF William Lynn, the department operates over 15,000 networks. The total number of named systems programs in 2009 was 2,190 (Air Force 465, Army 215, Navy 972 and Agencies 538). Each of these programs was further subdivided into subcontracts, some of which are legislatively dictated. Hardly any of the subcontracts share a common data dictionary, or data formats or software implementation codes.

The IT environment at the Defense Department is fractured. Instead of using shared and defensible infrastructure, over 50 percent of the IT budget is allocated to paying for hundreds and possibly for thousands of mini-infrastructures that operate in contractor-managed enclaves. Such proliferation is guaranteed to be incompatible and certainly not interoperable.

Over 10 percent of the total Defense Department IT budget is spent on cyber defense to protect a huge number of vulnerability points. The increasing amount of money spent on firewalls, virus protection and other protective measures are not keeping up with the rapidly rising virulence of the attackers.

---

[35] : http://www.afcea.org/signal/signalscape/index.php/2010/08/23/8000/; http://www.afcea.org/signal/signalscape/index.php/2010/08/23/8021/

Take the case of the Navy/Marine Corps Intranet, which accounts for less than 4.8 percent of Defense Department IT spending. The NMCI contains approximately 20,500 routers and switches, which connect to 4,100 enterprise servers at four operations centers that control 50 separate server farms. Since the NMCI represents the most comprehensive security environment in the Defense Department, one can only extrapolate what could be the total number of places that need to be defended. Vulnerability points include hundreds of thousands of routers and switches, tens of thousands of servers and hundreds of server farms. There are also over six million desktops, laptops and smart phones with military, civilian, reserves and contractor personnel, each with an operating system and at least one browser that can be infected by any of the 2,000 new viruses per day. From a security assurance standpoint, such proliferation of risks makes the Defense Department fundamentally insecure.

Defense Department leadership is aware that cyber operations are important. JCS Chairman Adm. Mike Mullen said that cyberspace changes how we fight. Gen. Keith B. Alexander, the head of the Cyber Command, said that there is a mismatch between technical capabilities and our security policies.

Meanwhile, the interconnectivity of Defense Department systems is rising in importance. For instance, the Navy's Information Dominance Corps views its information environment as being able to connect every sensor to all shooters. Information dominance makes no distinction between logistic, personnel, finance, commander or intelligence data because all of it must be available for fusing into decision-making displays. This calls for connectivity as well as real-time interoperability of millions of devices.

After decades of building isolated applications, the Defense Department has now arrived at an impasse with regard to cyber defenses just as the demand for enterprise-wide connectivity is escalating. Unfortunately, nobody in top leadership has identified the funded program that will remedy the inherent deficiencies in cyber defenses. Prior efforts to do that, such as the Joint Task Force for Global Network Operations (JTF-GNO) and the Joint Functional Component Command for Network Warfare (JFCC-NW) were disbanded. Right now, there are no adequate budgets in place for reducing the widely exposed "cyber attack vulnerability surface." As yet there is no unified enterprise system design or architecture that offers cyber security that works across separate Defense Department components at an affordable cost.

Defense Department IT budgets are now fully mortgaged to support ongoing operations and maintenance, while most large development funds are still paying for continuation of programs that were started years ago. With regard to these concerns, here are some ideas on what should be done:

- The Defense Department should proceed with the rapid consolidation of its communication infrastructure to generate cash that will pay for the merger of costly applications. SECDEF Robert Gates observed correctly on August 9 that "…all of our bases, operational headquarters and defense agencies have their own IT infrastructures, processes, and applications. This decentralization results in large cumulative costs, and a patchwork of capabilities that create cyber vulnerabilities and limit our ability to capitalize on the promise of information technology." Defense Department communications also cannot depend on the routers and servers that are a part of the public Internet. Instead, the department should switch to computing "on the edge" that utilizes government-controlled assets. Communication costs are the largest single component of the Defense Department's IT budget and can be reduced materially.

- The Defense Department should proceed with the consolidation of its servers and pack them through virtualization into a small number of fully redundant (and instant fail-over) data centers. Greater than 50 percent savings are available in operating costs, with payback periods of less than one year. Adopting platform-as-a-service cloud technologies will make that possible. Switching to network operated computing devices (thin clients) and to open source desktop software can also produce additional large savings.

- The Defense Department should complete its data standardization efforts that were started in 1992 and mandate compliance with an enterprise-wide data dictionary. It should proceed with the

standardization of meta-data definitions of all Defense Department data elements. The organization for accomplishing that is already in place.

- The Defense Department should dictate the acceptance of an all-encompassing systems architecture that would dictate Program Executive Officers (PEOs) how to acquire computing services and contractors how to build new application software. The current Defense Architecture Framework (DoDAF) as well as the OSD published architecture directives have not been accepted by the Services and should be superseded.

- From a cyber defense standpoint, the Defense Department should set up network control centers that would apply state-of-the art monitoring techniques for complete surveillance of all suspect incoming as well as outgoing transactions. One hundred percent end-to-end visibility of all Defense Department communications is an absolutely required capability for security assurance as well as for total information awareness.

The recent reassignment of the Network & Information Integration (NII) from the Office of the Secretary of Defense to the Defense Information Systems Agency (DISA) can be seen as an indication that a combination of policy and execution of enterprise-wide communications will be forthcoming. The Cyber Command now controls DISA. There is hope that DoD will finally have an organization that has the charter to deliver working cyber defenses.

However, the combination of NII, DISA, NSA and the Cyber Command is insufficient. Cyber defense inadequacies are embedded into the proliferation of the applications and into the fracturing of the infrastructure. They can be found in the absence of funding to launch a rethinking how to manage cyber defenses in the decades to come.

A different cyber security culture needs to be diffused throughout the Defense Department. It will have to view cyber defenses not as a bandage to be selectively applied to a patchwork of applications. The new cyber security must become an inseparable feature of every computer technology that enables DoD operations.

Information Dominance for War Fighters

According to DEPSECDEF William Lynn, the DoD operates more than 15,000 networks. These networks have no economies of scale. DoD networks do not meet minimum commercial standards for availability or connection latency. Despite what we tell ourselves, these networks do not possess assured security enforcement. There is no enterprise-wide configuration visibility. Children of DoD personnel enjoy better computer connectivity and functionality from their homes than what is available at most DoD installations.

DoD network interconnect more than 7 million information technology devices running over 7,000 major applications. These applications are mostly incompatible. It is rare for systems in finance, personnel or logistics to share even the most basic data elements.

Clearly, the world's largest and costliest networks are not configured to support the needs of our war fighters in the era of cyber warfare.

A New Way of Looking at War Fighter Support

The newly integrated OPNAV N2/N6 has defined "information dominance" as having four key principles:

1. Every platform is a sensor; 2. Every sensor is networked; 3. Every collector and sensor will be dynamically tasked and managed; 4. Every shooter must be capable of using target data derived from any sensor.

These principles have an important difference compared to other information-centric principles and frameworks that have been offered over the last two decades. This definition positions information dominance technologies as weapons that need to be integrated with the other weapons in the military arsenal. This acknowledges that cyberspace represents a military theater with is own sensors, platforms and war fighters. Computer technology and computer networks thus become merely a transport means for information. It is only when sensors are fused with information flows can personal intelligence become applied for making decisions.

But these principles also reassign information technology and information networks into an infrastructure role. They force DoD to revise its thinking about the conduct of its operations. No longer is it possible to depend solely on "network-centric" connectivity based on the Global Information Grid (GIG). Getting a dial tone is insufficient. Instead, we need a framework that delivers sensor-to-shooter useful information in real time. The need is for networks that are so predictable and so reliable that nobody needs to worry about technical performance of DoD's infrastructure. Instead, the military can concentrate on what is relevant for combat.

To create "information dominance" DoD must transform a loose set of independent networks into an interoperable, secure, and cost effective systems for collecting, processing, fusing, displaying text, graphics, video and scientific data. Today's DoD IT infrastructure consumes over 50% of the total DoD budget and employs over 300,000 IT support personnel. That is too labor intensive and highly error prone. Simultaneously decreasing costs, while achieving Google-like ease-of-use and delivering 99.999% uptime—roughly 5 minutes of downtime per year—requires shifting responsibility for managing systems away from information technologists (who are largely contractors) to war fighters who have to make the trade-offs between costs and network services. It requires the elimination of today's stove-piped solutions by imposing enterprise standards. It requires that DoD rethink its approach to IT governance so that IT meets user requirements, rather than today's approach of forcing users to keep accepting what IT can afford to deliver.

None of such thinking is new. There has been no shortage of ideas for more than 20 years. The problem is that despite an ample supply of concepts, DoD is nowadays not achieving information dominance than when I was the most senior OSD IT executive in 1990-1993.

The visions of Tuttle, Atwood, Cebrowski and Stenbit have not as yet materialized in the ways in which DoD is delivering information dominance. For further progress, the following issues must be resolved:

RELIANCE ON INTERNET NETWORKS

The majority of existing 15,000 DoD networks depend on the Internet. Even the GIG relies on services delivered by contractors who depend on links, routers and switches that are vulnerable to attacks. Though elaborate security precautions have been taken for security assurance, it is questionable whether the military can completely trust to convey critical traffic over the Internet from sensors to shooters without a compromise.

All critical communications as well as all messages that can reveal military intent must be conveyed over Internet networks, routers and switches that are owned and operated by the military. Companies such as Google, Akamai and banking firms already have such capabilities at costs that are affordable.

MANAGING NETWORK COSTS

Network costs are greater than IT costs. Over the last twenty years total DoD network costs have more than doubled while the cost of information operations was cut by more then a half and the costs of computer capital has been reduced by 80% for identical performance. At the same time the size of DoD manpower has been declining. By any measure of productivity spending does not show favorable gains in managing rapidly rising costs of personnel and contractors for systems maintenance and for information security because these costs are lumped into accounts where such costs are not identified.

The total DoD spending that is available for networks continues to decline because labor costs are escalating. For this reason the funding for information dominance will have to be extracted from personnel cost reductions. Except for a few recent instances of virtualization of data centers, the management of DoD processes does not favor manpower cuts. For cost reductions it will be necessary to proceed with the automation of the management of seven million desktops as well as the monitoring for improved security can be achieved through the adoption of "cloud computing", which offers quick paybacks. Contractors can operate such networks, but the network control centers and data centers must remain under the cognizance of DoD.

Business operations that account for almost a third of total costs should be also benchmarked against commercial firms. The information management costs of DoD are more than ten times greater than comparable expenses by the largest commercial firms. The cost of the enormous DoD infrastructure should benefit from the economies of scale.

INFORMATION GOVERNANCE

Based on Navy ratios, for every dollar spent on computers and communications (the information transport) there are over thirty cent spent on sensors. Real time as well as 99.999% reliability of the links between shooters and sensors will change the ways to manage tradeoffs for information dominance. The roles of the CIO (Chief Information Officer) will have to be subordinated to the war fighter command structure that starts with USSTRATCOM and flows through USCYBERCOM to military component CYBER commands. Funding should be then allocated on the basis of short-term changes in combat capabilities to support specific missions.

Information dominance is characterized by a very short time intervals allowed for making changes. The typical IT program has a development to fielding cycle of many years. An information superiority system must be adaptable for accepting major changes in a matter of days. Consequently, the acquisition processes cannot follow the patterns of procurement for weapons that have a very long life, but must suit programs that have a life expectancy of a week.

The funding and milestone control of information dominance programs will have to shift from compliance with directives that heretofore placed the acquisition organizations in the lead. Instead, the defense acquisition system that has been designed for airplanes and ships will have to develop new processes adapted to information management as required by Section 804 of the FY10 Defense Authorization Act. The current organizational walls between systems planning, systems development, systems acquisition and systems operation are not viable. The existing systems life cycle process imposes a rigidity of completely separate "silos" that prevent offering information services that can adapt rapidly.

LEGACY APPLICATIONS

There is no way of delivering information dominance with the current proliferation of tens of thousands of aged computer programs that house thousands of incompatible databases. Military services do not have the funds to support migrations to new standard applications are replacing and which are deeply embedded in the local bureaucratic processes. As an example, the Defense Integrated Military Human

Resources System (DIMHRS) failed to get DoD-wide acceptance from the services even after ten years of trying and untold hundreds of millions of dollars. Agreement on an over-arching DoD standard system was not feasible.

For rapid transition to information dominance DoD must follow commercial examples, such as in the case of VISA credit cards that connect, in real-time, hundred thousands of banks (each with different systems) with millions of credit card readers. VISA sets tight standards for centrally managed data and for central security authorizations so that translation tables can be constructed for instant interoperability between incompatible systems. DoD has experience with the use of such mediation programs, such as the Global Exchange Service (GEX). Real-time GEX can work provided bandwidth is available.

MOBILE COMMUNICATIONS

New networks must offer global mobile access to computing and to voice communications by means of portable devices such as smart phones. There should be no distinction or logical separation between messages received at the desktop, desk phone and any portable appliance, regardless where an individual may be located. A unified approach to secure mobile communications will require the DoD to install and to manage organic wireless connectivity as a shared enterprise-wide offering.

ACHIEVING INFORMATION DOMINANCE

The Navy has forged ahead by advancing information dominance as a concept to guide its operations in the era of cyber warfare. It has combined $5.6 billion of computer and network spending with the costs of $3.9 billion of sensors. The corresponding manpower costs are greater than $5 billion for a total FYDP budget in excess of $100 billion.

This combination of intelligence, IT and sensors offers a totally different way to manage information in the years to come. It is noteworthy that the chief of naval operations has approved the Information Dominance Corps Warfare insignia. This means that information warfare is now in the same categories such as surface warfare, aviators and submarines.

The DoD network of the future must be "shooter centric". They must support not only lethal combat but also associated tactical maneuvers. Such systems must be not only reactive but also anticipatory of the adversaries' countermeasures. "Network centric" systems are insufficient in assisting the military in the extraction of data from thousands of databases for delivery of improvised displays which are usable in real-time.

Every DoD shooter becomes the juncture for which networks, databases and applications must operate securely and reliably. Networks are designed only after the connections between the shooter and the relevant intelligence is established. The design of network starts with the enablement of people (shooters) rather than with the technical means for transporting information (e.g. network centric delivery).

Instead of a network centric Global Information Grid (GIG) time has come to augment this concept with the capacity to deliver Universal Interoperability Connections (UIC) that can adapt to future scenarios of warfare that may challenge DoD at least ten years ahead. To accomplish that time has come to move on from network centricity to shooter centricity.

DoD Culture for Cloud Computing

None of the 2.5 million military, civilian or reserve personnel in DoD care much about the technical details of computing. The users only wish to receive answers every time and fast. Requested information must be available regardless of the computing device they use. Responses must be secure. There should be no restrictions as to the place from where they communicate. Information must be available for people authorized to make use of what they receive.

The sources of information must include information received from people, from sensors or from public web sites. Information must be available to and from ground locations, ships, submarines airplanes and satellites. A user must be able to connect with every government Agency as well as with Allies.

What the DoD customer wishes to have is a "personal information assistant" (PIA). Such a device matches a person's identity. It is configured to adapt to changing levels of training. It understands what are the user's verbal or analytic skills. It knows where you are at all times. Any security restrictions are reconfigures to fit a user's current job. At all time every PIA is monitored from several network control centers that are operated by long-term DoD information professionals and not by contractors.

An example of such a device is illustrated below:

What a user sees on a display are either graphic Apps (applications) or choices of text. All that is needed from the user is the entry of authentication signatures.

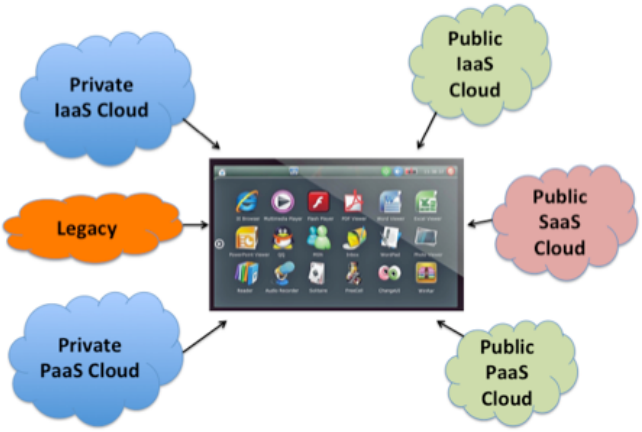
A DoD user will not know the location of the servers that drive their PIA. The user does not care what operating system the App is using or how a graphical display on the PIA are generated. What programming language is used is irrelevant. If a message is encrypted, the embedded security will authorize decryption. Although DoD has thousands of Apps, the user will have visibility of only those Apps that correspond to the situation for which use is authorized.

What the user then holds in his hand reflects what is a Software-as-a-Service (SaaS) cloud. SaaS is the ultimate form of cloud computing. It delivers of what the user can put to use without delays. It is a cheap computing solution for everyone, everywhere by any means.

However, SaaS is not the only model for cloud computing. There are private and there are public Infrastructure-as-a-Service (IaaS) clouds. There are also private and public Platform-as-a-Service (PaaS) clouds. There are also legacy applications that will be ultimately replaced, but will be meanwhile placed on the network as virtual computers.

Virtualized legacy Apps can be useful if encapsulated within IaaS. During the transition from legacy to

SaaS computing every PIA will be able to access all information. The result will be the DoD Hybrid Cloud. The ultimate DoD computing environment can be then represented as follows:



A transition into hybrid computing is a concept of what could be universally available to everyone in DoD. The centerpiece of the future of the DoD computing network is then the ubiquitous Personal Information Assistant, not the servers, not the networks and not the data centers. Servers, networks and data centers are necessary, but concentrating on them represents an obsolete way of thinking. What matters from now is the PIA, which is the computing appliance that delivers an all-inclusive computing presence. The user sees only the results. All the details how results are calculated remain invisible.

PIA-based hybrid networks differ from the client-server environment and even more from the mainframe data center environment. For instance:

In the client server environment there are multiple links from desktops to servers. There are estimated 15,000 networks in place in DoD for such connectivity. In the hybrid cloud environment there is only a single link from any PIA to every conceivable server.

In the client server environment each server-silo operates as a separate computing cluster. In a hybrid cloud configuration there are automated services that discover where the requested data is located.

In the client server environment there is a proliferation of separate databases with duplicate data. Almost every major application keeps storing a diversity of incompatible data because that is dictated by the way isolated applications are configured. This requires added software for the translation of formats. It makes real-time interoperability hard to achieve. In a hybrid cloud configuration there is a universal MetaData directory, which steers database accesses to servers that will provide automated responses.

The key issue facing DoD now is how to migrate from where DoD is now to where it should be, which is a more robust hybrid cloud environment. How fast can DoD transition into a hybrid cloud that delivers universal connectivity to every PIA? How can DoD change its architecture of information so that the technical details of information technology management will remain invisible to users?

The obstacles that block the transition from a fractured computer-centric asset acquisition environment to a universal user-centric architecture are cultural, not technological.

It is not the limitation of technology technologies that are holding up DoD progress. The technology of cloud computing is readily available. What needs to be overcome is the prevailing acquisition-centered culture. Instead of integrating systems for universal interoperability, DoD is still pursuing a policy of fragmentation of systems. Contracts are awarded that are dictated by acquisition regulations and not by operating needs.

The current process for acquiring IT is broken into six phases to contain risks. This separates planning, development, vendor solicitation, contracting for services, asset acquisition, operations and ongoing maintenance. To create a DoD system presently requires the coordination of dozens of organizations, multiple contracts and a large number of subcontracts. This results in an elongation of implementation schedules that are currently twice as long as the technology innovation cycle. Meanwhile projects are managed by short-term tenure of people. This will guarantee that changes in scope will creep in, budgets will rise and results will fall short of what was originally proposed.

The fundamental issue in DoD is not the supply of technologies but the capability of DoD to organize for making the transition into the hybrid clouds. What needs changing is not technology, but culture.

The roots of the current DoD IT culture can be traced to the Brooks Act, of 1965. This legislation reflected a mainframe view of computing. Its purpose was to limit the growth of data centers thereby to constrain costs. The Brooks Act was based on the assumption that if the General Service Administration could control the purchase of mainframe computers, all associated costs would be constrained. This 1965 idea that the closing data centers as a way of cutting costs has persisted to this day. The number of DoD data centers has meanwhile grown at an accelerated rate to 772 in recent years. However, this count has no significance as a cost management metric. Servers, with a capacity far exceeding what is now defined as a "data center", can be purchased for a fraction of mainframe costs. A culture that pursues numerical reductions in the number of data centers without an examination of associated cost reduction, security, application and support characteristics will simplify DoD information processing. This flaw in the Brooks Act thinking was recognized in the Clinger–Cohen Act of 1996.

Clinger-Cohen hoped to address rising costs of information technologies while the military was shifting to new concepts such as "information warfare". While the military was driving towards increased dependency on information technologies, the cost management culture set by Clinger-Cohen was lagging behind the military needs. Although Clinger-Cohen created the position of the Chief Information Officer (CIO) who was supposed to promote warfare-centric deployment of IT, the management of information technology continued to promote isolated projects as a way of minimizing risks.

The central CIO oversight was diluted when the role of acquisition personnel was enlarged. The acquisition corps, dominated by civilians with financial and not a computer science background, had every incentive to award contracts that followed regulations that were at least 20 years old. This fixed the DoD culture into client-serve concepts, which are alien to the ideas that are now getting implemented by leading firms such as Google, Amazon, Microsoft and IBM as well as thousands of commercial enterprises.

The unity of oversight that was advocated by Clinger-Cohen never happened. What we have now are the remainders of a culture that has been spending money on networks on the assumption that if you build the networks, integration must follow somehow. That is like building superhighways without worrying about the sources of traffic.

Clinger-Cohen had several consequences. The share of IT spending allocated to the DoD infrastructure, which supports an increasingly disjointed collection of systems, grew from 33% in 1992 to the present 57%. Since infrastructure reflects system overhead, less money would be now available for serving the warfighter's needs.

Under Clinger-Cohen there has been a pronounced shift of spending to contractors as well as to the legislatively mandated set-aside subcontracts. The best estimate of IT spending now managed by contractors is 76% of IT costs. There are well over 1,000 contracts with annual budgets of less than $1 million. The amount of documentation, testing and certification demanded by the stretched force of acquisition executives was now imposed as the preferred way of containing risks. The costs per function delivered by contractors now rose to a large multiple of commercial costs.

As the consequence of shifting to contractors there was depletion in the cadre of qualified military information professionals needed to provide user leadership. As an example, there are only two young flag level officers presently in the Navy identified as "information professionals". This can be contrasted with a long-tenure four-star flag Navy officer in charge of nuclear propulsion. He is supported by 12 flag officers. There are good reasons for the argument that in terms of complexity and scope the migration into the cloud environment may be comparable to the challenges (and obstacles) faced by Admiral Rickover. To support the future of computing in DoD requires a culture that views IT as a weapon and not as a back-office task that is best delegated to suppliers!

One of the consequences of Clinger-Cohen has been the shift of a large share of information technology costs from the Services to the Agencies such as DISA, DLA and DFAS. Agencies now spend twice as much ($14.6 billion) on IT than each of the Services individually (averaging $7 billion each). As result the culture favors the distribution of information technologies into 2,103 systems "silos". Due to funding limitations almost every system enclave must pursue architectures, applications, standards, software design methods and operating practices.

Small projects do not have sufficient money to fund increasingly complex security requirements. There is a rising dependency on the support from the Agencies who have no fiscal accountability for work done on behalf of customers.

Systems managers now concentrate on carving out the largest IT budget they can get rather than on keeping up with a rapidly changing technological environment. Clinger-Cohen left DoD with a culture that is over 20 years old just as information technologies are charging ahead with cloud concepts. Individual project managers do not have the funds to invest in innovation. They are just trying to manage ever smaller incremental projects.

A recent example of fracturing of systems into smaller components is discussed in a report by GAO-11-150 of March 2011. To increase the number of bidders for a major system, the tasks will be broken from the existing 3 contractual relationships to 21. Instead of the integration that is necessary for cloud computing to progress, one of the largest DoD projects will be headed in the opposite direction,

In FY11 the DoD information technology had rapidly expanded to a $36.3 billion, not including the costs of military and civilian personnel, which would add at least another $6 billion to the total cost of IT ownership. The per capita cost of IT spending now represents over 10% of payroll costs. It exceeds the expenses for most fringe benefits.

The DoD IT systems management culture has now arrived at a dead-end of a wrong street. What is then the way out? What cultural changes will be necessary to speed up the adoption of the cloud paradigm? What is the time urgency?

There are two variables that will influence decisions what to do next. The first one is the need to make vast improvement in the security in DoD operations to survive cyber attacks. The second, and perhaps more important is financial. To pay for better security and fund long overdue innovations new funds are needed. That can come not from larger budgets, but from savings that must be somehow extracted from current operations.

The fastest generator of DoD cash savings is to transfer the operation of "commodity computing" to the Software-as-a-Service clouds. Commodity computing includes: e-mail, collaboration, text processing, messaging, spreadsheets and calendars. It includes Voice over IP. Commodity computing also includes access to all forms of social computing such as YouTube, Facebook, MySpace Twitter and blogs.

Commodity applications consume a large share of the $4 billion per year FY11 network costs of DoD. Most of that is concentrated in DISA. It is feasible to use communication over the Internet as a secure

channel. There is no reason why such traffic should not be routed over the Internet instead of DoD operating enormously expensive links for commodity applications. Banking transactions, airline reservations and global trade are all conducted over an Internet that has been made more secure.

The technology for placing DoD commodity computing and communications into a SaaS cloud is available. SaaS services or licenses are available from competitive sources at a fraction of what it costs DoD. SaaS depends on Open Source applications, which further reduces license costs for proprietary software. 60-70% operating cost reductions are potentially available with only minimal up-front development expense.

Migration to SaaS must overcome many obstacles. First, DoD will have stand up an organization that will plan, manage and contract for commodity computing. Second, DoD components will have to commit to a uniform approach to network access authorization by everyone. Third, information security will have to be implemented consistently, following standards set by NSA. Fourth, every commodity computing service will had to be structured for interoperability across several SaaS platforms so that vendor lock-in cannot take place. Fifth, SaaS services will be delivered from redundant data centers. More than three data centers should handle every transaction for delivery of 99.9999% reliability. Sixth, SaaS would have to be distributed "to the edge" of DoD locations so that Google like latency can be realized. Seventh, the delivery of SaaS transactions will have to be monitored by means of automated network control centers.

SaaS can operate only as a component of a hybrid cloud configuration. Legacy, IaaS and PaaS clouds must supplement DoD computing and telecommunication during a transition that may be never completely finished. This will be discussed in the next Viewpoint article in September.

Reducing the Costs of Data Centers

According the Federal Chief Information Officer, the DoD was operating 772 data centers as of July 30. 2010.[36]

OMB defined a data center as any room that is greater than 500 square feet and is devoted to data processing. Kundra called for a 38% reduction in the number of data centers by 2015. Though such calls are driven by budget considerations, the metric of counting how many data centers can be eliminated is misleading. From a budget standpoint only the reductions in DoD's $36.3 billion FY11 IT expenses will matter.

Sun Microsystems, Hewlett-Packard, Dell, Microsoft and Silicon Graphics (SGI) now offer complete data centers in standard 40x8 ft. and 20x8 ft. shipping containers, occupying 320 and 160 square feet respectively. Such data centers have the capacity of up to 29.5 petabytes of storage and up to 46,080 CPU cores of processing power. Such data centers would not be included in any of the OMB statistics.

[36] http://cio.gov/documents/25-Point-Implementation-Plan-to-Reform-Federal%20IT.pdf

The current costs of petabyte files ranged from $117,000/petabyte to $826,000. It is declining steeply.[37] Therefore the petabyte costs of a single container – the single largest hardware cost - would be up to $24.4 million. The estimated 100 petabytes to support most initial DoD cyber operations would be close to $90 million, which is well within the range of affordability if one considers that the total DoD budget for IT Operations and Maintenance was $25.1 billion.

Perhaps the greatest omissions in the DoD data center count are the installations that are operated in the contractors' names. In addition, in DoD there are a large number of locations, which are not designated as data centers, but are nevertheless running many servers where each can cost less than $500/year. Small clusters of servers are costly because of their expense for operators, electricity and overhead. Excluding small locations from consolidation plans loses an additional savings potential.

Current server virtualization technologies offer a large multiple of the computing power than was available ten years ago. Consolidations through virtualization have so far achieved 10:1 reductions in the number of servers hosted in data centers. Just shrinking the number of servers will be sufficient to reduce the need for data center floor space. DoD Components may be therefore tempted to use server consolidations in lieu of data center reductions.

Counting the number of servers is not a good metric and should not be an index of data center size. Server consolidation without virtualization can replace poorly utilized $500/year servers with a poorly utilized $5,000/year mainframe unless the entire software environment in the data center is also overhauled. That will take additional effort than just facility consolidation.

Therefore, any plans for proposing consolidation of data centers to a much smaller number of secure sites must show how the total cost of ownership of Operations and Maintenance will be realized. What will be the net cash-flow break-even point for delivering the savings? Since DoD accounting does not show for depreciation expense, any cash-flow projections must also include the costs of new capital expense as well as any write offs for equipment that will be discarded. Plans that focus only on data center facility counts are insufficient.

The greatest gains in data center consolidations come not from a reduction in the hardware costs (<10% of operating costs) but from the cutting applications operations and maintenance expenses of personnel (>50% of cost) plus data center overhead (>20%). Every data center incurs the costs for software maintenance support, help desks manning as well as the expense for supervisory overhead. An excessive concentration on the shrinking of hardware costs as well as on power efficiencies and real estate charges will overlook what are the major savings opportunities, which are the expenses for direct and indirect labor.

[37] http://pstrassmann.blogspot.com/2010/07/petabyte-files-for-cloud-computing.html

Any discussion of data center consolidation must also include provisions for backup, fallback and service availability. Stand-alone, hard to defend and underfunded sites, with low capacity utilization, will result in service levels that are below acceptable levels for cyber operations, which are at least 5 minutes of total downtime/year and less than 200 milliseconds latency. Reconstitution of a failed data center, as is currently the case, cannot take place in hours and even days, which would be intolerable for cyber operations.

The primary objectives for data center consolidation should be therefore the improvement in reliability, performance and exceptional levels of information security. Cost reductions are necessary but only secondary.

The analysis of data center consolidation savings must include an evaluation of related communications costs. The speed of propagation of transactions over telecomm lines can be counted in a few milliseconds anywhere in the world. A reduction in the number of data centers will be feasible only if greater bandwidth is available to connect every major data centers with direct links over fiber optic channels rather than via Internet.

Any DoD commitment to reduce the number of data centers should be taken only after an examination of its consequences. For instance, all data centers have the characteristics that they support a dedicated number of customers. Each customer is then connected to their respective servers with unique Local Area (LAN) and Wide Area (WAN) connections. One cannot just rip out servers and mainframes without the careful re-engineering of whatever links are already in place. The existing LAN and WAN connections are supported by hundreds of small sub-contractors who provide maintenance and operating services, each using the best of custom-made technologies that they have found to be affordable. None of this connectivity is standardized. Anyone attempting data center consolidation cannot pick up the existing connecting links without a further examination and testing at ever point during the transition from many data centers to just a few.

In any of the 700+ DoD data center you will find multiple versions of operating systems, homegrown security appliances, unique communication processing methods, incompatible device interfaces and different network management consoles. It is unlikely that all of the procedures that are in place are adequately documented. Any data consolidation effort will have to develop a plan how such variety can be synchronized for operation in the new environment.

Resident contractors retain much of the know-how for operating the data centers, particularly under the condition of failure. Potentially, the DoD data centers can include up to 2,814 different server versions connected to up to 1,811 listed versions of customer clients that are managed by up to 1,210 versions of operating systems. If all of this is folded into a cloud, all of this diversity must be reduced first.

How much of this variability is present in the existing DoD data centers is not known but my guess is that our legacy systems remain a large repository of obsolete software or applications that have not been upgraded for years. Suppliers have also contracts that will either have to be liquidated or be paid off for their unexpired term. In any migration to a consolidated environment DoD will also have to deal with conditions where legacy operators have installed unique procedural steps to preserve their lock-up on the conduct of DoD's business.

Every data center contractor attempting to merge data centers into an Infrastructure-as-a-Service or a Software-as-a-Service cloud environment, as currently proposed by DISA, will have to first contend with the software complexity how to migrate out of an existing operating environment into a DISA cloud. DISA, or a cloud equivalent set up by the Army, Navy, Marine Corps or the Air Force, will have to then define a way of handling customers in tightly prescribed ways after any consolidation takes place as well as during every step taken during the intervening migration process.

DISA has now staked out a claim to become the leading provider of cloud computing services to the

Defense Department for both unclassified and classified data.[38]

This will require major new investments. Alternatives how to implement cloud services show at least 36 different variations. For instance, the Cloud Security Alliance (CSA) has published 192 cloud guidelines. The criteria include: Cloud Computing Architecture; Governance; Risk Management; Legal Matters; Compliance and Audit; Lifecycle Management; Portability and Interoperability; Operations; Business Continuity; Disaster Recovery; Incident Response; Remediation; Encryption; Key Management and Access Management.[39]

Any evaluation of cloud computing should consider technology acquisition from qualified cloud firms such as Amazon; AT&T; T Synaptic Hosting; BlueLock Virtual Cloud Computing; Enomaly; GoGrid; Google; Hosting.com; Microsoft Azure; NetSuite; Logica; Rackspace Cloud; RightScale; Salesforce.com; Terremark vCloud Express and Unisys Secure Cloud. Each of these offers combinations of features and functions that offer various degree of lock-up into their offerings. Which one of these technologies will offer to DoD a choice of moving smoothly computer processing from one cloud vendor to another is a matter that will require choices as to who offers the greatest application portability.

Before proceeding with a large number of data center mergers, particularly into DISA, planning must be devoted to the disruption that DoD's users will suffer during migration. Any plan to restructure data center operations should first start with investment into the engineering that would dictate what will be the standards and how the new DoD data center environment will function. Dictating the cuts in data centers without also detailing how the transitions will support desktop hardware, desktop software and communication processing is risky. Many of the millions of existing DoD desktops are custom-connected to what are purely local arrangements between the servers and the respective customers.

Discussion about data center consolidation must also include the organization and the set up of Network Control Centers (NOCs) for managing the interoperability and for assuring the security of the networks that connect all data centers. With the growth in the size of computer networks the importance of central monitoring of operations rises. There are already a large number of networks that manage hundreds or even thousands of servers. It is likely that DoD already operates about 100,000 servers. For example a 2010 server census shows the following number of servers: Intel: 100,000; OVH: 80,000; SoftLayer: 76,000; Akamai Technologies: 73,000; 1&1 Internet: 70,000; Facebook: 60,000; Rackspace: 63,996 servers, Google: >800,000 servers and Microsoft >300,000 servers. All of these complexes have in place sophisticated and highly automated NOCs and DoD would have to follow their commercial examples how to organize large networks.

Such large aggregations of equipment require real-time monitoring of security status, up-time, latency, capacity and response time. To take advantage of the limited technological expertise DoD will have to rely on Network Control Centers (NOC) for monitoring all networks as well as for managing all of the ten thousands of attached assets. In order to achieve interoperability and to conserve the services of scarce personnel a wide range of fully automated diagnostic routines will have to become standardized across all NOCs.

The tools that are available for NOC operations will become a major investment for DoD. Such tools will represent a major share of "intellectual capital" that will be necessary for the consolidation of data centers to take place. It will require also a tight integration between NOC operations and the Global Information Grid (GIG). Such coupling between the GIG and the NOCs will have to be included in the entire data center

[38] http://www.nextgov.com/nextgov/ng_20110103_7911.php?oref=topstory

[39] http://www.cloudsecurityalliance.org/

consolidation program. Ultimately, it will require moving the accountability for most of status monitoring from the individual data centers to the NOCs. The NOCs then become the first line of defense for assuring the security of cyber operations.

There are economies of scale available from the consolidating of data centers into complexes with over 500,000 sq. ft. of computer floor space. Huge data centers show a dramatic lowering in the costs of information processing, in decreases in manpower expenses while also increasing the reliability and latency of what would ultimately become one of many DoD's "server farms." There are further savings to be had from the lowering of the costs of electric power for running computers and for air-conditioning. The annual cost of electricity now exceeds the depreciation cost of computer hardware.

There are new firms that build and equip data centers. Commercial data center operators build efficient large facilities and then lease them either as totally dedicated facilities, or as "cages" available for partial occupancy. In the case of major companies, such as Apple, Microsoft or Facebook, the computer configuration in the "server farms" is standardized to meet the company's proprietary system requirements. For example, data center venture firm of Sabey is currently building a 350,000 sq. ft. data center for Dell. The firm of CoreSite offers several locations with finished "wholesale" data center space for users who seek turnkey space that can be deployed quickly. The Equinix operates 22 data centers in the USA, 7 in Europe and 5 in Asia. Some of the large Equinix data centers are larger than 200,000. Rackspace operates 9 data centers, which includes managed hosting. The firm of Interxion operates 28 sites in Europe.

Existing DoD datacenters have been constructed over the past thirty years in small increments, each funded by separate contracts. Even the DISA mega centers, authorized in 1992 by DMRD 918, do not have the scale that matches anything that is getting constructed now. DoD data centers do not reflect the economies of scale that have become available on account of optic communications and new virtual software. Meanwhile the cost of hardware has been shrinking at the rate of 18%/year, while the expenses for electricity and operating manpower have been rising to meet a growth in demand for services. The net result is that the existing proliferation of DoD data centers is not affordable.

The existing data centers have their origins in separate contracts, which dictated how computing facilities would be acquired and organized. It is also clear that the current number of data centers cannot support the increasing demands for security as well as for reliability without an increase in budgets. With capital costs for the construction of economically viable data centers, now approaching $500 million, it is unlikely that such capital would become available with the prospective budget cuts and the relatively short schedule that is available for DoD to arm itself for cyber operations.

The current FY10 O&M (operating and maintenance) of I.T. is $25.1 billion. This makes the leasing of DoD-dedicated data centers, constructed by any one of the many commercial firms, a manageable proposition. It may be the way in which data center consolidation into DoD clouds can be prepared for the time when information warfare will mandate a different way of configuring I.T. operations.

Optical Fiber or Wireless?

I discovered that Australia is engaged in a vigorous debate about a greater than $36 billion investment in broadband optical fiber. A new, state owned network monopoly, would deliver to every home at least 1 GB of bandwidth connectivity.

The Australian debates prompted me to look into the future of wireless bandwidth to examine the

optical fiber vs. 4G wireless tradeoffs.

The International Telecommunication Union defines 4G as a downlink speed of 1 gigabit/sec for stationary or slow moving users and 100 megabits/sec for when devices are traveling at higher speeds.

A 4G network is expected to provide all-IP based broadband to IP telephony, connectivity with laptop computers, access to wireless modems and support of smartphones. Current wireless carriers have nothing like that despite repeated claims of 4G. Technically, what carriers offer are pre-4G, or even 3G and a half capabilities. ITU lets carriers advertise LTE and WiMax Advanced as 4G because these networks are significantly faster than the established 3G technology, which runs at about 14.4 megabits/sec downlink. WiMax can deliver up to 70 Mbps over a 50Km radius.

When in place the 4G technology will be able to support interactive services like video conferencing (with more than 2 sites simultaneously), high capacity wireless Internet, and other communications needs. The 4G bandwidth would be 100 MHz and data would be transferred at much higher rates. Global mobility would be possible. The networks would be all IP and based on the IPv6 protocol. The antennas will be more capable and offer improved access technologies like OFDM and MC-CDMA (Multi Carrier CDMA). Security features will be significantly improved.

The purpose of 4G technology, based on a global WWWW (world wide wireless web) standard, is to deliver "pervasive networking", also known as "ubiquitous computing". The user will be able to simultaneously connect to several wireless access technologies and seamlessly move between them. In 5G, this concept will be further developed into multiple concurrent data transfer paths.

In the United States, the immediate challenge is finding the wireless spectrum. Recent tests in by LightSquared's ground-and-satellite LTE service found that it interfered with GPS signals. Therefore, the FCC is holding back on proceeding further.

Meanwhile the FCC is considering an interim deployment and operation of a nationwide 4G public safety network, which would allow first responders to communicate between agencies and across geographies, regardless of devices. The FCC released a comprehensive paper, which indicates that 10 MHz of dedicated spectrum currently allocated from the available 700 MHz spectrum for public safety will be used to provide adequate capacity and performance for special communications as well as emergency situations.

While the US is holding back on 4G South Korea has announced plans to spend US$58 billion on 4G and even 5G technologies, with the goal of having installed in S. Korea the highest mobile phone market share after 2012, with hope to set the international standards.

Japan's NTT-DoCoMo is jointly developing 4G with HP. A the same time Korean companies like Samsung and LG are also proceeding into 4G to gain global market share in advanced smartphones. Recently Japan, China and South Korea have started working on wireless technologies and they plan to set the global standards for 4G.

A 5G family of standards would be implemented around the year of 2020. A new mobile generation has so far appeared every 10th year since the first 1G system was introduced in 1981. The 2G (GSM) system rolled out in 1992. The,3G system, W-CDMA/FOMA, appeared in 2001. The development of 2G and 3G standards took about 10 years from the official start. Accordingly, 4G should start to be installed after 2011. There is no official 5G development project though this may take place only after a 2020 launch.

The development of the peak bit rates offered by cellular systems is hard to predict, since the historical bit rate development has shown little resemblance with the exponential function of time. The data rate increased from 1G (1.2 kbps) to 2G (9.6 kbps). The peak bit rate increased by a factor 40 from 2G to 3G for mobile users (to 384 kbps), and by a factor of 200 from 2G to 3G for stationary users (2 Mbps). The peak bit

rates are expected to increase by a factor 260 from 3G to 4G for mobile users (100 Mbps) and by a factor 500 from 3G to 4G for stationary users (1 Gbps).

Affecting the launch of 4G and 5G wireless technologies is the growth in mobile data traffic at CAGR of 92 percent between 2010 and 2015. Global mobile data traffic will grow three times faster than land based IP traffic from 2010 to 2015. Global mobile data traffic was 1 percent of total IP traffic in 2010, and will be 8 percent of total IP traffic by 2015.

SUMMARY

Whether Australia should proceed with digging trenches to every household to secure 1GB connectivity by 2015 is debatable. There are also political considerations.

First, the Australians will have to build a wireless network and erect wireless towers anyway. They will have to provide wireless connectivity for wireless traffic that is growing much faster than land-based traffic. Wireless assets will continue to require steady new capital investments as the S. Koreans, Chinese and Japanese forge ahead with 1GB wireless 4G service prior to 2015 and with 5G after 2020.

Second, the capital costs of any technology upgrades would be always much higher for landlines, based on fiber optic fiber, than on wireless towers. The operations and maintenance costs for last-mile copper circuits that will have to remain place exceed the capital costs of additional wireless towers.

Third, LTE and WiMax Advanced wireless already in place have the capacity to support over 14.4 megabits/sec downlink and up to 70 megabits/sec. I presently pay a premium for a 12 megabits/sec downlink high-speed connection to the Internet. That is completely satisfactory for my extensive Internet usage as well as movie downloads. What the Australian can do with a ready availability of over 14.4 megabits immediately is not clear.

Lastly, there is an issue of how to choose technologies that would lend themselves to market competition vs. monopoly operations. The capital cost for placing fiber optic cable is an investment that has a technology life of over 50 years. Fiber optic cable to every home suits a monopoly model of pricing, since the costs of the delivery of services are unrelated to the costs of capital investment. In contrast, once wireless towers are in place, they can host a variety of wireless providers at costs that can be adjusted to reflect geographic characteristics. There is no question that proceeding with wireless connectivity would allow diverse competitive offerings to co-exist.

With the prospects of a rapid evolution in wireless connectivity as well in consideration of the widespread dispersal of Internet users in Australia, it does not appear to make sense to commit immediately to expensive fiber optic circuits to support delivery of Internet services to the Australians.

Networks Without Satellites?

Every cyber operation contingency plan must include a case in which the use of most satellite connectivity will be lost. Whether such failure is caused by adversary intervention, malicious jamming or mechanical failure is irrelevant. Military operations depend on the availability of assured bandwidth. A contingency plan must also assume conditions of partial failure when only partial capacity is available for conveying minimum essential bandwidth. Partial communications must allow Network Control Centers to cut off low priority transactions, while keeping up SIPRNET or supporting only channels that deliver

mission-specific messages.

There are several options that could be considered as a replacement for satellite channels:

1. Use of high-altitude vehicles. An example of this solution is the Global Observer Unmanned Aerial Vehicle (GOUAV) or its successors, including solar powered planes. GOUAV has already conducted test flights. It will fly at an altitude of between 55,000 – 65,000 feet for 5 – 7 days. Longer times aloft are likely to be feasible in the future. That's above the weather and above sustained conventional airplanes. That height also helps due to the laws of physics, which allow aircraft at that altitude to cover a circular area on the surface of the earth up to 600 miles in diameter.

The goal for Global Observer is for payload capacity of >1,000 pounds. It uses liquid hydrogen fuel and fuel cells that drive 8 small rotary engines. It can handle communications intercepts over a wide area. In the future it will have the capacity to augment transmission of telecom bandwidth. It could be designed to act in lieu of GPS satellites. Multiple communication and remote sensing applications have already been demonstrated including high definition broadcast (HDTV) video, and third generation (3G) mobile voice, video and data using an off-the-shelf mobile handset.

The unit costs of loitering airplanes that act as store-and-forward stations make their deployment attractive. They are mobile and can be repositioned to avoid interference. Without crews they are easy to replace. They can be launched only as needed anywhere in the world when needed.

There are also other technologies available, such as the Boeing X-37B Orbital Test Vehicle that could be used to place in orbit steerable platforms for communications.

High-altitude vehicles (HAVS) offer many advantages over satellites, whether they are in low polar orbits (100 to 1,000 miles) or in stationary equatorial orbits of 22,300 miles above the earth. HAVS can be launched quickly to support a military mission as needed, rather than holding a persistent spot in space. HAVS can be steered and can be recovered for technology enhancement.

2. Use of a combination of ground based and wireless "mesh networks". Mesh networking is a type of networking where each node must not only capture and disseminate its own data, but also serve as a relay for other sensor nodes, that is, it must collaborate to propagate the data in the network. In effect a wireless mesh network replicates the structure of the land-based and hard-wired Internet.

A mesh network can be designed using a routing technique. When using a routing technique, the message propagates along a path, by hopping from node to node until the destination is reached. For insuring all its paths' availability, a routing network must allow for continuous connections and reconfiguration around broken or blocked paths, using self-healing algorithms. As a result, the network is reliable, as there is often more than one path between a source and a destination in the network, which can include both wired as well as wireless links.

The nodes on a mesh can be stationary, such as transmission towers, buoys or aerostats, or mobile such as ships or UAVs. The connectivity between mesh nodes can vary depending on bandwidth, antenna design and available power. For instance, High Frequency (HF) microware links can be set up so that a fleet of ships can be positioned in a way that assures delivery of transactions via mesh-like connections to their ultimate destination.

Mesh networks can be also deployed tactically. It is possible to deploy low cost unmanned helicopters (such as the Navy's MQ-8B Fire Scout) or long endurance helicopters (such as the Boeing A160 with 2,500 mile range). Swarms of these vehicles can be launched to act as store-and-forward communication nodes. At a moment's notice they can offer instant connectivity for rapid attack expeditionary forces.

Mesh networks have the advantage that they can be built as an extension of existing networks that use optical fibers. The construction of mesh network can be done as an evolutionary program, without obsoleting legacy networks already in place. In this way they can function as redundant hybrid mesh connections, with existing ground links or HF ship-to-ship and ship-to-shore links routing transactions through the most cost-effective path.

SUMMARY

The U.S. military cannot depend on satellites for support of its expeditionary forces. Even if degraded for transmission of only essential messages, multiple and fully redundant communications paths must become available. Unfortunately, the existing proliferation of network connections costs too much. It is also not interoperable. Its use of the available spectrum is profligate.

To make available communication options that dispense with most of the satellite links will require a different approach to network design. It will require funding that will not be available unless costs can be extracted from the existing communication arrangements (see diagram below of antennas on an aircraft carrier). Right now it is now apparent that the centerpiece of DoD's networks – the Global Information Grid (GIG) – has provided for such eventualities. Taking on the task for providing satellite-less bandwidth for its ships and expeditionary forces should be a priority tasks for the Navy's N2/N6 organization.

Open Flow Protocols

The Open Networking Foundation (ONF) has just been organized to create protocols that would make it possible for firms to control the processing of Internet transactions on switches and routers. ONF includes leading firms involved in Internet networking, such as Broadcom, Brocade, Ciena, Cisco, Citrix, Dell, Deutsche Telekom, Ericsson, Facebook, Force10, Google, Hewlett-Packard, I.B.M., Juniper, Marvell, Microsoft, NEC, Netgear, NTT, Riverbed Technology, Verizon, VMWare and Yahoo. [40] The new ONF protocols, named "Open Flow" will radically change how communication networks will operate in the future. This protocol will become an industry standard as an add-on feature to the existing IPv4 and IPv6 protocols. Open Flow code will be also embedded in network controllers, switches and routers. The first router that uses OpenFlow is already in place. Prototype installations have been in place at over dozen universities since 2008.

The objective of ONF is to make networks programmable in much the same way that individual computers can be programmed to perform specific tasks. This represent a major departure form the current approach in which the Internet switches and routers are pre-defined so that they cannot be modified to accommodate dynamic fluctuations as the traffic on networks keeps changing. OpenFlow focuses on controlling how packets are forwarded through network switches and routers.

In the past one of the key components of any system could not be programmed. That was the network that connected computing nodes. Under OpenFlow it will now be possible to customize networks to the applications that are actually being run.

[40] http://www.nytimes.com/2011/03/22/technology/internet/22internet.html?_r=1&partner=rss&emc=rss

OpenFlow protocols and associated software should open up hardware and software systems that control the flow of Internet data packets, systems that have until now been closed and vendor proprietary. This will cause a new round of innovation that will be focused principally upon the emerging computing systems, known as cloud computers, that require a variety of network services that currently are not available.

For instance, OpenFlow will permit setting up on-demand "express lanes" for voice and data traffic that is mission critical. Software will allow combining several fiber optic backbones temporarily for particularly heavy information loads and then have circuits automatically separate when a data rush hour is over. Another use of OpenFlow will be load balancing across an entire network, so that diverse data centers will be able to shuttle the workload so that performance does not deteriorate.

OpenFlow will be an open interface for remotely controlling the forwarding tables in network switches, routers, and access points. Based on such capabilities user firms will be able to build networks with a much wider scope, especially involving wireless communications. For example, OpenFlow will enable more secure default fail-overs, wireless networks with smooth handoffs, scalable data centers, host mobility, more energy-efficient allocation of resources and ready deployment of improvised new networks.

SUMMARY

OpenFlow warrants immediate attention even though large-scale implementation may be three to five years ahead. However, it will alter the architecture of networks, such as GIG, to a significant extent. The programmability of networks will change the role of the GIG from just a transmission medium to a component that becomes an active part of the design for DoD. Such planning should be starting soon. OpenFlow equipment will have to be provided for in acquisitions that will have a life of well over ten years.

How Secure is a Virtual Network?

The debate about the future of DoD networks and how that can be delivered over the Global Information Grid (GIG) revolves around the question whether the GIG could rely on Internet-based connectivity. The scope of the GIG is enormous - it encompasses over 10,000 routers and 10 million hosts, including wireless connections. It includes a wide range of nodes, link types as well as human portable and battery powered devices. The GIG provides capabilities from all operating locations (bases, posts, ships, camps, stations, facilities, mobile platforms, and deployed sites). The GIG also provides interfaces to coalition, allied, and non-DoD users and systems.

The GIG overarching policy makes clear that its objectives are all-inclusive of every application, anywhere. It does not accept the acquisition of IT capabilities as stand-alone systems. It rejects designs that are defined, engineered, and implemented one pair at a time – an approach that focuses on system or platform capabilities rather than on mission capabilities.

Instead, all DoD systems shall be based on a shared GIG. It will be based on a common, communications and computing architecture that provides a full range of information services, for all security classifications and information handling needs.[41]

[41] DoD Directive 8100.1, 11/21/2003

GIG data shall be shared and exchanged through common interoperable standards that will be based on the IPv6 common network protocol. It will allow all types of data to move seamlessly on the GIG's diverse transport layer, which includes landline, radio, and space-based elements. This means that every network link in DoD must be interoperable from a protocol standpoint. That is not the case at present. The diversity of the existing network protocol is not known.

The GIG supports mission critical operations. For complete security it must use IPv6 formats (e.g. IPSec). Connectivity to the GIG depends not only of land circuits and wireless links but also on Radio Frequency (RF) and satellite connections. The GIG must be a trusted network that comprises of units in the field (e.g. army companies or ships), which need to be seamlessly connected to the GIG even while they are mobile.

GIG will operate in accordance with common metrics, measurements, and reporting criteria.[42] Originally, the GIG was conceived as a federation; that is, ownership, control, or management of the GIG (people, processes, and hardware/software) was distributed throughout the DoD.[43] That approach did not work though it was reaffirmed by the Instructions from the Chairman of the Joint Chiefs of Staff in December 2008. Instead, the implementation of the GIG concept is now in the hands of USCYBERCOM.

In planning the evolution to the long-term GIG objectives does DoD really require a totally enclosed Intranet (such as NMCI) to assure its security in the interim? Is it possible to make the Internet sufficiently secure so that the costly acquisition of a variety of dedicated circuits, such as for the NGEN transition, is not necessary? Does it make sense for the individual services to continue with the contracting for dedicated networks that provide services to only a limited set of applications?

Though dedicated transmission lines can speed up communications and reduce latency of transmissions between major hubs, the costs of connecting all DoD locations must be planned as a part of the GIG and not on a stand-alone basis. The vulnerability of the Internet to security compromises (which includes corruption of LANs, WANs, intermediate switches and routers) is well understood. DoD will therefore have to resort to specially configured VPNs (Virtual Private Networks) to protect its transmissions. This cannot be done for only local networks, but must be engineered so that the protocols can be imposed universally.

DoD must develop and install DoD-specific VPN implementations because VPNs are a method for using the existing Internet infrastructure to provide secure access to all every IP addresses. This avoids the expense for dedicated implementing networks that carries only DoD traffic that work only for individual contracts. The public Internet offers an enormous redundancy with highly distributed links that will overcome local circuits failures that cannot be achieved economically by other means. Internet is more resilient against failure than any Intranet that could be designed, except at an exorbitant cost. However, any reliance on the Internet must be first engineered for enhanced security that would be approved by NSA and only then imposed as a uniform GIG solution.

A DoD version of VPN will encapsulate all transactions using NSA approved cryptographic methods between any points. Cryptography will then keep all data transfers private from intrusion by any internal or

[42] DoD Instruction 8410.02, 12/19/2008

[43] *cio-nii.defense.gov/docs/**GIG**ArchVision.pdf*

external other source. That will also safeguard against security breaches that could happen during a transmission until a transaction reaches its final destination where it can be decrypted.

There are several different classifications, implementations and uses for VPN solutions, which includes compliance with additional restrictions set by NIST and validated by the NSA. There are several standard protocols that assure how the "tunneling" of traffic will take place and how it can be inspected by DoD Network Control Centers. There are codes that will have to be added by DoD to support the end-to-end intrusion-proof procedures throughout the entire transmission sequence. The tunnel's termination point, i.e., customer edge, will finally offer the authentication of a legitimate recipient while still remaining subject to USCYBERCOM controls.

The most important requirement of a VPN are cryptographic protocols that block any intercepts and which allow sender as well as recipient authentication to preserve message integrity. This includes IPSec (Internet Protocol Security), which was originally developed as a requirement for IPv6. Until that protocol is implemented (see http://pstrassmann.blogspot.com/2011/02/are-ipv4-addresses-exhausted.html) the IPv4 Layer 2 Tunneling Protocol could be used as a substitute, though that is not recommended.

SUMMARY

VPNs play a central role in the (GIG), the combined network-of-networks being developed by not only DoD but also by other US government agencies to support the communication needs of the security, defense and intelligence communities. The GIG architecture can be viewed as having two main components, namely trusted edge networks and a large backbone core consisting of a combination of both trusted and untrusted network segments. In order to achieve privacy and integrity of the data crossing the backbone, edge networks must use consistent VPN gateway protocols to encrypt traffic as it passes through thousands of routers and switches.

VPN will reduce network costs because it avoids the need the dependence on dedicated lines that connect offices to private Intranets during the transition from the current state to where the GIG will ultimately provide for all connections.

Meanwhile, DoD will require the use of dedicated fiber optic connections primarily for back-up and fail-over traffic among data centers. DoD may also find it advantageous to acquire dedicated fiber optic links to servers "on the edge" as a way of reducing the number of "hops" that the public Internet imposes. Whether such connections are acquired for exclusive DoD uses is a matter of economics as well as of security. In any case, such links will all have to be subjected to the discipline dictated by standard GIG protocols.

Meanwhile, the DoD is struggling to assure its minimum acceptable network security. When asked, in Congressional testimony, how he would grade the U.S. military's ability to protect its networks, Gen. Keith Alexander, commander of U.S. Cyber Command, said he would give it a "C". For an essential combat capability nothing but an "A+" should be acceptable.[44]

When one examines the priority of all of the issues that affect the conduct of IT in DoD there is no question that proceeding with the implementation of the GIG is on the top of any list of actions that warrant the greatest attention.

[44] *http://defensesystems.com/articles/2011/03/17/cyber-command-head-rates-military-cyber-defense.aspx?admgarea=DS*

Network Control Center Monitoring

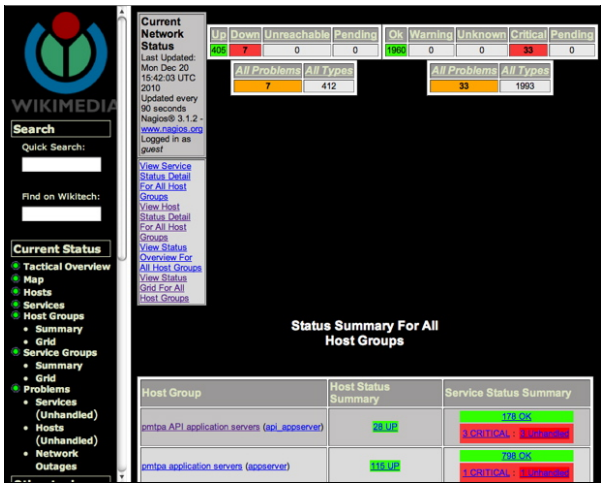
With the growth in the size of computer networks the monitoring of operations rises in importance. There are a large number of networks that manage hundreds or even thousands of servers.

For example a 2010 server census shows the following number of servers: Intel: 100,000; OVH: 80,000; SoftLayer: 76,000; Akamai Technologies: 73,000; 1&1 Internet: 70,000; Facebook: 60,000; Rackspace: 63,996 servers.[45] Large cloud services control enormous operations. Google has >800,000 servers and Microsoft has >300,000 servers.

Such large aggregations of equipment require the monitoring of up-time, latency, capacity and response time. To take advantage of technological expertise, control of fail-over in case of a defect and very large economies of scale all server operators rely on Network Control Centers (NOC) for monitoring networks as well as all of the attached assets.

The software used by large NOCs is always customized for proprietary configurations. However, there are numerous vendors that make NOC monitoring technology available. To illustrate the scope and some of the features of NOC monitoring we will use the Nagios Open Source software for WikiMedia. [46]

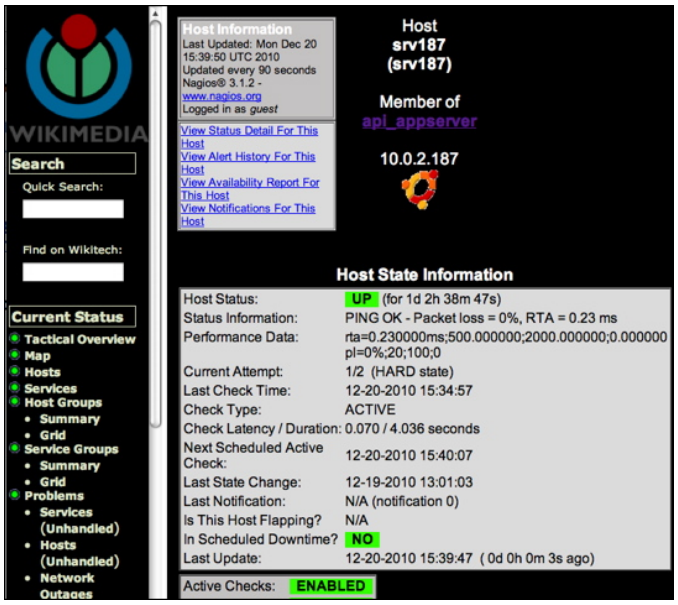
As shown below, WikiMedia had 405 servers up, seven down out of 1960 servers available but not used.



A wide range of diagnostic routines and statistics is available for operators for taking remedial actions. NOC personnel can narrow all reviews for a detailed examination of the status of each server:

[45] *http://www.datacenterknowledge.com/archives/2009/05/14/whos-got-the-most-web-servers/*

[46] *http://www.nagios.org/*



SUMMARY

The tools that are available for NOC operations will represent a major investment for DoD. In the case of NMCI these tools represented a major share of "intellectual capital" that the Navy had to pay for as a part of the migration to NGEN where NOC software will be government owned.

Any future DoD network will have to consider tight integration between NOC operations and the Global Information Grid. The personnel and the software in the NOC represent the first line of defenses for assuring the security of cyber operations.

PART 6: SECURITY

Einstein for Network Security

The EINSTEIN Program is an intrusion detection system that monitors the network gateways of U.S. government agencies for unauthorized Internet traffic. Einstein 1 examined network traffic while Einstein 2 looks at the content of incoming and outgoing transactions.[47]

In 2007 an upgraded version of Einstein 2 was required for all government agencies except for the Department of Defense as well as all intelligence agencies. That excludes 60% of all US IT spending. By 2008 Einstein was deployed in fifteen out of the nearly six hundred agencies. With such slow progress the Department of Homeland Security (DHS) has asked for $459 million for FY12 to include the installation of Einstein 3 and increasing agency participation. Congress my not, however, support enlarged Einstein funding.

Einstein is the result of the E-Government Act of 2002. It is under the management of DHS, which is responsible for the safeguarding all civilian agencies, which have over 2 million users. Einstein involves the centralization of all connections to the Internet in order perform consolidated real-time intrusion prevention for all incoming and outgoing communications. It supports 4 Federal Computer Incident Response Centers (FedCIRC).

Einstein 3 uses an intrusion prevention system to block all malware from ever reaching government sites. The technical problems with Einstein implementations are as follows:

1. While Einstein 2 is only partially implemented, the testing of Einstein 3 has not been implemented.
2. It is unlikely that Einstein 2 or 3 will have the capacity to defend against denial of service attacks (DoS). Criminal bot masters can now rent out as many as 5 million bots. Government cyber attackers can command more than that. Potentially, each bot can generate up to 10 MBs traffic. This could produce an onslaught of over 50,000 Terabytes/second on a single IP address. That is not scalable.
3. One way of detecting intrusion anomalies is through correlation. New intrusions are compared with prior cases. Unless supercomputers are employed for this purpose, Einstein does not have the capacity to make correlations for a network that serves two million users.
4. Einstein depends on the authentication of signatures from trusted as well as untrusted commercial sources. That is not acceptable.

[47] *Einstein, http://en.wikipedia.org/wiki/Einstein_(US-CERT_program)*

SUMMARY

It is unlikely that Einstein can be expected to protect the civilian sector of the government against cyber attacks. Current discussions promoting extensions of Einstein into the US critical infrastructure (electricity, energy, communications, etc.) have little merit.

From Network-centric to End-point Security

DoD has over 10,000 networks in place. These are subject to changing attacks. In addition there are thousands of roaming wireless users as well as millions of desktops, laptops and smart phones. These devices must be protected for assured security.

It is not feasible to protect all of these points of vulnerability during transmission, even with encryption. Along the way, from points of origins to point of destination, there are hundreds of routers and switches that can be compromised. Since networks are connected, huge amounts of effort must be invested to provide universal security for all communications.

With traffic encrypted at the transport or data layer, network-based inspection for compromises is unrealistic, uneconomic and cannot be implemented. Keeping all of the network devices secure is unmanageable under current budgetary and manpower limitations.

Shifting security controls to the endpoint makes it possible to inspect all traffic irrespective of the technologies that are in place. Therefore, in the case of DoD endpoint security becomes the most effective way of assuring secure delivery of all transactions. A diversity of threat countermeasures can be made available at the endpoints as contrasted with generic protection needed for all network levels.

Sophos Labs reports that there are more than 95,000 individual pieces of malicious code every day. A new infected Web page occurs every few seconds. The content-based detection techniques that have been used for the past 30 years as network-centric defenses are now becoming ineffective against the mass of malicious code. In contrast, at the endpoint the visibility of the applications, data, behaviors and system uses can be used to make better decisions and to achive better protection.

SUMMARY

The net effect of shifting from network-centric defenses to endpoint security makes it necessary for DoD to adopt private Platform-as-a-Services (PaaS) clouds as the architecture of information.

Individual firewalls and virus protection at the desktop, laptop or smart phone levels for protection is economically unaffordable. Endpoint security, at the PaaS server level, can manage thousands of virtual computers for security.

Secure Sign-on for Web-based Applications

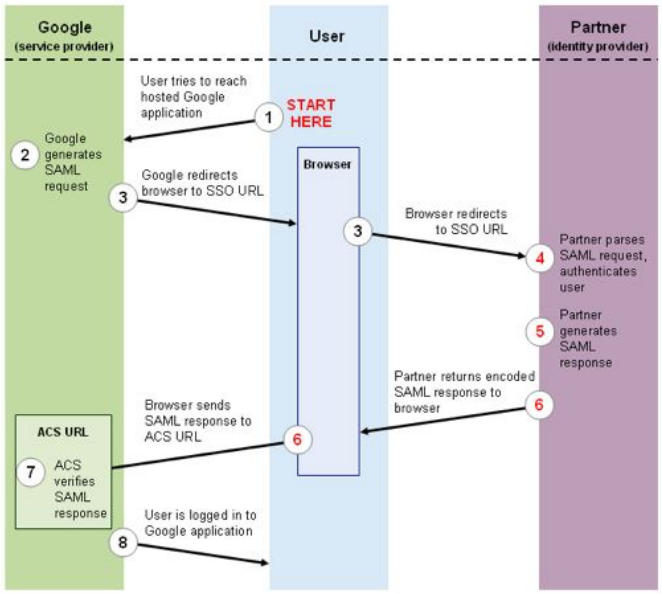
SAML stands for "Security Assertion Markup Language". It is an XML-based standard for communicating identity information between users and web applications service providers. The primary function of SAML is to provide an Internet Single Sign-On (SSO) for organizations looking to securely connect with Internet webs that exist both in the inside as well as on the outside of an organization's firewall.

Internet SSO is a secure connection where identity and trust is shared between a specified security certification Partner and a web an application provider. Such SSOs streamlines the access to applications by eliminating logins for individual applications. Logins are avoided by providing a persistent SSO. For instance, once a SAML SSO has been approved for a user to a Google SaaS service no further logins are required.

The Security Assertion Markup Language (SAML) is an OASIS (Organization for the Advancement of Structured Information) standard for exchanging authentication data. It uses security tokens to pass information about an end-user and an identity provider. IBM, Oracle, RSA, VMware, HP and Google support SAML.

How SAML authorizations are executed depends on how applications are hosted. Services vendors, and particularly various Infrastructure-as-a-Service (IaaS) or Platform-as-a-Service (PaaS), will offer different middleware versions for accepting SAML. This will limit the extent how applications can be relocated to other vendors.

The simplest use of SAML can be found for Google, which is the premier Software-as-a-Service (SaaS) offering:



The following describes the steps taken to obtain a SAML access authorization:

1. The user attempts to reach a hosted Google application.
2. Google generates a SAML authentication request.
3. Google sends a redirect to the user's browser. The redirect URL includes the SAML authentication that will be submitted to the Partner.
4. The Partner decodes the SAML request and extracts the URL for both Google and the user's destination URL. The Partner will then authenticate the user.
5. The Partner generates a SAML response that contains the authenticated user's username. This response is digitally signed with the partner's public or private RSA key.

6. The Partner encodes the SAML response and returns that information to the user so that the user's browser can forward that information to Google.
7. Google verifies the SAML response using the Partner's public key. If that is confirmed it redirects the user to the Google application.
8. The user is now logged to all Google SaaS applications.

SUMMARY

The step-by-step process described above represents only the simplest case of a SAML authorization process. In the case of Google this makes the use of Google applications not only rapid but also convenient as a preferred service.

From the standpoint of DoD the required SAML SSO authorizations can be specified for every user. Different security clearances can be attributed according to need-to-know uses.

The accreditation Partner should be organized within DoD. It should be also staffed by DoD manpower at Network Control Centers that track application access by each DoD person.

The current access authorization process in DoD is based on Common Access Card (CAC) readers. The user then logs to applications to obtain access privileges. There may hundreds of such separate logon procedures presently in place. Such logons were conceived when applications were developed in separate contracts. Single Sign On (SSO) is used only in application "silos". The multiplicity of SSO processes increases the costs as well as a vulnerability to security compromises.

The DoD should therefore consider adopting a customized version of SAML initially for all of its web applications and subsequently for all of its SaaS applications.

Client or Server Based Security

According to the DoD CIO the Secret Internet Protocol Router Network (SIPRNET) connects approximately two thousand DoD locations, with up to 500,000 users. Every SIPRNET connection is physically protected and cryptographically isolated. Each authorized user must have a SECRET-level clearance or higher.[48]

Following the unauthorized release of classified documents (e.g. Wiki Leaks) the Director of the Defense Intelligence Agency stood up an Information Review Task Force to assess the security of DoD SIPRNET data. The task force found that: units maintained an over-reliance on removable electronic storage media; the processes for reporting security incidents were inadequate and there was a limited capability to detect and monitor anomalous behavior (e.g. exfiltration of data).

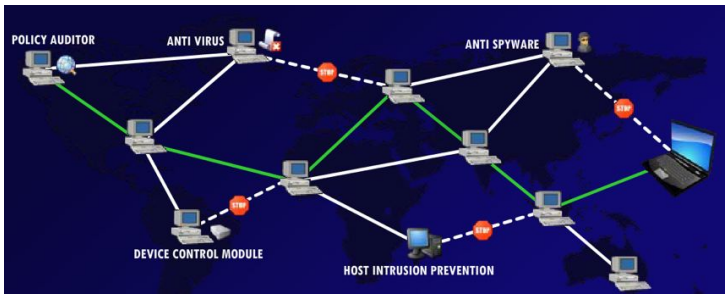
DoD is now proceeding with the installation of the Host Based Security System (HBSS) by June of 2011 provided by COTS vendors. This will provide for central monitoring and control over all computers and their configurations. HBBS includes a Device Control Module (DCM), which can be used to disable the

[48] http://hsgac.senate.gov/public/index.cfm?FuseAction=Hearings.Hearing&Hearing_ID=0c531692-c661-453a-bc97-654be6eb7d00

use of all removable media. 48,000 to 60,000 computers will be exempted from DCM restrictions and will be able to continue relying on removable media.

DoD will also continue to issue a Public Key Infrastructure (PKI)-based identity credential on a hardened smart card, which is more robust than the Common Access Card (CAC), which is used on unclassified networks. The PKI cards require positive identification of anyone who is any accessing data. This would be completed by mid 2013.

The key to HBBS – and what represents its weakness – is the ultimate dependency on a human policy auditor who can set up access restriction policies. The policy auditor will then receive real-time messages to aid in the recognitions of any actions outside of the limits set by policy.



Despite the strengthening of controls, the detection of insider compromises still depends on audits performed by human operators located at many separate SIPRNET locations.

The problem is how to identify selected events as "anomalies" security policies that would indicate questionable behavior. Though the implementation of HBSS offers a strengthening in the identification of information retrieval, ten thousands of individuals will still have the potential to ex-filtrate classified information by "write" actions that not unauthorized. How to identify such instances still remains an unresolved challenge. HBBS may provide more tools, but cannot prevent a Wiki-Leaks incident from happening again.

SUMMARY

The implementation of HBSS imposes on DoD operations large additional costs. The installation and maintenance of HBBS software is labor intensive. It is a task that adds to the work of hundreds of contractors that already maintain SIPRNET configurations. That cannot be done without more funding and without additional headcount.

Training is not trivial. HBSS requires at least a two-day course. To staff HBBS Policy Auditor positions, that watch operations around the clock, will require an incremental staff with a higher grade of skills. Personnel records will have to be expanded to include descriptions of what tasks an individual is permitted to perform and under what conditions, which is subject to frequent changes. The administration of such access privileges is labor-intensive. Administrative policies and processes will have to be put in place to determine who should (and should not) have access.

Clearly, HBSS is a costly short-term "patch" on an already overburdened system. Whether DoD will be able to staff and then deploy a sufficient number of Policy Auditors has not been as yet included in plans for FY12-13. How the auditors will be supplied with the intelligence that is necessary for the discovery of anomalies is yet to be established.

So far the HBBS fix has not examined how to evolve from the currently proposed short-term improvement to a longer-term solution. Instead of relying on each SIPRNET enclave to set up its own rules

for auditing its desktops and laptops, a "cloud" method for the management of secure networks offers many lower cost options.

The key to such an approach is to give up on adding more COTS software to already overburdened desktops. HBBS is a very expensive and manpower-intensive desktop-centric solution. It is unlikely to be implemented by FY13 on a timely basis because of budget and organizational limitations.

Instead of an HBBS add-on to ten thousands of existing desktops, security monitoring and audit should be relocated to a few hundred virtual servers on a secure private cloud. Central policy can be then administered more economically and consistently from a handful of Network Control Centers (NCC). Specialized headcount at the NCCs can operate with less manpower and with a much greater reliance on advanced diagnostic software.

Security should be added to a few servers, rather than to many desktops. Cloud servers can host sophisticated surveillance software more effectively and can be deployed much faster.

Applicability of the DISA DMZ

The Defense Information Systems Agency (DISA) has just announced the creation of a demilitarized zone (DMZ) for unclassified DoD applications. The objective of the DMZ is to control the access and improve security between the public Internet and Unclassified but Sensitive IP Router Network (NIPRNet). Implementation will take about two years. It is supposed to roll out across an estimated 15,000 DoD networks

In computer security, a DoD DMZ will be an isolated sub-network that will process all intra-enterprise transactions for estimated more than four million client computers before it will expose them to any untrusted networks such as the Internet.

The purpose of the DoD DMZ is to add an additional layer of security to DoD local area networks (LAN) and wide area network (WAN). An external attacker will then have access only to the perimeter defenses of the DMZ. This makes it necessary than none of the DoD 15,000 networks will have any computer ports –whether on client computers or on servers – exposed to access from the Internet with the exception of designated DoD web-based applications. It is expected that the DoD DMZ will deflect almost all of the known attack methods. However, it will still leave to human operators to discover and then to deal with any anomalies that are detected by monitoring software.

With progression and the evolution of cyber attack methods it is likely that there will be shift from software based and automatic detection methods to an increased reliance on human intelligence of the guardians of the DMZ.

Under conditions of a concentrated cyber attack the numbers of transactions that must be processed and then passed through the DoD DMZ can possibly approach ten thousands of events per minute. Therefore the capacity of a DMZ must be designed for handling exceptionally large amounts of peak traffic. On account of the increased complexity of zero-day attacks, this will place a burden on capabilities of the diagnostic methods that will be in place.

The servers most vulnerable to external attacks are those that provide services to users who engage in business outside of their local networks, such as e-mail, web and Domain Name System (DNS) routers. Because of the increased potential of these servers to being compromised, clusters would have to be placed

into their own sub-network in order to protect them if an intruder were to succeed in attacking them. Therefore servers within a DMZ will have to be assigned limited connectivity to designated servers within the internal network as an added precaution.

Communication with other servers within a DMZ may also have to be restricted. This will allow servers within the DMZ to provide services to both the internal and external network, while allowing the DMZ operators to cut off traffic when intervening controls indicate that the traffic between servers within the DMZ or with external networks has been compromised.

Simultaneously with the creation of a DMZ DISA is also implementing a DoD central command center (DCC). The DCC will provide continuous oversight of DISA's network as well as 13 subordinate regional operations centers. The center will employ a mix of 220 contractors, civilian employees and military personnel. The DCC is expected to be fully operational when DISA moves to Ft. Meade late in 2011.

SUMMARY

The construction of a DCC and the creation of a NIPR DMZ are milestone events in the creation of more defensible cyber operations for DoD. These are right moves, in the right direction. They are an indication that under the direction of USCYBERCOM the DISA organization is progressing in support of cyber operations.

It remains to be shown how technically effective will be the new DMZ. By creating one or more sub-networks that screen incoming and outgoing traffic DISA will be adding delays (latency) to all of its transactions. Transactions will be dropped and will therefore require a positive confirmation for critical messages, which will increase traffic volume.

Current NIPRNET e-mails already show delays, which will surely increase as additional layers of security monitoring are added. If the new DMZ is an add-on to the already existing security methods, the compounding effects are likely to slow down all traffic further.

The DCC or its subordinate points of control will have to deal with requests for access to Internet portals from NIPRNET computers via the DMZ. From an administrative standpoint the maintenance of a directory of permitted access privileges could represent a large workload. How the new DMZ will deal with SIPRNET communications, which can tolerate lesser latency, is not known. DISA will ultimately have to disclose the technical design of its DMZ and how it will handle peak loads. DISA will have to show how the DMZ will interact with already existing assurance software that is in place on existing networks.

Whether the penultimate Network Control Center (NOC) for DoD, now renamed DCC, can carry out the task of acting as the sentry of last resort for cyber operations remains to be demonstrated. The DCC will have to deal not only with the 13 subordinate regional operations centers that are under the control of DISA, but also with what is a large number of Component NOCs, each functioning under different concepts of operations and deploying different software.

Whether a workforce of only 200 has the capacity of coordinating the designs of multiple Component NOCs while also operating in a high alert mode 24/7 is open to questions. If the DCC is the hub of DoD-wide cyber operations, the presence of contractors is contrary to the objectives of cyber warfare to make it a combat capability of the USA.

- http://en.wikipedia.org/wiki/DMZ_(computing)
- http://gcn.com/articles/2011/01/07/disa-panel-dod-dmz.aspx?s=gcndaily_110111

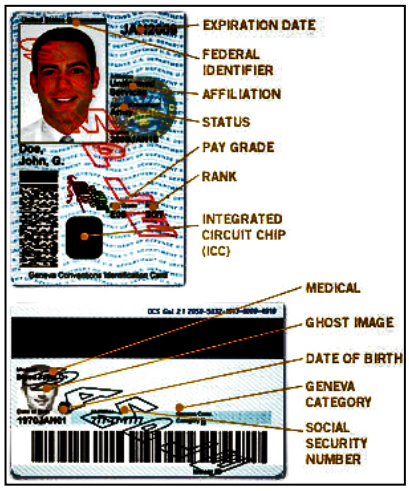
Access Authentication

Two-factor authentication is based on something a user possesses (such as a CAC card) and something a user knows (a password).

CAC stores 64KB of data storage and memory on a single integrated circuit chip. The CAC embeds a persons' Public Key Infrastructure (PKI) certificate (from the National Security Agency). It includes data storage, a magnetic stripe and bar codes. This enables cardholders to sign documents digitally, encrypt emails, and establish secure online network connections.

CAC authorizations originate from the Authentication Data Repository (ADR). ADR is part of the Defense Enrollment Eligibility Reporting System (DEERS), a service of the Defense Manpower Data Center (DMDC). The DMDC Identity Authentication Office (IAO) then provides web services to customers needing an authentication approval, which in turn must be then synchronized with Component human resources applications, which finally deliver the CAC.[49]

The CAC also requires a CAC-reader, which is attached to a computer or a smart phone device. CAC reader installation process is cumbersome. [50]



The information stored on a CAC cannot be used alone for access authorization without entry of a password. Since passwords can be cracked and are hard to revoke or invalidate, automatic password generation devices are preferred in most cases involving SECRET or high-level classification.

The preferred way for obtaining a password is to generate it by means of a security token. That is a physical device that makes up the second factor in a two-factor authorization method.

[49] http://www.cac.mil/Authenticating.html

[50] http://www.militarycac.com/files/SCR331FirmwareUpdateProcedure.pdf

Security tokens are used to confirm one's identity electronically. They have an internal battery that makes it possible to generate random password every sixty seconds. The system that confirms the token generated code must contain additional software for secure synchronization with data contained on the CAC. There is a great diversity in methods, device types as well as vendors who supply for authorization synchronization.[51]

SUMMARY

DoD access authorization methods are vulnerable. The actual revocation of a CAC card is not a real-time event and is not performed by the DMDC but by Components who rely on diverse and inconsistent personnel applications. Ideally, the revocation of a CAC card could be triggered from a Network Operations Center (NOC) instantly. However, the time elapsed from where the revocation is initiated to where it can be acted on is inconsistent with the risks of retaining access privileges for an unauthorized person.

The difficulty in achieving real time synchronization between the IAO, ADR, DEERS and the Component personnel systems is perhaps the primary reason why the dependability of access authorizations will remain a security risk. From a networking standpoint an on-line connection between IAO and a NOC is feasible. The greatest obstacle here is the continued absence of a DoD-wide integrated personnel database.

The management of virtual desktop and smart phone clients from data centers offers an opportunity to simplify the management of software that controls CAC readers. However, the greatest gains would accrue from enabling the real-time connectivity between NOC controls and the DEERS databases.

[51] http://en.wikipedia.org/wiki/Security_token

PART 7: DATA CENTER CLOUDS

Why So Many Data Centers?

The U.S, Government Accountability Office (GAO) in its March 2011 report GAO-11-318SP identified opportunities to reduce the potential duplication in government programs. These programs are usually managed by separate bureaucracies and will most likely operate separate databases, separate servers and possibly separate data centers or server farms.

We have extracted from the GAO listing the number of duplicate government operations as an indicator that can explain why the Federal Government has in place 2,094 Federal Data Centers according to the CIO of the Federal Government, Mr. V. Kundra.

Fragmented Food Safety: 15 Agencies

DoD Organization to Respond to Warfighter Needs: 31 Departments

DoD Organizations for Improvised Explosive Detection: Several

DoD Organizations for Intelligence, Surveillance and Reconnaissance: Numerous

DoD Organizations for Purchase of Tactical Vehicles: Several

DoD Organizations for Prepositioning Programs: Several

DoD Business Systems Modernization: 2,300 separate investments

Fragmented Economic Development Programs: 80 separate programs as illustrated in the following table:

Overlap and Fragmentation Among Selected Agencies Authorized to Fund Economic Development Activities

	Programs by agency				
Activity	Commerce	HUD	SBA	USDA	Total
Entrepreneurial efforts	9	12	19	12	52
Infrastructure	4	12	1	18	35
Plans and strategies	7	13	13	6	39
Commercial buildings	4	12	4	7	27
New markets	6	10	6	6	28
Telecommunications	3	11	2	10	26
Business incubators	5	12	—	3	20
Industrial parks	5	11	—	3	19
Tourism	5	10	—	4	19

Source: GAO

Note: Numbers of programs by agency do not total to 80 since an individual program may fund several activities.

Fragmented Surface Transportation Programs: 100 separate programs

Fragmented Federal Fleet Energy Programs: in 20 separate Agencies

Fragmented Enterprise Architectures: in more than 27 major Agencies

Fragmentation of Federal Data Centers: 2,094 data centers

Fragmentation Data on Interagency Contracting: Excessive duplication

Ineffective Tax Expenditures and Redundancies: 173 major programs

Modernization of Electronic Health Record: Multiple efforts in VA and DoD

Integration with Nationwide Public Health: 25 major systems

Biodefense Responsibilities: 12 Agencies

FEMA Operations: 17 major programs

Arms Control and Nonproliferation: Two separate bureaus

Domestic Food Assistance Programs: 18 programs

Homelessness Programs: Over 20 programs

Transportation of Disadvantaged Persons: 10 Separate Agencies

Number of Programs GAO Identified That Provide Transportation Services to Transportation-disadvantaged Persons, by Federal Department, as of October 2010

Federal department	Number of programs identified
Agriculture	2
Education	11
Health and Human Services	30
Housing and Urban Development	11
Interior	7
Labor	9
Transportation	7
Veterans Affairs	3
Total[a]	**80**

Source: Federal departments and GAO analysis of the Catalog of Federal Domestic Assistance (October 2010).

Employment and Training Programs: 47 Separate programs

Teacher Quality Programs: 82 Separate programs

Financial Literacy Efforts: 56 programs by 20 different Agencies

Farm Program Payments: Huge number of diverse programs

Improper Payments by 20 Agencies on 70 Programs (partial list):

Program	Agency	FY 2010 estimated improper payments
Medicare Fee-for-Service	Health and Human Services	$34.3 billion
Medicaid	Health and Human Services	$22.5 billion
Unemployment Insurance	Labor	$17.5 billion
Earned Income Tax Credit	Treasury	$16.9 billion
Medicare Advantage	Health and Human Services	$13.6 billion
Supplemental Security Income	Social Security Administration	$4.8 billion
Old Age Survivors' and Disability Insurance	Social Security Administration	$3.2 billion
Supplemental Nutrition Assistance	Agriculture	$2.2 billion

SUMMARY:

The 345 page GAO report, summarized only in part in this blog, offers a clue that simple data center consolidation, through virtualization of servers, is not easily accomplished. The various Agencies and bureaus that operate computers in support of their activities are often connected with legislation dictates that control how a program is executed. The idea of datacenter consolidation involves much more that applying simple technical solutions.

Cloud Computing for Business Applications

Popular slides in recent presentations keep announcing the arrival of a new computing environment – the "cloud" - that will make all applications work better, faster and cheaper. However, there are difference between generalizations and reality. Even at an optimistic 25% growth rate, the current cloud spending is less than 5% of total IT spending, though at that growth rate if may soon become one of the dominant forms for organizing information systems. Time has come for the DoD Business Mission to start picking the path to cloud operations in order to migrate from its current high cost and low performance environment.

The DoD FY10 IT cost of the Business Mission (exclusive of the payroll costs for uniformed and civilian personnel) is $5.2 Billion plus about one third of the costs of the communications and computing infrastructure adding another $5.4 Billion to total costs.

The scope of DoD Business Applications exceeds by a multiple of three the average IT budgets of the largest US corporate organizations. Consequently, DoD Business Operations must think about its future IT directions as operating a secure and private cloud that is managed organically by the DoD Business Mission in order to reap the benefits offered by the cloud environment.

Cloud computing comes in many forms. They are: Platform-as-a-Service (PaaS); Infrastructure-as-a-Service (IaaS) and Software-as-a-Service (SaaS). From the standpoint of the Department of Defense we will

discuss only offerings that can offer complete support of over 2,000 applications.

Business Operations cannot be attached to "public" clouds that are proprietary.

For instance, DoD cannot rely on the largest "cloud" service, such as the Amazon Elastic Cloud. Amazon offers computing capacity that is totally managed by the customer, which is operated as a "public cloud". Computing processing is purchased on demand. This makes Amazon an IaaS service. However, once you place your applications in the proprietary Amazon cloud, the ability to transfer the workload into a different environment would be difficult to extricate from.

The Google App Engine offers a (PaaS) service as a "public cloud", which is accessible to all. Google allows developers to build, host and then run web applications on Google's mature infrastructure with Google's own operating system. Google provides only a few Google managed applications.

Enterprise level computing by Salesforce.com now operates at a $1.4 billion revenue rate per year, with 2 million subscribers signed up for SaaS application services running in a proprietary PaaS environment. Salesforce offers only proprietary solutions and cannot be considered by DoD though this will most likely change as result of a recent partnership agreement with VMware.

There are several other cloud providers, such as Terremark Worldwide, that offer IaaS services, but in every instance leave it to customers to manage their own applications. These organizations qualify for DoD applications provided that would meet open source and security criteria.

Open Platform and Open Source

The Microsoft Windows Azure platform offers a PaaS environment for developers to create cloud applications. If offers services running in Microsoft's data centers in a proprietary .Net environment. Azure runs preferentially .Net applications, which are integrated into a Microsoft controlled software environment, but which can be defined almost entirely as a "closed" platform.

DoD Business Mission applications are running largely in a Microsoft .Net environment at present. The question is whether DoD will pursue cloud migration into a multi-vendor "open platform" and "open source" programming environment, or whether DoD will continue adhering to a restrictive Microsoft .Net? This matter will be discussed in a future blog that will deal with the economics of evolving from .Net legacy solutions.

The Defense Information Systems Agency (DISA), with the largest share of the DoD IT budget, has advocated the adoption of the open source SourceForge library in April, 2009 for unclassified programs. DISA's Forge.mil program enables collaborative software development and cross-program sharing of software, system components and services in support of network-centric operations and warfare. Forge.mil is modeled from concepts proven in open-source software development. It represents a collection of screened software components and is used by thousands of developers, taking advantage of a huge library of tested software projects. Forge.mil components are continually evaluated by thousands of contributors, including contributions to its library by firms such as IBM, Oracle and HP but not from Microsoft, which continues to control its own library of codes.

A DoD Memorandum of October 16, 2009 signed by the Acting DoD Chief Information Officer on "Clarifying Guidance Regarding Open Source Software (OSS)" defines OSS as software for which the human-readable source code is available for use, study, reuse, modification, enhancement, and redistribution by the users of that software." Accordingly, OSS meets the definition of "commercial computer software" and shall be given preference in building systems. With the announcement of Forge.mil DoD has started the process of adoption of open source computer code.

Implications

The migration of business applications into the cloud environment calls for a reorientation of systems development technologies in favor of running on "private" clouds while also taking advantage of "open source" techniques for maximum savings.

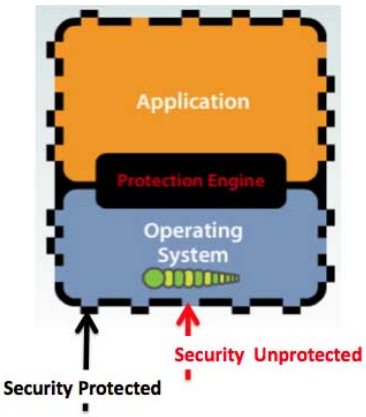
The technologies now available for the construction of "private" clouds will ultimately make it possible to achieve the complete separation of the platforms on which applications run from the applications themselves . In addition, the simplification that can be achieved through the sharing of "open" source code from the Forge.mil library makes it possible to deliver cloud solutions faster, better and cheaper.

Protecting Cloud Computing

SOURCE: http://pstrassmann.blogspot.com/2010/07/achieving-security-in-cloud-environment.html

The security of virtual computers can be achieved by means of application program interfaces that enable select partners (such as McAfee, RSA, Check Point, Symantec, Sophos and others) to install security products that will support virtual environments. The result is an approach to security that provides customers with a cloud-based approach for running secured applications.

The interoperability of hypervisors with the offerings of various security products makes it possible for third-party vendors to manage, through the hypervisor, the protection of virtual machines in a cloud. By this means the security applications can identify malware or denial of service attacks. Security vendors can also use the hypervisors to detect and eliminate intrusions that have unprecedented characteristics, while retaining a record of such attempts for taking corrective actions.



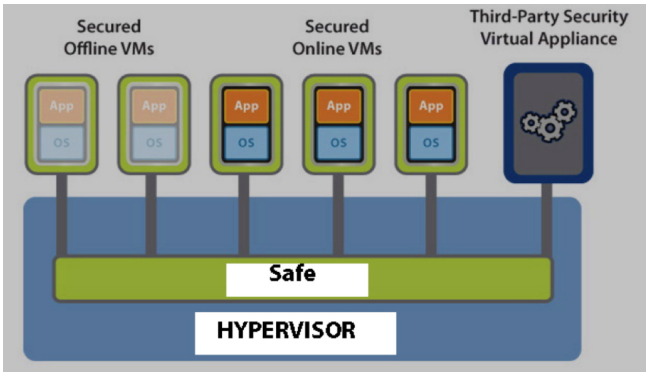
The virtualization technology program for security partners includes the sharing of open, interoperable and cross-platform technologies. These become affordable by providing a continued stream of innovative security solutions, which is spread over a large machine population. By deployment of security measures to the entire cloud of virtual machines customers can obtain lower costs and gain greater visibility at network control centers. By applying consolidated security techniques it is possible to fund sophisticated forensic analysis, which can be scaled over thousands of servers and millions of personal computers.

Virtualization of security cannot be simply appended to servers or desktop computers, as is currently the case when virus protection software and firewalls are installed individually. There will be always gaps in

the protective measures on account of the obsolescence of security software updates as well as the insufficiency of maintenance talent due to funding limitations. In most cases there will not be adequate personnel available for monitoring and then reacting to security incursions.

The intruders will be always seeking out unprotected gaps in protection. With millions of security incursions into the DoD networks per day, the number of potential out of control situations will overwhelm the defenders, unless the systems assurance designs offer well-staffed consolidation of surveillance.

Third party "Security Virtual Appliances" should be embedded within the hypervisor technologies. These appliances provide services such as antivirus, personal firewall, intrusion detection, intrusion prevention, anti-spam, URL filtering, and others. With the growing importance of cloud management of thousands of servers, under central control, it is important to realize that the implementation of security cannot be an afterthought. Security must be fused into the cloud design as it evolves into a comprehensive virtual machine infrastructure.



This arrangement allows the monitoring and enforcement of network traffic within a virtual datacenter to meet corporate security policies and ensure regulatory compliance. It enables the running of applications efficiently within a shared computing resource pool, while still maintaining trust and network segmentation of users and sensitive data as needed.

SUMMARY

Cloud computing, which can include thousands of servers, requires the full integration with the capabilities offered by vendor supplied security appliances. Such safeguards are expensive. They also require the vigilance of exceptionally well-trained personnel as well as the availability of an extensive suite of forensic tools.

On account of the huge costs for assuring the protection of DoD computing, the security of its 15,000 networks can be achieved only through the application of protective safeguards that operate in a cloud environment.

Transition to Platform-as-a-Service

The tight coupling that currently binds the architecture, the infrastructure, communications, databases, applications, security and desktops into over 2,200 unique "silo" projects must be separated into functional components. Right now every "silo" is the result of a contract where almost all of the software code is tightly

cobbled together into one-of-a-kind logical blob. As result the software is hard to maintain, applications are not interoperable and the data is not compatible.

According to Capers Jones, one of the foremost software experts, the applications that account for less than 20% of all software code, sit on more than 80% of the applications support infrastructure. That support infrastructure can be standardized by centralized maintenance. Individual customers will be able to modify applications but will not be allowed to alter the code of the infrastructure.

Only then will it be possible to re-assemble projects to fit into an enterprise architecture that is modular, interoperable, upgradeable, secure and inexpensive. Only then will it possible to place application-specific programs, without huge amounts of attached code, on top of a DoD enterprise standard infrastructure, defined as the DoD private Platform-as-a-Service (PaaS).

With PaaS set as the ultimate architectural objective for DoD computing, we must turn our attention to the most difficult challenge: how to migrate from thousands of incompatible legacy systems into a much simpler environment. That cannot be accomplished by retrofitting legacy systems with fixes, conversion routines, software bridges, emulations and patches. You can only rarely pretty it up legacy system look as if it were an interoperable PaaS. You can't make a silk purse out of a sow's ear. For instance, you cannot take a proprietary e-mail application, clean it up for incompatibilities and then claim it operates on an interoperable cloud.

The current emphasis on cost reduction in IT spending must subordinate any migration plans to generating short-term cash savings for the financing of PaaS operations. For example, one of the military service CIOs announced that he has been directed to cut IT expenses by 25% over the next five years. Consequently there will be no major funds available for making material conversions to a PaaS –based infrastructure. That is too hard and would take too long.

FY12-FY16 DoD IT budgets must nevertheless produce savings for funding cloud adoption investments. The present lack of funds is further aggravated by the rapidly rising costs for cyber security. GAO has just reported that the $3.6 billion in FY12 funds for cyber security may not be fully funded. Cyber security is now consuming 9% of total IT spending and is eating up most of the money that would be otherwise available for cloud investments. Spending on security will continue to grow and will have a higher priority than any moves into cloud computing. With a squeeze on IT budgets from where will the new investment funds come from?

At present, DoD spends 30% of its IT $36.5 billion budget on Investment. It spends 70% on Operations and Maintenance. The O&M amount is understated because it does not include the payroll of military and civilian personnel.

Prying money out of Investment funds to pay for PaaS is too difficult. Projects have multi-year duration. They are hard to truncate, especially the large ERP applications or NEXTGEN. There are also urgent immediate fixes needed to support warfare operations that cannot be deferred. Though some money could be obtained by eliminating redundant programs the pending IT budget shortfalls cannot be made up through a cannibalization of investment funds.

We must therefore turn to O&M funds as the primary "cash cow" for financing PaaS cloud migration. Somehow we need to extract out of $28 billion/per year of O&M the required cash to support migration to PaaS. Assuming level IT budgets for five years (FY12 through FY16) this represents a pool of $140 billion. From this amount we must find a way for squeezing out funds. We must find a way to collapse thousands of program "silos" into a handful of PaaS clouds.

PaaS clouds, when finally installed, will offer superior service levels, be more secure and operate at lower costs than the current collection of legacy systems. The issue is not what is conceivable, but how

sufficient cash can be collected in the next five years from legacy O&M operations. Is there sufficient time for making the necessary reinvestments so that DoD can continue operating without a rise in the IT budgets?

As the first step we need a business case for checking the financial feasibility of a PaaS. There are several Total Cost of Ownership (TCO) models available to make such calculations. For this article we will use the most mature cloud model (http://roitco.vmware.com/vmw/). It was derived from the Alinean Corp. where I was a founder and member of the Board of Directors. Alinean has also delivered TCO models to firms such as HP, Dell, CISCO, Microsoft, Citrix and IBM. [52]

For purposes of this paper an estimate of the five-year physical TCO costs was prepared for DoD's 4 million desktops and 200,000 servers. That TCO averages $2.5 billion/year or 7% of the total IT annual spending in FY11. This estimate excludes the costs of telecommunications and rising expenses for security. It does not include the manpower cost for contractors and military or civilian personnel. It does not reflect the excess costs for maintaining hundreds of marginal data centers. It does not reflect the expenses for acquisition personnel as well as the expenses for managerial overhead.

The largest component of the DoD's IT annual costs are the expenses for desktops, or $9.3 billion. This includes administrative support and the costs of downtime, which together exceed the costs of hardware and software.

The cost of servers of $1.9 billion per year is much less that the cost for the desktops. Though DoD is concentrating on server virtualization – which can bring down server costs by over 60%, the largest dollar gains that can be realized from the adoption of virtual desktops. The savings potential from concentrating on desktops can yield cash savings of up to $3.2 billion/year over a five-year period.

Estimated cash savings will be based exclusively on the five-year direct TCO costs. Additional cost reductions will be obtained in indirect support as a much smaller number of PaaS clouds can shrink expenses for the data center infrastructure.

The TCO model of physical IT costs shows that 83% of the total DoD costs are in desktop capital (CapEx) and operating (OpEx) expenses. If we wish to generate cash savings in support of migration to PaaS cloud service, the greatest opportunity should be found in reducing the TCO for desktops.

The virtualization of desktops, which shifts manpower costs from on-site or server-farms support to highly automated network control centers, offers the greatest savings by supporting over 12 desk virtual workloads per blade server. Long tenure network administrators can then manage standard desktop images on clusters of blade servers to streamline management, access control and provisioning of any desktop. Applying a conservative version of the DoD TCO model indicates that the five-year cost of 4 million desktops could be reduced from $46.7 billion to $30.8 billion using a conservative approach for pacing implementation in gradual increments with only 50% of virtual desktops converted at the end of the first year. 100% conversions would not be realized until at the end of a five-year program life.

After five years the cost of desktops would continue to shrink as devices are displaced by mobile wireless connections as well as by thin clients such as "pad" computers. With addition of desktops from the reserve forces, the National Guard, service academies and contractors additional savings can be realized.

[52] https://roianalyst.alinean.com/ent_02/AutoLogin.do?d=593411470991915416

As the control of desktops migrates to monitoring, from a few network control centers, more savings could be realized as existing server farms are consolidated through virtualization. The large capital expense for more powerful servers may require deferral of virtualization of servers towards the end of the five-year conversion period.

Desktop virtualization, the chosen cash-generation method for the next five years, improves business continuity and disaster recovery by means of automatic failover. An increased dependency on servers requires at least 99.9999% sever uptime for a server cluster, which can be operated with less reliable (and cheaper) individual servers.

The creation of virtual desktops delivers can be structure as a secure and centrally managed automated service instead of the current approach where desktops are administered by hundreds of local operators, who are mostly contractors. Most importantly, desktop virtualization enables access to applications from multiple locations and from multiple devices in contrast to current limitations that constrain access to "silo" defined networks.

Desktop virtualization eliminates planned and unplanned downtime for delivery of extremely high service levels because they can rely on server redundancy. As result the current large penalty of anywhere from 50 to 500 hours annual unavailability can be avoided and counted as savings. In addition, the load-balancing features of desktop virtualization make possible savings in storage capacity.

Desktop virtualization reduces capital and operating system costs because the workload peaks can be dispersed across geographically separate regions while improving capacity utilization. It reduces the need for most of the local IT administrative staff as well as the contractor overhead at hundreds of server farms. It centralizes the management of security, makes real-time surveillance affordable and speeds up the deployment of application upgrades and bug fixes. It makes it possible to manage globally all of the desktop assets behind DoD security firewalls.

The TCO calculations assume that the Microsoft desktop environment will persist. Upgrading from Windows XP to Windows 7 desktops can be included as a transition method, though the possibility of also moving to much cheaper Open Source "office" solutions is desirable. Open Source cloud computing allows DoD to place its operations with multiple competing vendors. Open Source applications programming interfaces (API) make it possible to switch applications across multiple vendors. Such capability is particularly attracting for making "commodity applications", such as e-mail, easily replaceable to the least cost vendor while providing added insurance against vendor specific security failures.

Added savings from Open Source "office" solutions are large. The increased rate of adoption of a variety of wireless "desktops" plus added security restrictions will steer DoD towards the adoption of identical cloud solutions PaaS solutions everywhere.

Perhaps the most important feature for enabling desktop migration is the ability to "encapsulate" legacy applications for migration into a standard PaaS environment. Encapsulation isolates applications from their underlying legacy environment including the operating system. Each legacy application can be packaged into a single executable code that runs completely isolated from all other applications and from every separate infrastructure in a properly structured PaaS cloud. This results in conflict-free execution of all devices regardless of their origin. It makes it possible to maintain applications independently of the underlying operating systems. This also permits the creation of standard desktops that would materially reduce the training costs of personnel that needs to deal with multiple applications. Standard desktop images can be than maintained independently of their supporting databases.

With "encapsulation" application packages can be simply re-deployed by moving individual icons to different Windows platforms this will eliminate costly recoding and testing. Afterwards applications are included as a part of a standard DoD PaaS virtual infrastructure. Irrespective of which of the many versions

of Windows an application was originally run it can be managed as a component of a PaaS virtual system environment.

Desktop virtualization breaks the links used by each contractor has wedged applications within contractor-controlled versions of various operating systems and within unique hardware. Virtualization eliminates the need to manage custom-fitted environments for each end-user devices separately. After desktop virtualization is in place a network control center can take over and to deliver and update every legacy desktop and applications in minutes, regardless of their original isolation. This makes the load balancing, testing, provisioning and support of applications and desktops much less costly to do. Such an approach can be applied across hundred thousands of desktops instead of custom managing each isolated project individually, irrespective whether they are in any of the versions of Windows PCs, Macs, thin clients or smart phones.

Desktop virtualization changes the way information security is implemented. Instead of installing anti-virus and anti-malware solutions on individual personal computers, great improvement in security assurance can be realized by offloading protection software and firewalls to centrally managed servers. Such a move will increase the consistency of security administration while speeding up the process of safeguarding personal computing. Though security safeguards are in place to protect encrypted networks, increased emphasis on end-point security at the personal desktop via centrally managed servers offers more effective protection of DoD cyber operations.

When fully implemented on a large scale, the annual TCO cost per seat has been quoted to be as low a $300, based on seven-year depreciation of de-featured devices. In this way mobile DoD personnel will be able to connect with their personal desktop from any place in the world, while keeping up consistent security access authorization. Desktop virtualization makes it possible to enforce compliance with centrally controlled security policies, encryption requirements, and access identification methods that confine any desktop to a defined security perimeter.

In a virtual desktop environment users can work with their virtual desktops via LAN or WAN landline connections or over wireless links. Mobile users should be able to roam to any geographic location and receive instant access to their personal screen though any network, which includes a secure version of Internet. Users can then use their own applications, while keeping sensitive data secure and protected.

Desktop virtualization also makes it possible to work offline, such as during airline travel or while on a military mission. Consequently the virtual desktops offer a seamless and completely scalable user experience that is superior to what is available at present. DoD should be able to standardize on similar client computing platforms so that equipment can be re-used (re-purposed) instead of getting junked when it loses local utility. When each platform is tracked with globally traceable RFID tags, the multi-million inventory of computing devices will make is possible to manage over $28 billion worth of capital assets.

Centrally managed virtualized desktops make it possible to extend the management of local physical assets to third-party support contractors that could include access to public cloud providers to process workloads that do not require compliance with DoD security requirements. This can be done without sacrificing control over security policies or administrative privileges, using centrally managed oversight. Support contractors would have no control over user authorization or user network access.

Virtual desktops are only a part of a greater puzzle how DoD can migrate to its ultimate objective how to operate on private PaaS clouds. In the forthcoming article in the SIGNAL View Point the question of how to migrate "commodity computing" into Open Source cloud computing will be discussed.

Performance Indicators

Network systems are like icebergs. Less than 10% of their volume is visible to the user of an application. Almost all of the hidden code, measured in hundred thousands of lines of logic, is invisible in the operating system, the data base management software, in security safeguards and in communication routines. The problem with such software is that for each application – and DoD has more than 7,000 major software projects – contractors will develop the hidden coding to suit separate requirements. For instance, the operating systems – even from the same vendor – will have sufficient variability not to be reusable. Contractors will then add special purpose software routines from different vendors as custom "glue" to make the software code function. Contractors will also patch in custom code to make an application to survive stringent testing requirements.

Such results are hugely expensive and hard to maintain. Applications that are developed separately will not share most of the common 90% of the code that remains submerged within the information infrastructure. Network systems will not be interoperable, except through additions of software connections that increase the costs, reduce performance and increase the risks of malfunctions.

To deal with the software iceberg we must revise our approach to software design so that a shared infrastructure makes it possible for DoD to concentrate on <10% of code that drives applications rather than on the >90% of code that constitutes the software infrastructure. To achieve such a change calls for re-examination of the organization of software.

A Technology View of Cyber Operations

Every transaction involved in cyber operations must ultimately communicate in the form of physical bits (0's or 1's). Every question launched must originate and then be returned from an application, as illustrated in Figure 1:

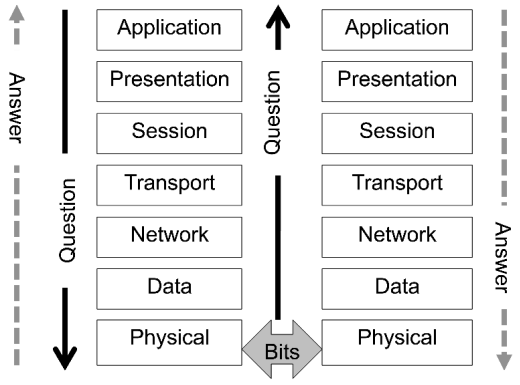


Figure 1: Flow of Information According to OSI

For questions and answers to be converted into streams of physical bits calls for a seven-layered process, each controlled by standards, which define how the respective layers connect. These standards are described by an international standard, the Open Systems Interconnection Model (OSI).

Bandwidth for the passing of physical bits between layers is defined as "return latency" and is measured in microseconds depending on priority and on different ways how to complete a transaction. How the delays in information flows are achieved is then a matter of tradeoffs across each of the OSI layers. Custom-made or improvised OSI connections will increase costs and the latency of a system.

For DoD to migrate to high performance cyber operations requires a design that allows for the sharing of at least three of the OSI layers for Physical, Data, Network and Transport. These layers may account for as much as three quarters of infrastructure code that is written for each stand-alone application.

MEASUREMENTS BASED ON OSI LAYERS

The Open Systems Interconnection (OSI) layers will be used to define measurements for cyber operations. The layers must function as a whole for the successful delivery of results. Except in cases that call for real-time (combat) responses DoD components should field only applications that are using OSI layers that are shared as a DoD enterprise infrastructure service.

MEASUREMENTS FOR SIGNAL TRANSMISSION

The Physical OSI Layer (Layer 1) defines the electrical and physical specifications for components from which networks are constructed. This includes cable specifications, hubs, repeaters, network adapters, bus adapters as well as any devices that convey electronic signals.

Measurements call for capacity mapping that describes every element of the physical layer, defined as to its location description and capacity. Continuous monitoring of capacity, at the circuit level, keeps track of the cyber operations, such as traffic rerouting or instant detection of unauthorized accesses. Configuration mapping displays all connections to and from every circuit. Configuration information is needed to track progress of every transaction, such as the number of "hops" from every source to every destination. Configuration databases that are protected by security measures must list the logical connectivity between network components, including origin and termination points. Logical links are necessary to identify paths for process fallback and for recovery of failed processing.

Measurements include identification of the conversions between digital data and any incompatible signals (such as analogue) transmitted over communications channels. This is critical for tracking translations of legacy data.

Measurements for Data Links

The Data Link Layer (Layer 2) provides the functional and procedural means to the transfer data between networks and to detect and to correct errors that may occur in the Physical Layer.

Measurements require tracking of all LAN connections used for network capacity determination, for network simplification or for identification of alternative paths for passing packets of data under condition of failure. Included is the tracking of all WAN connections used for network management, including re-routing of traffic under conditions of failure. A WAN registry identifies circuits used for diverting communications under peak-load conditions.

MEASUREMENTS FOR NETWORK LAYER

The Network Layer (Layer 3) provides the functional and procedural means of transferring data from a source to a destination while maintaining a specified quality of service.

Tracks all addresses defined as the number of IP's on the entire network, including every address such as desktops, laptops, smart-phones and RFIDS (Radio-frequency Identification). The registry of IP's is managed in real-time and is the main indicator of the size of the network managed by cyber operations. Router IP addresses specify the number and location of routers such as their function, capabilities and processing capabilities. Routers perform network functions such as reassembly packets and the reporting of delivery errors. Routers send data throughout the extended network and make connections possible through the TCP/IP protocol.

MEASUREMENTS FOR TRANSPORT LAYER

The Transport Layer (Layer 4) provides transparent transfer of data between all points participating in the cyber operations.

Measurements require the evaluation of transport uptime, which is the % of hours of scheduled connectivity minus hours of unavailability divided by hours of scheduled connectivity, measured over a one-year period. The unavailability of every link is tracked and recorded in a number of redundant NOCs (Network Operations Center). Individual downtime statistics cannot be averaged but must be displayed in terms of the number of IP addresses than cannot be served, such as any unavailability in excess of one minute.

Measurements of the transport layer define computing nodes as either redundant virtualized resources or as clustered resources.

MEASUREMENTS FOR SESSION LAYER

The Session Layer (Layer 5) controls the connections between computers. It establishes, manages and terminates the connections between the local and remote application.

Measurements keep track of architectures, such as the Service Oriented Architecture, which is defined by the number of reusable components that are available for applications. The total number of reusable and certified software components divided by the total number of components in use is used to measure the pervasiveness of SOA services. Measures include network service statistics such as:

1. Number of legacy applications as related to the total number of applications. This metric is used to measure the extent to which legacy applications have not been integrated into cyber operations.
2. Number of virtual servers with cached services. Cyber operations depend on virtual servers that deliver applications to "the edge" of networks for low latency processing.
3. Number of data dictionary services. Describes the number of unique Meta Data and data dictionary services available from COIs (Communities of Interest).

MEASUREMENTS FOR PRESENTATION LAYER

The Presentation Layer (Layer 6) is responsible for the formatting of information for display. Syntactical differences in inputs to the presentation layer will be reconciled by means of dictionaries that trace the differences in data representation to the point of original data entry. Measures include network service statistics such as:

1. The number of applications that use encrypted coding. Measures the extent to which applications are delivered in the approved encrypted formats.
2. The number of applications that rely on data warehouses for support. Measures the use of data dictionaries to assure consistent syntax.
3. The number of portals for unencrypted access to the public Internet. Provides a method for bypassing cyber operations for access to public network services. The portal blocks transfer of transactions or files to and from Internet to cyber operations.

MEASUREMENTS FOR APPLICATION LAYER

The application layer is the OSI layer, which is closest to the end user. The user interacts directly with applications. This layer interacts with software that implements end-to-end communication. Governance

rules may allow the use of locally managed databases provided they are not connected to cyber operations.

Measurements include access milliseconds counting from the send command to receipt of output in excess of defined delays. Latency is measured in comparison with all active IP addresses and is not averaged but counted as the number of incidents.

SUMMARY

Cyber operations networks must have end-to-end visibility, measurement and control of every keyboard associated with every IP address. This visibility should be present not only at the highly automated Network Control Centers, but also as status displays offered for each local command.

Cyber operations are not comparable to commercial systems such as those for Google, Wal-Mart or Bank of America. None of these systems are subjected to information warfare attacks. We must view DoD cyber operations as having a high security design for the OSI layers in its infrastructure. DoD designs must be based on parameters that far exceed whatever may be acceptable in commercial systems.

It may take 10-20 years for DoD to change its current disjointed software to a shared infrastructure where the code residing in OSI layers will be measured and shared. Budget realities will dictate that DoD components will have to execute such transition largely within existing budgets while the scale of demand for services will rise. This will require the automation of all network metrics in order to cut the operating and maintenance costs that dominates DoD networks nowadays.

As the costs of computing hardware shrink to less than 8% of total I.T. spending, the funding of network centric systems will have to come from cost reductions in software by sharing across applications and from reductions of operating personnel.

The existing development, operating and maintenance costs for the DoD infrastructure are prohibitive. They absorb roughly a half of all I.T. budgets. The acquisition of cyber operations must be driven by elimination of redundant systems as well as by sharing common OSI software layers. The performance measures described in this paper can be viewed as the direction for architecting investments in the years to come.

Path to the Cloud

You should not transport a tribe from the Amazon jungle into an apartment house in Chicago and expect life to continue as before. You should not lift thousands of applications that populate DoD and place them into clouds without interrupting operations during the transition.

DoD applications are run by hundreds of contractors, in over five hundred data centers and using thousands of networks. Programs run on hundred thousand servers that have different versions of operating systems, a wide variety of security measures and unique solutions in making network connections. There is a proliferation of inconsistent databases. Programmers have spun custom-made feeds among applications for updating information.

Cloud computing is based on the pooling of resources to create centrally managed infrastructures that can be deployed as services. That can be accomplished only through standardization of software and the computing environment. If you wish to move from DoD's fractured and diverse legacy environment into a high performance cloud it will take effort, money and disruption in operations to accomplish that.

The path to DoD clouds should be evolutionary. It should be accomplished by making technically safe moves, but only if gains are made at each step. You cannot relocate natives into the "promised land" without doing that gradually.

THE FIRST STEP TOWARDS THE CLOUD IS VIRTUALIZATION.

At least a 15:1 consolidation of servers can be realized by means of virtualization of computing capacities. The servers can then share machine cycles in a pool rather than processing applications individually. Such sharing also applies to combinations of disk storage that house DoD databases.

Virtualization of servers, along with the installation of monitoring and control software, can deliver more than seventy percent reductions in capital costs for equipment if new servers are acquired. If only existing servers are pooled, there will be no reduction in capital costs except that spare capacity can be now used for greater reliability by means of backups and fail-over capacity.

The big savings from cloud virtualization come from the reductions in personnel and in lower costs of electricity. This yields reductions of more than thirty percent in operating costs. The break-even point for taking this approach can be less than two years.

Virtualization enables organizations to take advantage of multicore microprocessors to allocate computing power as the demand for processing cycles rises. Such utilization of capacity is achieved by means of automated resource allocations. Since electric power for computers and for air conditioning now costs more than equipment depreciation, virtualization will cut IT power usage by up to fifty-five percent.

Virtualization can also help in the reduction of disk storage by up to fifty percent by managing disk space for multiple applications simultaneously rather than one application at a time. That enables administrators to defer purchases of disk files.

Virtualization brings visibility, control and scalability to networks, allowing administrators to move beyond per-host network configurations to managing network clusters for fulfillment of Service Level Agreements (SLAs). That makes possible the enforcement of uniform network policies for migration across several computer clusters or over geographically dispersed data centers. Whereas a single administrator can support up to 150 servers one operator can manage well over 2,000 servers in an automated data center. A significant reduction in the number of operators takes place in the control of large server complexes. These are jobs now held by a large number of small local contractors.

Virtualization of servers also allows the improvement in data center performance so that over a million desktops can be converted from "thick" to "thin" clients. Such move will deliver large savings and short-term payoffs.

DoD Components may proceed with the virtualization of their server clusters. This will require the standardization of hypervisor software, which will set the stage for proceeding with the next step.

THE SECOND STEP TOWARDS THE CLOUD IS TO ASSURE BUSINESS CONTINUITY.

Organizations can now reduce the cost of achieving close to 100.0% uptime as well as operate with a much smaller number of data centers. Performance and scale of operations are necessary for critical applications in cyber operations. This is accomplished through instant disaster recovery and though fallback

capabilities that are applied for dependence on a smaller number data centers that are geographically dispersed, but connected with dedicated fiber optic circuits.

Properly configured cloud environments eliminate unplanned as well as planned outages. Backup and recovery are provided to prevent data loss, which includes built-in de-duplication features to reduce data storage requirements.

Workloads are distributed to at least one back-up site but preferably to more. For instance, in the case of Google the processing from a failed site takes place in at least two other data centers that are connected by dedicated high-capacity optical circuits.

Data centers can be now connected with fiber-optic lines so that computing capacity can be shared across a large network, thus delivering additional large savings.

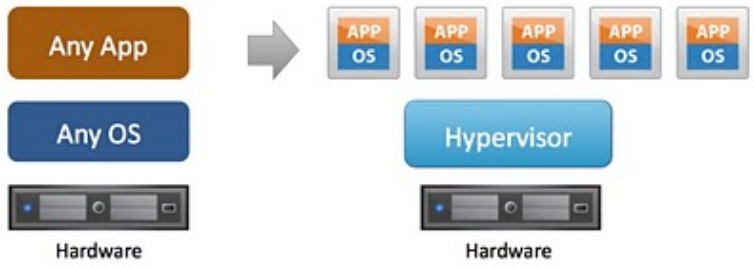
THE THIRD STEP TOWARDS THE CLOUD IS STREAMLINING OF OPERATIONS.

Administrative personnel can be reduced dramatically for the management of geographically distributed development support and computing capacity. Systems development and testing personnel can be concentrated at a few sites. The standardization of cloud operations makes it possible to specialized programming staffs to concentrate on maintenance of software, while keeping personnel housed in close proximity for improved cooperation.

Automated management of the production environments includes the control of the capacity, latency and uptime of a diverse mix of applications. Operating systems, security software and specialized utilities are managed across multiple data centers. By automating managerial tasks cloud computing will simplify the provisioning of a wide range of service offerings. This ensures the delivery of service levels regardless of the physical infrastructure, diversity of computing devices or the configuration of networks. The central control staff can then move work from one virtual infrastructure to another without the need to reconfigure or customize security policies.

Streamlining of operations includes the encapsulation of legacy applications into a centrally managed environment. Diverse applications can be placed into executable packages that run completely isolated from each other. The virtualization layers map the physical hardware resources into virtual machine resources. Each virtual machine will have its own CPU, memory, disks, and I/O devices for each separate legacy application. In this way applications transported from servers that have been discarded will now have the full equivalent of a standard x86 machine with Intel and AMD processors along with its version of the obsolete Windows or Linux host operating systems.

In virtual operations the encapsulated hardware support is provided by means of inheritance from the legacy host operating system.



Migrating individual applications to run on top of a hypervisor makes it possible to place different versions of the Windows or Linux operating systems to run conflict-free on the same server.

Once legacy applications are deployed in the virtual cloud environment, individual application packages can be relocated to different virtual computers, eliminating costly recoding and testing. Subsequently, when time and funds are available, the legacy applications can migrate into successor environments. At that time conversion can begin migration of legacy applications into the new environment.

The interim placement of diverse legacy applications on a shared hypervisor offers the following:

1. Delivers uniform application access to all users while administering compliance with uniform security policies;

2. Eliminates additional server hardware for support of different operating systems.

3. Converts legacy applications for support by different operating systems versions without the need to re-code, retest and recertify.

4. Streams applications from a shared network drive with no intermediate servers or client software to install.

5. Controls storage costs by providing a higher level of utilization.

6. Allows over-allocation of storage capacity for increased storage utilization, enhanced application uptime, and simplifies storage capacity management.

The relocation of legacy applications into a virtual data center is an evolutionary step. It will deliver cost savings that will continue delivering results until such time when legacy applications are phased out. The placement of legacy applications in a virtual data center should be seen as a path for greater interoperability of data through consolidation of databases.

THE FOURTH STEP TOWARDS THE CLOUD IS THE DELIVERY OF SERVICES.

Individual customers acquire the capacity to realize the benefits of secure computing while maintaining security, compliance with OSD policies and reducing costs. Central management of cloud operations depends on large collections of management and automation tools. Such tools allow the pooling of resources, administration of computing resources, and the provisioning of self-service menus for end users who have non-technical skills.

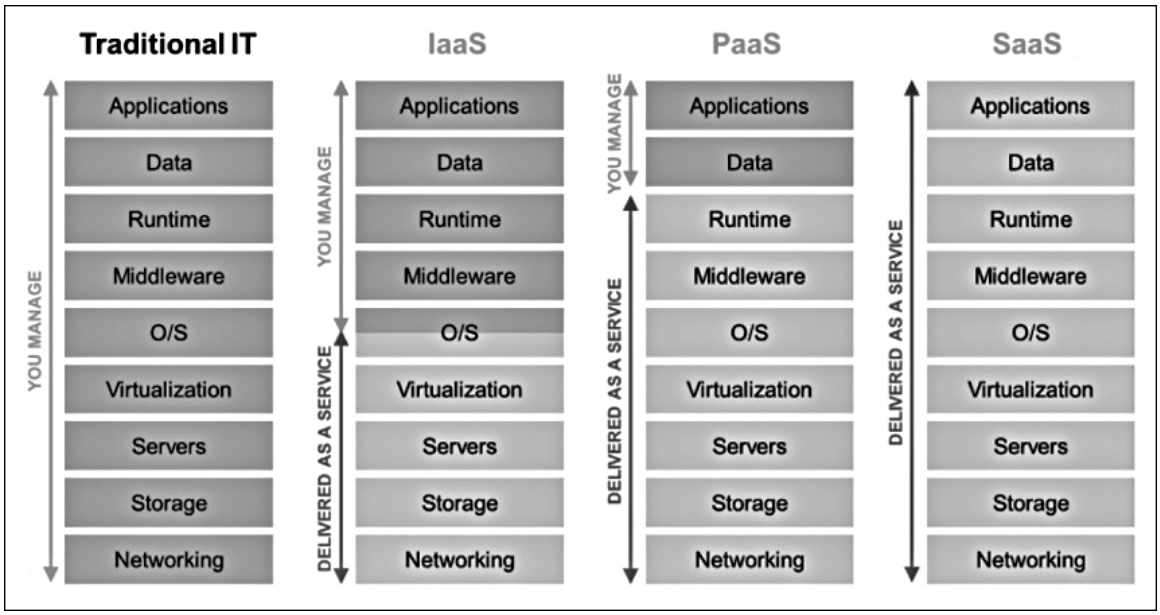
Cloud support tools make it possible to customize IT resources to custom-fit with business needs. They enable the deployment of cloud computing as a utility-like service where the customer does not need to have a high level of technical expertise.

By adopting cloud computing customers can stop viewing IT as a fixed annual budget, which in DoD takes a long time to negotiate and then to authorize. Instead, customers can pay for computing as a metered expense, as needed and whenever used.

From the standpoint of economics, the adoption of cloud computing removes the capital costs as well as user technology risks from DoD and passes it on to a utility that is optimized for the management of obsolescence. Long-term efficiencies will move IT assets to a cloud provider who must operate IT as a business. It is the cloud operator who is responsible for deciding where pooled capital assets can be either leased or acquired or outsourced for the least cost. As result the elaborate acquisition process that presently burdens Program Executive Officers (PEOs) with the responsibility for securing bids is removed. PEOs will focus on applications, which have a much shorter acquisition cycle.

Software can be offered and maintained on the cloud. The customer pays only for their use. That is

Software-as-a-Service (SaaS). The customer can also host own software on the cloud. That becomes Infrastructure-as-a-Service (IaaS) where the customer pays for the delivery data processing. Lastly, the customer can use the cloud as a computing facility that delivers complete results. That is the Platform-as-a-Service (PaaS). It is likely that DoD Components will tend towards the PaaS solution to maintain control over data and applications (Source: The Economics of Cloud for the Public Sector, Microsoft, 2010):



DoD operations in a cloud – whether as a public or a private cloud – must be accessible by means of the Open Virtualization Format (OVF) communications. This standard opens access to a catalogue, which describes what services are available, what is the quality of the offerings and what capacity as well the security is available. It will be the purpose of a DoD-wide catalogue of services for Components to create and maintain their own cloud offerings that can be made available subject to policy and security restrictions.

IT personnel without expertise in programming and even technology-savvy business users will ultimately have the skills necessary to use cloud services because access will be standardized. Most requests will be made by using graphic "buttons" to retrieve answers from the cloud.

Bringing the business user closer to the IT provisioning process offer significant improvements in the management of business systems as compared with the current practice where a contractor must be engaged to thread their way through non-standard versions of networks and through unique access procedures. With the separation of the cloud infrastructure form applications computing needs can be met without the lag time and complexities that have so far been associated with obtaining results from the computer applications.

Diverse physical infrastructures that are listed in a DoD cloud catalogue can be viewed, in the future by departments that need not know about each other except for data that is available for shared use. The result of the evolution towards a cloud can be seen either as a DoD proprietary and secure private cloud or as a public cloud that offers services with different security features. Such arrangement makes it possible to enable a wide range of commercial service firms for delivery of services to DoD where they can be re-assembled as a offering from several sources, including third party applications.

For example, an organization, such as the Navy, can set up a Private Secure Cloud within its own infrastructure that supports its critical applications. For rapid transition they could also choose a hybrid solution by engaging SaaS vendors for generic applications such as e-mail, calendars, collaboration utilities, group communications, documents, spreadsheets, presentations, videos, slide shows and project

management. As long as other organizations, such as the Army and the Air Force remain interoperable by means of cloud standards, the entire Department of Defense can be viewed as an interoperable network.

DoD Components will not always be self-contained. They will have to draw on commercial suppliers' Public Clouds, which can provide to DoD complementary capabilities such as access to Public Clouds offered by FedEx, Wal-Mart, banks, travel agencies, airlines, food suppliers and health providers. Ultimately there will be thousands of public clouds available in the computing landscape that will be needed to support DoD operating requirements.

Security comes will come in the form of several offerings that can certify security assurance for every connection. A variety of commercial security tools are already available for use as virtual appliances or services. In the case of every connection to an external cloud DoD Components will have to ensure that every one can be protected and isolated an integral part of DoD's virtualization infrastructure.

With the availability of a set of standard DoD cloud management tools, the computer industry has entered into a new era how to organize computing. One can compare the recent availability of such tools with the introduction of the pervasive Microsoft Windows Operating System in 1981. Over a period of over twenty-five years the Microsoft OS enabled users to abstract the management of personal computers from individual involvement with the technically difficult management of increasingly complex hardware. The dominance of the Microsoft OS is now vanishing because the dominance of the personal computer in systems architecture is disappearing. The new era of computing is based on the dependency of systems on Internet-related networks. Microsoft's OS is becoming displaced. The cloud computing infrastructures are making it possible to abstract the management of hardware and software from direct user interventions. Systems dependency is moving from personal computers to network provided service.

Over the next decade the current focus of DoD IT executives will pass from the management of physical assets to an emphasis on the choice of cloud services. DoD executives will concentrate on taking advantage of the concept of cloud computing, which views all computing not as dedicated assets, but as accessible utility which delivers computer capacity as well as application services. It is a utility that is based on demand-driven variable costs. The new cloud management tools should be therefore seen as a meta-operating system. That will make it possible for any DoD component to draw on all available IT resources.

DoD IT operations will ultimately split into the physical side and the user side, with separate organizations and budgets for each. The physical side will continue to own and operate the private cloud hardware and communications infrastructure anchored on consolidated DoD data centers. For example, the Defense Information Systems Agency (DISA), or whatever will emerge from USCYBERCOM, will act in that role for DoD.

A new DoD cloud utility will deliver secure raw computing capacity which will be available as a commodity offering with guaranteed levels of security, service levels and transaction pricing. Local management will continue to be responsible for applications, LANs and desktops.

The user side will increasingly use self-service tools to deploy applications from published catalogs based on policies, service levels, and pricing. Users will be able to choose between internal and externally hosted offerings, for a hybrid solution, on the basis of competitive pricing that will lower the cost computing because of the huge economies of scale that cloud services are enjoying. DoD needs to catch up with that.

Cloud Standards

Cloud computing is pursuing a proprietary approach to promote its offerings. Firms do not offer readily portability or cloud solutions from one vendor to another. The current tendency is to offer as little interoperability as possible. Cloud computing, especially IaaS and PaaS, can be compared to a hotel where guests can check it but find it difficult to ever check out.

However, some progress is getting made in view of several organizations that have now established cloud standardization as an objective.

The Distributed Management Task Force (DMTF) so far has made greatest progress. It has created an Open Virtualization Format (OVF). OVF provides a way for moving virtual computers from one hosted platform to another. ANSI recognizes the OVF as a standard. It is also under consideration by ISO. IBM, Microsoft, Citrix and VMware lead the DMTF. It has a large number of global supporters

The IEEE has in place two working groups. Each has so far published only Draft Guides. Work will not be completed for at least two years.

The Open Grid Forum is attempting to create an Open Cloud Computing Interface, which is work in process.

The Organization for the Advancement of Structured Information Standards (OASIS) has two technical committees working on cloud standards, with no results so far.

SUMMARY

With the exception of DMTF, with limited applicability, the progress made so far on cloud standardization has not resulted in interoperability across diverse cloud offerings.

What matters now is the "eco structure" that surrounds various cloud vendors which includes software firms and cloud providers. The rapidly expanding cloud provider industry (software plus service offerings) supports primarily dominant vendors. Service providers continue to be dispersed, but the concentration in cloud software clearly identifies VMware (with 70% market share) and Microsoft (with 23% market share) as the leaders.

From the standpoint of DoD support of VMware as the de facto standards appears to be the safest approach to pursue for the time being.

Open Source Platform

The consolidation of data centers into service clouds will reduce the capital and operating costs in data centers. However, cloud solutions do not cut all of the total life cycle costs of IT. They do not improve interoperability across different clouds. They do not increase the rate of innovation of applications. They do not strengthen security and assure a high level of availability. They do not offer a solution to the proliferation of chaotic IT applications.

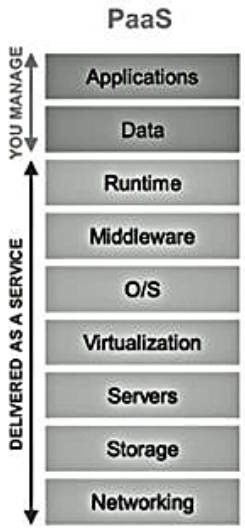
Data center hardware depreciation accounts for not more than 10% of total annual cost of information technologies. A rationalization of information processing by means of virtualization does not attack most of the operating costs. These are the costs of customer-operated devices, applications development, application maintenance, administrative management, costs of user support and communications expenses.

Implementing information-as-an-Infrastructure (Iaas) will take care only of hardware costs, expenses for electricity and the payroll of support personnel. Although IaaS is the first step how to migrate into cloud computing it delivers only a portion of potential savings.

By far the largest IaaS firm in the world is Amazon EC. It has estimated 2011 revenue of about $500 million. Other IaaS firms in the world do not exceed Amazon revenues. With total global IT revenues of $3.4 trillion a policy that concentrates exclusively on the savings from IaaS clouds cannot catch up with the market.

Software-as-a-Service (Saas) offerings are attractive. If they could be scaled up to include a wide range of applications SaaS could solve many of the prevailing IT problems. However the SaaS approach to cloud computing requires an unusually high level of uniformity in meeting customer needs. Applications are dictated completely by the supplier of the SaaS services. A total reliance on only pre-fabricated applications is not feasible. The enormous customer diversity will always limit what can be ever delivered through standard offerings. The largest SaaS firm, SalesForce has revenues of only $300 million. Although it is growing rapidly, it cannot catch up with the size of the market.

The future of IT calls for the adoption of different methods how to cut the costs of the IT infrastructure and to simplify IT operations. The solution is Platform-as-a-Service (PaaS).



A PaaS the customer's unique data and applications are placed on top of a standard (prefabricated) computing "stack". The "stack" takes care of all of the infrastructure services such as communications and security. The customer needs to worry only about applications and data. The rest is taken care of without further intervention from a customer.

PaaS requires the placement of a standard software overlay on top of the existing IT coding "stack" acting as an intermediary between what belongs to the customer's and what belongs to the cloud. The intermediation layer allows for the diversity of IT solutions to function while residing on a standard "stack" that supports a wide range of applications.

The PaaS arrangement makes it also necessary for applications and data to be constructed using a cloud specified development "framework". Such standardization is necessary to enable applications to relocate easily from one PaaS cloud to another PaaS cloud, whether these clouds are private or public. With such standardization applications and data can relocate to take advantage of competitive offerings from various PaaS suppliers.

SUMMARY

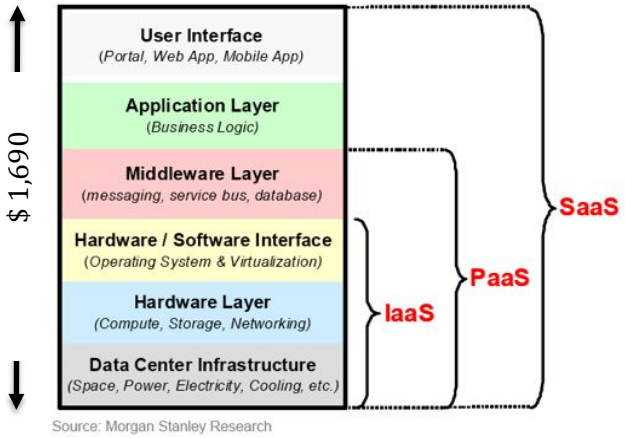
To prevent PaaS firms to offer cloud solutions that capture customers by means of proprietary runtime and middleware solutions it will be necessary to institute to define the interoperability across several PaaS services. That can be done by government regulation or through the adoption of open source interfaces that can be tested for compatibility.

Achieving PaaS interoperability through government regulations is not likely to happen. The PaaS technologies are global whereas the reach of governments is national. The ability of customers to migrate from one PaaS vendor to another PaaS vendor is required to preserve competitive services, to prevent cartel-like hold on services and to enable smaller firms to participate in offering various forms of cloud offerings.

It will take government support of dominant software to prevent a monopolistic lock-in by major telecommunication firms. In various places it is the national communication carriers that are losing their voice business to the Internet. These firms are now seeking new opportunities by shifting to PaaS services that could offer much large profit margins. It should be the policy of the government to insert itself into the business of cloud firms to assure that competition will prevail.

Long Path to Cloud Computing

The best way to understand cloud computing is to view it from the standpoint of a "workload stack" for structure all systems. Various vendors have organized computing differently, but all can be represented by means of a six-layer structure. The ultimate scope of cloud computing, when all information technologies would become potentially available as a service, would be $1,690 billion (in 2010 IT costs):



How far is cloud computing in 2011 from where it would have a major share of total IT spending? Before we can answer that question, it would be useful to examine a history of how various generations of computing organized their workload stacks.

In the period from 1955 to 1975 vendors, lead principally by IBM, wedged the layers of the entire stack into a tightly coupled set. Each stack layer would be dictated by IBMs decision what hardware and software to use. Any attempts to introduce an element of choice into operations, such as introducing other vendor's disk drives or printers, was allowed only after regulatory intervention.

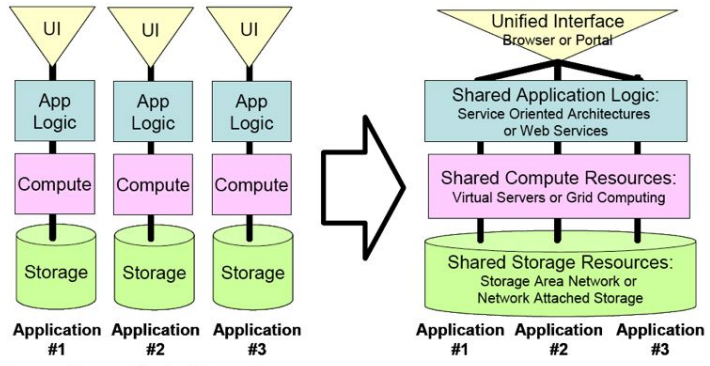
From 1975 there has been a gradual loosening of IBMs control over the entire stack. Microsoft managed to tear the layers apart by creating a separate server stack to support personal computers. The Microsoft stack was different from the server stack that was running in the data center or at a "server farm". This resulted in what is known as the client/server architecture, which is dominant to this day.

The problem with the separate client/server arrangements is a growing inefficiency. Millions of separate server computing stacks was created that each had individual computing, storage and networking layers. The server stacks were almost always incompatible and could not share hardware layers across several stacks such as computing power, disk drives or communications switches. As result the millions of server stacks were poorly utilized. There was no way how separate computing assets could share the capacity of different computing devices.

The client/server arrangement also suffered from a diversity of operating systems. There was a proliferation of "middleware" how various developers installed software that was placed between the operating systems and the applications.

When the hardware layer plus the hardware/software interface plus the middleware layers were installed for diverse applications, the number of combined choices became very large. A customer was forced to break up the enterprise into separate computing enclaves (also called "silos"), which were neither interoperable nor capable of sharing processing capacity. For instance, in the Department of Defense there are well over 10,000 separate computing enclaves that collectively have an excessive amount of unshared capacity. Each enclave was built to accommodate expected peak workloads. The combined capacity far exceeded what was needed. The total capital costs were excessive. The total operations expense for maintaining individual enclaves were a large multiple of what would be otherwise sufficient.

On account of the existing inefficiencies cloud computing should be seen first and foremost as a way of combining individual application stacks into assets where capacity can be shared. The second objective of cloud computing is to aggregate capacity so that storage resources, compute resources, applications and the interface can be managed separately and independently. This means that every shared layer in the enterprise stack have the capacity to be engineered for its own optimization:



Source: Morgan Stanley Research

Storage resources would be shared across all applications. Compute resources would be shared across all applications. Application logic would be shared across all software programs. User interfaces would be common across all applications.

Best practices for the management for storage, compute resources, application logic and user interface would be able to change vendors and operating practices without the need to tightly couple each stack layer with its proximate layers. Systems would be modular, which means that changes in technologies could be made incrementally rather than having to restructure an entire stack, which is the prevailing practice.

There are vendors that accomplish the separation through "public cloud" services or as "private cloud" services. There are hundreds of firms now offering various combinations of cloud services. Different firms may pursue diverse approaches of how to organize for delivery of cloud computing stacks.

ASSESSMENT OF CURRENT STATUS

The ultimate objective of cloud computing is SaaS. This gives a vendor complete control over every layer of the stack. The most aggressive SaaS firm is SalesFoce. It summarizes its offering by stating that it has "no hardware, no software, no headaches" by concentrating exclusively on the delivery of a full suite of Customer Relationship Management (CRM) methods. A wide range of out-of the-box applications is available. These are custom-tailored for the needs of specific vertical markets or business processes. SalesForce 2011 revenue is $1 billion, which represents less than 0.1% of the global total SaaS services spending.

There are hundreds of cloud companies that offer services that imitate the SalesForce model. In each case these are relatively small niche firms that provide access to specialized software where a customer does not wish to manage their own computer organization. Nevertheless, the pure SaaS business typified by SalesForce has taken only an significant share of the IT business.

A more ambitious effort is the recent commitment by Apple to restructure itself to operate as somewhat of a SaaS set up. Apple can do that because it has been always successful by managing its entire computing stack. It makes its proprietary hardware, offers a proprietary operating system as well as a proprietary browser. It originates its own applications while those provided from other vendors have to conform to its tight protocols. With the projected 2011 revenue of $100 billion and a very high growth rate Apple is the most successful IT firm in the world. However, its focus is limited by concentrating primarily on consumer computing.

The Apple cloud strategy is evolutionary. It shows a high executive level commitment to cloud computing. It will migrate from its proprietary strength in customer-owned hardware and software to relocate shared storage resources to the cloud. Even then, the Apple share of SaaS computing will represent only a small share of its revenues that will be derived primarily from the sale of hardware devices and not from cloud fees.

Microsoft has already declared its plans to move into cloud computing by offering a mix of IaaS and PaaS services. They have so far shown only a limited executive commitment to its Azure cloud offering. It is not as yet apparent whether Microsoft can aspire to become a significant cloud services provider. Microsoft depends on revenues from a limited set of application suites and from selling operating systems that are increasingly getting squeezed out by open source offerings. The prospect of Microsoft cloud computing taking a significant share of total IT spending is uncertain.

What will be the Its Gmail offering has over 200 million customers on the Google cloud. Its Android operating system controls the major share over smart-phone devices in the world. However, Gmail nor Android are offered by Google at no cost and therefore cannot be counted as a profitable venture. Its

ChromeBook is potentially a pure SaaS device. However, a comprehensive strategy is yet to emerge from Google that would reveal a business case of how to migrate into a cloud environment.

By far the largest IaaS provider is Amazon.com. The best revenue number available for its EC2 cloud service is $500 million for 2010. These services can be best seen as a form of outsourcing of servers from companies that do not wish to own computing assets to Amazon on a pay-for-use basis. Though there many other IaaS small firms, their aggregate revenues do not add up to a significant share of total IT spending.

Base on reviewing the major cloud firms, such as Apple, Microsoft, Google and Amazon, the prospects of large gains from cloud computing are not significant as yet. Despite enormous media attention, public cloud computing so far takes only a minute share of total information spending.

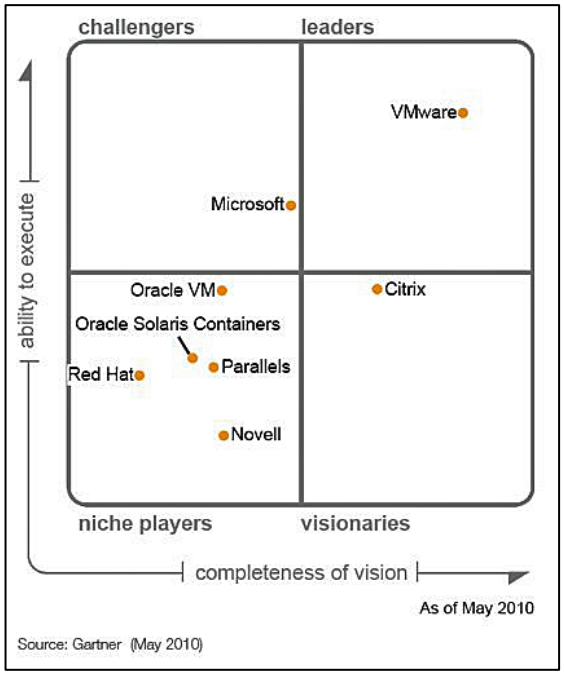
However, the story may be different in the case of private clouds, where individual firms are proceeding at a rapid pace to virtualize their data centers. Almost every CIO in the world is proceeding with a cloud project. What progress firms are making in moving into private clouds or what share of their workload is shifted to public clouds has not been reported.

We conclude that the adoption of cloud computing still remains only a distant goal. It may take decades before cloud computing will manage a large share of the world's IT spending.

Benchmarking Cloud Services

Virtualization makes it possible to run multiple operating systems on a single server. It reduces capital costs by increasing the efficiency by requiring less hardware while decreasing the number of administrating personnel. It ensures that applications will perform with the highest availability and performance. It enables business continuity rough improved disaster recovery. It delivers high availability throughout the datacenter. It improves desktop management and desktop control with faster application deployment and fewer support calls due to application conflicts.

The leading firm in virtualization is VMware. The leading consulting firm, Gartner, offers the following comparisons of the relative strength of companies that sell virtualization software licenses:



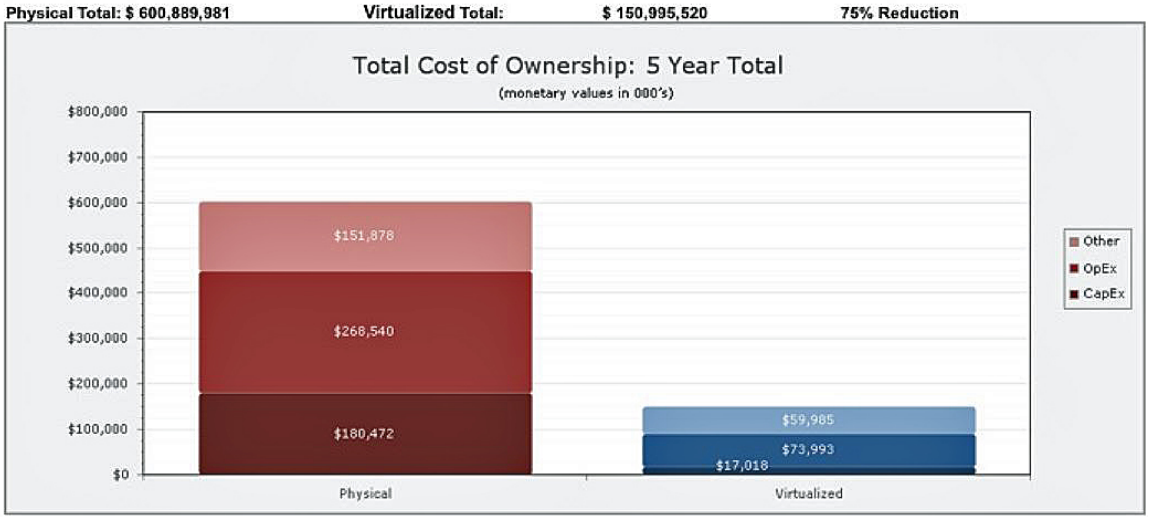
Terremark is a leading VMware customer. They publish prices for using its VCloud Express. Terremark is $350 million provider of IT infrastructure services with twelve datacenters in the United States, Europe and Latin America. It can be used as a benchmark for making comparisons with other cloud data center rates (such as from DISA). The unit list prices are as follows:

Resource	Price
System Storage	$0.25/month per GB
Additional Storage	$0.25/month per GB
Public IP Addresses	$0.01/hour per IP
Internet Services	$0.01/hour per service
Internet Bandwidth	$0.17 per transferred GB
Servers (4 Virtual Processors; 0.5GB	$0.045/hour
Servers (4 Virtual Processors; 2 GB)	$0.161/hour
Servers (4 Virtual Processors; 8 GB)	$0.602/hour

What are then the potential savings that can be gained from the virtualization of 10,000 servers? The Total Cost of Ownership (TCO) program will be used to make the computations. [53]

[53] http://roitco.vmware.com/vmw/. This calculator is the result of work done by the Alinean Corporation. Strassmann was founder and member of the original Board of Directors.

Virtualization enables a substantial reduction in the number of servers, with consequential reductions in operating manpower, energy and infrastructure costs. A 75% cost reduction, over a five-year period is achievable, with break-even time of less than one year.



The largest cost reductions are realized from the elimination of Capital Expense (CapEx) costs, as servers with more core and more processors replace configurations that cannot operate at high levels of utilization since they cannot share computing power. Operating Expense (OpEx) reductions are found almost entirely in the reduction of personnel. "Other" costs reductions come from substantial reductions in the costs of electricity in support of computers and air conditioning.

			Total 5 Yrs	
		Physical	Virtualized	% Reduction
CapEx		$ 180,471,670	$ 17,017,575	91%
OpEx		$ 268,540,189	$ 73,993,365	72%
Other		$ 151,878,121	$ 59,984,580	61%
	Total	**$ 600,889,981**	**$ 150,995,520**	**75%**

A further breakdown of cost reductions is shown in the following table. There is also a reduction in the number of desktops, since the personnel headcount is also cut.

			Total 5 Yrs	
		Server	Desktop	Total
CapEx	Client HW + MS VECD	$ 0	$ 0	$ 0
	Server HW	$ 156,912,817	$ 0	$ 156,912,817
	Storage HW	$ -3,579,275	$ 0	$ -3,579,275
	Networking & Security HW	$ 8,756,000	$ 0	$ 8,756,000
OpEx	Infrastructure Admin Productivity	$ 51,938,892	$ 102,356	$ 52,041,248
	Power & Cooling	$ 93,407,699	$ 0	$ 93,407,699
	Rack Space & Office Space	$ 59,875,200	$ 0	$ 59,875,200
Other	Planned Downtime	$ 34,818,600	$ 0	$ 34,818,600
	Unplanned Downtime	$ 337,920	$ 0	$ 337,920
	Business Downtime	$ 49,103,183	$ 0	$ 49,103,183
	Total	**$ 451,571,036**	**$ 102,356**	**$ 451,673,393**

Power and cooling energy savings are highlighted in the following table:

Energy Savings	Over 3 Years	Over 5 Years
Power and Cooling Energy Savings (Kilo-Watt Hours)		
Extension of Unvirtualized Environment	514,500,161	901,832,206
Transition to Virtualized Infrastructure	8,801,672	15,441,083
Reduction	-505,698,490	-886,391,123
Net impact of Desktop Client transition	0	0
Reduction	**-505,698,490**	**-886,391,123**

SUMMARY

Server virtualization savings are attractive, though they represent only a stage in the process of migrating operations into a cloud. Without a plan that would also coordinate the re-alignment of client devices, increasing security and up-time reliability the virtualization savings would include an element of risk that is not present in ongoing operations. However, the most important issue concerns the decision how to configure the new data processing environment. Virtualization would reduce the number of servers from 10,000 to only 521.

Where and how the new computers would be located requires a rethinking about the management of much smaller data centers operated by a much smaller complement of operating personnel.

Server virtualization should be therefore seen not as merely as a technical means to achieve the consolidation of computing but primarily as a managerial challenge how to start the migration into the cloud environment.

PART 8: LEGACY APPLICATIONS

Modular Development is Not the Answer

The "25 Point Implementation Plan to Reform Federal Information Technology Management" issued by the U.S. Chief Information Officer on December 9, 2010 states that "… OMB found that many IT projects are scheduled to produce the first deliverables years after work begins, in some cases up to six years later. In six years, technology will change, project sponsors will change, and, most importantly, program needs will change. Programs designed to deliver initial functionality after several years of planning are inevitably doomed."

The House Armed Service Committee has further elaborated on the current situation in DoD:[54]

1. Only 16% of IT programs are completed on time and on budget.
2. 31% are canceled before completion.
3. The remaining 53% are late and over budget, with the typical cost growth exceeding the original budget more than 89%.
4. Of the IT programs that are completed, the final product contains only 61% of the originally specified features.

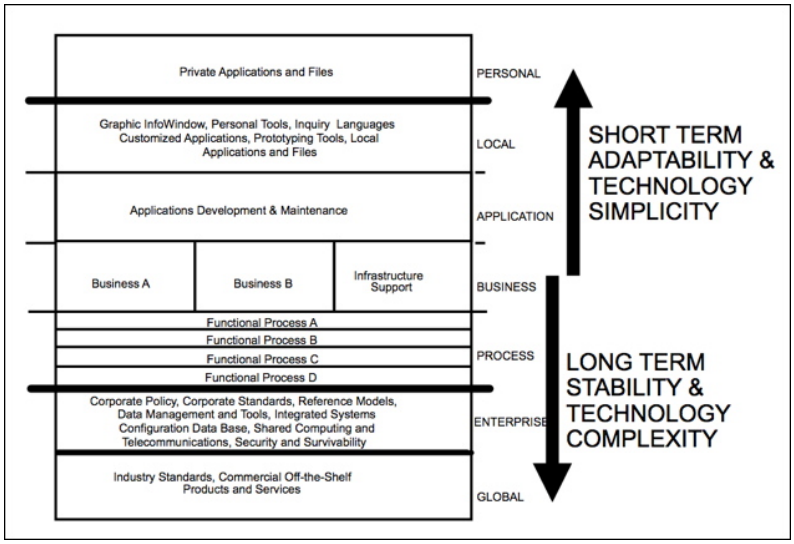
To deal with this situation the Office of the Federal CIO advocates the adoption of a modular approach to systems development, defined as "… delivery in shorter timeframes … based on releases of high level requirements and then refined through an iterative process." This would require deploying systems in release cycles no longer than six to twelve months, with initial deployment to end users no later than 18 months after the program is authorized.

The problem with such guidance is the disregard of typical time-lines for information technology projects. Constructing the Metadata for DoD databases is continuous. It is never finished and requires continuous investments that forever.

The time to develop a communications infrastructure for DoD takes decades. It may then take more than several decades to migrate into a new communications environment. DoD has core business applications, such as in Finance or Human resources that remain in place with little change for many years. Tactical applications have an extremely short life, which may be only a few hours while a mission lasts. Attempting to solve DoD's development problems with rapid paced modular development does not recognize that enterprises need to solve some long-term problems so that very short-term solutions can be implemented.

[54] *www.esi.mil/Uploads/HASCPanelReportInterim030410%5B1%5D.pdf*

The following diagram illustrates the differences in the timing of programs:



Global: Global standards take a very long time to evolve. DoD must select only a minimum of technical standards and enforce them across all applications. Emphasis placed on assuring interoperability with the least migration costs. Proprietary solutions must be avoided.

Enterprise: Enterprise standards must be followed. Control over databases, over shared communications, security and survivability should never be included as an integral part of a development program. Enterprise directions should hold steady for decades. A DoD cloud is clearly a shared enterprise program, not a functional investment.

Process: Functional processes, especially in business applications, should not be managed as modular releases, but planned and funded as multi-year programs. Core features of functional processes should be extracted by Services and Agencies as "plug-in" add-ons.

Business: Businesses should be built as applications that have only unique and specific uses. There should not be multiple logistics, personnel or financial systems within a Service or an Agency.

Application: Application Development and Maintenance can be decentralized to meet local needs. Standard modular code should be used to compose programs that take advantage of an in-place infrastructure, thus minimizing the amount of investment that is required to deliver results. It is only here that the six to twelve month modular development schedule would apply.

Local: Local applications should be composed from parts available from the Enterprise, Functional and Business levels. Depending on the capabilities of the DoD networks, local applications should be assembled in a very short time, often in hours.

Personal: Personal applications should be totally separated from the DoD business to protect privacy. They should be subject to only records management controls.

SUMMARY

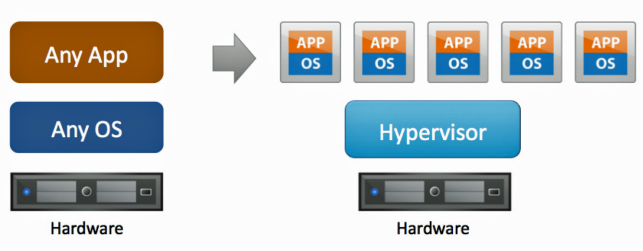
DoD projects that last more than six year, or be terminated only to be restarted again reflect current program management practices. What we have are programs that are attempting to develop their own unique infrastructure, with little dependence on Enterprise or Process services. Such an approach is expensive and time consuming. This situation is compounded by a limited enforcement of global standards.

DoD programs should not be cut into modules that are launched incrementally. Programs should be fitted into an architecture in which complexity is solved by means of programs that have multi-year stability. That would leave to short-term execution modular efforts that depend on the simplicity of using minimal amounts of code.

Legacy Applications in Virtual Environment

Legacy applications can be encapsulated into executable packages that run completely isolated from each other in a virtual environment on data center servers. The virtualization layer maps the physical hardware resources to the virtual machine's resources, so each virtual machine has its own CPU, memory, disks, and I/O devices, and is the full equivalent of a standard x86 machine with Intel and AMD processors as well as with most Windows or Linux host operating systems.

In virtual operations the hardware support is provided by means of inheritance from the legacy host operating system.



Migrating individual applications to run on top of a hypervisor makes it possible to place different versions of the Windows or Linux operating systems to run conflict-free on the same server. Once the legacy applications are deployed in the virtual environment, individual application packages can be moved to different virtual computers, eliminating costly recoding and testing. After that the existing applications can migrate into successor environments in order to start conversion or conversion of previously disjointed or incompatible systems.

The placement of diverse legacy applications on a shared hypervisor offers the following advantages:

• Delivers uniform application access to all users.

• Eliminates the need for additional server hardware in support of different operating systems.

• Converts legacy applications for support by different operating systems versions without the need to recode, retest and recertify.

• Streams applications from a shared network drive with no intermediate servers or client software to install.

• Controls storage costs by providing a higher level of utilization.

• Allows over-allocation of storage capacity for increased storage utilization, enhanced application uptime, and simplifies storage capacity management.

• Lowers capital and operating expenditures by reducing disk purchase while reducing power and cooling cost.

SUMMARY

The encapsulation of legacy systems, with subsequent migration into a virtual environment is the next step after server consolidation. It should be seen as another phase of cloud formation.

The relocation of legacy applications into a virtual data center should be seen as an evolutionary step. It will deliver cost savings even after the legacy system will continue intact until such time when it is finally phased out.

The placement of legacy applications in a virtual data center should be seen as a way for aiming for the ultimate achievement of greater interoperability of data, communication links and application logic.

Integration of Legacy Applications

"Data management technologies are fundamental to the creation of applications, and with the rise of virtualization and cloud computing, the manner in which applications need to access data is evolving. Cloud computing is a distributed deployment model, and for that reason, caching and data accessibility are of far greater strategic importance than before.

Although there have been a sea of changes in the software industry over the last 30 years, there has been no major change in data management since the introduction of the relational database system (RDBMS) in the 1970's. The world has changed drastically since then. We have orders of magnitude more data, arriving at much faster rates, from more sources. Applications that depend on this data have proliferated, reflecting the needs of the business to have faster and more ready access to information. The relationships among those applications have grown as one business process affects another, requiring the applications to share the data in real-time.

Modern relational databases have resolved many of the problems that they either introduced or suffered from in the early stages. They now provide mechanisms for dealing with high availability, clustering and fault tolerance. They can replicate data to peer databases around the world. However, a few problems remain. Firstly, relational databases are a good way to achieve data integration but are poor at achieving process integration. Secondly, using features such as 'triggers', they may be able to detect 'events' (changes in data that some application may be interested in) but they are traditionally poor at distributing events back out to the client tier. And thirdly, they do not store, nor present data to the client in a 'ready-to-use' format for most of the applications. There are multiple layers of translation, transformation, memory mapping and allocation, network I/O and disk I/O that need to occur in order for the simplest of queries to return the simplest of responses. As our use of the RDBMS has grown over time, we have come to depend on them to share data, but they were really only designed to store data.

In an attempt to break down stovepipe systems, there has been a move to Service Oriented Architectures (SOA). SOA helps organizations achieve reuse of individual components of a business process, and makes it easier to adapt their overall processes to align with changing business needs. SOA enables organizations to quickly build new business workflows. However, SOA still fundamentally leaves business processes as stovepipes and it operates on a basic assumption that the components are completely independent. SOA does not address the issue of the real-time interdependencies on the data that the processes share.

In an attempt to get a comprehensive view of data, large organizations are building data warehouses and online/real-time dashboards, so that senior management can see the big picture and drill into critical details. Most dashboard and/or data warehouse solutions pull a copy of the operational data together (usually into a new dimensional format), leaving the original data in place. The operational applications cannot advantage of this combined data view. Data warehousing does not do anything to solve the real-time data interdependencies between applications where business processes intersect. The missing link is 'data awareness'.

Consider as an example the way that mission-planning applications (such as JTT or JMPS – Joint Mission Planning System) depend on data from Battle Damage Assessment (BDA) systems, Enemy Order of Battle (MIDB), situation reports, etc. The process flow from the mission planner's perspective and how potential changes to sources he works with impact the work and the mission.

1. The mission planner(s) start work design missions to destroy enemy targets (bridges, bunkers, SAM batteries, etc.).

2. They pull in data from other systems, BDA, MIDB. Whether they use an SOA based process or not has no real impact in the result. Only on how tightly coupled one system is to another

3. If one second later there is an update to the BDA system or MIDB, the mission planner is left unaware. He continues to plan to destroy a target that may already be destroyed, or plan a mission with inadequate resources due to a change at the target location (new SAM battery, additional enemy forces, etc).

4. The mission planner(s) pull in data from others systems as a final check before releasing the plan. They make adjustments to the plan and release it for execution.

5. If one second later there is an update to the BDA system or MIDB, the mission planner is unaware. The executor of the mission will have to deal with it at run-time. Rerouting to another target, hitting the wrong target or encountering unexpected enemy resistance.

How could this be different? The next generation in data management combines: Distributed Caching; Messaging & Active Event Notification; Active/Continuous Querying; Traditional Querying; Support for users/applications on disadvantaged or periodically disconnected networks; High Availability and some

degree of Fault Tolerance

The interdependency between applications on data and changes to data has serious impacts on mission critical processes. The current way in which data management is done in enterprise applications is over 40 years old and just can not provide many of the critical features needed to build today's high performance, cross organization applications. It is time to consider enhancing your systems data management ability.

Arguably one of the most significant developments in the history of data management came with the invention of the relational database (circa early 1970's). With traditional database access, queries are run against the database, result sets are returned, and work then begins from the returned information.

If new data arrives in the database a microsecond later that would have changed the results set, life is tough. You work with the data you have and maybe synchronize with the database before you finish your analysis, planning, or other work. But once again, the data could change right after you finish. What can you do?

If only the database could call you back, based on your query and show the data changes that would have caused that query to have a different result set. That is exactly what happens with new database software. It acts like a tap on the shoulder to alert people when queries they made have changed results.

The new software enables the creation of applications that work both in garrison and in the field. It has built in support for applications that are not always on the network and/or need to work over distressed networks. During the Trident Warrior military exercise the Navy experienced a 90% reduction in the bandwidth used by a Command and Control (C2) application on a ship built with new software. Additionally, that ship experienced a network outage, during which the application continued to function (although it did not receive new data). When the network was functioning again, the ship received new data and the latest update for each piece of stale data.

Customers experience applications that run 4 to 10 times faster on the average; Speeds online transaction processing and analysis systems in the financial industry by as much as factor of ten; Speeds up complex long running scientific computing jobs on a data grid; Order of Battle data access for complex unit subordination sped up data access times from previous two to twenty minutes (depending on the size of the nations forces) to sub-second; Reduce application footprints – instead of running faster shrink footprint; Instead of running an application faster, that speed can be used to run the application on less CPU's, and less database resources. In turn that often means reduced costs for other software licenses. Overall the result is a significant savings in cost, and power to deploy a system. With today's edge environments stressed for electricity, the new software can help address that issue. It also supports the green initiatives in government.

The Global Command and Control Systems new Common Operating Picture application, Agile Client, uses advanced software to provide high performance service oriented architecture, where users can dynamically subscribe to near real-time track management data from multiple sources, and view that data live on 3-D or 2D surface. Agile Client enables data fusion from multiple sources. It supports DIL environments.

In the defense and intelligence sector, the new software provides four fundamental virtues: Data Awareness, Support Disconnected Operations and Distressed Networks Increase Performance, Reduce Application Footprint and Real Time view of operational data. In essence, it is was able to pull in all the data streams produced by the military, manage that data in-memory so application could provide a window into that data for the military and achieve all of this with guaranteed low-latency, fault tolerance and high throughput. Additionally the software can provide a unified view of data across datacenters with high throughput and low latency.

SUMMARY

Real time integration of transactions from diverse legacy applications will be dictating most of the investments in command and control of military operations. Because of its critical importance in achieving the sensor-to-shooter integration for Information Dominance, this blog has extracted most of the text from the Gemfire Company (a division of VMware) relevant text.

Migration of Legacy Systems

As IT infrastructures become more complex close to 70% of a typical IT operating budget will be spent on maintenance. This will leave only limited money available for new development and for capital purchases.

What is needed is a method how to extract systems from existing inefficiencies. A way must be found how to place all systems into a cloud environment where computing resources can be delivered without money spent on keeping up with a proliferation of options.

Enabling IT as a cloud service will give to a business what it needs while retaining control and improving security. Applications and virtual machines can be provisioned automatically while resources are pooled and then delivered on-demand according to business policies. This makes it possible for a firm to build a self-managing virtual infrastructure while cutting costs.

Encapsulation makes it possible for virtual machines to be portable into Infrastructure-as-a-Service (Iaas) or Platform-as-a-Service (PaaS) cloud environments. A complete copy a legacy system, regardless how organized, can be then relocated from a dedicated computer into cloud pooled computing. A virtual machine is a software container that bundles or "encapsulates" into a complete set of virtual hardware resources. It includes the operating system plus all applications and utilities inside a software package without making any changes in the programs.

The term "legacy application" is a misnomer. Legacy applications are simply software that still works but runs inefficiently because it does not share the ability to participate in the sharing of resources, including security appliances. The purpose of inserting encapsulated legacy into a cloud is to achieve combinations of assets for maximum efficiency. Legacy applications, such as written in COBOL, that are still working can continue participating in a wide range of services that are available on a cloud that is offered by a particular vendor. Companies will not have to rewrite program code or to created new programming interfaces just to get to an application to operate in a private or a public cloud. Encapsulation of applications allows the re-use of legacy systems with only minor changes. The cloud platform will continue supporting identical interfaces so that applications can move instantly and seamlessly into a chosen cloud environtment without interoperability problems.

SUMMARY

The Department of Defense can realize currently mandated rapid cost reductions only by materially cutting back the number of data centers. This makes it necessary to depend on virtualization of computing assets to improve capacity utilization.

However, such consolidations will not deliver the expected savings unless applications are also migrating from poorly utilized servers to virtual cloud operations that can show much higher levels of

efficiency.

The encapsulation of as-is legacy applications and then relocating them into an IaaS cloud environment is the best way of reducing operating costs in the immediate future. Large savings will accrue not only from greater efficiencies in capacity utilization but also from a reduction in the operating manpower. A small number of people will be needed to manage operations that are simpler and smaller.

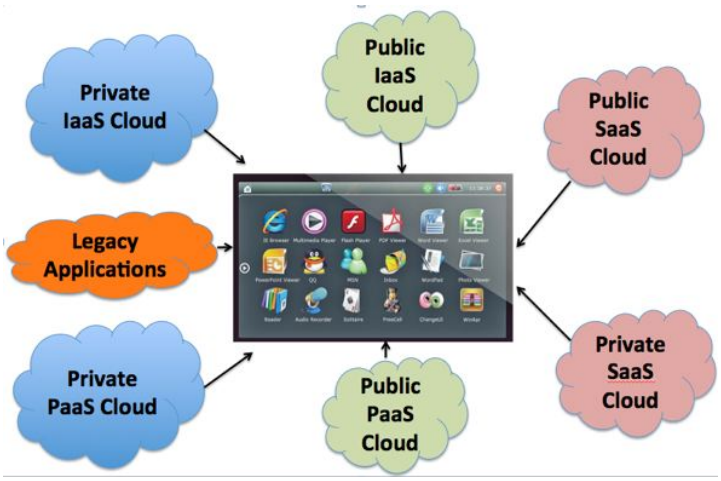
After the legacy applications have been herded into consolidated data centers the work of systems designers to work on further integration can then begin.

The Future of Cloud Computing

Customers do not care much about the technical details of computing. They only wish to receive answers every time and fast. Requested information must be available regardless of the computing device they use. Responses must be secure. There should be no restrictions as to the place from where they communicate. Information must be available for people authorized to make use of what they receive. The sources of information must include information received from people, from sensors or from public web sites. Information must be available to and from ground locations, ships, submarines airplanes and satellites. A user must be able to connect with every commercial enterprise on the Internet.

It is the objective of the cloud architecture of the future to totally separate customer's computer appliances from any of the technical housekeeping details that currently consume huge amounts of time by customers as well as of support staffs. There is no doubt that a number of firms will operate in this mode within ten years.

The greatest challenge for cloud computing will be its ability to gain access to every application whether it is a legacy or a new application. The future of cloud computing is the hybrid environment where a variety of services are accessible from any computer appliance. The following illustrates the scope of hybrid clouds:



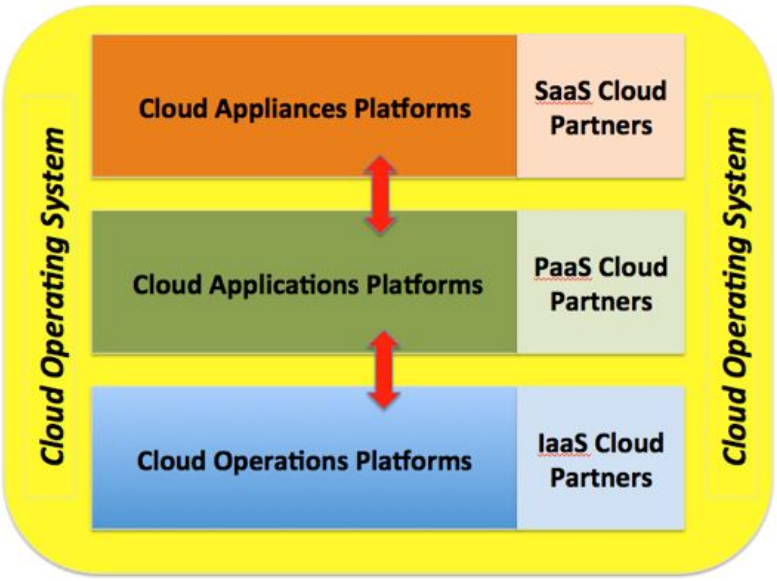
Such arrangement makes it possible to use any application, regardless where or how deployed. Applications would not require separate procedures for gaining access or for obtaining separate security

permissions. Everything that is either device specific or location unique remains under control of a Cloud Operating System (COS) that is not visible to the customer.

COS manages the selection of the source of applications, load balancing, back-up of services, user access privileges and the recognition of a customer's device. All this must be done without a user having to sign into to different servers, logging into to different applications, identifying of different user devices or signing in with different passwords.

What the customer wishes to have is a "personal information assistant" (PIA). Such a device matches a person's identity. It is configured to adapt to changing levels of training. It understands what are the user's verbal or analytic skills. It knows where you are at all times. Any security restrictions are reconfigured to fit a user's current job. At all time every PIA is monitored from several network control centers. From a catalogue of available services the customer finds what they need and by a single click can obtain the desired service. All of the "housekeeping" is taken care of by the COS without any user intervention.

A user must be able to obtain instantly services from a diversity of cloud operations platforms, each hosting a broad range of applications. The COS software must be able to channel a user's request to a diversity of sources. Such flexibility is necessary to assure the portability requests across any platform, retrieval of data from any application and compatibility with every conceivable appliance, as illustrated below.



The COS of the future differs from the current Operating Systems (OS) such as Windows or Linux. The present OS manages how vendor-defined applications are integrated with dedicated servers. The future COS will manage widely different software-defined environments across the entire "stack" of services. This diversity will include diverse hardware, diverse applications and diverse computer devices.

The key to the deployment of COS is the establishment of a user's personal "Cloud Identity." Such also security offers catalogues to show what services are available on private and public clouds. The catalogues enable customers to take advantage of services from the public cloud while maintaining the security and control that is necessary for access to private clouds. With single-sign on security available, the progress to cloud computing can than accelerate.

SUMMARY

COS is not a figment. It represents a series of evolutionary software offerings that are emerging to

dominate the way firms will invest in information technologies.

The integration of three completely separate but interoperable tiers of cloud computing (Operations, Applications and Appliances) becomes the way for planning the architecture of the information of the future. It will be the availability of completely new software that will make such integration feasible.

Google Docs a Step Into the Cloud

Unisys partnering with Google and Section 8(a) contractors (Tempus Nova and Acumen Solutions) will deliver Google cloud-computing services to the General Services Administration (GSA). Their 15,000 employees will switch from desktops and laptops hosted on local servers to network-hosted applications operated from Google data centers. GSA is among the first federal agencies to move into cloud computing. [55] Competitors for the contract were also Microsoft and IBM who lost for unspecified reasons.

The Unisys fixed price contract is $6.7 million or $90/employee seat/year. GSA will still have to provide client computers, plus whatever LAN connectivity to the Internet is needed even though the high responsiveness of Google will make possible the replacement of high cost "fat" clients with low cost "thin" clients.

Anyway one accounts the costs, GSA's IT budget will realize a huge reduction in operating expenses and in the elimination of most of its IT support personnel. GSA will also avoid substantial future capital investments for servers. In return GSA will receive e-mail, word processing, spreadsheets, presentation slides preparation means, collaboration applications as well as a wide range of diverse services that Google makes available at no cost. The applications migrated into the Google cloud represent the majority of GSA's computing needs.

The source selection documents for choosing Google are not available and therefore we cannot say whether it was the price, the migration cost or security requirements that were the basis for the vendor selection. We only know that the major objection to the engagement of Google came from Microsoft who were offering their online Business Productivity Only Suite consisting of the Microsoft Exchange Online for email and calendaring; Microsoft SharePoint Online for portals and document sharing; Microsoft Office Communications Online for presence availability and Office Live Meeting for web and video conferencing. Microsoft's argued that the existing interoperability between Microsoft applications and the diverse GSA applications were not easily interoperable, if at all.

[55] http://www.gsa.gov/portal/content/208417

The differences between Google and Microsoft applications will be hard to ever reconcile. Google offers solutions that are based on open source programs, using published Apps APIs (Application Program Interfaces).

An examination of the interoperability between Google documents, spreadsheets and presentation software and Microsoft Office applications were found to be compatible. There appears to be no valid reason whey Google Apps cannot coexist with any other Microsoft linked applications that remain hosted on GSA servers.

Microsoft's applications are tightly wedged into their Operating system environment. From a security standpoint, Microsoft application software, operating system and browser are also under continual attack. Hundreds of bugs are discovered every year. Until the fixes are installed, there is always a time during which Microsoft programs remains vulnerable. No such vulnerabilities have been as yet attributed to Google.

SUMMARY

The large savings available from the migration to Google Apps for the most frequently used GSA workloads offer a relatively easy and fast path into cloud computing. Other applications, such as database intensive uses, can be scheduled for transfer later or remain hosted on clouds that specialize in applications such as Oracle database services. There is no reason to suppose that GSA cannot operate in the future in a hybrid cloud environment where a part of applications are run by Google, some are run on other clouds and some remain hosted within GSA.

An important consideration in the choice of Google is the opportunity for GSA to disentangle itself from largely total dependence on Microsoft. GSA would now have an opportunity to choose from diverse computing clouds where interoperability can be competed primarily for the least cost as well as the highest levels of security.

PART 9: OPERATIONS

Desktop Virtualization

Desktop virtualization has extraordinary payoffs that could cut total DoD IT spending by up to 12%. Depending on legacy configurations there are numerous approaches that are available to achieve that rapidly. It is not a "bridge too far". The technology is mature. It is a path that has been already paved by thousands of commercial firms.

Proceeding with desktop virtualization calls for altering the IT infrastructure, which establishes how data centers connect via communication networks to millions of various user devices. It calls for an architecture that is extensible to meet the diverse needs of the Army, Navy, Marine Corps and the Air Force. Projects to install desktop virtualization must enable a migration path from the costly "as-is" configurations to what will evolve into a low budget "to-be" environment.

Desktop virtualization can potentially reduce the Department of Defense IT spending by large amounts. The total population of Department of Defense client computers is over three million client computers. Applying desktop virtualization to this population delivers operating savings as well as capital cost reductions.[56]

To illustrate: One million virtualized desktops can deliver five year cost reductions of $5.2 billion with a payback in 1.2 years. With DoD FY10 Operations & Maintenance costs of $25 billion, that results in potential savings of up to 12%. However, such cuts can take place only while also reaping additional savings from the virtualization of servers at data centers where the virtual clients are hosted. That will be featured in the March issue of SIGNAL.

Picking the right combination of hardware and software for desktop virtualization should be funded as a major project coordinated under the direction of a Network & Information Integration executive now in DISA. It will require oversight from USCYBERCOMM because desktop security is a critical key for assuring DoD security.

The primary objective for such projects should be to drive for immediate improvements in security as well as for producing major cost reductions in FY12-FY14. A mature and readily available technology makes that possible.

The goal of desktop virtualization is to create an environment that is far less complex than what is currently in place. DoD can meet the threats of cyber warfare only through a far greater simplification of its fractured networks which desktop virtualization will make possible. In this regard DoD CIOs would be well advised to adopt the motto from a recent letter by Ray Ozzie that "complexity kills".[57]

---

[56] http://pstrassmann.blogspot.com/2010/10/savings-from-desktop-virtualization.html

[57] http://www.digitaltrends.com/computing/microsofts-ray-ozzie-on-the-future-of-computing-and-end-of-pcs

The existing 193(!) DoD security Directives and Policy Memoranda are not executable.[58] The prevailing proliferations of data centers (there are 772 large data centers in DoD), 15,000 networks, multiplicity of operating systems as well the diversity in ways how desktops, laptops, cell phones and other devices connect has resulted in complexity that makes DoD systems unaffordable as well as dysfunctional.

DoD desktop virtualization must be based on a standardization of technologies that manage user clients. Program Executive Officers (PEOs) cannot continue to examine which one of the five hypervisors; 2,814 servers; 1,811 desktop client device versions and 1,210 implementations of operating systems can fit DoD needs.[59] Only a limited set of technology options can be chosen to accomplish the desired objective of simplicity.

DoD cannot afford debugging and maintenance of Microsoft's 280 known operating system versions.[60] Network operations centers will never have sufficient staffs to cope adequately with HP's 144 server options. Choices about hypervisors, data center servers, client devices and operating systems must be therefore planned as a fully compatible set that can be controlled easily and cheaply.

The implementations of desktop virtualization from the leading firms such as Citrix, Microsoft and VMware plus hardware vendors such as IBM, Dell, HP and Oracle will reflect a wide variety of features and capabilities. Hardware will differ with regard to storage requirement, peripherals and the number of servers needed to support virtual desktops. It may take 50 to 250 virtual clients to obtain support from a single server at the data center.

CHOICE OF HYPERVISOR SOFTWARE

The choice of hypervisor software is the single most important decision before DoD can proceed with desktop virtualization. Microsoft Hyper V; Intel's Intel VT-x; AMD's AMD-V; Citrix XenServer and VMware ESX are the leading software firms that produce hypervisors. There should be a preferred choice which hypervisor will offer the least operating cost for DoD in the long run.

There are major differences in how vendors insert hypervisor software between a microprocessor and an operating system. There are type #1 hypervisors (or native, bare metal) that run directly on the host's hardware. In this way the hypervisor controls the hardware separately from the guest operating systems. There are type #2 hypervisors (or hosted) that run inside the operating system. How security software interacts with types #1 or #2 hypervisors should be one of the major distinctions that must be evaluated and benchmarked before deciding which hypervisor type to adopt except that from an engineering standpoint the "bare metal" solution will be always more reliable.

Hypervisors have numerous resellers. Each includes customization of interfaces and defines the ways a hypervisor will be deployed. PEOs will have to evaluate how to pick from the numerous competing features. This will require benchmarking how a hypervisor will function in a specific DoD environment. Unverified vendor claims cannot be trusted when choosing software that is as pervasive as a hypervisor.

---

[58] http://www.wired.com/images_blogs/dangerroom/2010/10/ia_policychart-1.jpg

[59] http://pstrassmann.blogspot.com/2010/10/complexity-of-desktop-virtualization.html

[60] This includes multiple versions and various releases of Windows 2000, Windows 7, Windows 95, Windows CE, Windows NT, Windows XP, Windows Vista, Windows Server 2003, Windows Server 2008, etc.

CHOICE OF CLIENT HARDWARE

Fat clients (or thick clients) are computers, which provide stand-alone functionality that is independent of a central server except for access to the Internet by means of a browser. However, there are fat clients that can be scaled down so that they depend entirely on a central server. Depending on features the number of configurations that can be managed from a data center can be large. Using de-featured fat clients makes sense when migrating legacy applications from fat clients to centrally managed virtual computers. There are fat computers can be made to act as virtualized desktops starting at $400.[61]

A thin client does as little processing as possible, if any at all. It relies on accessing a data center server each time input data needs to be processed. Thin clients are available from firms such as Wyse, Devon, HP, Dell, Pano, Astec and IGEL. The total count of thin devices that can work with virtual computers at central locations is 381. For instance, Wyse offers seven Linux versions, eleven Windows CE versions and fourteen Windows XP options. The cost of these devices ranges from $99 to $646.

From a business standpoint DoD should not seek short-term savings from a major reduction in capital costs when making a switch from fat to thin virtual desktops. The primary purpose for swapping hardware should be in the material improvement of systems security. The long-term payoff from desktop virtualization accrue when capital cost savings continue beyond the three to four year fat client replacement cycle. The depreciation of properly configured thin virtual clients can extend to more than seven years. In addition there are immediate major savings from cutting rapidly rising costs for electricity. If existing legacy computers can be virtualized and then reused the savings can be larger.

The overwhelming cost advantage from any conversion will be in the reduction in ongoing operating costs. Operating savings came from a rapid decline in support personnel. In the case of the Department of Defense that comes from contractor labor, which are mostly small local firms.

The purpose of desktop virtualization is to increase to the availability, reliability and scalability of computing, including the support of portable devices, especially under combat conditions. Shirt-pocket communicating devices now cost less than a round of artillery ammo. Therefore, portable devices should be viewed as disposable items whenever they fail and are hard to fix.

Virtualization of clients increases uptime by offering instant fail-over to other devices without requiring reconfiguration. All it takes is re-booting of a device, which can be performed for multiple sites simultaneously. Multi-computer rebooting at different geographical location is useful whenever a military unit redeploys. That assures practically no downtime as well as zero data loss while keeping military personnel connected.

Client virtualization also increases the reliability of communications by making it possible to switch from hand-held communicators to laptops and vice versa. Such interchangeability can be controlled from network control centers and not from local sites that could be at risk. It is plausible that control of on-shore networks could be managed for more effectively from offshore Navy platforms.

Virtualization increases scalability for rapid access to computing services. A Marine Corps expeditionary force can redeploy from shipboard without delays. Applications and devices could be reassigned using only a minimal staff that would depend on fully automated diagnostic software.

[61] http://www.dell.com/us/p/inspiron-560/fs

CHOICE OF DATA CENTER SERVER SOFTWARE

Vendors can stream applications from data center servers to desktops while simultaneously enforcing consistent security. Virtual data center computers allow the deployment of patches and image updates while retaining each user's settings undisturbed. Virtualization guarantees that all updates are applied uniformly while also synchronizing the image on each client with a master copy. Desktops can be restored to their latest state whenever that becomes necessary, such as in cases of a local failure or when a user is disconnected while moving around.

Virtualization software has the capacity to deliver multiple types of desktops. These can be either "persistent" (everyone receives the identical desktop automatically) or "non-persistent" (each user has an option of setting up their own formats).

To standardize the training of military personnel most virtual desktops will have to remain "persistent". Access to applications then becomes a routine procedure that never changes regardless whether the user relocates from using a smart phone to operating a laptop.

The DoD training commands, and not the IT staffs, will have to devote much attention to the design, layout, symbols, icons and colors of visual images that are placed on every screen. The designs and particularly the graphic "buttons" (the "App" icons) should be always consistent in serving different types of use regardless of whether they appear on desktops, laptops or smart-phones. The Army Corps of Engineers will see different screen templates than members of the Special Forces though each of their screens will be composed from a library of DoD standard "Apps". When desktop technologies and displays are acquired the PEOs must recognize that the worth of a display to a user will always exceed the costs of delivering the information technologies.

Thousands of virtual desktop "Apps" can be managed and audited from consoles at the network control centers from where operators will master a comprehensive view of the entire desktop infrastructure and applications – including separate personal desktop "Apps" for social computing or for NIPRNET or for SIPRNET. In all cases it will be necessary to put in place fenced partitions between what are the DoD communications and what are the messages to and from private desktops. Desktop virtualization will always have to isolate securely all military communications from social communications and vice versa.

Desktop virtualization makes it possible to connect social communications only to their respective designated servers. This makes it possible to instantly switch a user from a private social desktop to the NIPRNET desktop without ever compromising military communications.

Instant deactivation of a desktop for security reasons must be executed without delay whenever the security status of a person changes. At present too much time elapses between the time an employee is dismissed, when termination is recorded in one of the many personnel files and when the Common Access Card (CAC) is deactivated. For security reasons CAC deactivation should take not longer than a second.

The software vendor supplying desktop virtualization must be able to deliver identical experiences across diverse end-user devices, including Microsoft Windows, Apple MacOS, zero clients, thin clients, data kiosks or any newly announced computing platforms such as the iPhones, iPads or Android devices. Anti-virus and malware scanning should be installed as a centrally administered security protection measure. Firewall appliances should be managed centrally to for sharing the costs of security safeguards.

Centralized security measures ensure that in cases a person uses multiple computing devices (which in the future will be increasingly the case) each client will operated under identical levels of security protection as well as compliance with IT policies. Central control may also assign different levels of security protection depending on a person's temporary exposure to risks.

Some desktop virtualization software allows network administrators to control "clone" computers to enable the migration of operating systems such as to Microsoft Windows 7. This is important for dealing with legacy application during the transition to a "thin" client environment. Such adaptations are necessary because the migration to a fully virtualized desktop environment will take at least seven years.

Virtualization allows users to check out their virtual desktop from the data center and take it on the road. Upon return they can then re-synchronize their virtual desktop with their virtual computer. This is useful if communication links are intermittent, such as in the case of computers on ships or on submarines.

Desktop virtualization devices can be refreshed every time a device is restarted. In this way the staff at a network control center can redirect a desktop to a different location or to any alternate device in the world. This feature is needed in the Department of Defense, where personnel is continually moving around and must maintain connectivity without impairing mobility.

CHOICE OF PROTOCOLS

The desktops hosted in the datacenter must connect with a remote screen, a keyboard, and a mouse. The connecting display protocol between the desktop and the data center server then defines the quality of the end-user experience such as the resolution of windows, how fast the scrolling takes place or whether access to high-resolution video is possible. Such protocol must be able to deliver simple displays to task workers as well as complex multi-media images to power users.

IT organizations have always had a problem with what display protocol to use when communicating with a variety of endpoint devices. In most cases such protocols were proprietary, which required additional contractor efforts to achieve the compatibility with installed hardware. There are over 60 software protocols for handling the connections between the desktop platforms and their corresponding servers.[62] They differ in license fees, encryption options as well as in their audio and video quality. They differ whether they support secure connectivity to Linux, Mac OS, Microsoft Windows, Blackberries, Apple IOS or Android. In addition there are over 1,200 versions of almost entirely proprietary host operating systems that connect servers to a desktop platforms. Each data center is may operate its unique versions of connection protocols.[63]

The large combination of desktop connectivity protocols inhibits the enterprise-wide adoption of desktop virtualization. To accommodate the prevailing diversity DoD has broken up its WANs and LANs into hundreds of contractor-administered enclaves. Diverse local implementations promote the acquisition of additional hardware and software because desktop connectivity will work only by spending more money on software links.

Acceptance of virtualization requires connectivity that is independent of proprietary desktops or of server software. Network clients must be able to hook up to a shared DoD network anywhere without requiring software fixes. The end-user must be also able to see their persistent displays anywhere in the world.

A new standard now offers connectivity based on the Internet Protocol (IP), which is independent of either sender or receiver.[64] The PC-over-IP (PCoIP) procedure has been accepted by Dell, HP, IBM, Devon,

[62] http://en.wikipedia.org/wiki/Comparison_of_remote_desktop_software

[63] In a October 1, 2010 memorandum Vivek Kundra, the Federal Chief Information Officer, noted that DoD operated 772 large data centers.

[64] http://en.wikipedia.org/wiki/PCoIP

Fujitsu, Juniper, Oracle, Cisco, Samsung, Wyse and WMware. PCoIP is the right start for achieving vendor independent communications between servers and clients.

SUMMARY

The purpose of desktop virtualization is to free IT management from more than three decades of labor-intensive client computing that was device-centered and not network-centered. DoD should now embark on a direction that will shift the support of user computing to enterprise "clouds" that can support client computing from a much smaller number of DoD data centers over the network to a much larger number of thin and zero client end user devices..[65]

The savings from desktop virtualization are attractive. The technology for installing it is mature. Thousands of commercial firms have demonstrated how to do that successfully. There is no reason why the DoD should not proceed with desktop virtualization without further delay.

Developing a DoD Infrastructure

The best estimate of the costs of the DoD infrastructure is $19.5 billion out of total FY10 IT spending of $33.7 billion, or 57.9%.[66] This does not include the costs of military and civilian personnel that support the infrastructure.

As compared with commercial practice, the amount of money spent by DoD on its infrastructure is clearly excessive. There are no standards for networking, interoperability or security. Each of the >2,700 major applications ends up building its own infrastructure, which adds to the cost of every application and elongates its implementation. Local variations in the infrastructure increases the risks that projects will over-run budgets and will end up incomplete. Too many infrastructures increase the exposure to security vulnerabilities.

Compliance with elaborate policies, guidelines and instructions that dictate how systems are built and operated are unlikely to give assurance that a system will be delivered on time, on budget and with all of the features that the users requested. There is a long list of GAO reports that attest to the consistent failures of

[65] It is now possible to place 25 petabytes and 46,000 CPU cores in a 20'x40' shipping container. That capacity represents a significant share of DoD computing needs.

[66] http://www.whitehouse.gov/omb/e-gov/

programs, each of which followed the thousands of documents required to comply with practices dictated by DoD Directives 5000.1 and 5000.2.[67]

Programs do not fail because they do not comply with concepts dictated by DoD policies. Failures are a consequence, not a cause. Failures are attributable to a flaw in policy that does not recognize that the engineering for acquiring an infrastructure is fundamentally different from the engineering of applications.

Most IT programs start broken because of unrealistic budgets and overly optimistic schedules too ensure program approval. When schedules are not met there is no penalty unless the program manager breeches a major program milestone, which can be fixed by giving up on capabilities.

Letting each program manager pick own technical specifications for applications as well as for the infrastructure leads to massive sub-optimization, higher costs as well as in a loss of interoperability and adaptability to further changes. There is no reward or penalty for a program manager who to consider anything beyond defined deliverable at the next checkpoint review. By the time a delayed program is handed over from development to the acquisition process, most of the requirements will be obsolete. As result operating costs will increase and the quality of what is delivered will be far below commercial practices.

DoD programs are characterized with a requirement creep that grows as program schedules slip. Adding on "features" during program execution usually adds to more problems. Only short-term and incremental implementations are a remedy, but this can be accomplished only if the program managers are not burdened with the task of delivering a new infrastructure.

When systems are finally installed, even after compliance with OSD dictated standards, there will be additional local modifications. Contractors can interpret standards that allow for contractor lock-in. Such interpretations are sufficiently different to impose additional costs on the infrastructure and inhibit interoperability.

The problem is that the current acquisition system does not recognize that the development and the acquisition of an infrastructure is a long-term process, extending over decades. Once an infrastructure is in place the fielding of an application can become a short-term process that can range from days (for special inquiries) to months (for basic new features). The amount of work a contractor has to do will be substantially reduce because the amount of code expended on building a custom infrastructure vastly exceeds the amount of code needed for applications.

With regard to operations, the construction of data center or server farms assets is also a long-term process, possibly extending over decades. Once an operating infrastructure is already in place any local improvisations (such as setting up an "edge" server on a truck) can be accomplished instantly.

The fundamental deficiency in the management of DoDs IT is the absence of a recognition that there are IT elements that take a very long time to put into place. There are also IT capabilities that can be acquired very quickly. For example, equipped only with a credit card, a customer can sign up to the Amazon EC cloud, check out a server and start testing a new application in only 30 minutes! This is possible because Amazon has spent years implementing a standard infrastructure that a customer does not have to worry about.

The current acquisition process is based on a weapons-centric concept of development. It tears apart systems planning, systems development and systems implementation (by multiple contractors) after a lengthy acquisition process is completed and fully tested. Long-term components of an application are implemented

[67] http://www.dtic.mil/whs/directives/corres/pdf/500001p.pdf

simultaneously with short-term features. In terms of the amount of work to be completed, documented and tested the long-term effort will always exceed what can be instantly compiled from readily available software components. Unfortunately, such compilations will not happen because DoD conceives most of its programs as a unique assembly of custom-designed parts.

The chain of actions for any program to succeed is also constrained by over thousand directives, memoranda and committee decisions. The single largest flaw in this sequence if the mandated hand-over from users to acquisition, which resides in a completely different organization and which is run by different rules than program development. Although an elaborate acquisition oversight process in place it is flawed because it focuses on best procurement terms rather than on what is finally delivered as the best operational performance.

Closely related to these difficulties is the lack of funding stability due to the annual appropriation process. This often leads to program delays and slips that quickly lead to major over-runs.

As result DoD then ends up buying thousands of applications where disparate infrastructures try to deliver efficient operations. What DoD actually receives will end up costing too much. It will take too long and will not perform to satisfy the users current needs.

The security of the DoD infrastructure must be seen as a warfare capability and not as overhead cost. Therefore, the acquisition of the DoD infrastructure should adhere to a centrally directed set of technical specifications (performance, interoperability, resiliency, security). Budget limitations and an overwhelming emphasis on assuring information security dictates that this can be accomplished best by an organization such as USCYBERCOM. Therefore, it must be controlled by a combatant command and not as a staff agency function.

How long-term infrastructure needs relate to tactical short-term applications is a matter of policy and governance. Ultimately this translates into control over funding. The estimated $300 billions spent to date on the Global Information Grid (GIG) is an example how it is possible to operate a program that keeps laboring on a mission without connecting with applications pursued by individual program managers.

There is also a major concern about the serious lack of technical and managerial talent in the government, which results in contractors favoring the fracturing of programs into a large number of separate contract vehicles. If DoD pursues the consolidation of the currently highly diverse DoD into commercially funded cloud operations that would make it possible to bring a managed infrastructure within the capabilities of available talent.

Program Managers of infrastructure projects should be on long-term assignments. The current practice of short-term rotations leads to lots of changes and stops and starts. Program cost accounting and progress schedules are only indicators of how well the engineering and the organization of work is working. It takes more to measure of a PEOs success than what is accomplished during the acquisition phase. What matters are the follow-on operations and maintenance costs that will as well as the delivery of capabilities that, with enhancements and upgrading, will make a system viable for many decades after the PEO has moved on. Life-cycle operations and maintenance costs will always exceed the acquisition costs by a large multiplier!

The current acquisition review and oversight process has too many cooks trying to avoid risk in every application rather than letting the program manager in charge of the infrastructure decide how to manage risk. In sumary: You do not need to build a new kitchen every time you wish to cook dinner!

Systems Reliability

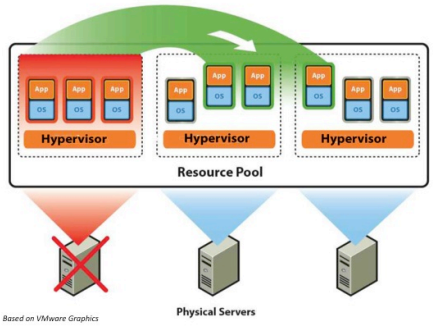
NGEN has been conceived as the backbone network for supporting the Navy's Information Dominance Corps (IDC). IDC networks will transmit sensor-to-shooter communications, globally. Therefore NGEN will have to operate with 100.00% reliability which means zero unavailability to connect keyboards from personal computers to the network hosted databases. All of that will have to be accomplished with a latency time that is at least comparable to Google' end-to-end responses measured in microseconds.

Will NGEN achieve the IDC ultimate uptime and latency objectives? It does not appear to be so. The original contract with EDS called for the following Service Level Agreement (SLA) uptimes such as: Critical services with an average >99.7% service level, which is 26-hours/year of downtime per client device; E-mail would be delivered, on the average, 99% of time with at least 4 minutes of delay for 88 hours/year per client device.

To keep track of compliance with uptime SLAs can be done automatically as a byproduct of information assurance. A counter can keep track of the availability of each of the 400,000 computers on NMCI in order to support critical services for over ten million hours/year for the Navy and the MC.

Setting annual averages for IDC service levels is misleading. What matters is uptime availability of the individual devices that personnel use whenever needed. Averages hide large variability. In any moment there will be many client devices that will be out of commission, regardless whether that is due a failure of a desktop, a server or a communication channel. Unless NGEN tracks the availability of each individual client device, in real time, along with a situational awareness of local conditions any contractual compliance report is not only without value but will also give commanders a false sense of the trustworthiness of NGEN. IDC warriors will always have to give second thoughts to the possibility that NGEN may not be completely reliable.

The only solution that has merit for the IDC is to implement NGEN as a totally dependable network. There is no way of accomplishing this objective except by means of redundancy. Instead of expensive desktops NGEN should depend on an excess population of thin clients plus replaceable smart phones for all communications. These devices are location agnostic and can work from any location. Instead of fixed connections to dedicated servers NGEN should depend on a pool of virtual devices that can relocate a workload automatically and in seconds.



Instead of depending on fixed network connections NGEN should depend on virtual networks that can set up connections to data centers by means of multiple paths.

SUMMARY

NGEN should be constructed as an inter-connected and completely redundant network that is for all practical purposes failure free. The economics of proceeding in this manner is attractive. Savings in labor will offset the costs of added technologies.

The greatest gain comes from the capacity to support IDC missions where the reliability of communications is now an integral part of warfare operations.

Uptime Performance Metrics

The reliability of end-to-end transaction processing for cyber operations is perhaps the single most important metric that dictates the design of networks. Under the conditions of warfare, seconds, not minutes will matter.

Network designers must reach agreement on the measurement of uptime and transaction response latency. When providing for interoperability, the reliability of a network cannot be isolated to the Army, Navy, Marine Corps or the Air Force. Under conditions of information operations, the uptime of a DoD network will the response time from every participating network.

The calculation of network uptime using undefined averages will be misleading. Is uptime averaged over minutes, hours or days? Is uptime measured at the user's keyboard or at the data center? Will uptime be measured in the number of transactions that exceed a defined standard, or is uptime expressed as the number of transactions that are below a stated threshold? Or, will the network operators resort to surveying a random sample of users for an indication of user satisfaction. If so, will the sample be taken at a maximum peak load time or during average business hours?

The following illustrates a statistically valid approach to measuring uptime:

Time	12:05 PM	12:10 PM	12:15 PM	12:20 PM	12:25 PM	12:30 PM	12:35 PM	Total 30 Min Performance
Number of Seats Active	275,002	259,755	244,532	229,298	254,321	263,325	212,431	1,738,664
Seats without Reponse in 30 Seconds	1,003	369	1,589	4,800	320	4,350	850	13,281
Actual Non Performance	0.36%	0.14%	0.65%	2.09%	0.13%	1.65%	0.40%	0.76%
SLA Non-Performance Standard	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%
Performance (Green=OK)	OK	OK	OK		OK		OK	OK

1. The time interval over which the measurement of uptime is taken is defined. This could be in seconds, for missile tracking to hours for inventory evaluation. In our case the downtime increments are in five minutes.

2. The number of transactions (users or seats in this case) that do not receive a response in a defined time interval. That could be more than 200 milliseconds (in the case of a Google search) to less than five minutes when downloading geographic data.

3. The SLA non-performance standard is defined not as a usually misleading 99.5% uptime of hours per years, but at 0.50% non-performance downtime that occurs over five minutes.

4. System performance can be then viewed as the frequency of five minute failures (Green or Red), or as an aggregation measured over a 30 minute period.

When designing for network reliability in cyber operations one must consider whether the network has a single point of failure or whether it is redundant. Cascaded single points of transaction processing show the following downtimes averaged over one year:

Single Point of Failure - System Reliability (Key Performance Parameter)	Single Point of Failure - Downtime/Year - Minutes	Single Point of Failure - Downtime/Year - Hours
0.95	26,280	438.0
0.97	15,768	262.8
0.99	5,256	87.6
0.999	526	8.8

If cyber operations will use a redundant design (two identical system processes running fully redundant transactions) then the overall system reliability will show large improvements in uptime. Automatic fail-over rates can assure near zero failures. This should be pursued for critical applications. Virtualization makes fail-over economically feasible:

Single Point of Failure - System Reliability (Key Performance Parameter)	Redundant Point of Failure - System Reliability (Key Performance Parameter)	Redundant Point of Failure - Downtime/Year - Minutes
0.95	0.0025	1,314.000
0.97	0.0009	473.040
0.99	0.0001	52.560
0.999	0.000001	0.526

The time to restore a failed system is difficult to predict, especially on ships. The only solution is to use end-to-end network redundancies of critical network services.

SUMMARY

Service Level Agreements used in diverse DoD contracts are not consistent in their definitions as well as calculations of uptime/downtime metrics. With increased dependency on multi-Component interoperability it is necessary to standardize the evaluation methods that would make it possible to predict the reliability of complex networks.

Measuring Transaction Latency

The goal of Information Superiority is to deliver a capability where: Every data collector makes information available, in real time, for use by all other /warfare/ nodes. Every sensor is connected via standard interfaces to make local data globally accessible.[68] What metric will confirm that Information Superiority is achieved?

The simple answer is that DoD will have to operate a network where the latency (defined as request-to-response time) will always meet the required response time within tightly defined statistical control limits. In warfare situations the required latency will be dictated by the speed of the response to react to a threat. In an administrative situation the latency will be dictated by the workflow of business processes.

Unfortunately, in DoD there is no difference between the latency of warfare and the processing of administrative transactions. They are intermingled on the same networks. Only networks that are dedicated to the control of weapons can be exempted from DoD network latency standards.

In any discussion of DoD networks the latency metric must become one of the primary criteria for the design of circuits, data centers, security and applications software. The latency of an end-to-end system is the result of the interactions of every one of its components. For this reason the current separation between the management of systems acquisition and on-going operations is not tenable. Acquisition and operation or interlinked by performance and cannot be treated separately.

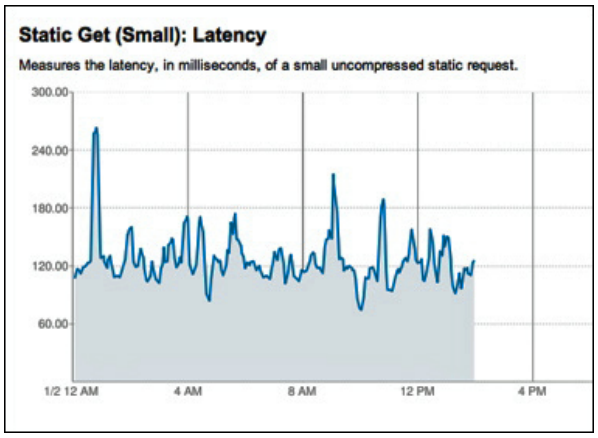
The adoption of standard latency metrics for all of DoD is mandatory. With thousands of applications, operating in thousands of networks, connecting hundreds of data centers and millions of personal computers the system responses will be paced by the latency of the slowest application.

Standard DoD latency metrics must mimic best commercial practices. Since the scope of Google is comparable to that of DoD an illustration is in order.

Google shows latency results for various grades of service in one-minute increments:[69]

[68] http://www.insaonline.org/assets/files/NavyInformationDominanceVisionMay2010.pdf

[69] http://code.google.com/status/appengine/detail/serving-java/2011/01/02#ae-trust-detail-static-get-small-nogzip-java-latency



The latency in this case varies from 70 milliseconds to one instance of a 240 milliseconds peak, with a median of 120 milliseconds (0.120 seconds). That is comparable to the speed of a keystroke. For all practical purposes that can be considered to be near instantaneous.

Latency speed has now become one of the key parameters for network architects and designers. Both Google and Face Books consider the speeding up latency as one of the prime influences in improving user acceptance. The race for speeding up latency is particularly present in the financial services industry where direct optical links have been constructed to reduce latency of financial trades down to 13 milliseconds.[70]

SUMMARY

The current DoD practices either overlooks or minimizes the importance of measuring transaction performance. For instance, one of the largest DoD systems is satisfied by reporting only multi-day "user satisfaction" indicators, based on random sample questionnaires. Such measurements are unsatisfactory because latency is an instant event. It must be measured using a statistically 100% valid set that includes every instance.

The transaction latency will have to become one of the key design and operating parameters of cyber operations if DoD networks will meet information warfare objectives.

Comparing VISA vs. DoD

Analysis of commercial operations offers interesting insights how DoD information technologies could possibly become more efficient. Although VISA is completely different from DoD, there are nevertheless differences that can explain why VISA and DoD budgets are so widely different. As DoD will be looking for cost reductions in I.T. spending there are lessons to be learned from VISA operations that could possibly have merit in planning for DoD improvements.

[70] http://www.nytimes.com/2011/01/02/business/02speed.html

VISA makes available data about its "Network, EDP and Communications" costs. [71] For one year, ending on September 30, 2010, the total expenses for I.T. were $425 million. VISA information technology expenses are therefore only 1.3% of the total cost of DoD information technologies.

Although VISA processes are much different than DoDs, from the standpoint of speed, security, reliability, flexibility and scalability the VISA operations can offer useful lessons how to design and how to manage a large information enterprise.

Here are the major differences between VISA and DoD:

1. VISA operates globally from three data centers, DoD from 772.

2. VISA data centers are redundant and provide for fail-over in real time. Most DoD data centers are not backed up.

3. VISA network uptime is close to 100.0%. DoD uptime availability is not measured.

4. VISA manages the software and configuration management for the entire world from only two locations. DoD does that from at least 2,200 separate projects.

5. VISA provides a global infrastructure and leaves to individual financial institutions to manage their operations and input terminals as long as they conform to centrally dictated standards. DoD is reported to have 15,000 communication infrastructures, each of which is attempting to achieve complete integration down to desktops, laptops and smart phones.

6. There are only two carefully managed software updates for the VISA infrastructure per year. DoD software updates are as needed, whenever and wherever that is affordable.

7. A single VISA executive group controls VisaNet budgets and priorities in quarterly reviews. In DoD the management over budget is widely dispersed so that planning, development, testing, installation and operation is separate both in organization and in timing.

VISA can deliver a formidably collection of services for a fraction of DoD costs because its organization and its concept of operation is completely different.

The following statistics illustrates what VISA delivers for the money it spends: [72]

1. Every day, VISA processes up to 1.8 billion credit card entries and has the capacity of handling over 20,000 transactions per second. The number of DoD daily transactions not more than a tenth of this amount.

2. VISA accepts cards at 1.7 million locations. DoD supports not more than a tenth of this.

3. VISA processes entries for 15,700 financial institutions. The DoD network interfaces with not more than a tenth of that.

---

[71] http://investor.VISA.com/phoenix.zhtml?c=215693&p=quarterlyearnings

[72] http://corporate.VISA.com/about-VISA/technology-index.shtml

4. VISA processed at peak time more than 200 million authorizations per day. The peak load on DoD, under warfare conditions, is unknown but would not be comparable.

5. VISA operates globally from three synchronized data centers linked by 1.2 million miles of optical lines. The DoD GIG does not permit real time synchronization of data centers because it has limited capacity for that.

VISA shows the following operating characteristics:

Fast – On average, transactions are processed in less than a second. This includes providing business-critical risk information to merchants and banks. DoD applications will average a latency that is much greater. DoD latencies are not measured and not tracked.

Secure – VISA employs multiple defense layers to prevent breaches, combat fraud and render compromised card data unusable. These defense layers include data encryption, network intrusion detection and neural network analysis. Real-time risk scoring capabilities are the result of more than 30 years of monitoring transaction patterns and applying sophisticated risk management technologies during the authorization process. Risk analysis methods detect unusual spending patterns and flag possible fraud in real time. These examine 40 separate transaction aspects and 100 or more known fraud patterns which they weigh against a profile of all of the cardholder's transactions from the last 90 days. The result is an instantaneous rating of a specific transaction's potential for fraud, which is then passed to the card issuer. The card issuers, based on their own proprietary criteria, decide to accept or to decline transactions. DoD does not have the forensic assets in place to apply "artificial intelligence" screening methods either to infiltration or exfiltration of its traffic.

Reliable – VISA runs multiple redundant systems to ensure near-100% availability. A self-correcting network detects transmission faults and triggers recovery. For DoD a real time redundancy is not affordable. Up-time reliability is not measured. In fact, standards for up-time reliability measurement and reporting do not exist.

Flexible –VISA supports a diversity of payment options, risk management solutions and a number of different information products and services. This includes more payment methods as well as a choice of access and controls. In DoD the GIG is only a telecommunications carrier, with limited capacity. The GIG does not include a capacity to vary its functionality.

Scalable – VISA processed over 92 billion transactions per year, each settled to a choice of currencies such as penny, peso, ruble or yen. This is accomplished in over 50 languages. On a peak single day last year, VISA processed more than 200 million authorization transactions. VISA stress tests show the capacity to process close to a billion transactions per day. DoD network scalability is fractured and therefore has a very limited capacity.

VISA authorization transactions can be complex. The following is a simplified description of the authorization and payment processes. VISA offers to Issuers a wide range of collection plans and features, such as customer loyalty programs, which add more steps to the following sequence:

1. The Cardholder swipes a credit card into millions of VISA-compatible card readers or accounting machines. Hundreds of different manufacturers make these devices, each with different software. These devices are located even at the most remote locations in the world.

2. The authorization transaction is checked, secured and encrypted by the Merchant's software.

3. It is passed to the Acquirer — usually a merchant's bank — where the Cardholder's account is credited after checking and verification using bank-specific software.

4. The Acquirer reimburses the Merchant instantly after verifying the authorization request. The purchase is authorized at the point of sale.

5. The encrypted authorization is then passed from ten thousands of Acquirers to one of three VisaNet global data centers where every authorization transaction is subject to further risk analysis, security verification and protection services.

6. VisaNet then passes the authorization transaction to hundreds of Issuers, which are the Cardholder's bank. The issuer collects from the Cardholder's account by withdrawing funds if a debit account is used, or through billing if a credit account is used. After the funds are successfully transferred, the approved transaction is returned to its origin where it would be displayed on different formats.

7. If the Cardholder's account is overdrawn, the sequence of the entire process is reversed and the credit authorization is withdrawn.

The entire workflow of credit card authorization from start to finish takes place over the public Internet, or over dedicated optical lines, in encrypted format using the "tunneling protocol" in conformity with VISA dictated standards. By using "tunneling" the VisaNet can receive and transmit over incompatible trusted networks, or provide a secure path through untrusted networks.

In the case of DoD applications it is impossible to track, evaluate or measure end-to-end performance. The DoD architecture has not been designed for assigning separate and distinct roles to the required standards, to the functions of the infrastructure, to the roles of enterprise systems and to the missions that have been delegated for completely decentralized control.

SUMMARY

VisaNet is not just a network service. It can be best described as a global cooperative organization that reaches directly into each of its 15,700 financial institutions with software upgrades, standards enforcement, compliance verifications, security assurance and diagnostic help. VisaNet is a confederation of banks for VisaNet voluntary participation since competitive offerings are also available.

Perhaps the most important single insight to be gained from the VISA environment is a focus on applying systems engineering to the credit card network in its entirety from points of entry to the processing of authorizations in banks. VISA views its business as an integrated continuum that requires continuous tuning as technologies, features and networks change. For instance, VISA tracks the latency (response times) and up-time availability in every link. VISA deploys network engineers who work closely with application designers and data center operators to shave microseconds from transactions.

Perhaps the greatest economies of scale are gained from a complete centralization of control over the management of the software infrastructure of VisaNet. While leaving the complete management of banking

software in the hands of each of the 15,700 financial institutions, Visa continuously implements enhancements to its global payment network from a central location. There two major system upgrades each year for the entire network. Each of these upgrades is a carefully choreographed event, which involve the collaboration with each of the financial institutions, merchants and processors around the world. An average system upgrade requires some 155,000 person-hours. In each case there are up to 100,000 lines of code changed, creating 50,000 application upgrades each year.

The VISA approach is different from current DoD practices where the severance between the developers, infrastructure operators and the managers of the client environment takes place without synchronized integration of every part.

The VISA operates in close coordination between IT management and business executives. Business managers control the budget and dictate how to make trade-offs between schedule, cost and features. In VISA computer networks are treated as an integrated and seamless workflow that is continually maintained and upgraded. In contrast, the DoD approach is to tear asunder planning, engineering, software implementation, testing, installation, infrastructure operations and data processing. Nobody is in charge of the entire workflow from conception to the delivery of results.

DoD is trying to create and manage something that is fundamentally an inseparable process. DoD systems are a collection of subdivided efforts that are time-separated into contractually organized parts. Such an approach is not affordable any more.

PART 10: SEMANTIC SOFTWARE

Semantic Web for Information Dominance

SOURCE: http://pstrassmann.blogspot.com/2010/06/semantic-web-for-navy-information.html

The information requirements for the Information Dominance Corps, which combines the Navy's intelligence and information technology capabilities, will create an unprecedented increase in the demand for information services:

1. Navy forces will be connected into a single, global network for afloat, ashore and space. 2. Every Navy platform will function as a data collector. 3. Every data collector will make information available, in real time, for use by all other nodes. 4. Every sensor will be connected via standard interfaces to make local data globally accessible. 5. Every shooter will have the capacity to compile, assess and exploit data from any sensor or data repository. 6. All data is will be universally discoverable, accessible and secure.
2. Translating these requirements into an operating environment will: 1. Require every Navy sensor to be interconnected (such as radar, UAVs, intelligence sources, satellites, and observation sources). The estimated number of such sensors is at least 10,000. 2. Generate, on the average, at least ten transaction/minute, which suggests at least six million transactions/hour. 3. Retrieve and store electrical signatures, text and video with an average of at least 2 Megabytes per transaction. This would generate a stream of data totaling of at least 12 thousand terabytes/hour, or 300 petabytes/day. At present (2009), Google processes about 25 petabytes/day. With the cost/petabyte declining 25-30% year (a 30 fold decline over ten years) one can project Google-like systems operating well in excess of the range projected for the Navy. 4. Display to at least 50,000 shooters simple graphic displays extracted from the shared global files. Such data extraction would require a latency of not more than a quarter of a second, while assuring 100% network reliability achieved through multiple redundancies of data centers and communications links.

To link the shooters to the data cannot use Google like key-word extraction methods. Only a semantic web, in which the computer network relates the relevance of data to a shooter's local situation can deliver what is necessary for meeting information dominance requirements.

The deluge of video data from these unmanned aerial vehicles, or UAVs, is likely to get worse. By next year, a single new Reaper drone will record 10 video feeds at once, and the Air Force plans to eventually upgrade that number to 65. Chief of the Intelligence, Surveillance and Reconnaissance Division of the U.S. National Geospatial-Intelligence Agency, projects that it would take an untenable 16 000 analysts to study the video footage from UAVs and other airborne surveillance systems. (http://spectrum.ieee.org/robotics/military-robots/the-uav-data-glut).

The semantic web makes it possible for computers to understand what can be extracted from huge files in the context of a shooter's unique inquiry. The key to such a capability is the availability of machine-readable metadata that provide the logical tags for connecting related information. This makes it possible for automated agents to search and then display information from globally distributed databases.

SUMMARY

The stated Navy Information Dominance vision calls for the delivery of the most ambitious operational concepts ever conceived, anywhere. None of the existing commercial designs, such as Google, are comparable in scope.

The systems planners for the Information Dominance capabilities should now consider proceeding with cloud designs that will function according to the stated vision.

Starting with virtual servers, virtual desktops, data virtualization and network virtualization will place the Navy on a path that may take at least a decade to achieve.

Semantic Web for Navy Dominance

It is the objective of Information Dominance Corps (IDC) to manage a global network that delivers instant integration of military data across several separate specializations such as geographic, intelligence, logistics, manpower as well as information about blue or red forces.

These objectives create an unprecedented demand for the retrieval of unrelated data from sources that are diverse and not interoperable. Such data is now stored in files that have inconsistent coding. The existing files are organized in contract-mandated projects that answer only inquiries that are limited to their respective enclaves. For answers that combine weapons, geography or logistics the Navy/MC analysts must surf through several databases, which are neither synchronized nor compatible.

Presently the Navy/MC has to depend on human analysts to use judgment in the interpretation of scattered facts. That is not easy because the examiners of the data have to deal with different vocabularies, undocumented data definitions and dissimilar formats. Therefore an enormous effort is expended in the cross-referencing of disparate data repositories and to reconcile data sources that describe the identical event, but are coded differently. With the inclusion of ten thousands of sensors and with the presence of thousands of computing devices in the global Navy/MC network, the number of analysts that would be required for sifting through all this data would exceed whatever is manageable and surely affordable.

To overcome manpower limitations in the future the IDC will have to resort to semantic web technologies to assemble and correlate data that would support operating needs. The semantic methods are techniques that rely on the extraction of the meaning of data from their related context. Such context is obtained by appending to each original data source a long list of related information. These are named to be data "ontologies".

Ontologies are formal statements that describe the knowledge about a particular data element. The texts of ontology statements are annexed to their respective data in a standard format. In this way they become readable as computer-addressable data entries. As result all data files end up as strings of ontologies that are attached to their respective data sources, which reveals the logical relationships. This arrangement makes it possible for computers to search and retrieve relationships and connections to data sources. It connects the scattered "dots" of seemingly random military data. It reveals what is the hidden meaning of transactions that are under examination.

In a mature semantic web gigabytes are devoted to associate ontology statements for descriptions of only a few bytes of original data. The adoption of ontology-based semantics requires the construction of computing facilities that house huge amounts of computing and storage capacity. The handling of such enormous amount of data requires data centers that possess economies of scale in capital cost while

conserving energy that would otherwise swamp most of the available generating capacity. Such data centers can cost as much as a billion dollars.

Ontologies can be generated automatically by browsing through logically related information in multiple databases searching for numerical information but primarily for unformatted text that has been placed on disks in a narrative format. Indexing text by some sort of a numerical coding schema is not of much use. Indexing relies on pinpoint identification of each data element either from its numerical value or from words used as keywords. Index methods are precise, but cannot discover relationships that have not been previously tagged. They are useless in the case of foreign languages or with new vocabularies the people invent every day.

The difference between the index and the semantic methods is that data retrieved by index methods must depend on human intervention to extract knowledge out of a huge number of possible choices. For semantic extractions the available data is examined by computers and only then presented as a small number of results for further examination by human operators.

The purpose of the semantic web is to make it possible for the IDC network to connect useful information from ten thousands of databases automatically. The war fighters can be then shown what possible actions they could take. With the adoption of semantic methods IDC will not be looking for thousands of uncorrelated search results, as is the case right now. It would receive answers in the form of a few priority-ranked findings.

THE IDC NETWORK

The IDC computing environment should consist of a distributed but highly redundant global network. Various nodes of this network should collect information from every platform that acts as a data collector, such as desktops, laptops, smart-phones, battlefield texting communications, unmanned aircraft video images, satellite pictures and radar tracking. A selection from this data would become available to appropriate persons, since the network would possess situational awareness about each war fighter.

The ultimate objective of all these extractions is to endow everyone with the capacity to compile, assess and exploit information in support of decisions. Only a semantic approach in which the computer network relates data to its local situation can deliver that. The semantic approach makes it possible for computers to "understand" what is dispersed among widely distributed files. Only machine-readable data can be used sift through every file that could possibly reveal what is otherwise hidden. It will be only by means of automated software agents that IDC analysts will be able to support information dominance.

Ultimately, the data collected by IDC will require the recovery and storage of information from ten thousands of connected devices. This data would be placed in petabytes (thousands of terabytes) files, growing into exabytes (thousands of petabytes) in less than a decade. It would require the offering of high reliability levels - 100.0% with automatic fail-over - when supporting combat. All of the data, in different data centers, would have to be instantly accessible - in less than 250 milliseconds - for the retrieval from multiple files. This would make IDC information universally discoverable, accessible while maintaining assured levels of security.

The IDC network requirements are demanding. They exceed, by a wide margin, what are the existing Navy/MC capacities. The Initial Operational Capability (IOC) would call for processing of over hundred thousand transactions per second. The capacity for handling these transactions would have to grow exponentially with time because it would be carrying high-bandwidth graphic, image and color video. Such transmissions consume multiple megabytes of carrier capacity per transaction. Consequently the bandwidth to and from the IDC channels would have to be ultimately measured in terms of thousands of gigabyte (GB) per second.

After the receipt of the raw data into the IDC files, linked super computers would have to screen the inputs for further analysis. Computer software would then be deployed to pre-process inquiry patterns in order to identify standard queries so that typical questions can be answered without delay. One of the liabilities of semantic methods is the enormous amount of computation that is required to deliver useful results. The pre-processing workloads on the IDC super computers would vastly exceed what is needed for the handling of simple messages.

The projected size of IDC data files that support semantic processes is likely to exceed current Navy/MC space by a large multiplier. At maturity it would require storing a stream of data totaling at least a thousand terabytes/hour or more than 20 petabytes per day which is comparable to the processing load of Google. Google and IDC only differ in that the Navy/MC requires higher system uptime in order to support warfare conditions.

SEMANTIC TECHNOLOGIES

The key tools for constructing and using the semantic web are the Extensile Markup Language (XML), the Resource Description Framework (RDF) and the Web Ontology Language (OWL). The management of these standards is under the guidance of the World Wide Web Consortium (W3C- http://www.w3.org/), which has been led by Tim Berners-Lee, the inventor of the World Wide Web. The term "Semantic Web" refers to the W3C's vision of how data should be linked on the web. Semantic Web technologies are methods that enable people to create data-collections, build vocabularies and write rules for the handling of related data. These techniques are now labeled as Web 3.0 solutions and consist of:

1. XML (and related standards), which is the protocol for recording data for web accessibility. It is the format in which all data are recorded.

2. RDF (and related standards) is the model for assuring data interchange on the Web. RDF facilitates data merging and correlation even if the underlying recording schemas differ. RDF supports learning about data recording patterns over time without requiring the data identification to be changed. RDF forms graph views of recorded information, which is useful in presenting easy-to-understand relationships among data sources.

3. OWL is a family of languages for authoring ontologies. OWL would represent the knowledge about the events and their respective relationships as they apply to IDC operations. They form an added layer of meaning on top of the existing web service protocols. Although many of the OWL descriptions can be obtained by automatic means that use mathematical algorithms, ultimately it will take a human analyst to note what are the applicable IDC relationships. This can be done only if everyone shares a common vocabulary for describing what is the definition of shared knowledge for the IDC enterprise. However, there are already commercial software packages available that support the formation of OWL-complaint semantic relationships, which should speed up the adoption of these methods.

After sufficient experience is accumulated by means of automatic data mining of transactions many of the ontology templates can be reused so that the labor cost of maintaining the semantic web can decrease. The IDC databases can be then organized as specialized web services such those that produce information for target selection. Such services can then fuse data from dozens of sensors, latest geographic images as well as data about available weapons and could be deployed aboard ships.

SUMMARY

The semantic web should be viewed as the latest extension to the current web. The semantic web advances searching methods from inquiries that are based on structured data to producing results that answer uncorrelated questions even if they are in the form colloquial sentences. The semantic web should be

therefore seen as an enhancement to the already existing methods that are available for accessing information over the Internet.

Semantic methods overcome the present limitations of separate and disjointed web pages that cannot be readily collected for the assembly of enterprise-wide information except through human intervention. The semantic web advances the IDC from connecting web pages by means of the analysts' eyes to connecting the underlying data by means of computers. It advances IDC analysts from sifting through piles of computer listings to using computers to identify a few possible answers.

Data ontologies will become the method for applying semantic-based applications to IDC operations within the next decade. The enormous expansion of IDC data, especially with the sharing of sensor, logistic and personnel information, will make the semantic-based retrievals of information an absolute economics necessity.

Ultimately, ontologies will form the foundation on which other advanced methods, such as "fuzzy logic", artificial intelligence, neural networks as well as heuristic searching can be adopted. Those are reasons why the use of the semantic web should be seen only as another but very important stepping stone in the evolution of computer-based reasoning that cannot be delayed.

PART 11: DATA & IT MANAGEMENT

The Merits of Storage Virtualization

SOURCE: http://pstrassmann.blogspot.com/2010/06/merits-of-storage-virtualization.html

Storage virtualization is an abstraction (separation) of logical storage from physical storage so that it may be accessed without regard to physical storage or technology management methods.

Storage virtualization makes possible the identification, provisioning and management of data storage at multiple locations. Storage virtualization can extract data from completely different applications as if they were a single, consolidated resource.

Managing disk storage was once simple. If we needed more space, we got a bigger disk drive. For reliability we developed RAID, network-attached storage and storage-area networks that required the installation of more complex data center storage management processes.

The latest answer to this dilemma is storage virtualization, which adds a new layer of software between storage systems and servers. Network management no longer needs to know where specific drives are located. The management of partitions or storage subsystems, or the identification of where data resides now becomes a task for the virtualization software. Administrators can identify, provision and manage distributed storage as if it were a single, consolidated resource available across an entire network.

Availability increases with storage virtualization, since applications aren't restricted to specific storage resources. Data access is thus insulated from failure of a particular disk capacity. This automates the expansion of storage capacity and reduces the need for manual provisioning in supporting a specific applications. Storage resources can be updated on the fly without affecting application performance, thus reducing and even eliminating downtime.

Storage virtualization operates as an intermediate layer. It becomes the primary interface between servers and storage. Servers see the virtualization layer software as a single storage device, while all the individual storage devices see the virtualization layer as their only server. This makes it possible to group storage systems—even devices from different vendors and different data base software solutions—for unified management.

Storage virtualization shields servers and applications from hardware or software changes to the storage environment, letting users easily hot swap a disk. Data copying and data-backups are also managed at the virtualization layer. For instance data replication, whether for snapshot or disaster recovery to different data centers can be handled by the virtualization system, as a background task, with a shared management interface.

Because data can be moved at will, vulnerable data centers or outdated storage capacity can be moved to the best storage devices. The virtualization software or device is responsible for maintaining a consistent view of all the mapping information for the virtualized storage, which defines meta-data and is stored as an overarching mapping table. Storage virtualization can be structured to look up metadata in the virtual disk space for data discovery and for de-duplications.

Technology Maturity

Storage virtualization can be considered a mature technology with F5 Networks and Citrix ranked as the leaders in the ability to execute. There are other vendors such as 3PAR, Compellent Technologies, DataCore Software, Hitachi Data Systems, IBM, StarWind Software and Violin Memory. Storage virtualization has origins in 1980's but has become widely adopted with the large-scale virtualization of data center servers.

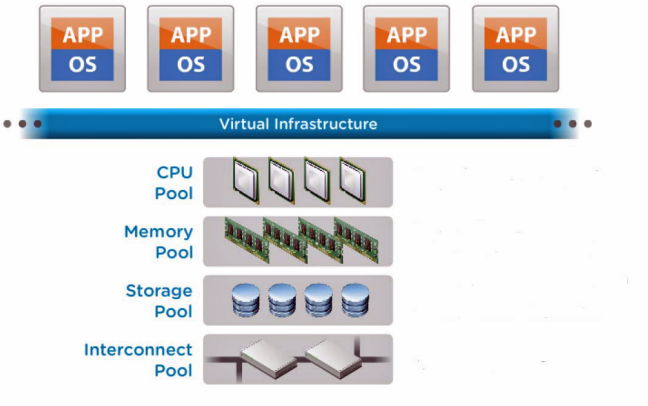
Implications

Storage virtualization, in addition to offering savings in the utilization of disk capacity, enables improved real-time interoperability of data extracted from dissimilar data sources. DoD can now migrate its data files, which are presently wedged into incompatible servers, to an environment where pooled data services and universally shared metadata become economically feasible.

Enterprise Data Base on the Cloud

SOURCE: http://pstrassmann.blogspot.com/2010/07/constructing-enterprise-data-base.html

The ultimate objective of cloud computing is to separate the CPUs, memory, storage and communication technologies from their respective applications. The objective is to allow the creation of shared pools that can be re-organized for achieving high levels of utilization.



When each application owns its own technologies the customer will acquire excess capacity for peak loads. With separate virtualization of technologies that is not necessary. Sharing of resources will deliver reductions in capital costs as well as cuts in operating expenses. Such pools can support thousands of applications and can be managed with fewer people.

The greatest cost advantages will be derived from the creation of storage pools. The need for added disk capacity is rising faster than for other assets while disk capacity utilization is declining. The technical means for creating a storage pool is accomplished by means of virtual disks. These are stored as files on the

hypervisor.

One of the key features of Type 1 hypervisors is the encapsulation of legacy applications. This means that the complete legacy files can be migrated, copied, moved, de-duplicated and accessed quickly. Since an entire disk partition is saved as a file, virtual disks are easy to back up, move, and copy.

The bare-metal hypervisor architecture for managing storage pools permits near-native virtual machine performance as well as reliability and scalability without the need for a host operating system. Virtual machine disk files offer access to data while giving to administrators the flexibility to create, manage and migrate virtual machine storage as separate, self-contained files. Redundant virtual disks eliminate single points of failure and balance storage resources. This allows the clustering of files and enables accessing several files concurrently.

Many of the available hypervisors are certified with storage such as systems from Dell, EMC, Fujitsu, Fujitsu Siemens, HP, Hitachi Data Systems, IBM, NEC, Network Appliance, StorageTek, Sun Microsystems and 3PAR. Internal SATA drives, Direct Attached Storage (DAS), Network Attached Storage (NAS). Both fibre channel SAN and iSCSI SAN are supported. This provides the means for providing infrastructure services such storage migration, distributed resource scheduling, consolidated backup and automated disaster recovery.

All files that make a virtual environment consolidate data in a single logical directory managed by means of a meta-data directory. With automated handling of virtual machine files, the management system provides encapsulation so that it can easily become part of a disaster recovery solution. Conventional file systems allow only one server to have read-write access to the same file at a given time. By contrast, enterprise storage allows multiple instances of virtual servers to have concurrent read and write access to the same resources. Virtual enterprise files also utilize the journaling of meta-data changes to allow fast recovery across these multi-server resource pools. Snapshot features are available for disaster recovery and backups.

SUMMARY

The migration of data files from individual applications into the virtual disk environment should be seen as a way to deliver storage pools into a cloud environment. Initially, the linking between applications and corresponding data will be closely coupled. However, through conversion software (such as provided by firms comparable to Informatica and AbInitio), a clean separation of data from applications can be achieved ultimately.

The consolidation of all business applications data files into a DoD business repository, controlled by a single meta directory, will materially reduce the costs of DoD business systems operations. Over $20 billion/year is spent in DoD running applications that consume machine cycles in exchanging each other's data files. A consolidated data file will eliminate much of that and make a SOA (Service Oriented Architecture) possible.

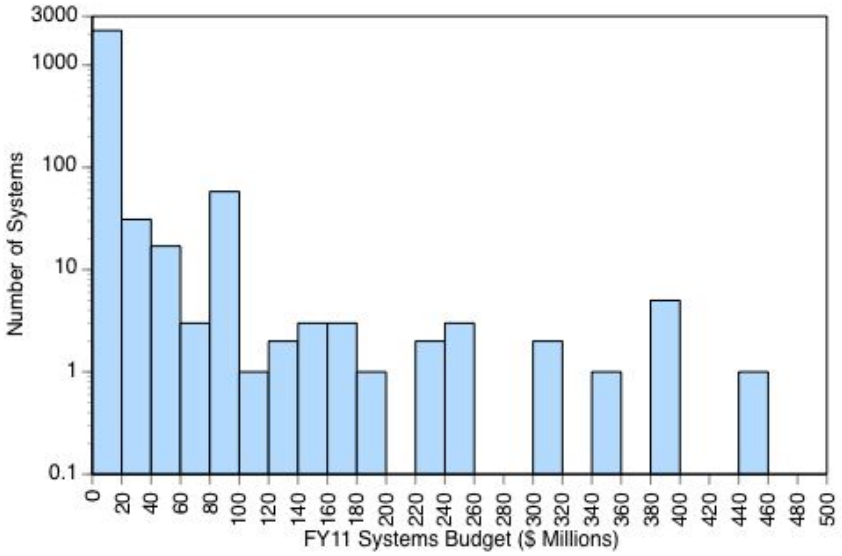
How to Fit DoD Into a Cloud?

It is a policy now that DoD should start migrating into an architecture that is based on cloud-like concepts. There are leaders who have suggested that transporting DoD application into a Google-like environment would solve persistent problems with security, interoperability and cost.

Before anyone can subscribe to a cloud-centered policy it may be useful to examine the current state of DoD applications. What we have now is fractured and often broken, non-standard, custom-fitted and without a shared data environment. To move thousands of such systems into an ordered Platform-as-a-Service or Infrastructure-as-a-Service will require an enormous expenditure. It will require restructuring how systems are designed and how they can run.

To gain a better understanding where we find ourselves in a constrained budgetary environment it may be useful to examine what are the "as-is" conditions of existing systems before making "to-be" projections. What is needed is funding commitments and a realistic time-line.

The OSD Deputy Chief Management Officer has defined the scope of work that needs to be done.[73] It includes a list of 2,319 systems costing $23.6 billion, which represents 69% of total DoD IT spending.[74] The following statistic shows the number of projects plotted against their respective FY11 budgets:[75]



94% of systems have budgets of less than $20 million. Most of these applications are categorized either as "legacy" systems to be phased out, or "interim" systems, which have limited life expectancy. There are 1,165 such systems, or 55% of the DoD total, which are obsolete and require ultimate replacement. How DoD systems are now fractured is best illustrated in the following table:

[73] http://dcmo.defense.gov/etp/FY2011/home.html (Master List of Systems)

[74] DoD IT spending excludes the costs of the military and civilian IT workforce

[75] There are only eight projects reported with budgets >$500 million

Number of Systems	Financial Management	Human Resources	Other Systems	Total
Army	96	252	398	746
Navy	92	110	270	472
Air Force	42	102	344	488
Agencies	106	201	306	613
Total	336	665	1,318	2,319

This table shows that about half of all systems are in Financial Management and in Human Resources. The Business Transformation Agency, now discontinued, had spent six years attempting to unify these applications. A large number of independent Agencies now control a large number of systems, which will make planning for consolidations into a cloud environment difficult.

SUMMARY

The current proposals to ultimately merge 1,165 obsolete systems into 1,154 "core" systems may not be executable. The problem lies not in the proliferation of systems (each with many applications) but in the contractual structure for small systems, each with its unique infrastructure. Most of the current 2,319 have been built one contract at a time over a period of decades. Limitations on funding will insure that each of these systems will have a unique configuration of operating systems, application codes, communication management and data base access. With over 76% of software maintenance and upgrades in the hands of contractors while DoD oversight experiences a managerial high turnover any interoperability can be achieved only by means of expensive add-ons.

There is no reason why the Army should not operate with more than 252 human resources systems and even more applications. There is no reason why the Navy does not need more than 92 systems with hundreds of diverse applications to manage is financial affairs. More 60% of the current costs for Operations & Maintenance are consumed in communications, data management and security infrastructure. If DoD can acquire a Platform-as-a-Service or an Infrastructure-as-a-Service capability, the Army, Navy, Air Force and Agencies will be able to construct many inexpensive (and quickly adaptable) applications. These would be then placed on top of a much cheaper standard, shared and secure infrastructure.

DoD should not proceed with the re-writing of existing applications or with the consolidation of existing systems into a smaller number of systems. Instead, DoD should proceed with the separation of its infrastructure and data management from applications. The direction should be: thousands of applications, but only a few cloud infrastructures!

Protecting Databases with a Software Vaults

The innermost cores of DoD systems are the databases. In cyber attacks viruses can be implanted in applications, denial of service can block networks or malicious code can produce false results. In all of such cases there are ways how reconstitution can take place. However, if an adversary degrades a shared database from which applications draw data, recovery is difficult. A petabyte database may contain tens of millions of data elements, which are updated at millisecond speeds. If the database integrity attack is designed to be gradual and progressive the users will be receiving results that only few will question. At a critical point the users will stop trusting their computer screen and resort to other means how to improvise what to do without computer aid.

The greatest threats to DoD cyber operations are not external attacks but the subversion from an insider. Whatever may be the motivation for perverting a major database is immaterial. Whether it is malicious, or from a disgruntled employee or from an enemy operative is immaterial. What will ultimately matter is that a critical moment an act of cyber warfare will disable warfare operations.

A number of database vendors offer "Database Vault" protection software. The question is whether this offers safeguards in cases where personnel performing data base administration (DBA) tasks are a threat.

When databases are corrupted the greatest risks do not arise from technical failures against which elaborate safeguards are known to exist. In protecting databases the most important question is who is in charge of real-time policing of the actions taken by DBAs? What is the chain of responsibility for real-time countermeasures? What is the role of the auditors and administrators to see to it that adequate safety processes are in place? What are the separate chains of command through which the various actors in ensuring database security report?

The problem of safeguarding databases is compounded by the fact that database software is one of many applications that run in the same datacenter environment. Numerous versions of Unix and Windows will be accessing the identical database. The DBA administrators, the auditors, the oversight administrators and a wide number of final users will be running millions of queries per hour that access a shared database. The users will only retrieve data from databases, which are "owned" by the DBAs. However, it is possible that without appropriate safeguards there may be many individuals that will have access to a database. Unless such access is controlled and fully accounted for there will be an exposure that damage to the database could come from many sources.

There are many methods for attacking databases. One of these, and perhaps the most persistent means is through a "trojanization" of code. To trojanize a software product, one of the diverse and high turnover contractor employees doesn't even have to actually write an entire backdoor into an application. Instead, the malicious developer could purposefully write code that contains an exploitable flaw, such as a buffer overflow, that would let an attacker take over the machine when a database application has to be restarted. Effectively, such a purposeful flaw will act just like a backdoor. If the trojan sneaks by the DBA, the developer would be the only one who knows about the hole. By exploiting that flaw, the developer could control any systems using such code at a time of their choosing. For this reason the DBA will have to see that the interactions between applications and the database are totally isolated.

The DBA will always have access privileges and attach a debugger to any database process and record all operations, reset function and modify the ways the database system works. There are many libraries and tools to do it. Each vendor provides own proprietary tools for that purpose. There are also numerous open source and proprietary software available for extracting data or for modifying data bases such as DUDE (Database Unloading by Data Extraction).[76]

Under extraordinary circumstances the DBAs must be able to recover and reconstitute data files. That can be complex and time-consuming especially if the database was encrypted. Such recovery must take place under the surveillance by qualified personnel. An attacker can always induce a system failure and apply changes to the database software during recovery without making it a suspicious act.

[76] http://www.ora600.nl/introduction.htm

A Database Vault (DBV) is designed to provide a separate layer of protection around a database application.[77] Its purpose is to prevent access to data especially from highly privileged users, including Data Base Administrators (DBAs) and application owners.

A DBV introduces into the database environment the ability to define data domains, to specify applicable command rules, to assign who can access data, what data they can access, and the specific conditions that must be met in order to grant such access.

Databases consist of data domains, which define collections of data in the database to which security controls must be applied. They can consist of database objects such as a single table or multiple tables, procedures, programs, an entire application, or multiple applications. In an enterprise scenario, for example, data domains separate the data used by one department from that used by another. In the case of DoD the Army, Navy, Marine Corps and Air Force would have respective DBAs define and control the assignment of domains.

A DBV defines the rules and control processes how users can execute data base management statements, including within which domains and under what conditions they may do so. Command rules leverage individual or combinations of factors, such as identifying individuals and their access characteristics, in order to restrict access to data. Built-in factors include authentication method, identification type, enterprise identity, geographic origin, language, network protocol, client IP, database hostname, domain, machine, and others. In addition to these, custom factors can be defined. Restrictive factors can be assigned to all users, including DBA's. Multi-factor authentication rules are supported. For example, a certain action could be restricted to being allowed only from a specific IP address within a specified time range.

To protect the database from even high privileged users such as DBAs, the vault includes a definition of the separation of duty in which the DBV is separated from DBA functions. The database vault information itself is protected by its own secure domain, which prevents tampering and therefore must be kept on physically separate servers. The database vault software requires that the DBV managers assume the responsibility for the creation of all new data domains in the database. This will then override all existing accounts with the create user privilege.

Finally, a built-in reporting mechanism provides reports, including those that detail as to who has access to what data and if there were any attempted violations.

SUMMARY

It has been reported that DoD presently operates over 750 data centers and runs thousands of applications. There must be thousands of databases that pass information back and forth for interoperability. The risk that at least one of these databases can become a source of infection to others remains as a security exposure. Software protection, such as properly administered "vaults" can offer great protection in such cases. However, the software "vault" cannot be a part of an application but must be a DBA responsibility.

The personnel with DBA responsibilities will remain the holders of the most critical role in managing DoD information assets. Though there is a long list of security measures that all must be in place to assure that appropriate processes are in place, the elimination of compromises from an insider cannot be dictated by procedures. The security of DBAs, as well as of all personnel auditing the functions of the DBA, deserves at least SECRET level of clearance. TOP SECRET clearance is warranted for all DBA related positions

[77] http://products.enterpriseitplanet.com/security/security/1146070533.html

involving warfare actions. Whether such increases in security can be achieved with the current DoD reliance on contractor personnel is questionable.

The current proliferation of diverse database cannot be fixed in the short run. The best one can do is to reduce the number of DBAs in DoD, to assure their security clearance, to increase the number of Civil Service civilian personnel and to conduct oversight of database surveillance procedures exclusively through military officers.

A DoD enterprise level database will be accessed by possibly hundreds of applications originating from diverse Components. It will contain petabytes of data that are updated and accessed in real time. There is no question that such a database will become the target of choice for any cyber attack. Consequently, extraordinary precautions will be taken to offer protection from any unauthorized access to specific data domains, whether they come from external or internal sources.

Creating a Database Vault protection mechanism must be mandatory for mission critical cases. By this means DoD obtains not only an assured layer of protection but also creates a well-defined separation between the roles of the DBV, the DBA and the auditor or the supervising military personnel. All reporting of violations of restrictions occurring in the Database Vault would have to be routed as secure messages directly to those who are accountable for the data vault. Under defined conditions all alterations to the database could be then restricted automatically until human intervention authorizes what steps can be taken next.

PART 12: OPEN SOURCE SOFTWARE

Secure Workstation for System Development

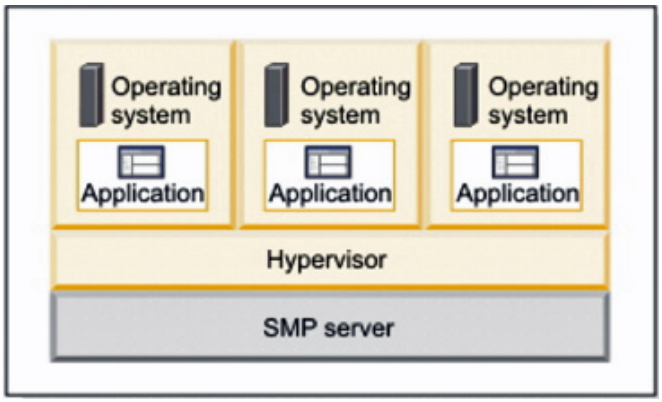
SOURCE: http://pstrassmann.blogspot.com/2010/07/secure-workstations-for-systems.html

A virtual workstation enables the development, testing and deployment of diverse applications without changing equipment at a customer's site. This is accomplished by adding a hypervisor, as virtualization software. Desktops, laptops or smart-phones thus become virtual workstations with the capacity to perform a large variety of tasks. The following are the functions of virtual workstations that operate as the platforms for systems development:

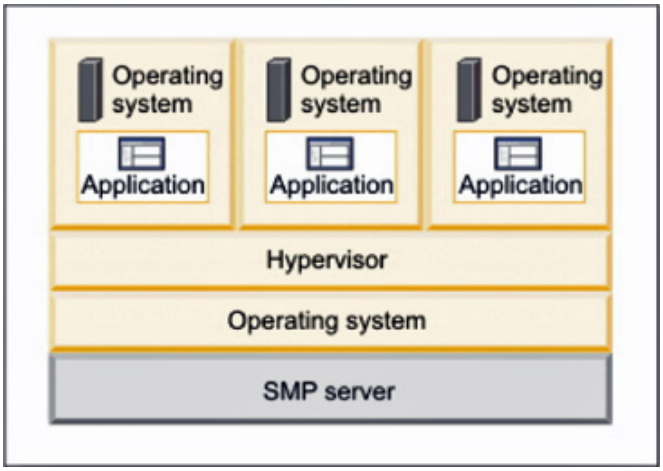
1. Test applications, with different levels of security on the identical desktop, using Linux or Windows but without rebooting.
2. Experiment and test a combination of new versions of security proposed safeguards on separate and isolated virtual computers without the need to acquire separate computing devices.
3. Deploy different combinations of browsers and third party security appliances for examination how they interact with different applications. Assure the elimination of conflicts arising from new software patches.
4. Validate if there is interference between security software and various versions of browsers, operating systems and proprietary application development tools. The number of cases that need testing could be in thousands.
5. Demonstrate how the performance of the security software will affect proposed computing configurations, multi-core processors or virtual disks. This includes the verification of encryption codes.
6. Run demonstrations of prototype versions of applications, which includes systems assurance.

The development environment for a secure workstation requires the creation of fully isolated and secure virtual machines that encapsulate an operating system and its applications. The virtualization layer must map the physical hardware resources to the virtual machine's resources, so each virtual machine has its own CPU, memory, disks, and I/O devices, and is the full equivalent of a standard x86 machine.

Virtual workstations can operate either as a Type 1 (or native, bare metal) or as a Type 2 (or hosted) hypervisors. The difference is that Type 1 runs directly on the host's hardware to control the hardware and to monitor guest operating systems whereas Type 2 runs within a conventional operating system environment. The following figure shows one physical system with a type 1 hypervisor running directly on the system hardware, and three virtual systems using virtual resources provided by the hypervisor.



The following figure shows one physical system with a type 2 hypervisor running on a host operating system and three virtual systems using the virtual resources provided by the hypervisor.



The virtual workstation will run on any standard personal computer and will be the equivalent of a full PC, with full networking and devices — each virtual machine has its own CPU, memory, disks, I/O devices, etc. This allows the capacity to run on the supported guest applications such as Microsoft Office, Adobe Photoshop, Apache Web Server, Microsoft Visual Studio, Kernel debuggers, as well as all security software provided by vendors such as McAfee, RSA, Check Point, Sophos and others.

SUMMARY

The development environment for secure systems requires the capacity to test and validate complex interactions between hardware, operating systems, applications and a variety of security offerings. A very large number of possible combinations must be tested not only to verify compliance with required functionality but also to assure operational viability. A virtual workstation has the capacity to assure the exploration of a large number of security features so that project schedules can be accelerated.

By using virtual workstations developers can check the acceptability of available security options in a non-homogeneous environment.

Apache Hadoop

Apache Hadoop is open source software for consolidating, combining and analyzing large-scale data. Apache Hadoop is a software library that supports distributed processing of vast amounts of data (in terabytes and petabytes) across huge clusters of computers (thousands of nodes). It scales up from single servers to thousands of machines, each offering server localized computation and storage. Rather than rely on hardware to deliver high-availability, the software is designed to detect and handle failures at the application layer. It delivers a service for computer clusters, each of which may be prone to failures.

Relational data base software excels in storing workloads consisting of structured data. Hadoop solves a different problem, which is fast, reliable analysis of structured data as well as unordered complex data. Hadoop is deployed along legacy IT systems to combine old data with new incoming data sets.

Hadoop consists of reliable data storage using the Hadoop Distributed File System (HDFS). It uses high-performance parallel data processing using a technique called MapReduce.

Hadoop runs on commodity servers. Servers can be added or removed from a Hadoop cluster at will. A Hadoop server cluster is self-healing. It can run large-scale, high-performance processing jobs despite of system changes.

Dozens of open source firms participate in the upgrading and maintenance of Hadoop/MapReduce. Critical bug fixes and new features are added to a public repository, which is subject to rigorous tests to ensure software reliability. All major cloud services firms that offer computing services already employ Hadoop/MapReduce.[78]

A Map/Reduce job splits input data into independent chunks, which are processed as separate tasks in a completely parallel manner. The Map/Reduce software sorts the outputs of the individual "maps" on separate servers, which are then fed into the reduce process. The software takes care of scheduling tasks, monitoring progress and re-executing any failed tasks.

The compute nodes and the storage nodes are identical. The Map/Reduce framework and the Hadoop Distributed File System run on the same set of servers. This configuration allows Hadoop to schedule tasks on the nodes where data is already present, resulting in high bandwidth across each cluster.

The Map/Reduce framework consists of a single master JobTracker and of separates TaskTrackers for each cluster-node. The master is responsible for scheduling the jobs' component tasks on the individual servers, monitoring them and re-executing any failed tasks.

Applications specify the input/output locations and supply the map of how a job is processed. This reduces processing overhead via implementations of all connecting interfaces. These, and other job parameters, the comprise configuration management for each application.

SUMMARY

The masses of data, such as is currently tracked at multiple DoD network control centers, cannot be analyzed by existing relational database software. In addition, access to multiple web sites to extract answers to customized queries requires a new architecture for organizing how data is stored and then extracted.

[78] *http://wiki.apache.org/hadoop/PoweredBy*

The current DoD incoming traffic is too diverse. It shows high real time volume peak loads. The text, graphics and video content are unstructured. They do not fit the orderly arrangements for filing of records into pre-defined formats. The bandwidth that is required for the processing of incoming messages, especially from cyber operations and from intelligence sources, calls for the processing of data in a massively parallel computer in order to generate sub-second answers.

The conventional method for processing information, such as used in the existing Enterprise Resource Planning systems, rely on a single massive master database for support.

A new approach, pioneered by Google ten years ago, relies on Hardoop/Map Reduce methods for searching through masses of transactions that far exceed the volume of transactions currently seen in the support conventional business data processing.

With the rapid expansion of wireless communication from a wide variety of personal devices, DoD messages subject to processing by means of massive parallel computers will be exceeding the conventional workload of legacy applications.

DoD is now confronted with the challenge of not only cutting the costs of IT, but also with the task of installing Hardoop/Map Reduce software in the next few years. In this regard the current emphasis on the reduction in the number of data centers is misdirected. The goal for DoD is to start organizing the computing as a small number of massive parallel computer networks, with processing distributed to thousands of interconnected servers. Cutting the number of data centers without a collateral thrust for software architecture innovation may be a road that will only increase the obsolescence of DoD IT assets as Amazon, Baidu, Facebook, EBay, LinkedIn, Rackspace, Twitter and Yahoo forge ahead at an accelerating pace.

Meanwhile DoD is wrestling how to afford funding the completion of projects started after FY01. DoD must start carving out a large share of its $36 billion+ IT budget to make sure that FY13-FY18 investments can catch up with rapid progress now made by commercial firms.

After all, DoD is still spending more money on IT than any one else in the world!

Open Source Frameworks

A software framework is a reusable set of programming methods and program libraries that will produce a standard structure for coding an application. A software framework will produce applications that run in a specified environment.

A specialized version of frameworks is the Web application framework. This can be applied for the development of websites, web applications, and web services. Web application frameworks are becoming the dominant method for delivering computing services.

Programmers find easier to create code when using a standard framework, since this defines how the underlying code structure is organized. Applications can inherit code from pre-existing classes in the framework.

The most widely adopted framework is the Microsoft Framework that has contributed to the widespread adoption of Windows applications. A variety of frameworks are also available from Apple, Oracle (Application Development Framework), Mozilla and Linux. The limitation of these frameworks is their proprietary characteristic. The proliferation of existing frameworks reinforces the writing of code that

reinforces a customer's adherence to vendor-specific solutions.

The most recent innovation is the availability of "open source" frameworks. These make it possible to redeploy code, such as Java, to run code for applications that run on a wide range of platforms. [79]

Cloud Foundry is an open platform service from VMware. It can support multiple frameworks, multiple cloud providers, and multiple application services all on a single cloud platform. It is an open Platform-as-a-Service (PaaS) offering. It provides a method for building, deploying, and running cloud apps using the following open source developer frameworks: Spring for Java applications; Rails and Sinatra for Ruby applications and Node.js as well as other JVM for Grails applications. Cloud Foundry also offers MySQL, Redis, and MongoDB data services. This is only the initial list of open source frameworks. Other frameworks are expected to offer different tools or templates.

Cloud Foundry takes an open approach to connecting with a variety of cloud offerings. Most PaaS offerings restrict a developer to proprietary choices of frameworks and infrastructure services. The open and extensible nature of Cloud Foundry means that developers will not be locked into a proprietary framework or a proprietary cloud such as, for instance, Microsoft Azure or Amazon EC2.

VMware believes that in the cloud era this maps to flexibility and community participation. With this fundamental belief, VMware is open sourcing the Cloud Foundry application execution engine, application services interfaces and cloud provider interfaces.

Cloud Foundry offers an application platform, which includes a self-service application library, an automation engine for application deployment and lifecycle management, a scriptable command line interface (CLI), integration with development tools to ease development and deployment processes. It offers an open architecture for quick development framework integration, application services interface and cloud provider interface.

Cloud Foundry is ideal for any developer interested in reducing the cost and the complexity of configuring programs as well as runtime applications. Developers can then deploy applications that have been built with the aid of open source frameworks without requiring modification to their code when applying their code to different cloud environments.

Cloud Foundry allows developers to focus on applications, and not on hardware or on middleware. Traditional application deployments require developers to configure and patch systems, maintain middleware and worry about network connections. Cloud Foundry allows developers to concentrate on the business logic of applications. Consequently applications can be tested and deployed instantly.

VMware now operates an open-source community site, CloudFoundry.org, where developers can collaborate and then contribute to individual Cloud Foundry projects.

SUMMARY

The Open Source Cloud Foundry is a dramatic innovation. It is based on the concept that in cloud computing there must be complete separation between the underlying hardware, the intervening software (which includes Operating Systems) and the application logic. Applications should be able to run on any

[79] http://pstrassmann.blogspot.com/2011/05/springsource-development-framework-for.html

hardware, regardless of vendor origin. Applications must function regardless of the platform such as desktops, laptops or cell phones. Applications must be accepted by any operating system or any middleware regardless of the way it has been configured.

The objective of the Cloud Foundry is to deliver open source frameworks that will make it possible to run universally accessible Platform-as-a-Service clouds. In the future there will be a large variety of PaaS vendors who will distinguish themselves by offering different service level agreements.

If DoD wishes to reduce its cost by adopting cloud services it will have to re-examine how software is generated, tested and then deployed. DoD will have to adopt open source frameworks for developing applications.

Development Framework for Java Code

SpringSource is the leader in Java application infrastructure and management. It provides a complete suite of software products that accelerate the entire build, test, run and revisions management of the Java application lifecycle. SpringSource employs leading open source software developers who created and now drive further innovations. As results Spring has become the de facto standard programming model for writing enterprise Java applications programming code.

SpringSource also employs leading thought leaders within the Apache Tomcat, Apache HTTP Server, Hyperic, Groovy and Grails open source communities. Nearly half of the Global 2000, including most of the world's leading retail, financial services, manufacturing, healthcare, technology and public sector clients are SpringSource customers.

Millions of developers have chosen the Open Source Spring technologies to simplify Java code development and to dramatically improve productivity and application quality. These developers use Spring because it provides a centralized configuration management and a consistent programming model for: declarative transactions; declarative security; Web Services creations and persistence integration. Unlike the complex and hard to use Java Enterprise Java Bean (EJB) platform, Spring enables all of these capabilities to be applied in a consistent manner that simplifies and partially automates the production of reliable application code.

Spring provides the ultimate programming model for modern enterprise Java applications by insulating business objects from the complexities of platform services. It manages application component management, enables Java components to be centrally configured and linked —resulting in code that is more portable, reusable, testable and maintainable.

SUMMARY

The traditional application development process involves linking Java components manually. Such approach runs into difficulties as the size of the application code and its complexity grows. Writing, tracking and testing such code is hard and difficult to reuse. Spring performs many of the previous

manual coding tasks automatically. Developers can concentrate on business logic and code, without having to worry as much about the software infrastructure that supports the running of the application programs. This is critical when the Java code is used to add features to web sites that are viewed by millions of customers.

As the dependency on cloud computing increases, DoD must pay attention how applications are written and maintained. Software routines have to be interoperable across a wide range of applications. Adherence to Open Source frameworks is necessary to assure that applications can be relocated between data centers either to assure backup or to change contract relationships.

DoD must impose standards how Java code is written. DoD must dictate how such code is deployed for operations and maintenance. The current contract separations between applications planners, application developers and operators - dictated by acquisition rules - is not viable.

Platform for Cloud Applications

The User-centric User Interface (UCUI) software will make it possible to access Software-as-a-Service (SaaS) services with only major Internet browsers. Such approach will enable organizations to centrally manage the provisioning of diverse applications, while applying open standards to security and access controls.

The UCUI software increases the security of using SaaS applications. Users will have a single login available across multiple devices, with self-service accesses to a corporate repository that offers industry standard SaaS applications. This grants access to multiple web applications such as SalesForce.com, Facebook, Google, WebEx and others. With the evolution of cloud computing, hundreds of firms will offer off-the shelf SaaS applications.

At present the access to SaaS requires separate access authorizations and software fixes for configuration alignment. That is hard to do, especially for integration with existing systems that already reside on private clouds or continue to operate as legacy applications. It is the purpose of UCUI software to manage such integration.

Users are also bringing diverse devices to the workplace. Systems managers must now manage multiple access protocols and conversion software routines to enable legacy devices to extract useful information from

any SaaS offerings. It is the purpose of the UCUI software to accept a variety of protocols from all of the devices already in place.

The UCUI is a hosted service that enables organizations to centrally manage the access as well as the usage of different SaaS applications in a seamless continuum. IT management can therefore extend the users' enterprise identity from a private cloud to the public cloud while simplifying the processing of applications in real time. This is ten supported by strong policy management on security restrictions as well as by consistent activity reporting.

The purpose of the UCUI is to offer a single display for managing user access, identity and security across multiple business apps and multiple cloud environments. It is independent from the Microsoft Active Directory. This should be seen as an evolutionary step for migrating from the proprietary Microsoft to an open source environment.

To assure security the user centric platform will have to implement the Security Assertion Markup Language (SAML) [80] and the Open Authentication (Oauth) codes. [81]

The UCUI will ultimately bridge the gaps between the private DoD clouds and the public SaaS clouds. The UCUI can be then deployed in an evolutionary manner without time-consuming integration efforts while also reducing security risks from multiple access locations. Most importantly, the addition of SaaS services to the portfolio of DoD hosted applications will materially reduce DoD infrastructure costs at a time when budgets are shrinking.

SUMMARY

Today's DoD workforce expects access to their data anytime from anywhere. Therefore the workforce will have to be turning to SaaS applications to meet rising needs and to cut operating costs.

An increased dependence on SaaS will most probably have to be met by offerings available from commercial clouds that have been modified to meet DoD's tight security requirements. With rising budget restraints this appears to be the most effective path for migrating DoD systems to increased reliance on cloud services.

Cloud Standards

For further progress of the development of cloud operations it will be necessary to establish standard that will assure that the cloud environment is interoperable. It is the sense of this blog to describe the formats

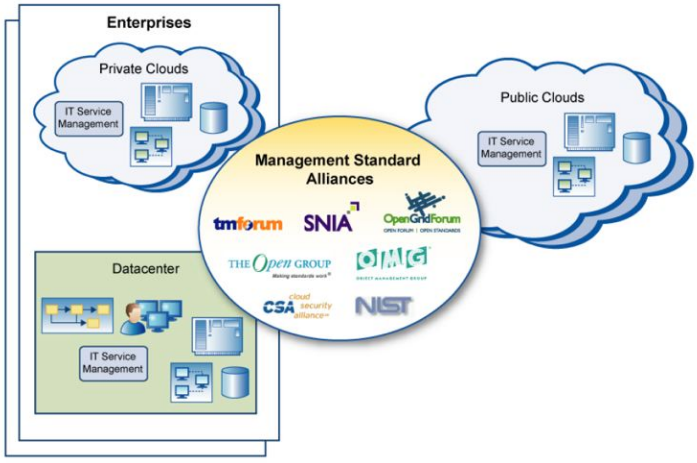
[80] http://pstrassmann.blogspot.com/2011/05/secure-sign-on-for-web-based.html

[81] http://pstrassmann.blogspot.com/2011/05/applying-open-authentication-oauth.html

that will assure a customer that any application, once placed in the cloud environment, will be transportable to any other cloud environment.

Standards are critical with increasing pressure to ensure that cloud technology investments remain viable for years to come. Standards allow CIOs to select products that best suit their needs today–regardless of vendor–while helping to ensure that no proprietary constraints arise when new systems are put in place in the future.

The goal of any Open Cloud Standards is to enable portability and interoperability between private clouds within enterprises, hosted private clouds or public cloud services.



The prime mover of cloud standards is the **Cloud Security Alliance** (CSA), which includes CSA Security Guidance for Cloud Computing, CSA Cloud Controls and CSA Top Threats.[82] There is also a CloudAudit Group associated with CSA. It is a volunteer cross-industry effort on cloud, networking, security, audit, assurance and architecture backgrounds.

The **Distributed Management Task Force** (DMTF) has published the **Open Virtualization Format** (OVF), which is a defined standard for the portability and deployment of virtual appliances. OVF enables simplified deployment of virtual appliances across multiple virtualization platforms. A virtual appliance is a virtual machine image designed to run on any virtualization platform. Virtual appliances are a subset of the broader class of software appliances, which eliminate the installation, configuration and maintenance costs associated with running complex software.

OVF is a common packaging format for independent software vendors (ISVs) to package and securely distribute virtual appliances, enabling cross-platform portability. By packaging virtual appliances in OVF, ISVs can create a single, pre-packaged appliance that can run on customers' virtualization platforms of choice.

The **International Organization for Standardization** (ISO) manages the **Study Group on Cloud Computing (SGCC**), which provide a taxonomy and terminology Cloud Computing and assesses the current state of standardization in Cloud Computing.

[82] https://cloudsecurityalliance.org/

The **European Telecommunications Standards Institute** (ETSI) has formed a **TC Cloud Interest Group** to help with the formulation of standards for cloud computing. So far its work is based entirely on US standards bodies.

The **National Institute of Standards and Technology** (NIST) role in cloud computing is to promote the effective and secure use of the technology within government and industry by providing technical guidance and promoting standards. NIST's work to date has focused on the definition and taxonomy of cloud computing. Such definitions serve as a foundation for our upcoming publications on cloud models, architectures, and deployment strategies. So far no formal software protocols have been issued.

The **Open Grid Forum** (OGF) has organized the **Open Cloud Computing Interface Group** (OCCI) defines general-purpose specifications for cloud-based interactions. The current OCCI specifications consist of three procedural documents.

The **Object Management Group** (OMG) has formed the **Open Cloud Consortium** (OCC) which has generated set of scripts which generate large, distributed data sets suitable for testing and benchmarking software designed to perform parallel processing on large data sets.

The **Organization for the Advancement of Structure Information Standards** (OASIS) has formed the **OASIS Cloud-Specific Technical Committees,** which address the security issues posed by identity management in cloud computing.

The **Storage Networking Industry Associations** (SNIA) has created the **Cloud Storage Technical Work Group** for the purpose of developing SNIA Architecture related to system implementations of Cloud Storage technology.

The **Open Group** has formed the **Cloud Work Group** to create a common understanding among buyers and suppliers of how enterprises of all sizes and scales of operation can include Cloud Computing technology in their architectures to realize its benefits.

The **TM Forum** is an industry association focused on enabling IT services for operators in the communications, media and cloud service markets. It operates the **TM Forum's Cloud Services Initiative** to promote the use of cloud standards**.**

The Institute of Electrical and Electronics Engineers (IEEE) is in the process of organizing the IEEE P2301 group that will produce a guide for cloud portability and interoperability profiles.

SUMMARY

I have counted over twenty organizations involved with the standardization of how to manage cloud computing. So far, only one formal standard that addressed virtualization methods has been published, though not widely adopted.

It is clear that the various standards organizations are lagging far behind the actual work done by cloud vendors to buttress their competitive positions as proprietary (or hard to dislodge) suppliers. From the standpoint of enabling portability of applications across various cloud service offerings we do not find any evidence that vendors would be relying on standards bodies to make customers interoperable.

As was the case with Microsoft thirty years ago, interoperability of the cloud environment will not be achieved by means of public standards. It will be delivered by means of adoptions that will be implemented by a few of the strongest software vendors.

Open Source Platform

Transferring applications to a cloud offers enormous cost reductions. It also can be a trap. Placing applications on an Infrastructure-as-a-Service (IaaS) requires an elaborate process to provision computing power (servers, files, operating systems). After placing an application on IaaS it becomes wedged into a unique software environment. For all practical purposes applications cease to be transportable from one IaaS to another IaaS. The most widely used IaaS, the Amazon Web Services operates in this mode. There are hundreds of other cloud services that also operate in this manner. IaaS is useful in offering raw computing power but it is not flexible how it can be redeployed when conditions change.

Applications can be also placed in a Platform-as-a-Service (PaaS) cloud. All you have to do is to comply with the specific Application Interface (API) instructions and your application will run. Google, Microsoft Azure, a version of Amazon PaaS as well as other cloud services work in this way. After applications are placed in a particular cloud environment they must comply with a long list of required formats. For instance, various PaaS vendors may limit what software "frameworks" can be applied. Such "frameworks" include reusable libraries of subsystem offered by software tools such as Ruby, Java, Node.js and Grails. Some PaaS vendors may also restrict what operating systems (such as which version of Microsoft OS) can be installed. Consequently, PaaS applications will not be always transportable from one cloud vendor to another.

To support the future growth in cloud computing customers must be able to switch from one cloud vendor to another. What follows is restricted to only PaaS cases. This requires that cloud operators must offer the following features:

1. The interface between customer applications and the PaaS must be in the form of Open Source middleware, which complies with approved IEEE standards. Standard Open Source middleware will allows any application to run on any vendors' PaaS cloud. Regardless how an application was coded it will remain transportable to any cloud, anywhere.

2. The isolation of the customer's applications from the PaaS software and hardware will permit the retention of the customers' intellectual property right, regardless of which cloud it may be hosted.

3. Certification by the cloud vendor to customers that that applications will remain portable regardless of configuration changes made to PaaS. This includes assurances that applications will retain the capacity for fail-over hosting by another PaaS vendor.

4. Assurance that the customers' application code will not be altered in the PaaS cloud, regardless of the software framework used the build it.

This week, VMware introduced a new PaaS software offering called Cloud Foundry. It is available as open source software. It provides a platform for building, deploying and running cloud apps that make it possible for cloud vendors to comply with the four features listed above. Cloud Foundry is an application platform, which includes a self-service application execution engine, an automation engine for application deployment, a scriptable command line interface and development tools that ease the applications deployment processes.

Cloud Foundry offers developers the tools to build out applications on public clouds, private clouds and anyplace else, whether the underlying server runs VMware or not.

Cloud Foundry is the first open PaaS that supports services to cloud firms such as Rackspace or Terremark. Cloud Foundry can be also deployed behind firewalls for enterprises can run this software as a private cloud. There is also a version of Cloud Foundry, the "Micro Cloud", which can be installed on a

personal lap top so developers can write code themselves, and then push to whichever cloud they choose. "Micro Cloud" should be therefore understood as a single developer instance of Cloud Foundry.

Cloud Foundry aims to allow developers to remove the cost and complexity of configuring infrastructure and runtime environments for applications so that they can focus on the application logic. Cloud Foundry streamlines the development, delivery and operations of modern applications, enhancing the ability of developers to deploy, run and scale applications into the cloud environment while preserving the widest choice of public and private clouds.

The objective is to get an application deployed without becoming engaged in all kinds of set-ups, such as server provisioning, specifying database parameters, inserting middleware and then testing that it's all set up after coordinating with the data center operating personnel to accept new run-time documentation. The Cloud Foundry offers an open architecture to handle choices of developer frameworks. It accommodates choices of application infrastructure services. It enables the choosing from a variety of commercially available clouds.

Cloud Foundry overcomes limitations found in today's PaaS solutions. Present PaaS offerings by commercial firms are held back by limited or non-standard support of development frameworks, by a lack in the variety of application services and especially in the inability to deploy applications across diverse public and private clouds.

SUMMARY

It is increasingly a prerequisite for modern software development technologies to be available as open source. DoD memorandum of October 16, 2009 offers guidance a preferred use of open source software in order to allow developers to inspect, evaluate and modify the software based on their own needs, as well as avoid the risk of lock-in. Cloud Foundry is now an open source project with a community and source code available on www.cloudfoundry.org. This provides the ultimate in extensibility and allows the community to extend and integrate Cloud Foundry with any framework, application service or infrastructure cloud. It includes a liberal licensing model encourages a broad-based community of contributors.

Cloud Foundry takes an Open Source approach to PaaS. Most of such vendor offerings restrict developer choices of frameworks, application infrastructure services and deployment clouds. The open and extensible nature of Cloud Foundry means developers will not be locked into a single framework, single set of application services or a single cloud. VMware will offer Cloud Foundry as a paid, supported product for customers as well as provide the underlying code so developers can build their own private clouds. VMware will also offer Cloud Foundry as a PaaS service in combination with a recently acquired data center in Las Vegas that presently runs data back-up services for over million customers.

Cloud Foundry allows developers to focus on applications, not machines or middleware. Traditional application deployments require developers to configure and patch systems, maintain middleware and worry about network topologies. Cloud Foundry allows you to focus on your application, not infrastructure, and deploy and scale applications in seconds. In the future interoperability of applications across several PaaS firms will matter to more and more companies especially those starting new systems. Flexibility to choose from a range of available PaaS service will become one of the most compelling factors behind the choice of trusting any one firm with the custody of its data processing in a cloud environment.

Any DoD plans to migrate systems into a PaaS environment will henceforth have to consider whether Cloud Foundry software, or a similar offering yet to come, will be in place to assure application portability.

PART 13: CYBER ISSUES

DoD Social Media Policy

The Defense Department has just reauthorized, for another year, the social media guidelines. [Directive-Type Memorandum (DTM) 09-026]. Accordingly, the NIPRNET will continue to be configured for easy access to insecure Internet-based offerings for several millions of computing devices.

This will include access to social media such as YouTube, Facebook, MySpace, Twitter and Google Apps. The DTM states that DoD commanders and Agency heads will continue defending their computers against all malicious activity.

Without prescribing how malware defenses will be applied there is a question how effective is a generic DTM, which allows the widespread use of social media, but without specific guidelines how to defend the networks.

The widespread use of social media cannot be stopped or curtailed any more. In remote locations and on long rotations, the network time spent on social media can exceed the traffic for conducting DoD business operations. For troop morale the free access to social media is a necessity.

Without a defined policy how to assure security, social media will continue to make the DoD networks insecure. To demonstrate this vulnerability we will use the most pervasive social media, Facebook, to illustrate what are the dangers to NIPRNET.

According to data from security company BitDefender, there's harmful content behind about 20 percent of posts on Facebook news feeds. BitDefender said about 60 percent of attacks on Facebook stem from threatening third-party apps.[83] Most of the infectious software originates from thousands of independent developers who often sell such software for a fee. By clicking on infected links users risk having all sorts of viruses downloaded to their computers.[84]

People who are tweeting can install from their friends Facebook accounts a variety of bots. These bots have access to all of the data of anyone connected to a hacked account. Facebook accounts can then be linked with more people in a social circle - opening up new opportunities for identity fraudsters to launch further attacks.[85]

---

[83] http://www.pcmag.com/article2/0,2817,2373281,00.asp

[84] http://www.bbc.co.uk/news/technology-11827856

[85] http://blogs.computerworld.com/17418/security_warnings_whether_or_not_you_plan_to_drink_and_drive_a_keyboard_this_weekend

In late October, a particularly malicious piece of malware called Koobface resurfaced on Facebook. Like the original strain of the Koobface virus is spread via Facebook messages. The messages usually have clickable topic lines like "Is this you in the video?" or something similar. When a user clicks on such message, they are brought to a third party site where a link is waiting. Open the link and their computer will turn into a zombie that can be commanded to execute more damaging procedures.

SUMMARY

With hundreds of data centers and thousands of servers the attacks transmitted through social media cannot be stopped any more. What is required now is a policy that dictates the technical means for isolating such attacks.

Social media transactions should be completely isolated and segregated. User displays should be partitioned to communicate all public Internet traffic exclusively with dedicated severs. In this way infected communications will be shuttled into partitions from where a further propagation of malware will not affect the conduct of DoD operations. However, such solutions will require a major overhaul how networks are organized.

The use of social media by DoD personnel cannot be stopped. What is needed is an architecture that will allow the separation of the insecure from the secure environment for an assured safeguarding security.

Anomalies in Social Computing

According to the National Security Agency, in 1928, Secretary of State Henry Stimson, closed down the Department's intelligence bureau. His rationale was that "Gentlemen do not read other gentlemen's mail."

We have now a comparable situation in the Department of Defense. New policies and guidance have been issued that declare, in effect, that well-behaved gentlemen and gentlewomen should abstain from reading potentially toxic attachments to social computing messages.

Such policies and guidance do not promote the security of defense networks and should be therefore modified.

THE DEPUTY SECRETARY OF DEFENSE MEMORANDUM

The Deputy Secretary of Defense issued a policy for guiding the uses of Social Networking Services in a Directive-type Memorandum of February 25, 2010. The memorandum acknowledges that "… SNS capabilities as integral to operations across the Department of Defense using the Non-Classified Internet Protocol Router Network (NIPRNET)." There are at least five million computing devices connected to the Department of Defense networks.

This policy is deficient in that it does not address the danger of allowing access to web services, such as social computing, that can insert malicious software attachments to any message. Such insertions from the Internet, if opened, can then compromise the security of computing devices on numerous networks.

The DEPSECDEF generic policy states that: "commanders shall defend against malicious activity" and "commanders shall deny access to sites with prohibited content, such as pornography, gambling, hate

crime activities." Unfortunately, none of this can be executed with the existing manpower. It cannot be enforced using the available technical means.

Browsers exist in every personal computer. They can connect to millions of web pages without anyone in the DoD having the capacity to restrict access to every potential source of malware. Without enforcement there will be always web pages from where a military or civilian person can download computer code that subsequently trigger attacks that can be launched from the inside of the NIPRNET.

Even with firewall and anti-virus protection, which is always imperfect, there will always be web pages capable of delivering malware to DoD. This is because the malware will always be technically superior to any institutional defenses, which are administered by overworked, understaffed and under-resourced personnel. Therefore DoD cannot and should not depend on blocking of known sites and certainly not on malware protection safeguards managed by error-prone people.

The Air Force Public Affairs Agency Guidance

In November 2009, the Air Force Public Affairs Agency released Version 2 of the guidance for using LinkedIn, YouTube, Flickr, Facebook, MySpace, and other social media sites.

The Air Force offers rules for gentlemanly conduct in posting social media entries: Do not post classified information; Replace all errors; Readily admit mistakes; Use best judgment in whatever your post; Avoid offensive language; Abstain from violation of privacy; Never, but never lie.

The problem with the Air Force guidelines is that they do not acknowledge the danger of picking up code that is toxic. Although an attachment may appear to be harmless, it can contain harmful code. A click will unpack a hidden program that can be lodged where it can do the greatest harm either immediately or eventually whenever it becomes unleashed.

Clever "social engineering" of incoming messages will aggravate such perils. Social media reveal much information about sources. Private information makes it possible for an attacker to construct a plausible message that will be opened without further examination.

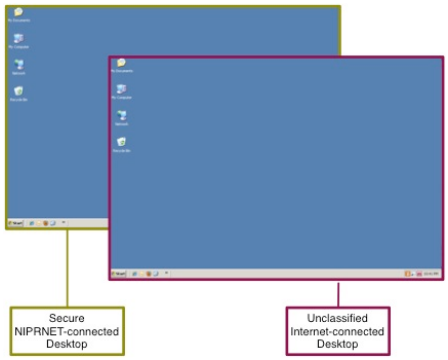
The existing DoD policies that promote the use of social media may continue, but must also include enhancements that provide for the complete separation of secured NIPRNET desktops from the capacity to access the unprotected Internet without acceptable restrictions.

Offering to the military and to the civilians separate but different desktops, displayed on an identical computing device by means of virtualization is now feasible and represents mature commercial practices. This approach is also affordable, especially in the case of thin clients where such approach offers opportunities for achieving quick as well as major cost reductions.

There is no reason why the existing DoD policies should not be revised through the introduction of more advanced technical means that will manage automatically how the general access to social computing can be achieved with assured safety.

Thin Client Case

A person with a ".mil" address walks up to a thin client anywhere in the world and logs in to the DoD NIPRNET "Secure Desktop" using a Public Key Infrastructure (PKI) access card, plus biometric ID. A thin client then presents a menu of available virtual computers to connect to. The choices will include secure NIPRNET-connected desktops, as well as insecure desktops connected to the Internet, as illustrated below:



One can choose more than one of the available options, keeping them open in multiple windows and switching among them. Each of the windows can run in an overlapping mode, or take over the whole screen. Alternating between windows does not require the rebooting the computing device.

It is not possible to transfer files from a "Secure Desktop" to the "Publicly Connected Desktop." It is not possible to cut and paste from a secure window to a public Internet window. Data transfer is limited to keyboard entries and to mouse movements.

An insecure device such as a digital camera or a "thumb drive" can be connected to the thin client's USB port. However, this port is only active when the "Publicly Connected Desktop" is in the foreground.

All communications from a "Publicly Connected Desktop" pass through a separate security gateway, where they are automatically inspected for policy compliance and will be logged in (for compliance with records management policies) for further examination. Accesses to all physical media (hard disks, flash drives, CD/DVDs or USB ports) are disabled meanwhile although authentication can be obtained from a Network Control Center in exceptional cases. When the desktop is switched from the "Publicly Connected Desktop" back to the "Secure Desktop" the USB port that communicates directly with the Internet is deactivated and reverts to security compliance that is governed by DoD policies.

The servers that run the "Secure" or "Publicly Connected" desktops do not ever combine secure and insecure Virtual Computers into a pool, since servers with different levels of security are always isolated. This can be accomplished by resorting to the use of "hypervisors" that can separate secure and insecure desktops on the same physical server. Hypervisors are good at creating a secure isolation of applications and operating systems from the underlying "bare metal" microprocessors. Hypervisors are the means for achieving the "virtualization" of servers and for delivery of a high level of security assurance such as guaranteeing anti-malware protection.

DESKTOP, LAPTOP AND NETBOOK CASES

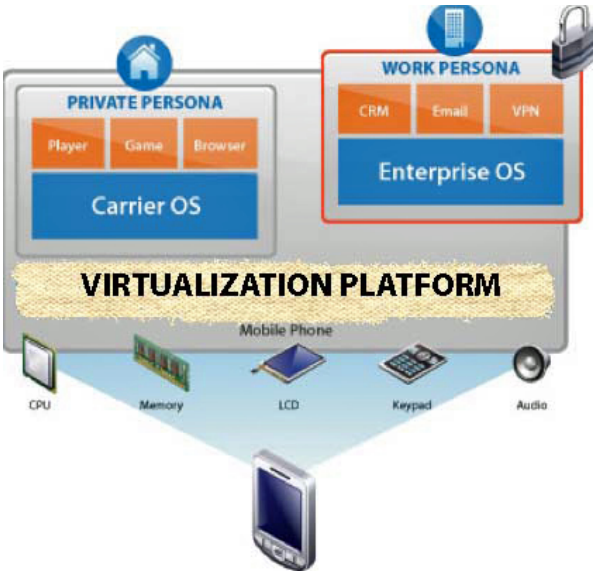
A person with a ".mil" address has a personally issued device such as a desktop, laptop or smartphone and logs in to the "Secure Desktop" using a PKI access card, plus biometric ID. The computer boots either into a secure Operating System or to a secure hypervisor, which have been hardened against tampering and eavesdropping. For instance, the hypervisor ensures that the disk images of protected Virtual Computers are encrypted and that there are no means of transferring data into or out of the protected environment.

After that the procedure is identical with Case 1 above, although some of the Virtual Computer logic will be stored locally and some parts will be stored on DoD servers in DoD data centers in order to improve response time.

Downloads from public Internet sites remain on the "Publicly Connected Desktop" but cannot be extracted or copied into the secure desktop. When the "Secure Desktop" is in the foreground, all input/output actions are restricted by NIPRNET security policies.

Whenever connecting through the "Publicly Connected Desktop" its settings are either reset to a like-new condition or can be refreshed according to practices managed by a Network Control Center.

To accomplish the separation between the Private Personal desktops and the Work Personal desktops calls for the placement of isolated logical windows on top of the Virtualization Platforms both at the desktop devices as well as at servers located at the data centers. By far the most secure and least expensive way of achieving this it by resorting to the use of thin clients for social computing:

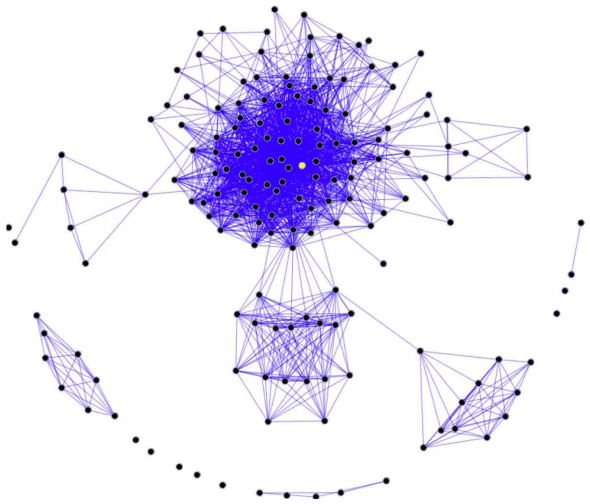


Though DoD work is protected against security intrusions because any virus or botnet conveyed over the Internet lands up in a completely isolated virtual server, the outbound traffic from the Private Personal computers is open to compromise from unauthorized disclosures. In the case of potential security compromises from insiders, DoD remains completely unprotected.

Giving access to social computing therefore calls for the complete tracking of all transactions. Such monitoring must account for every social computing message. Forensic methods can be then used to identify incidents for the apprehension and ultimately as evidence used for the prosecution of security violators.

The monitoring of social computing messages will be taking place at network control centers equipped with automated software that would reduce the workload on the surveillance staffs. Peak load transactions of social computing (including reserves, contractors and dependents) could approach 100,000 messages per hour. Without massive automation combined with a security schema that permits the correlation of message patterns over an extended time period the monitoring of social computing is not manageable.

There are a number of forensic tools available identify security anomalies, such as shown below. The isolated connections would receive the highest priority for added surveillance.



SUMMARY

The authorization of even restricted social computing access to the toxic Internet, without interrupted monitoring, is a risk that should not be tolerated.

Social Computing is Asymmetric

SOURCE: http://www.afcea.org/signal/signalscape/index.php/2010/07/14/7490/

The Defense Department is spending $3.2 billion/year on information technology to secure networks against incoming malware. Meanwhile, it spent hardly any money to protect against outgoing compromising data from insiders. Nobody seems to care much about the prevention of exfiltration of information.

Time has come to recognize that cyber security has to deal with unequal amounts of inbound and outbound traffic. Our enemies can gain more credible information from easily available disclosures from inside sources than from encrypted data that must be mined through firewalls, virus protection and filtering. That is why the imbalance between the expensive defenses against incoming intrusions vs. the puny amounts spend to deter outgoing leaks can be labeled as asymmetric.

By far the greatest source of information leakage from the Defense Department is via social computing such as through YouTube, Facebook, MySpace Twitter and blogs. The OSD policy on social networking of February 25, 2010 makes such activity "integral to operations across DoD". It orders the re-configuration of the NIPRNET to provide access to Internet-based capabilities from all components. But the "how" of implementing that was left without any guidance on how to arrest the revealing of military information. In short, the current OSD policy has opened the gates to the loss of intelligence to close a billion people now engaged in social computing. A well-informed source tells me that about 20 percent of all Defense traffic is in conducting social communications through public sites, which are unprotected as well as potentially toxic.

A recent incident demonstrated that outsiders could use the social media to extract DoD information. A phony "Robin Sage", easily masquerading as an employee of the Naval Network Warfare command, was able to accumulate in a few months 300 friends on LinkedIn, 110 on Facebook and had 141 followers on Twitter. She connected with the Joint Chiefs of Staff, the CIO of the NSA, an intelligence director for the

U.S. Marines and the a chief of staff for the U.S. House of Representatives. In all communications there were clues that "Robin" was a fake. In one case "Robin" duped an Army Ranger into "friending" her. The Ranger inadvertently exposed information about his coordinates in Afghanistan with uploaded photos from the field that contained GeoIP data.

Here is another case of disregarding elementary security, which disregarded the asymmetric effects of cyber security. It is a case in which I was involved. A bank's currency trading system was very secure. In its operations it followed best practices and was often praised as an exemplar of good risk management. All of the money transfers—sometimes in hundreds of millions of dollars in a matter of an hour—were securely executed without ever having a problem. The computers, the data center and the transmission lines were locked-down securely.

Yet, suddenly, there was a problem: A large sum of money ($80 million) disappeared in a matter of seconds. When we finally walked through all of the scenarios, the problem was that although the currency applications were absolutely secure, the maintenance programmers (who were supporting money transfer applications) were communicating by open e-mail about software fixes and the next software release. The e-mails were mostly about project management housekeeping, such as when you run the tests and when you do a software update. The e-mails therefore flagged when the money systems were most vulnerable. By keeping track of the programmers' chatter over e-mail the attackers knew exactly when, for a few seconds, the system was naked.

When verifying cyber security, the number one rule is that the attackers will first devote their time not on attacking a target directly. Devoting efforts to seek out locations of maximum vulnerability will always take precedence. Therefore, I favor managing social media on the NIPRNET against potential exfiltration as a priority. Unchecked outgoing traffic will always leave military information vulnerable.

Method for Assuring Social Computing

The total population of Internet users is 1.6 billion. The majority of users engage in social computing where there are numerous on-line services that offer opportunities for sharing information. At present there are 156 social computing sites but their number is growing to meet increasingly diverse interests. Sites with more than 15 million registered users include Digg, FriendFinder, Facebook, Flixster, Flickr, Friendster, Habbo, LinkedIn, MyLife, MySpace, Orkut, Plaxo, Twitter, YouTube, UStream and Wiki. These services have a total membership of 1.4 billion as of September 2009.[86]

Military and civilian personnel are now relying on social media for personal communications as well as for sharing information that covers social and military topics. The problem with these services lies in their vulnerability to security breaches. Social networks rely exclusively on the public Internet, which was conceived forty years ago as an insecure network for academics. There is nothing that DoD can do about that except by elaborate and costly protective overlays that guard over half a billion daily transactions. Every social network exposes DoD to hostile attacks that can be used as a conduit for toxic software that can infect every computer device.

[86] http://en.wikipedia.org/wiki/List_of_social_networking_websites

When the Marine Corps confronted the prospect of persistent cyber corruption via social media its only recourse was to prohibit their uses. Though such resolute action is commendable this prohibition is not enforceable. Meanwhile, all other DoD components remain aware of the risks of continuing with the rising dependence on social computing. Studies have been launched even though there is universal acknowledgement that the dependence on social computing cannot be stopped.

DoD policy makers have now three options:

1. Accept all of the risks that are endemic to social computing by hoping that prophylactic measures such as virus protection software and firewalls will limit the damage. Such expectations are unrealistic. The increasing capabilities of attackers cannot be overcome. Our enemies will continue tracking social communications to extract intelligence about operations.

2. Do not accept the risks of social computing except where social computing can be practiced on isolated computers that use circuits for dedicated access to the public Internet. This is a workable but expensive and hard to manage solution, especially on ships or in the battlefield. It will not deal with the problems of intelligence leakage. It will also deprive DoD of the value that social computing brings to the contemporary military culture.

3. Alter the current designs of systems so that social computing becomes a controlled and integral part of all communications. This option can be secure, enforceable and less expensive provided that DoD is ready to change the architecture of its systems.

What follows are cases favoring the adoption of a revised approach to social computing using technical means for achieving information security.

SOCIAL COMPUTING BASE CASE

A user walks up to any "thin" or "zero" client and logs in using approved authentication. The "thin" client does not have a disk drive and uses only a browser to communicate with the server that houses the virtual desktops. A "zero" client does not have a microprocessor, does not have a disk drive, has no software, has no drivers and cannot be patched. It is therefore totally secure, since it acts only as a frame buffer.

The "thin" and "zero" clients obtain menus of virtual desktops from virtual servers on redundant data centers with zero downtime and with Google-like latency. These menus include access to social computing.

The soldier or sailor can choose any of the virtual desktop options, keeping them open in multiple windows. The windows are labeled to indicate the security of each site. Server security will prohibit cutting and pasting from a social computing to a secure desktop.

Users can connect devices such as iPods or memory cards to a USB port, but the port is only active when a desktop is activated. This makes it possible to upload personal photos or movies to social networks. Whenever a user switches to a secure virtual desktop, the USB port is deactivated. The net result is that each soldier owns a fully portable as well as a totally isolated open access virtual computer for personal use.

The hosting of servers that support social networking should be a service that is built into the DoD's Global Information Grid (GIG). A library of standardized templates is then accessible. These templates are reset to a like-new condition at the beginning of each session in order to preserve security. Virtual desktops based on these templates may also be issued to a solider, and maintained as personal workstations without being erased between session.

Every communication from virtual client must include a personal certificate of authenticity and must pass through security gateways, which log transactions into permanent and de-duplicated storage. Suspicious records then become available for forensic examination by security personnel using business intelligence software. All transaction, and particularly those coming from already tagged Internet connections will be automatically inspected at one of the Network Control Centers (NOC) or an Anti-terrorist Unit by a battery of semantic filters that scan for possible discrepancies. Any compromises of security restrictions or any detected anomaly in text is flagged while communications are either cut off or passed on for monitoring at higher levels of security.

Conventional anti-malware software runs within the desktop's OS and can be compromised by user activity. Virtual desktops run on a server, which is capable of monitoring the desktops for malware from a privileged process, which cannot be compromised by any user activity. Any unauthorized modification to the OS kernel running any virtual desktop immediately halts execution of that desktop. This gives malware protection that is much more effective, more universally applied, and easier to manage.

Virtual desktops can be transferred between physical desktops, laptops or smartphones or between one datacenter and another at a different location. When a virtual desktop is transferred an expiration time limit is issued which allows the virtual desktop to be used offline, such as in the case of air travel or on a combat mission. If the time limit expires the virtual desktop is scrambled. While a virtual desktop is checked out it is periodically synchronized with the copy on the server so it can be completely recovered in case of loss.

SOCIAL COMPUTING LEGACY CASES

Migration to the social computing base case will take many years to accomplish. There are applications as well as computers that are already in place and that must continue with legacy programs until they can be fitted into the new architecture. Meanwhile, the safeguarding of connections to the public Internet requires protective measures. This cannot be delayed and therefore desktop virtualization must proceed immediately.

Potential cost reductions should steer the priority with which the migration of legacy systems is planned. The largest payback comes from the conversion and then from the consolidation of servers into redundant "private clouds". Whether that is accomplished through government owned facilities or by renting the cloud infrastructures on a per transaction basis from several cloud vendors is a matter of the speed how fast cost reductions can be delivered. Such choices will also require an ability to relocate the virtual servers across the clouds for the interoperable pooling of assets to assure fallback for failure-proof continuity of operations.

The placement of all applications into dedicated Internet servers consolidates desktops from legacy computers. Bug fixes, software updates and virus protection can be now administered in a uniform way throughout the network. Centralization of controls makes possible how communications to and from social networks is performed. It reduces the cost of client upgrades since social applications will be sharing vastly more economical pools of processing, disk storage, communication and energy consumption services. After a persons' Internet connection is placed under the control of a hypervisor (software that allows multiple operating systems to run on a host computer concurrently) it can be relocated regardless where a person is located geographically or organizationally.

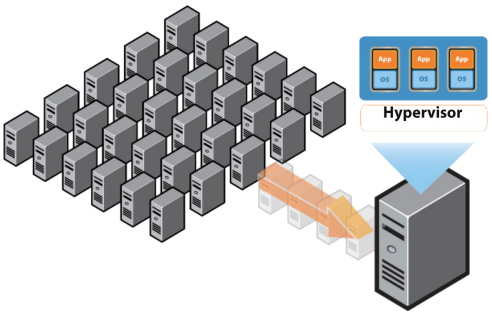


Figure 1 – Consolidation into Virtual Servers

The placement of social computing under the direct control of a virtual desktop will dispense with the dependency on proprietary versions of operating systems (OS). In many cases, a soldier needs just an ordinary web browser and nothing else. Social computing can then rely on a fraction of the OS code, thus reducing licensing costs as well as eliminating labor-intensive support. After clients for social applications are placed on a consolidated virtual computer, they can be further hardened against tampering. The National Security Agency (NSA) has shown that the bloated OSs, which must support an excessive number of features and functions, are the largest targets for software attacks.[87]

CONCLUDING OBSERVATIONS

There is nothing to prevent soldiers from engaging in social computing on public networks using their own computers while paying for their own access privileges. Personal privacy allows that though there are cumbersome regulatory, legal and other restrictions that limit such uses. As costs of client devices drop the soldiers' personal social computing will come to rely on network connections purchased from commercial information services providers. In such cases messages will have to carry an address, which is different from .mil. Such connectivity will be prohibited at sensitive military sites and therefore social computing will have to rely exclusively on circuits provided by DoD.

The DoD has the unquestionable authority to control all Internet transactions that originate from every .mil Internet Protocol (IP) address. A recent pronouncement from the Office of the Secretary of Defense to pursue a "balanced" approach to social computing is inconclusive. It leaves huge gaps for tens of thousands of transactions a day that can bypass all protective measures.

DoD must take decisive remedial steps to achieve positive controls over all social computing transactions originating from the toxic Internet. The risks are too great to accept insufficient safeguards.

New Roles for CIOs

On August 8, 2011 the Director of OMB issued a memorandum for the purpose of enlarging the roles of the government's Chief Information Officers. Its objective is to change the roles of Agency level CIOs from just policymaking portfolio management for all IT.

[87] http://cs.unomaha.edu/~stanw/papers/csci8920/losco

Does the OMB memorandum change the roles of the CIO?

The OMB memorandum adds to the CIO responsibilities, as defined by the Clinger-Cohen act of 1996, a recommendations that CIOs should work with the Chief Financial Officers and Chief Acquisition Officers as well as with the Investment Review Boards (IRBs). Such coordination should have the goal of terminating one third of all underperforming IT investments by June 2012. Though this objective useful, it cannot be construed as an enlargement of the CIOs role as portfolio manager for all of IT. The job of eliminating underperforming systems was always one of the principal CIO tasks.

The OMB memorandum adds to the CIO responsibilities the mission of managing "commodity IT". CIOs are advised to pool agency purchasing power to improve the use of commodity IT. For instance this concerns dealing with e-mail, collaboration tools, human resources or administration. To achieve that, CIOs should rely on "enterprise architectures" and the use shared commercial services instead of standing up separate services. Although these recommendations are commendable, the government does not have an enterprise architectural design in place as yet. It has been unsuccessful in organizing efforts in which commercial shared services are used in the government. Though efforts have been made to organize pooled e-mail service in the Army, Congress has denied such funding. In the absence of establishing methods for pooling "commodity IT" funds, this enlargement is not executable.

The OMB memorandum adds to the CIO responsibilities a "program management" mission. This is largely as a personnel management function. In the absence of administrative rules it is not apparent how a CIO, without authority, can conduct annual performance reviews of component CIOs. He cannot be accountable for the performance of IT program managers, especially where such personnel reports to Acquisition officers. There is no way how CIOs can carry out the OMB dictated "program management" responsibilities to enlarge their authority.

The OMB memorandum adds to the CIO the primary responsibility for implementation of information security programs that support the assets and missions of an agency. Such authority is subject to an examination of implementation in "CyberStat" sessions conducted by the Department of Homeland security. In the absence of a qualified staff or funding needed to carry out such responsibility this enlargement of CIO responsibilities is lacking an understanding how security responsibilities are managed in agencies such as DoD, which accounts for more than a half of total government IT spending. The proposed enlargement in security matters not sufficiently explained to be credible.

The OMB memorandum requires the CIOs to participate in cross-agency portfolio management through the Federal CIO Council (CIOC). The objective would be to reduce duplication of IT spending across agency boundaries. In the absence of changes in the budgeting processes it is not clear how the CIOC can take actions that would pool agency funds into a multi-agency program. The CIOC is a committee without fiscal power. It cannot be seen as the basis for the enlargement of the powers of a CIO.

SUMMARY

Memorandum M-11-29 from the Director of OMB makes an attempt to increase the power of the federal CIOs. However, the memorandum lacks substance. Other than increasing the coordination between CIOs there is no evidence that the powers of the CIO would change in any way.