The 2012
Cloud Networking Report

Part 2: Data Center LANs

By Dr. Jim Metzler
Ashton Metzler & Associates
Distinguished Research Fellow and Co-Founder
Webtorials Analyst Division

Platinum Sponsors:

cisco

ipanema Technologies

NETSCOUT.

Gold Sponsors:

A10 Networks

Alcatel·Lucent Enterprise

Aryaka

Blue Coat

agility made possible™ ca technologies

certeon Accelerate Your Business

crossbeam

NEC Empowered by Innovation

radware

VISUAL NETWORK SYSTEMS

Produced by:

Webtorials

Executive Summary

The **2012 Cloud Networking Report** (The Report) will be published both in its entirety and in a serial fashion. This is the second of the serial publications. The first publication in the series described the changes that are occurring in terms of how cloud computing is being adopted, with a focus on how those changes are impacting networking. The topics of the subsequent publications of The Report are:

- Software Defined Networks
- Wide Area Networking
- Management

The Report will also be published in its entirety and there will be a separate executive summary that covers the totality of The Report.

One goal of this publication is to provide a very brief overview of how data center LAN technology and design has evolved and to identify the factors that are currently driving the vast majority of IT organizations to rethink how they design their data center LANs. Another goal of this publication is to provide insight into the technologies and design choices that IT organizations are making. The third and primary goal of this publication is to describe the data center LAN architecture and technology options that either are currently available in the market or are likely to be available within two years.

Given the breadth of fundamental technology changes that are impacting the data center LAN, this section is very technical.

The Emerging Data Center LAN

First and Second Generation Data Center LANs

As recently as the mid 1990s Local Area Networks (LANs) were based on shared media. Throughout this report these shared media LANs will be referred to as First Generation LANs. In the mid 1990s, companies such as Grand Junction introduced Ethernet LAN switches to the marketplace. The two primary factors that drove the deployment of Second Generation LANs based on switched Ethernet were performance and cost. For example, performance drove the deployment of switched Ethernet LANs in data centers because FDDI, which was the only viable, high-speed First Generation LAN technology, was limited to 100 Mbps whereas there was a clear path for Ethernet to evolve to continually higher speeds. Cost was also a factor that drove the deployment of Ethernet LANs in data centers because FDDI was fundamentally a very expensive technology.

A key characteristic of Second Generation data center LANs is that they are usually designed around a three-tier switched architecture comprised of access, distribution and core switches. The deployment of Second Generation LANs is also characterized by:

- The use of the spanning tree protocol at the link layer to ensure a loop-free topology.
- Relatively unintelligent access switches that did not support tight centralized control.
- The use of Ethernet on a best-effort basis by which packets may be dropped when the network is busy.
- Support for applications that are neither bandwidth intensive nor sensitive to latency.
- Switches with relatively low port densities.
- High over-subscription rate on uplinks.
- The separation of the data network from the storage network.
- VLANs to control broadcast domains and to implement policy.
- The need to primarily support client server traffic; a.k.a., north-south traffic.
- Redundant links to increase availability.
- Access Control Lists (ACLs) for rudimentary security.
- The application of policy (QoS settings, ACLs) based on physical ports.

Drivers of Change

One of the key factors driving IT organizations to redesign their data center LANs is the requirement to support the growing deployment of virtual servers. With that in mind, The Survey Respondents were asked to indicate the percentage of their company's data center servers that have either already been virtualized or that they expected would be virtualized within the next year. Their responses are shown in **Table 1**.

Table 1: Deployment of Virtualized Servers *N = 112*

	None	1% to 25%	26% to 50%	51% to 75%	76% to 100%
Have already been virtualized	18%	30%	25%	16%	11%
Expect to be virtualized within a year	11%	28%	24%	25%	12%

The way to read the data in **Table 1** is that in the current environment only 18% of IT organizations have not virtualized any data center servers and that within a year, that only 11% of IT organizations will not have virtualized any of their data center servers.

As pointed out in Virtualization: Benefits, Challenges and Solutions[1], server virtualization creates a number of challenges for the data center LAN. One of these challenges is the requirement to manually configure parameters such as QoS settings and ACLs in order to support the dynamic movement of VMs. In order to quantify the extent to which IT organizations move VMs between physical servers, The Survey Respondents were asked to indicate whether they agreed or disagreed with the statements in the left hand column of **Table 2**.

Table 2: Movement of VMs *N = 265*

	Agree	Disagree
We currently manually migrate VMs between servers in the same data center	66%	34%
We currently automatically migrate VMs between servers in the same data center	55%	45%
We currently manually migrate VMs between servers in disparate data centers	48%	52%
We currently automatically migrate VMs between servers in disparate data centers	26%	74%

The data in **Table 2** indicates the great interest that IT organizations have in moving VMs between physical servers. However, as will be described throughout this section of the report, moving VMs between physical servers can be very complex.

[1] http://www.webtorials.com/content/2010/06/virtualization.html

Manually configuring parameters such as QoS settings and ACLs in order to support the dynamic movement of VMs is not the only challenge that is associated with server virtualization. Other challenges include:

- **Contentious Management of the vSwitch**
  Each virtualized server includes at least one software-based virtual switch (vSwitch). This adds yet another layer to the existing data center LAN architecture. It also creates organizational stress and leads to inconsistent policy implementation.

- **Limited VM-to-VM Traffic Visibility**
  Traditional vSwitches don't have the same traffic monitoring features as do physical access switches. This limits the IT organization's ability to do security filtering, performance monitoring and troubleshooting within virtualized server domains in both private, public and hybrid clouds.

- **Inconsistent Network Policy Enforcement**
  Traditional vSwitches can lack some of the advanced features that are required to provide the degree of traffic control and isolation required in the data center. This includes features such as private VLANs, quality of service (QoS) and sophisticated ACLs.

- **Layer 2 Network Support for VM Migration**
  When VMs are migrated, the network has to accommodate the constraints imposed by the VM migration utility; e.g., VMotion. Typically the source and destination servers have to be on the same VM migration VLAN, the same VM management VLAN and the same data VLAN.

Server virtualization, however, is not the only factor that is causing IT organizations to redesign their data center LANs. The left hand column in **Table 3** contains a list of the factors that are driving data center redesign. The center column shows the percentage of The Survey Respondents who in 2011 indicated that the corresponding factor was the primary factor that is driving their organization to redesign their data center LAN. The right hand column shows the percentage of The Survey Respondents who recently indicated that the corresponding factor was the primary factor that is driving their organization to redesign their data center LAN.

Table 3: Factors Driving Data Center LAN Redesign *N = 265*

Factor	% of The Survey Respondents In 2011	% of The Survey Respondents in 2012
To reduce the overall cost	24.6%	20.8%
To support more scalability	20.8%	9.1%
To create a more dynamic data center	12.6%	10.2%
To support server virtualization	12.1%	14.0%
To reduce complexity	5.3%	12.5%
To make it easier to manage and orchestrate the data center	13.0%	14.3%
To support our storage strategy	3.4%	3.4%

Table 3: Factors Driving Data Center LAN Redesign *N = 265*

Factor	% of The Survey Respondents In 2011	% of The Survey Respondents in 2012
To reduce the energy requirements	1.0%	0.8%
Other (please specify)	3.4%	9.1%
To make the data center more secure	3.9%	6.0%

The data in **Table 3** indicates that a broad range of factors are driving IT organizations to re-design their data center LANs. There is, however, significant overlap between some of the factors in **Table 3**. For example, there is significant overlap between creating a more dynamic data center and supporting server virtualization. There is also significant overlap between reducing complexity and making it easier to manage and orchestrate the data center. Combining the factors that overlap indicates that:

The primary factors driving IT organizations to re-design their data center LAN is the desire to reduce cost, support server virtualization and reduce complexity.

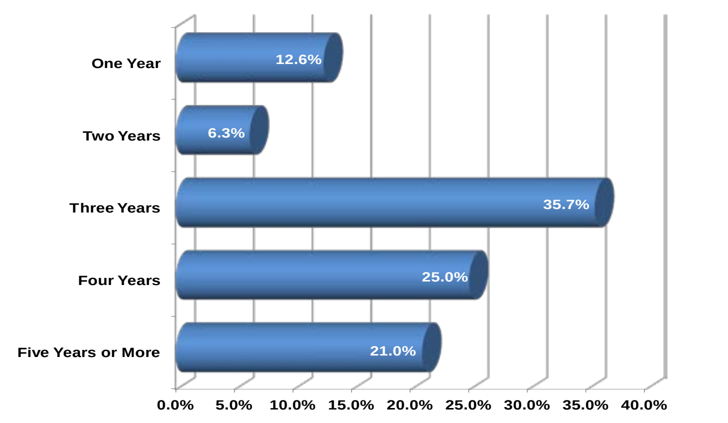
The conventional wisdom in the IT industry is that the cost of the power consumed by data center LAN switches is not significant because it is a small percentage of the total amount of power that is consumed in the typical data center. There is the potential for that situation to change going forward as 10 Gbps, 40 Gbps and 100 Gbps LAN interfaces will potentially consume considerably more power than 1 Gbps LAN interfaces currently do. As such, a requirement of third generation data center LAN switches is that the amount of power that they consume is only marginally more than what is consumed by second generation data center LAN switches and that these switches provide functionality to intelligently manage the power consumption during off peak hours.

Third Generation Data Center LAN Architecture and Technology Options

During the transition from First Generation LANs to Second Generation LANs there was considerable debate over the underlying physical and data link technologies. Alternative technologies included Ethernet, Token Ring, FDDI/CDDI, 100VG-AnyLAN and ATM. One of the few aspects of Third Generation Data Center LANs that is not up for debate is that they will be based on Ethernet. In fact, the Third Generation LAN will provide the possibility of leveraging Ethernet to be the single data center switching fabric, eventually displacing special purpose fabrics such as Fibre Channel for storage networking and InfiniBand for ultra low latency HPC cluster interconnect.

Many of the technologies that are discussed in this chapter and in the chapter on Software Defined Networks are still under development and will not be standardized for another year or two. In order to understand whether or not IT organizations account for emerging technologies in their planning, The Survey Respondents were asked to indicate their company's planning horizon for the evolution of their data center LANs. To avoid ambiguity, the survey question stated "A planning horizon of three years means that you are making decisions today based on the technology and business changes that you foresee happening over the next three years." Their answers are shown in **Figure 1**.

Figure 1: Planning Horizon for Data Center LANs



The data in **Figure 1** indicates that almost 75% of IT organizations have a planning horizon of three years or longer. Since most of the technologies discussed in this chapter will be standardized and ready for production use in three years, that means that the vast majority of IT organizations can incorporate most of the technologies discussed in this chapter into their plans for data center LAN design and architecture.

Below is a discussion of some of the primary objectives of a Third Generation Data Center LAN and an analysis of the various alternatives that IT organizations have relative to achieving those objectives.

Two Tier Data Center LAN Design

There are many on-going IT initiatives that are aimed at improving the cost-efficiency of the enterprise data center. This includes server virtualization, Services Oriented Architecture (SOA), Web 2.0, access to shared network storage as well as the implementation of HPC and cluster computing. In many cases these initiatives are placing a premium on IT organizations being able to provide highly reliable, low latency, high bandwidth communications among both physical and virtual servers. Whereas the hub and spoke topology of the traditional three-tier

Second Generation LAN was optimized for client-to-server communications that is sometimes referred to as *north-south* traffic, it is decidedly sub-optimal for server-to-server communications, which is sometimes referred to as *east-west* traffic.

One approach for improving server-to-server communications is to flatten the network from three tiers to two tiers consisting of access layer and aggregation/core layer switches.

A two-tier network reduces the number of hops between servers, reducing latency and potentially improving reliability. The typical two-tier network is also better aligned with server virtualization topologies where VLANs may be extended throughout the data center in order to support dynamic VM migration at Layer 2.

As discussed below, two tier networks require switches that have very high densities of high-speed ports and a higher level of reliability to protect the soaring volumes of traffic flowing through each switch. As is also discussed below, the requirement for increased reliability and availability creates a requirement for redundant switch configurations in both tiers of the network.

High Port Density and Port Speed

The network I/O requirements of multi-core physical servers that have been virtualized are beginning to transcend the capacity of GbE and multi-GbE aggregated links. As the number of cores per server increases, the number of VMs per physical server can increase well beyond the 10-20 VMs per server that is typical today. With more VMs per server, I/O requirements increase proportionally. Thankfully, the traditional economics of Ethernet performance improvement[2] is falling into place for 10 Gigabit Ethernet (10 GbE). As a result, Third Generation data center LAN switches will need to support high densities of 10 GbE ports to provide connectivity for high performance virtualized servers, as well as an adequate number of 10 GbE ports and 40 GbE, plus 100 GbE ports when these are available and become cost-effective for data center applications. These high-speed ports will be used for multiple purposes, including connecting the access switches to the core tier.

As noted, second generation LAN switches had fairly low port density. In contrast:

The current generation of switches has exploited advances in switch fabric technology and merchant silicon switch-on-a-chip integrated circuits (ICs) to dramatically increase port densities.

Modular data center switches are currently available with up to 768 non-blocking 10 GbE ports or 192 40 GbE ports. The typical maximum port density for TOR switches which are generally based on merchant silicon, is 64 10 GbE ports (or alternatively 48 10 GbE ports and 4 40 GbE ports). Today, high-speed uplinks are often comprised of multiple 10 GbE links that leverage Link Aggregation (LAG)[3]. However, a 40 GbE uplink typically offers superior performance compared to a 4 link 10 GbE LAG. This is because the hashing algorithms that load balance traffic across the LAG links can easily yield sub-optimal load distribution whereby a majority of traffic is concentrated in a small number of flows. Most high performance modular switches

[2] Ethernet typically provides a 10x higher performance for a 3-4x increase in cost. This is an example of how Moore's Law impacts the LAN.

[3] www.ieee802.org/3/hssg/public/apr07/frazier_01_0407.pdf

already have a switch fabric that provide 100 Gbps of bandwidth to each line card, which means that as 40 GbE and 100 GbE line cards become available, these can be installed on existing modular switches, preserving the investment in these devices. Most vendors of modular switches are currently shipping 40 GbE line cards, while 100 GbE line cards will not be widely deployed until 2013 or later due primarily to economic considerations. Currently, most 100 GbE deployments have restricted to service providers, such as Internet exchanges.

In the case of stackable Top of Rack (ToR) switches, adding 40 or 100 GbE uplinks often requires new switch silicon, which means that at least some of the previous generation of ToR switches will need to be swapped out in order to support 40 GbE and, at some future date, 100 GbE uplink speeds.

High Availability

As previously noted, IT organizations will be implementing a growing number of VMs on high performance multi-core servers.

The combination of server consolidation and virtualization creates an "all in one basket" phenomenon that drives the need for highly available server configurations and highly available data center LANs.

One approach to increasing the availability of a data center LAN is to use a combination of redundant subsystems within network devices such as LAN switches in conjunction with redundant network designs. A high availability modular switch can provide redundancy in the switching fabric modules, the route processor modules, as well as the cooling fans and power supplies. In contrast, ToR switches are generally limited to redundant power supplies and fans. Extensive hardware redundancy is complemented by a variety of switch software features, such as non-stop forwarding, that ensure minimal disruption of traffic flow during failovers among redundant elements or during software upgrades. Modular switch operating systems also improve availability by preventing faults in one software module from affecting the operation of other modules. Multi-chassis Link Aggregation Group is described below. Implementing this technology also tends to increase availability because it enables IT organizations to dual home servers to separate physical switches.

Alternatives to the Spanning Tree Protocol

The bandwidth efficiency of Layer 2 networks with redundant links can be greatly improved by assuring that the parallel links from the servers to the access layer and from the access layer to the core layer are always in an active-active forwarding state. This can be accomplished by eliminating loops in the logical topology without resorting to the Spanning Tree Protocol (STP). In the current state of evolution toward a Third Generation data center LAN, loops can be eliminated using switch virtualization and multi-chassis LAG (MC LAG) technologies, which are described below. Another approach is to Implement one of the two emerging shortest path first bridging protocols, TRILL and SPB, that eliminate loops and support equal cost multi-path bridging. TRILL and SPB are also described below.

Switch Virtualization and Multi-Chassis Link Aggregation Group

With switch virtualization, two or more physical switches are made to appear to other network elements as a single logical switch or virtual switch, with a single control plane.

In order for multiple physical switches to form a virtual switch, they need a virtual switch link (VSL) or interconnect (VSI) that supports a common control plane and data flows between the members of the virtual switch. In redundant configurations, connections between end systems and virtual access switches and between virtual access switches and virtual aggregation switches are based on multi-chassis (MC) link aggregation group (LAG) technology[4], as shown in **Figure 2**. MC LAG allows the links of the LAG to span the multiple physical switches that comprise a virtual switch. The re-convergence time associated with MC LAG is typically under 50 ms., which means that real time applications such as voice are not impacted by the re-convergence of the LAN. From the server perspective, links to each of the physical members of a virtual access switch appear as a conventional LAG or teamed links, which means that switches can be virtualized without requiring any changes in the server domain.

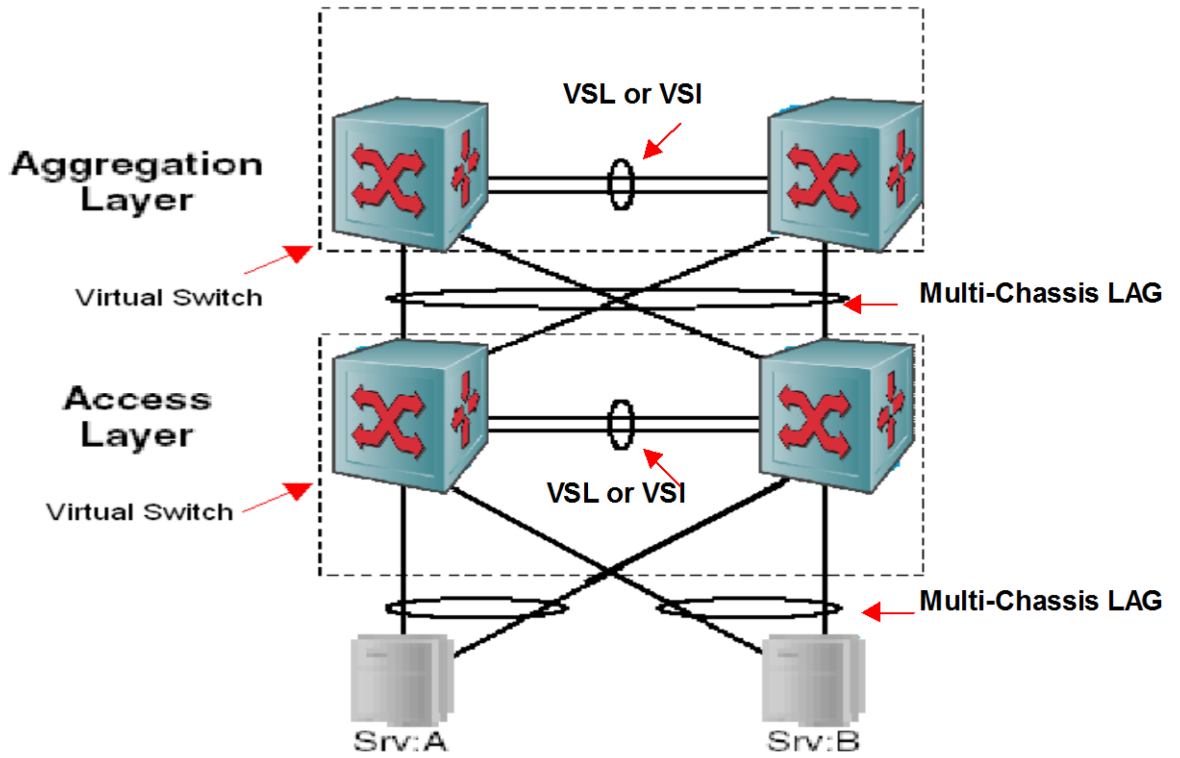
The combination of switch virtualization and multi-chassis LAG can be used to create a logically loop-free topology

This means that data center LANs can be built without using the spanning tree protocol (STP) and first hop router redundancy protocols (e.g., VRRP). This is important because these protocols prevent all available forwarding resources in a redundant network design from being simultaneously utilized.

In **Figure 2**, loops are eliminated because from a logical perspective, there are only two switches with a single LAG from the server to the access switch and a single LAG from the access switch to the aggregation switch. The traffic load to and from each server is load balanced across the two links participating in the multi-chassis LAG connecting each server to the virtual access switch. Therefore, both server connections are actively carrying traffic in both directions rather than being in an active state for some VLANs and in an inactive state for others. In the same fashion, traffic between the access virtual switch and the aggregation virtual switch is load balanced across all four physical links connecting these devices. Both physical switches participating in the aggregation layer virtual switch are actively forwarding traffic to the network core that is not shown in **Figure 2**. The traffic is load balanced via the LAG hashing algorithms rather than being based on VLAN membership, as is the case with more traditional redundant LAN designs. The virtual switch not only improves resource utilization but also enhances availability because the relatively long convergence times of STP topology calculations are circumvented. Virtual switch technology also simplifies management because multiple physical switches can be managed as a single entity.

[4] http://en.wikipedia.org/wiki/Link_aggregation

Figure 2: Switch Virtualization and Multi-Chassis LAG



Most vendors of data center switches support switch virtualization and MC LAG in their ToR and modular switches, and these technologies are fully utilized in the two-tier LAN designs that they are currently recommending to enterprise customers. As a result, most two tier LAN designs being proposed by vendors will not be based on STP for loop control. There are some differences among vendors in the VSL/VSI technology and in the LAG hashing algorithms. For example, some vendors of stackable ToR switches take advantage of the stacking interconnect as the VSL/VSI link, while other vendors will use 10 GbE or 40 GbE ports when available for VSL/VSI. From the server perspective, most LAG implementations conform to the IEEE 802.3ad standard. However, LAG hashing algorithms are outside the 802.3ad standard and more sophisticated hashing algorithms can provide for some differentiation between LAN switches by improving load balancing across the MC LAG links. In addition, there are some differences in the number of ports or links that can participate in a LAG. Some vendors support up to 32 links per LAG, while 8 links per LAG is the most common implementation.

Currently MC Lags are based on proprietary implementations that have a variety of different names. As a result, MC LAG interoperability between switches from different vendors cannot be expected. Most vendors recommend MC LAG 2 tier topologies similar to the one shown on **Figure 2**. MC LAG are generally not recommended in configurations with more than two aggregation switches, such as large 2 tier fat tree topologies.

SPB and TRILL

It must be noted that two-tier LANs and switch virtualization are far from the final word in the design of data center networks. Standards bodies have been working on technologies that will allow active-active traffic flows and load balancing of Layer 2 traffic in networks of arbitrary switch topologies. TRILL (Transparent Interconnection of Lots of Links) is an Internet Engineering Task Force (IETF) standard for a Layer 2 shortest-path first (SPF) routing protocol for Ethernet. The TRILL RFC (RFC 6325) is currently supported by some vendors as part of their proprietary Layer 2 fabric implementations. However, most of the current implementations of TRILL are based on pre-standard drafts in combination with added proprietary features and are not interoperable. In the future, vendors that provided early support for TRILL are likely to offer two versions: openTRILL which is strictly standards compliant and interoperable and a proprietary fabric solution based partly on TRILL.

Shortest Path Bridging (SPB) as defined in IEEE 802.1aq is a competing standard for equal cost multi-path bridging Ethernet fabrics. There are two variants of SPB: SPBM where packets are encapsulated at the edge using 802.1ah MAC-in-MAC frame formats and SPBV where packets are tagged with 802.1D/802.1ad tags. Three switch vendors (Avaya, Alcatel Lucent, and Huawei) have demonstrated interoperability with SPBM.

With either TRILL or 802.1aq SPB, it would be possible to achieve load-balanced, active-active link redundancy without having to resort entirely to switch virtualization, MC LAG, and VSL/VSI interconnects. For example, dual homing of servers can be based on MC LAG to a virtual access switch comprised of two physical access switches, while the rest of the data center LAN is based on TRILL or SPB.

There is currently considerable debate in the industry about which is the best technology – TRILL or SPB. While that is an important debate:

> ***In many cases, the best technology doesn't end up being the dominant technology in the marketplace.***

TRILL and SPB have some points of similarity but they also have some significant differences that preclude interoperability. Both approaches use IS-to-IS as the Layer 2 routing protocol and both support equal cost multi-path bridging, which eliminates the blocked links that are a characteristic of STP. Both approaches also support edge compatibility with STP LANs. Some of the major differences include:

- TRILL involves a new header for encapsulation of Ethernet packets, while SPB uses MAC-in-MAC Ethernet encapsulation. Therefore, TRILL requires new data plane hardware, while SPB doesn't for Ethernet switches that support 802.1ah (MAC-in-MAC), 802.1ad (Q-in-Q) and 802.1ag (OAM).

- SPB's use of MAC-in-MAC Ethernet encapsulation eliminates the potential for a significant increase in the size of MAC address tables that are required in network switches.

- SPB forwards unicast and multicast/broadcast packets symmetrically over the same shortest path, while TRILL may not forward multicast/broadcast packets over the shortest path.

- SPB eliminates loops using Reverse Path Forwarding (RPF) checking for both unicast and multicast traffic, while TRILL uses Time to Live (TTL) for unicast and RPF for multicast.

- TRILL can support multi-pathing for an arbitrary number of links, while SPB is currently limited to 16 links.

- TRILL is supported by vendors with large market share in LAN switching. SPB is currently supported by vendors with a relatively small market share.

- With TRILL, Layer 2 network virtualization is limited to 4K VLANs, while SPBM supports a 16 million virtual network service instances via its 24 bit I-SID field in the encapsulating header.

- SPBM can also support Layer 3 network virtualization as described in an IETF draft (IP/SPBM)

- SPB is compatible with IEEE 802.1ag and ITU Y.1731 OAM which means that existing management tools will work for SPB, while TRILL has yet to address OAM capability.

- SPB is compatible with Provider Backbone Bridging (PBB), the protocol used by many service providers to provide MPLS WAN services. This means that SPB traffic can be directly mapped to PBB. Also, virtual data centers defined with SPB can be mapped to separate traffic streams in PBB and given different QoS and security treatment.

In the future TRILL and SPB should have major implications for data center LAN designs and most of the larger switch vendors are well along in developing switches that can support either TRILL or SPB and network designs based on these technologies. It may well turn out that two-tier networks based on switch virtualization and MC LAG are just a mid-way point in the evolution of the Third Generation LAN.

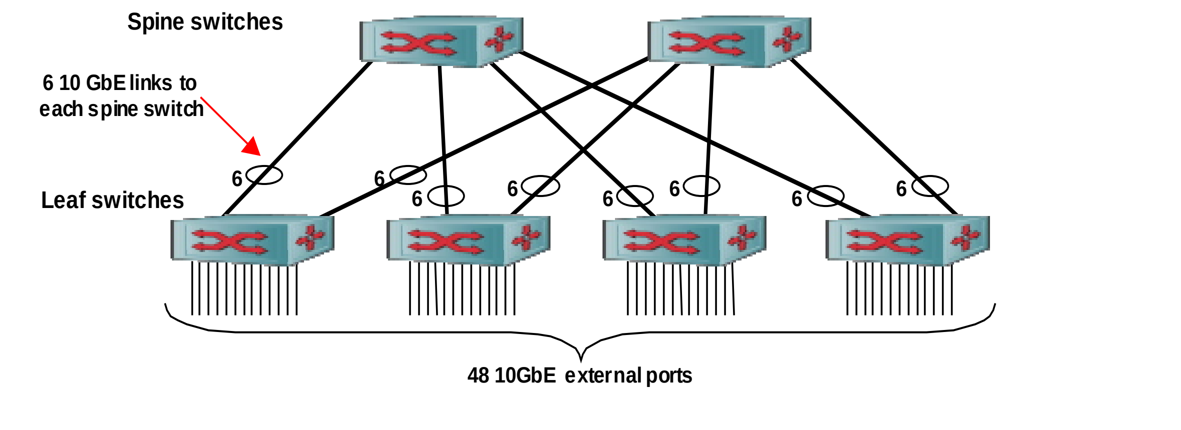
With technologies like TRILL and SPB, the difference between access switches and core switches may shrink significantly.

As a result of TRILL or SPB, the switch topology may shift from a two-tier hub and spoke, such as the one in **Figure 2**, to a highly meshed or even fully meshed array of switches that appears to the attached devices as a single switch. TRILL and SPF bridging can support a variety of other topologies, including the fat tree switch topologies[5] that are popular in cluster computing approaches to HPC. Fat trees have also gotten a lot of attention as a topology for highly scalable data center LANs, such as Cisco's FabricPath and Juniper's QFabric. Fat tree topologies are also used by Ethernet switch vendors to build high density, non-blocking 10 GbE switches using merchant silicon switch chips. This trend may eventually lead to the commoditization of the data plane aspect of Ethernet switch design. **Figure 3** shows how a 48 port 10 GbE TOR switch can be constructed using six 24-port 10 GbE switch chips. By increasing the number of leaf and spine switches, larger switches can be constructed[6]. A number of high density 10 GbE switches currently on the market use this design approach.

[5] www.mellanox.com/pdf/../IB_vs_Ethernet_Clustering_WP_100.pdf

[6] The maximum density switch that can be built with a two-tier fat tree architecture based on 24 port switch chips has 288 ports.

Figure 3: TOR Switch Fat Tree Internal Architecture



A discussion of the alternatives to STP amongst six of the primary data center LAN switch vendors can be found at Webtorials[7].

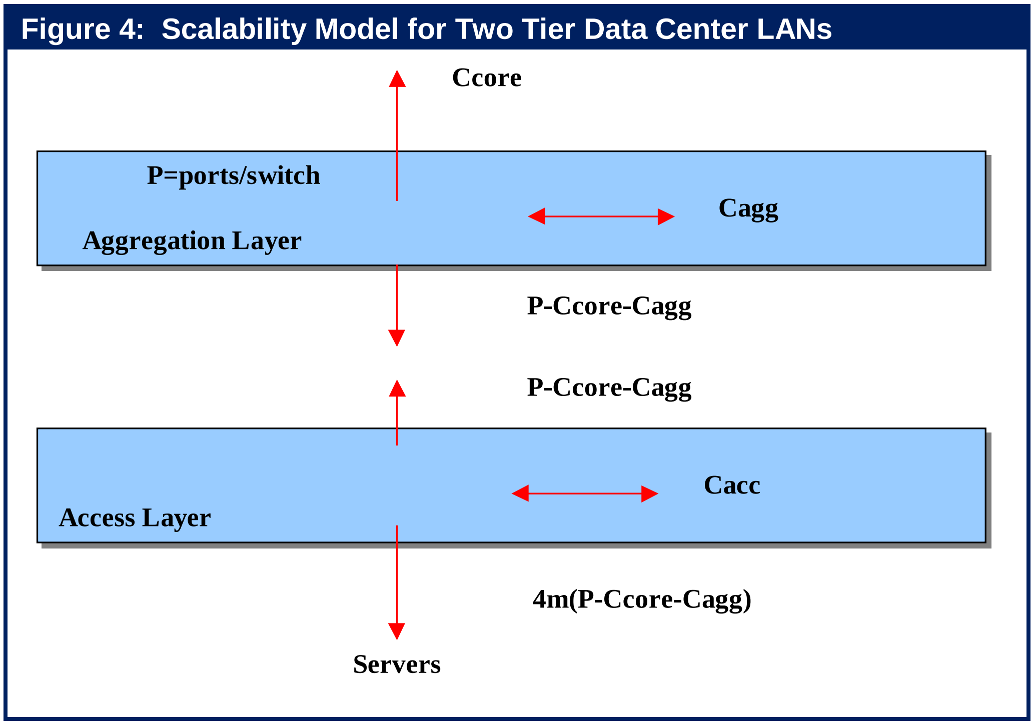
Scalability of Two Tier LAN Designs

The scalability of a LAN architecture is determined by the number of server ports that can be supported with a given level of redundancy and over-subscription at different points within the LAN topology. Many data center LANs being deployed today are based on a two tier design that provides high levels of redundancy and low over-subscription levels for server-to-server traffic. Two tier LAN designs are frequently implemented with Top of Rack (TOR) access switches in conjunction with chassis-based aggregation switches. The aggregation switches are connected to the LAN core and to the Internet, but all the server-to-server traffic within the data center flows only through the two tiers of access and aggregation switches.

Figure 4 shows a general model for two tier switched LANs that takes into account both connections for redundancy and connections to the LAN core. It is assumed that all servers are attached to the access/TOR switches via 10 GbE ports. Any inter-switch links at the access layer are assumed to be 10 GbE, and all other inter-switch links (i.e., inter-aggregation, access-to-aggregation and aggregation-to-core) are assumed to be 40 GbE. If a given model of switch does not yet support 40 GbE, a LAG with four 10 GbE member links could be substituted. It should be noted that as previously mentioned a 40 GbE link is preferable to a LAG of four 10 GbE links because having a single 40 GbE link avoids the issues that can occur when attempting to load balance traffic that consists of a small number of high volume flows.

[7] http://www.webtorials.com/content/tls.html

Figure 4: Scalability Model for Two Tier Data Center LANs



Definition of Symbols

P: The number of 40 GbE ports per aggregation switch

m: The effective over-subscription ratio

S: The number of aggregation switches

Ccore: The number of 40 GbE ports per aggregation switch that are used to connect to the LAN core

Cagg: The number of 40 GbE ports per aggregation switch used to connect to other aggregation switches

Cacc: The number of connections between TOR switches

P – Ccore – Cagg: The number of 40 GbE ports per aggregation switch available for connections to the access layer

4 x m x (P-Ccore-Cagg): The number of 10 GbE access layer ports that are available for server connection per aggregation

4 x S x m x (P-Ccore-Cagg): For two tier LAN design with multiple aggregation switches, the number of available server ports

This model can be applied equally well to two tier LANs based on MC LAGs and two tier fat trees. The model focuses on P, the number of 40 GbE ports per aggregation switch and the number of ports required to make connections both within and among network tiers.

In the model, *Ccore* is the number of 40 GbE ports per aggregation switch that are used to connect to the LAN core, *Cagg* is the number of 40 GbE ports per aggregation switch that are used to connect to other aggregation switches (e. g., for ISL/VSL). There may also be 10 GbE inter-switch links within the access/TOR tier to support virtual switch/router functions such as multi-chassis LAG (MLAG) or VRRP.

The access/TOR switches may be oversubscribed with more switch bandwidth allocated to server connections vs. the amount of bandwidth that is provided from the access tier to the aggregation tier. The over-subscription ratio is given by the following ratio:

The amount uf bandwidth allocated to server access / The amount of bandwidth allocated to access-to-aggregation connectivity.

A typical high density TOR switch has 48 10 GbE ports for server connectivity and four 40 GbE ports for inter-switch connectivity. Where servers are single-attached to these TOR switches, m is equal to (48 x 10)/(4 x 40) = 3. Where the servers are dual-attached to a pair of TOR switches with active-passive redundancy, m = 3, but the effective over-subscription ratio is 1.5:1 because only one of the pair of server ports is active at any given time. Where the servers are dual-attached to a pair of TOR switches with active-active MC LAG redundancy, the requirement for inter-switch connections (*Cacc*) between the TOR switches means there are two fewer 10 GbE ports per TOR switch available for server connectivity and the over-subscription ratio is equal to m = (46 x 10)/(4 x 40) = 2.88

As shown in **Figure 4**, the number of 40 GbE ports per aggregation switch that is available for connections to the access layer is equal to P-Ccore-Cagg and the number of 10 GbE access layer ports that are available for server connection per aggregation is equal to 4 x m x (P-Core-Cagg). For a two tier LAN design with multiple aggregation switches, the number of available server ports is 4 x S x m x (P-Core-Cagg), where S is the number of aggregation switches.

It should be noted that the model presented in **Figure 4** is based on having a single aggregation switch, and the factor S needs to be included to account for an aggregation tier with multiple aggregation switches. For an MC LAG 2 tier network S is generally limited to 2. For fat trees, the number of aggregation switches, or spine switches, is limited by the equal cost forwarding capabilities (16 paths is a typical limit), as well as the port density P. The port configuration of the access/TOR switch also imposes some limitations on the number of aggregation/spine switches that can be configured. For example, for a TOR switch with 48 10 GbE ports and four 40 GbE ports the number of 40 GbE aggregation switches is limited to four. Scaling beyond S=4, requires both a denser access switch with more 40 GbE ports and more 10 GbE port as well to maintain a desired maximum over-subscription ratio. The ultimate fat tree scalability is attained where the 10 GbE/40 GbE access switch has same switching capacity as the aggregation/spine switches.

With these caveats, the model takes into account redundancy and scalability for various Layer 2 and Layer 3 two-tier network designs as summarized in **Table 4**.

Table 4: Scalability of Two Tier 10/40 GbE Data Center LANs

Parameter	2 Tier L2	2 Tier Layer 2 MC LAG	2 Tier Layer 2 Fat Tree	2 Tier Layer 3 Fat Tree
Redundancy	none	Full	full	Full
Ccore	variable	Variable	variable	variable
Cagg	0	ISL/VSL 2 per agg switch	0	0
Cacc	0	active/passive server access: 0 active/active: 2 per TOR	active/passive server access: 0 active/active: 2 per TOR	active/passive: 2 per TOR active/active: 2 per TOR

Max 10 GbE server ports	4Sm(P-Ccore-Cagg)
	S=1
		4Sm(P-Ccore-Cagg)
		S=2
			4Sm(P-Ccore-Cagg); S = # of aggregation switches	4Sm(P-Ccore-Cagg); S = # of aggregation switches

Scaling	Larger P, m	Larger P, m	Larger P,m,S	Larger P,m,S

As highlighted in **Table 4**, the only way that the scalability of the data center LAN can be increased is by increasing the:

- Number of aggregation switches
- Number of 40 GbE ports per aggregation switch
- Level of over-subscription

As stated earlier, a typical initial design process might start from identifying the required number of server ports, the required redundancy, and an upper limit on the over-subscription ratio. As shown in **Figure 5**, calculating the required number of 40 GbE ports per aggregation switch to meet these requirements is accomplished by inverting the scaling formula. An IT organization could use the following process to utilize the formula:

1. Determine required number of server ports
2. Select the desired network type from Table 4. This will determine Cagg
3. Select a access/TOR switch model. This together with the network type will determine Cacc and m.
4. Select the desired Ccore. This will determine over-subscription ratio for client/server traffic via the core
5. Calculate the required port density of the aggregation switch using the following formula:

Figure 5: Required Aggregation Switch Port Density

$$P=((\#\text{ of server ports})/4Sm)+Ccore+Cagg$$

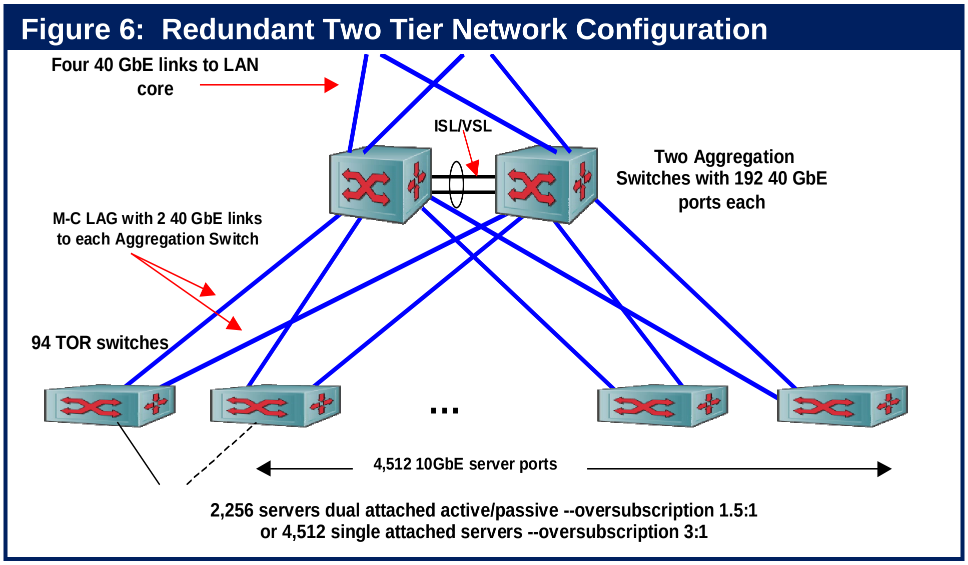
To exemplify the formula shown in **Figure 5**, consider the following network parameters:

The number of servers ports = 4512
Network type; MC LAG
m = 3
S = 2
Ccore = 2
Cagg = 2

The formula in **Figure 5** indicates that in order to support the indicated network parameters, an aggregation switch with 192 40 GbE ports is required.

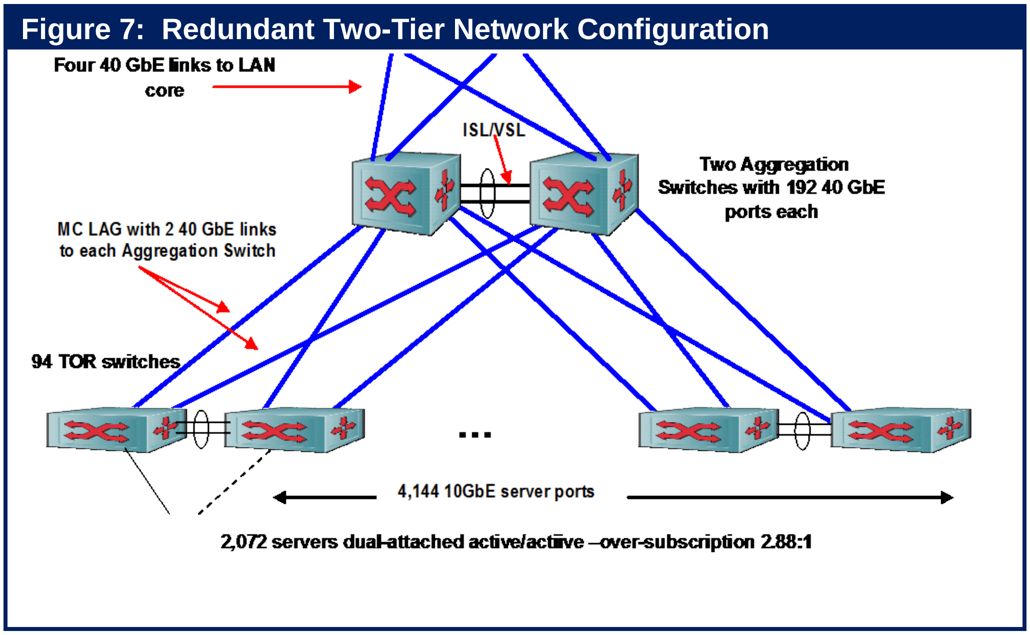
Figure 6 shows an example of a data center network that provides fully redundant Layer 2 server-to-server connectivity based on 94 TOR switches, each having 48 10 GbE ports and 4 40 GbE ports plus a pair of high density aggregation switches with 192 40 GbE ports each. The topology is an MC LAG Layer 2 network with oversubscribed TOR switches. Each of the 2,256 servers is connected to two TOR switches in an active/passive mode. The same configuration could also support 4,512 single-attached servers. With active/passive redundancy, the over-subscription of access switches for server-to-server traffic is 1.5:1.

Figure 6: Redundant Two Tier Network Configuration


Four 40 GbE links to LAN core
ISL/VSL
Two Aggregation Switches with 192 40 GbE ports each
M-C LAG with 2 40 GbE links to each Aggregation Switch
94 TOR switches
...
4,512 10GbE server ports
2,256 servers dual attached active/passive --oversubscription 1.5:1
or 4,512 single attached servers --oversubscription 3:1

For active-active server connectivity, each pair of TOR switches would need to be configured as a virtual switch with a pair of inter-TOR 10 GbE links for the ISL/VSL connectivity required for the virtual switch, as shown in **Figure 7**. This would reduce the number of servers per TOR switch from 24 to 23 and the number of dual-attached servers to 2,072. With active/active redundant MLAG server connectivity, the over-subscription ratio for server-to-server traffic is 2.88:1.

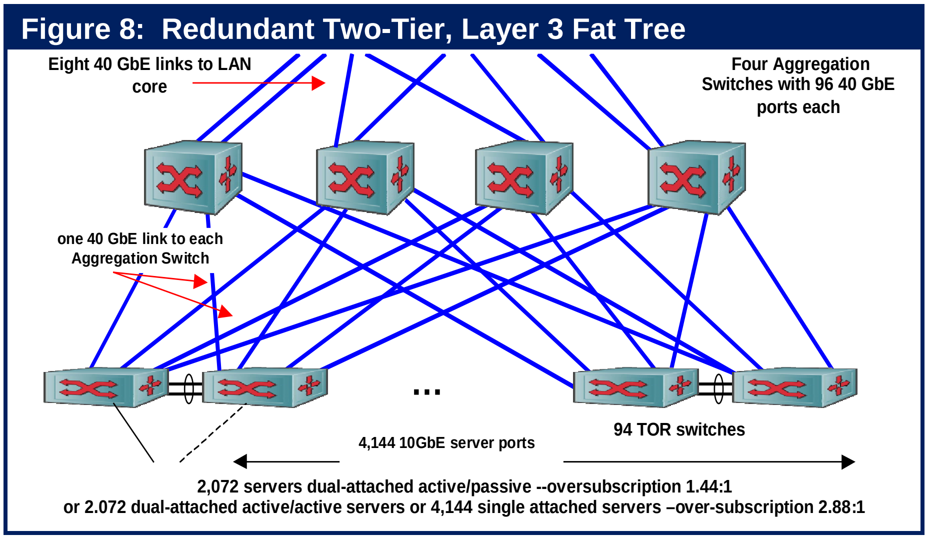
Figure 7: Redundant Two-Tier Network Configuration


Four 40 GbE links to LAN core
ISL/VSL
Two Aggregation Switches with 192 40 GbE ports each
MC LAG with 2 40 GbE links to each Aggregation Switch
94 TOR switches
...
4,144 10GbE server ports
2,072 servers dual-attached active/active –over-subscription 2.88:1

Building a comparable network with essentially the same number of 10 GbE server ports and similar over-subscription ratios using similar TOR switches and an aggregation switch with half the density (i.e., 96 40 GbE ports) requires some design changes. Comparing the two designs provides an illustration of the effect that the density of the aggregation switch can have on the network design and the resulting TCO.

One possibility would be to build a Layer 2 fat tree network using four aggregation switches in the spine/aggregation layer and the same number of TOR switches (94) as the leaves/access switches. However, most TOR switches do not yet support Layer 2 equal cost multi-path forwarding alternatives other than with some form of MC LAG. One workaround is to move the Layer 3 boundary from the aggregation switch to the TOR switch and build a Layer 3 fat tree with OSPF ECMP providing the multi-path functionality. **Figure 8** shows what this could look like. Here the ISL links are only at the TOR level rather than the aggregation level and the

Figure 8: Redundant Two-Tier, Layer 3 Fat Tree


Eight 40 GbE links to LAN core
Four Aggregation Switches with 96 40 GbE ports each
one 40 GbE link to each Aggregation Switch
...
4,144 10GbE server ports
94 TOR switches
2,072 servers dual-attached active/passive --oversubscription 1.44:1
or 2.072 dual-attached active/active servers or 4,144 single attached servers –over-subscription 2.88:1

server connection can be made active/active without affecting the topology. With active/passive redundancy, the over-subscription of aggregation switches for server-to-server traffic is 1.44:1, while with active/active redundant server connectivity, the over-subscription ratio is 2.88:1. Note that Layer 2 and Layer 3 fat trees based on switches with the same port densities at the aggregation and access levels have the same physical topology.

If a TCO comparison is made of the two networks shown in **Figure 7** and **Figure 8**, some of the differences to consider are:

- Capex and Opex differences with four switches vs. two at the aggregation level, including switch cost, power capacity requirements, rack space requirements, annual power, annual routine administration, and annual service contract costs
- Difference in the number of server ports per TOR
- Differences in over-subscription ratios to the core
- Eight links vs. four links to the LAN core needed for redundancy
- Administrative cost and complexity differences with 98 Layer 3 devices if the fat tree is implemented at Layer 3 vs. two Layer 3 devices with MC LAG.

In addition, in a Layer 3 fat tree, there is a requirement for a Layer 2 over Layer 3 network virtualization to enable VM migration across Layer 3 boundaries

This example shows some of the complexities that can be encountered in comparing the TCOs of competing data center switching solutions that are based on switches of different port densities, as well as somewhat different functionality.

Network Support for Dynamic Creation and Movement of VMs

When VMs are migrated between servers, the network has to accommodate the constraints imposed by the VM migration utility; e.g., VMotion. Typically the VM needs to be on the same VLAN when migrated from source to destination server. This allows the VM to retain its IP address which helps to preserve user connectivity after the migration. When migrating VMs between disparate data centers, these constraints generally require that the data center Layer 2 LAN be extended across the physical locations or data centers without compromising the availability, resilience and security of the VM in its new location. VM migration also requires the LAN extension service have considerable bandwidth and low latency. VMware's VMotion, for example, requires at least 622 Mbps of bandwidth and less than 5 ms of round trip latency between source and destination servers over the extended LAN[8].

The data storage location, including the boot device used by the virtual machine, must be accessible by both the source and destination physical servers at all times. If the servers are at two distinct locations and the data is replicated at the second site, the two data sets must be identical. One approach is to extend the SAN to the two sites and maintain a single data source. Another option is to migrate the data space associated with a virtual machine to the secondary storage location. In either case, there is a significant impact on the WAN.

As noted earlier, the requirement to support the dynamic creation and movement of VMs is one of the primary factors driving IT organizations to redesign their data center LANs. As was also noted earlier, the requirements for VM migration within VLAN boundaries have provided a major impetus for flattening the LAN with two-tier designs featuring Layer 2 connectivity end-to-end. Extending VLANs across the data center requires configuration of 802.1Q trunks between the intermediate switches, which can be a labor intensive task. With other forms of network virtualization (discussed in a later section of the report) virtual networks can be created without reconfiguration of intermediate switches.

Many of the benefits of cloud computing depend on the ability to dynamically provision VMs and to migrate them at will among physical servers located in the same data center or in geographically separated data centers. The task of creating or moving a VM is a relatively simple function of the virtual server's management system. There can, however, be significant challenges in assuring that the VM's network configuration state, including VLAN memberships, QoS settings, and ACLs, is established or transferred in a timely fashion. In many instances today, these network configuration or reconfigurations involves the time-consuming manual process involving multiple devices.

Regulatory compliance requirements can further complicate this task. For example, assume that the VM to be transferred is supporting an application that is subject to PCI compliance. Further assume that because the application is subject to PCI compliance that the IT organization has implemented logging and auditing functionality. In addition to the VM's network configuration state, this logging and auditing capability also has to be transferred to the new physical server.

The most common approach to automating the manual processes involved in VM provisioning and migration is based on communication between the Hypervisor Management system and the switch element management system (EMS) via APIs supported by both vendors[9]. This type of solution is commonly referred to as Edge Virtualization.

When a Virtual Machine is created or when the movement of a VM is initiated, the Hypervisor manager signals to the EMS that the event is about to occur and provides a partial VM network profile including a virtual MAC, VLAN memberships and the target hypervisor. Based on existing policies, the EMS extends the VM network profile to include appropriate QoS and security parameters such as ACLs. The EMS can then determine the target hypervisor's access switch and can configure or reconfigure it accordingly. Where VLANs need to be created, the EMS can also create these on the uplinks and neighboring switches as appropriate. In a similar manner, when a VM is deleted from a hypervisor, the EMS can remove the profile and then prune the VLAN as required. All of these processes can be triggered from the hypervisor.

[8] http://www.vce.com/pdf/solutions/vce-application-mobility-whitepaper.pdf

[9] While this approach is the most common, some vendors have alternative approaches.

Most data center switch vendors have already implemented some proprietary form of VM network profile software, including linking their switches to at least one brand of hypervisor. Some differences exist between the range of hypervisors supported and the APIs that are used. Distribution of VM network profiles is only one of many management processes that can benefit greatly from automation, so it would benefit IT departments to develop expertise in open APIs and powerful scripting languages that can be exploited to streamline time-consuming manual processes and thereby reduce operational expense while improving the ability of the data center to dynamically reallocate its resources in response to changes in user demand for services.

Another approach to edge virtualization is the Distributed Virtual Switch (DVS). With DVS, the control and data planes of the embedded hypervisor vSwitch are decoupled. This allows the data planes of multiple vSwitches to be controlled by an external centralized management system that implements the control plane functionality. Decoupling the data plane from the control plane also makes it easier to tightly integrate the vSwitch control plane with the control planes of physical access and/or aggregation switches and/or the virtual server management system. Therefore, DVS can simplify the task of managing a large number of vSwitches, and improve control plane consistency, in addition to providing edge virtualization in support of VM creation and mobility.

The DVS is a significant improvement over earlier hypervisor vSwitches, but retains a number of characteristics of vSwitches that may be of concern to network designers, including:

1. The vSwitch represents another tier of switching that needs to be configured and managed, possibly requiring an additional management interface. This can partially defeat an effort to flatten the network to two–tiers.

2. The vSwitch adds considerable complexity, because there is an additional vSwitch for every virtualized server.

3. vSwitch control plane functionality is typically quite limited compared to network switches, preventing a consistent level of control over all data center traffic

4. As more VMs per server are deployed, the software switch can place high loads on the CPU, possibly starving VMs for compute cycles and becoming an I/O bottleneck.

5. VM-VM traffic on the same physical server is isolated from the rest of the network, making these flows difficult to monitor and control in the same fashion as external flows.

6. The vSwitch functionality and management capabilities will vary by hypervisor vendor and IT organizations are increasingly deploying hypervisors from multiple vendors.

IEEE 802.1Qbg is a standard that addresses both edge virtualization and some of the potential issues with vSwitches. The standard includes Edge Virtual Bridging (EVB) in which all the traffic from VMs is sent to the physical network access switch. If the traffic is destined for a VM on the same physical server, the access switch returns the packets to the server over the same port on which it was received. The shipping of traffic from a VM inside of a physical server to an external access switch and then back to a VM inside the same physical server is often referred to as a hair pin turn or reflective relay. With Edge Virtual Bridging, the hypervisor can be relieved from all switching functions, which are now performed by the physical access network. With EVB, the vSwitch can perform the simpler function of a Virtual Ethernet Port

Aggregator (VEPA) aggregating hypervisor virtual NICs to a physical NIC . Basic EVB can be supported by most existing access switches via a relatively simple firmware upgrade.

The IEEE 802.1Qbg standard includes some additional protocols that standardize the switch side of edge virtualization. The additional protocols Edge TLV Protocol and VSI Discovery and Configuration Protocol (VDP) support edge virtualization where the Layer 2 configuration of the network to support VM creation and migration is automated. Using VDP, the target switch can be informed of the imminent VM deployment, allowing the target switch to be properly configured in advance of VM creation or movement Therefore, Qbg provides a standards-based alternative to proprietary approaches to edge virtualization via integration between switch management systems and hypervisor management systems. A companion effort, the IEEE's 802.1BR Bridge Port Extension is defining a technique for a single physical port to support a number of logical ports and a tagged approach to deal with frame replication issues. Port Extension is used in fabric extenders for blade servers and rack mounted servers as an alternative to blade server switches and full function ToR switches.

Vendors of data center switches are expected to provide some level of support for 802.1Qbg Some vendors may focus on either EVB or edge virtualization, while others will support the full range of Qbg capabilities. Some vendors may also offer DVS implementations that support Qbg-based edge virtualization.

Network Virtualization

Within the IT industry, the phrase *network virtualization* is used in a wide variety of ways. In order to eliminate confusion and ambiguity, The Survey Respondents were told that "Network virtualization is the creation of multiple logical networks that share a common physical network in a manner that is somewhat analogous to how multiple virtual machines share a common physical server. While techniques such as VLANs have been available for a long time, emerging technologies such as VXLAN, NVGRE and Software Defined Networks are enabling new forms of network virtualization."

The Survey Respondents were then given a set of possible actions and were asked to indicate which of the actions best describes their organizations approach to these new forms of network virtualization. Their responses are shown in **Table 5**.

Table 5: Status of Network Virtualization *N = 307*

Action	Percentage of The Survey Respondents
We have already implemented network virtualization	23.8%
We are interested in network virtualization, but we will not likely take any steps towards implementing it for at least a year	17.6%
We are in the process of evaluating network virtualization	16.3%
We are not currently taking any steps towards implementing network virtualization, but are likely to in the next twelve months	12.7%
Don't know	10.1%
We currently have no interest in network virtualization	9.1%
We are in the process of testing network virtualization	8.8%

Table 5: Status of Network Virtualization *N = 307*

Action	Percentage of The Survey Respondents
Other	1.5%

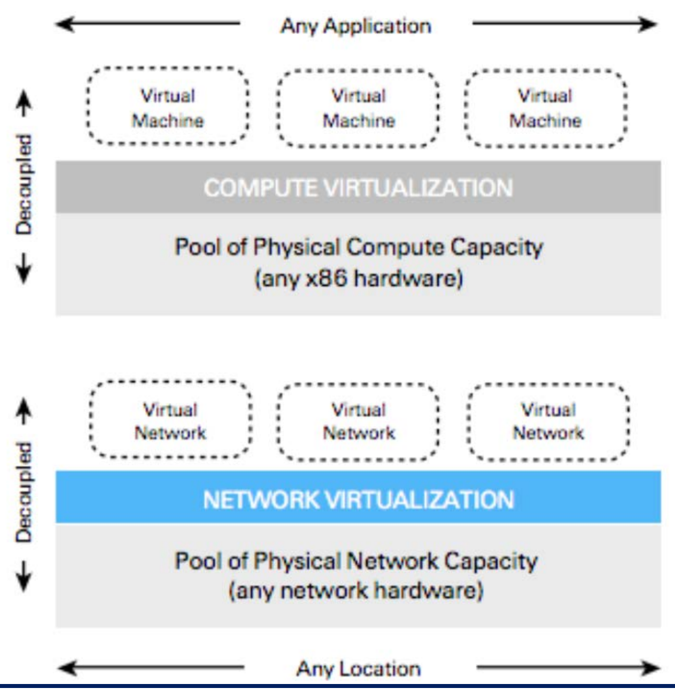
One conclusion that can be drawn from the data in **Table 5** is that:

> ***There is very strong interest on the part of IT organizations to implement network virtualization.***

In addition to 802.1Qbg there are a number of emerging and proposed standard protocols that are focused on optimizing the support that data center Ethernet LANs provide for server virtualization. Several of these protocols are aimed at network virtualization via the creation of multiple virtual Ethernet networks that can share a common physical infrastructure in a manner that is somewhat analogous to multiple VMs sharing a common physical server, as shown in **Figure 9**.

Figure 9: Network Virtualization

Source: VMware



Most protocols for network virtualization are based on creating virtual network overlays using tunneling/encapsulation techniques. The protocols that provide network virtualization of the data center include VXLAN, NVGRE, STT, and SPB MAC-in-MAC. SPB is already a IEEE standard, while it is likely that only one of the other proposals will achieve IETF standard status.

<u>Traditional Network Virtualization</u>

One-to-many virtualization of network entities is not a new concept. The most common traditional applications of the virtualization concept to networks are VLANs and Virtual Routing and Forwarding (VRF) instances.

VLANs partition the Ethernet network into as many as 4,094 broadcast domains as designated by a 12 bit VLAN ID tag in the Ethernet header. VLANs have been a convenient means of isolating different types of traffic that share the same switched LAN infrastructure. In data centers making extensive use of server virtualization, the limited number of VLANs can present problems, especially in cases where a large number of tenants need to be supported, each of whom requires multiple VLANs. Extending VLANs across the data center via 802.1Q trunks to support VM mobility adds operational cost and complexity. In data centers based on Layer 2 server-to-server connectivity, large numbers of VMs, each with its own MAC address, can also place a burden on the forwarding tables capacities of Layer 2 switches.

VRF is a form of Layer 3 network virtualization in which a physical router supports multiple virtual router instances, each running its own routing protocol instance and maintaining its own forwarding table. Unlike VLANs, VRF do not use a tag in the packet header to designate the specific VRF to which a packet belongs. The appropriate VRF is derived at each hop based on the incoming interface and information in the frame. An additional requirement is that each intermediate router on the end-to-end path followed by a packet needs to be configured with a VRF instance that can forward that packet.

Network Virtualization with Overlays

Due to the shortcomings of the traditional VLAN or VRF models, a number of new techniques for creating virtual networks have emerged over recent years and months. Most of these network virtualization techniques are based on tunneling/encapsulation to construct multiple virtual network topologies overlaid on a common physical network. A virtual network can be a Layer 2 network or a Layer 3 network, while the physical network can be Layer 2, Layer 3 or a combination depending on the overlay technology. With overlays, the outer (encapsulating) header includes a field (generally 24 bits wide) that carries a virtual network instance ID (VNID) that specifies the virtual network designated to forward the packet.

Virtual network overlays can provide a wide range of benefits, including:

- Support for essentially unlimited numbers of virtual networks (24 bits equates to 16 million virtual networks)
- Decoupling of the virtual network topology, service category (L2 or L3), and addressing from those of the physical network. The decoupling avoids issues such as MAC table size in physical switches.
- Support for virtual machine mobility independent of the physical network. If a VM changes location, even to a new subnet, the switches at the edge of the overlay simply update their mapping tables to reflect the new location of the VM. The network for a new VM can be be provisioned entirely at the edge of the network.
- Ability to manage overlapping IP addresses between multiple tenants.
- Support for multi-path forwarding within virtual networks

The main difference between the various overlay protocols lies in their encapsulation formats and the control plane functionality that allows ingress (encapsulating) devices to map a frame to the appropriate egress (decapsulating) device.

VXLAN

Virtual eXtensible LAN (VXLAN)[10] virtualizes the network by creating a Layer 2 overlay on a Layer 3 network via MAC-in-UDP encapsulation. The VXLAN segment is a Layer 3 construct that replaces the VLAN as the mechanism that segments the data center LAN for VMs. Therefore, a VM can only communicate or migrate within a VXLAN segment. The VXLAN segment has a 24 bit VXLAN Network identifier. VXLAN is transparent to the VM, which still communicates using MAC addresses. The VXLAN encapsulation is performed through a function known as the VXLAN Tunnel End Point (VTEP), typically a hypervisor switch or a possibly a physical access switch. The encapsulation allows Layer 2 communications with any end points that are within the same VXLAN segment even if these end points are in a different IP subnet. This allows live migrations to transcend Layer 3 boundaries. Since MAC frames are encapsulated within IP packets, there is no need for the individual Layer 2 switches to learn MAC addresses. This alleviates MAC table hardware capacity issues on these switches. Overlapping IP and MAC addresses are handled by the VXLAN ID, which acts as a qualifier/identifier for the specific VXLAN segment within which those addresses are valid. The VXLAN control solution uses flooding based on Any Source Multicast (ASM) to disseminate end system location information.

As noted, VXLANs uses a MAC-in-UDP encapsulation. One of the reasons for this is that modern Layer 3 devices parse the 5-tuple (including Layer 4 source and destination ports). While VXLAN uses a well-known destination UDP port, the source UDP port can be any value. As a result, a VTEP can spread all the flows from a single VM across many UDP source ports. This allows for efficient load balancing across LAGs and intermediate multi-pathing fabrics even in the case of multiple flows between only two VMs.

Where VXLAN nodes on a VXLAN overlay network need to communicate with nodes on a legacy (i.e., VLAN) portion of the network, a VXLAN gateway can be used to perform the required tunnel termination functions including encapsulation/decapsulation. The gateway functionality could be implemented in either hardware or software.

VXLAN is the subject of a IETF draft supported by VMware, Cisco, Arista Networks, Broadcom, Red Hat and Citrix. VXLAN is also supported by IBM. Pre-standard implementations in hypervisor vSwitches and physical switches are beginning to emerge.

NVGRE

Network Virtualization using Generic Router Encapsulation (NVGRE) uses the GRE tunneling protocol defined by RFC 2784 and RFC 2890. NVGRE is similar in most respects to VXLAN with two major exceptions. While GRE encapsulation is not new, most network devices do not parse GRE headers in hardware, which may lead to performance issues and issues with 5-tuple hashes for traffic distribution in multi-path data center LANs. The other exception is that the current IETF NVGRE draft does not address the control plane question, leaving that for a future draft or possibly as something to be addressed by (Software Defined Networking) SDN controllers. Some of the sponsors of NVGRE (Microsoft and Emulex) expect that some of the performance issues can be addressed by intelligent NICs that offload NVGRE endpoint processing from the hypervisor vSwitch. The intelligent NICs would also have API interfaces for

[10] http://searchservervirtualization.techtarget.com/news/2240074318/VMware-Cisco-propose-VXLAN-for-VM-mobility

integration with overlay controllers and hypervisor management systems. Emulex has also demoed intelligent NICs that offload VXLAN processing from the VMware Distributed Switches.

STT

Stateless Transport Tunneling (STT) is a third overlay technology for creating Layer 2 virtual networks over a Layer 2/3 physical network within the data center. Conceptually, there are a number of similarities between VXLAN and STT. The tunnel endpoints are typically provided by hypervisor vSwitches, the VNID is 24 bits wide, and the transport source header is manipulated to take advantage of multipathing. STT encapsulation differs from NVGRE and VXLAN in two ways. First, it uses a stateless TCP-like header inside the IP header which allows tunnel endpoints within end systems to take advantage of TCP segmentation offload (TSO) capabilities of existing TOE server NICs. The benefits to the host include lower CPU utilization and higher utilization of 10 GbE access links. STT generates a source port number based on hashing the header fields of the inner packet to ensure efficient load balancing over LAGs and multi-pathing fabrics. STT also allocates more header space to the per-packet metadata, which provides added flexibility for the virtual network control plane. With these features, STT is optimized for hypervisor vSwitches as the encapsulation/decapsulation tunnel endpoints.

The STT IETF draft sponsored by Nicira does not specify a control plane solution. However, the Nicira network virtualization solution includes OpenFlow-like hypervisor vSwitches and a control plane based on a centralized network virtualization controller that facilitates management of virtual networks.

Shortest Path Bridging MAC-in-MAC (SPBM)

IEEE 802.1aq SPBM uses IEEE 802.1ah MAC-in-MAC encapsulation and the IS-IS routing protocol to provide Layer 2 network virtualization and VLAN extension in addition to the loop-free equal cost multi-path Layer 2 forwarding functionality normally associated with SPB.. VLAN extension is enabled by the 24 bit Virtual Service Network (VSN) Instance Service IDs (I-SID) that are part of the outer MAC encapsulation. Unlike other network virtualization solutions, no changes are required in the hypervisor vSwitches or NICs and switching hardware already exists that supports IEEE 802.1ah MAC-in-MAC encapsulation. For SPBM, the control plane is provided by the IS-IS routing protocol.

SPBM can also be extended to support Layer 3 forwarding and Layer 3 virtualization as described in the IP/SPB IETF draft using IP encapsulated in the outer SPBM MAC. This draft specifies how SPBM nodes can perform Inter-ISID or inter-VLAN routing. In addition, IP/SPB also provides for Layer 3 VSNs by extending Virtual Routing and Forwarding (VRF) instances at the edge of the network across the SPBM network without requiring that the core switches also support VRF instances. VLAN-extension VSNs and VRF-extension VSNs can run in parallel on the same SPB network to provide isolation of both Layer 2 and Layer 3 traffic for multi-tenant environments. With SPBM, all the core switches (starting at the access or aggregation switches that define the SPBM boundary) need to be SPBM-capable. SPBM hardware switches are currently available from Avaya, Huawei, and Alcatel-Lucent.

A discussion of network virtualization would not be complete without at least a mention of two Cisco protocols: Overlay Transport Virtualization (OTV) and Locator/ID Separation Protocol (LISP). OTV is optimized for inter-data center VLAN extension over the WAN or Internet using MAC-in-IP encapsulation. It prevents flooding of unknown destinations across the WAN by advertising MAC address reachability using IS-IS routing protocol extensions. LISP is an encapsulating IP-in-IP technology that allows end systems to keep their IP address (ID) even as they move to a different subnet within the network (Location). By using LISP VM-Mobility, IP endpoints such as VMs can be relocated anywhere regardless of their IP addresses while maintaining direct path routing of client traffic. LISP also supports multi-tenant environments with Layer 3 virtual networks created by mapping VRFs to LISP instance-IDs.

Another way to implement network virtualization is by implementing a Software Defined Network (SDN). SDN is the subject of a subsequent section of The Report.

Network Convergence and Fabric Unification

In contrast to Second Generation Data Center LANs:

> ***A possible characteristic of Third Generation Data Center LANs will be the convergence of block-level storage and data traffic over a common high-speed Ethernet data center switching fabric.***

This unified fabric offers significant cost savings in multiple areas including converged network adapters on servers and a reduction in rack space, power and cooling capacity, cabling, and network management overhead.

Traditional Ethernet, however, only provides a best effort service that allows buffers to overflow during periods of congestion and which relies on upper level protocols such as TCP to manage congestion and to recover lost packets through re-transmissions. In order to emulate the lossless behavior of a Fibre Channel (FC) SAN, Ethernet needs enhanced flow control mechanisms that eliminate buffer overflows for high priority traffic flows, such as storage access flows. Lossless Ethernet is based on the following standards, which are commonly referred to as IEEE Data Center bridging (DCB):
.

- **IEEE 802.1Qbb Priority-based Flow Control** (**PFC**) allows the creation of eight distinct virtual link types on a physical link, with each virtual link mapped to an 802.1p traffic class. Each virtual link can be allocated a minimum percentage of the physical link's bandwidth. Flow is controlled on each virtual link via the pause mechanism which can be applied on a per priority basis to prevent buffer overflow, eliminating packet loss due to congestion at the link level. In particular, block-level or file-level storage traffic on one of the virtual lanes can be protected from loss by pausing traffic on one or more of the remaining lanes.

- **IEEE 802.1Qau Congestion Notification (CN)** is a traffic management technique that eliminates congestion by applying rate limiting or back pressure at the edge of the network in order to protect the upper network layers from buffer overflow. CN is intended to provide lossless operation in end-to-end networks that consist of multiple tiers of cascaded Layer 2 switches, such as those typically found in larger data centers for server interconnect, cluster interconnect and to support extensive SAN fabrics.

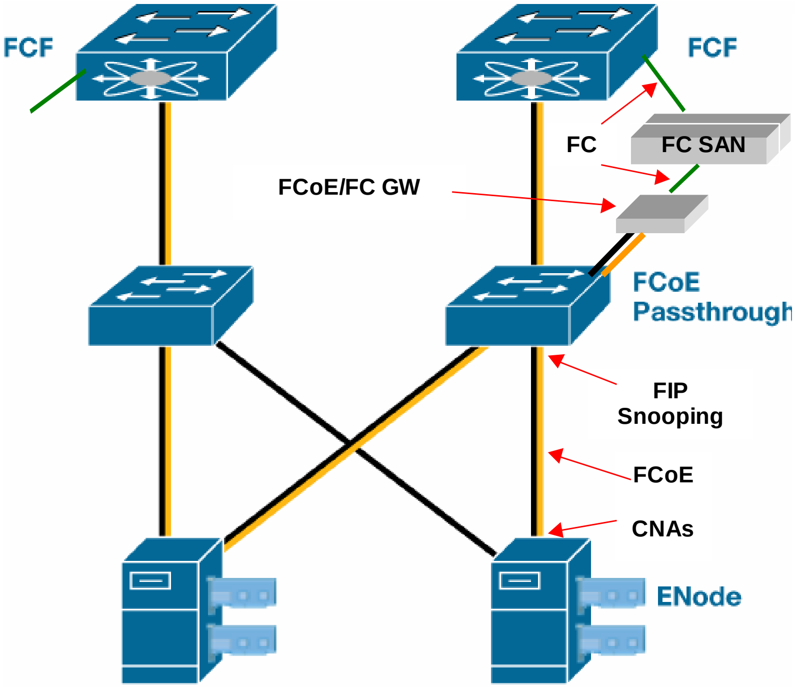
- **IEEE 802.1Qaz Enhanced Transmission Selection (ETS)** specifies advanced algorithms for allocation of bandwidth among traffic classes including the priority classes supported by 802.1Qbb and 802.1Qau. While the queue scheduling algorithm for 802.1p is based on strict priority, ETS will extend this by specifying more flexible drop-free scheduling algorithms. ETS will therefore provide uniform management for the sharing of bandwidth between congestion managed classes and traditional classes on a single bridged network. Priorities using ETS will coexist with priorities using 802.1Qav queuing for time-sensitive streams. **The Data Center Bridging Exchange (DCBX)** protocol is also defined in the 802.1Qaz standard. The DCBX protocol is an extension of the Link Layer Discovery Protocol (LLDP) that allows neighboring network elements to exchange request and acknowledgment messages to ensure consistent DCB configurations. DCBX is also used to negotiate capabilities between the access switch and the adapter and to send configuration values to the adapter.

DCB Lossless Ethernet will play a key role in supporting Fibre Channel over Ethernet (FCoE) technology that will allow the installed base of Fibre Channel storage devices and SANs to be accessed by Ethernet-attached servers with converged FCoE network adapters over the unified data center switching fabric. DCB will benefit not only block-level storage, but also all other types of loss and delay sensitive traffic. In the storage arena, DCB will improve NAS performance and will make iSCSI SANs based on 10/40/100 GbE a more competitive alternative to Fibre Channel SANs at 2/4/8/16 Gbps. In order to take full advantage of 10 GbE and higher Ethernet bandwidth, servers accessing iSCSI storage resources may also need intelligent converged NICs that offload iSCSI and TCP/IP processing from the host.

Fibre Channel over Ethernet (FCoE) is an industry standard that is being developed by the International Committee for Information Technology Standards (INCITS) T11 committee.

The FCoE protocol specification maps Fibre Channel upper layer protocols directly over a bridged Ethernet network. FCoE provides an evolutionary approach to the migration of FC SANs to an Ethernet switching fabric while preserving Fibre Channel constructs and providing reliability, latency, security, and traffic management attributes similar to those of native FC. FCoE also preserves investments in FC tools, training, and SAN devices; e.g., FC switches and FC attached storage. Implementing FCoE over a lossless Ethernet fabric requires converged server network adapters (e.g., CNAs with support for both FCoE and IP) and some form of FC Forwarding Function (FCF) to provide attachment to native FC devices (FC SAN switches or FC disk arrays). FCF functionality can be provided by a FCoE switch with both Ethernet and FC ports or by a stand alone gateway device attached to a FCoE passthrough switch, as shown in **Figure 10**.

Figure 10: FCoE Converged LAN Source: Cisco Systems



As shown in **Figure 10**, End Nodes (servers) don't need to connect directly to a FCF capable switch. Instead the FCoE traffic can pass through one or more intermediate FCoE passthrough switches. The minimal requirements for a simple FCoE passthrough switch is support for lossless Ethernet or DCB. The FCoE Initialization Protocol (FIP) supports handshaking between a FCoE End Node and an FCF in order to establish and maintain a secure virtual FC link between these devices, even if the end-to-end path traverses FCoE passthrough switches. For DCB passthrough switches that support FIP Snooping, the passthrough switches can inspect the FIP frames and apply policies based on frame content. FIP Snooping can be used to enhance FCoE security by preventing FCoE MAC spoofing and allowing auto-configuration of ACLs.

As this discussion illustrates:

There are several levels of support that data center switch vendors can provide for FCoE.

For example:

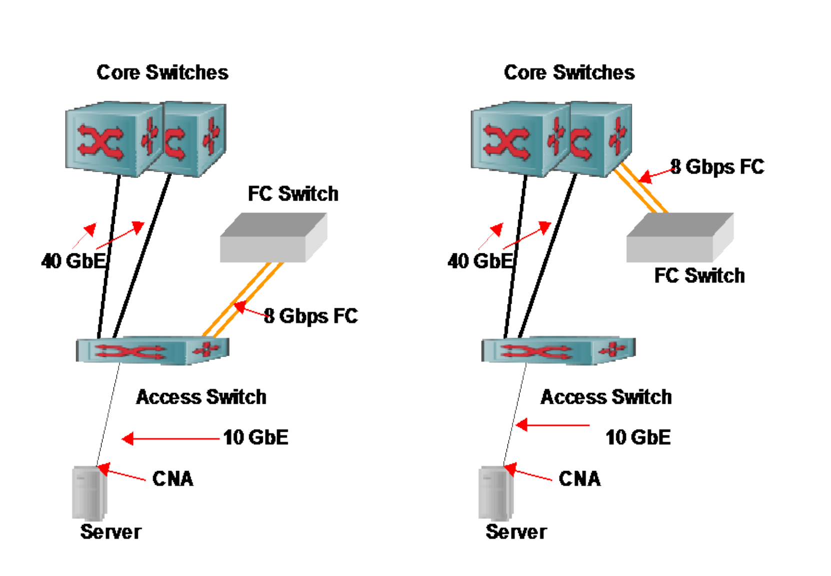
1. The lowest level of support is FCoE passthrough via lossless Ethernet or DCB alone.
2. The next step up is to add FIP Snooping to FCoE passthrough switches.
3. A third level of support is to add standalone FCF bridges/gateways to front end FC SAN switches or disk arrays.

4. The highest level of support is to provide DCB and FIP Snooping for FCoE passthrough switches and also to provide FCoE switches that incorporate FCF ports, creating hybrid switches with both DCB Ethernet and native FC ports.

Most vendors of Ethernet data center switches that don't also have FC SAN switches among their products are planning FCoE support at levels 1, 2, or 3 described above. In fact, most of these Ethernet-only vendors are considerably more enthusiastic about iSCSI SANs over 10/40/100 GbE than they are about FCoE.

The primary drivers of FCoE are the vendors that offer both Ethernet and FC products.

Figure 11: FCF Support Options



These are the vendors that are already shipping lossless 10 GbE Ethernet switches and hybrid lossless 10 GbE/FCF switches. Even among the vendors providing early support for FCF there are some significant differences, as shown in **Figure 11**.

The left side of the figure shows single hop FCoE with the FCF function integrated into the access switch. It would also be possible to use intervening FCoE/FCF gateways, either standalone or incorporated in the FC switch, which would be connected to the access switch via 10 GbE, making the access switch an FCoE passthrough switch, as shown in the previous figure. The advantage of single hop FCoE is that the storage traffic doesn't compete for bandwidth in the uplinks or the core switches and the core switches aren't required to support DCB or FIP Snooping. The right side of the figure shows multihop FCoE with the FCF function integrated into the core switch, and the access switch in FCoE passthrough mode. Again it would be possible to use FCoE/FCF gateways, either standalone or incorporated in the FC switch, connected to the core switch via 10 GbE. FC SANs and disk arrays connected at the core offer the advantage of a more centralized pool of storage resources that can be shared across the data center LAN.

Security Services in Virtualized Data Centers

As pointed out in the first section of The Report, security is generally considered by enterprise IT departments to be the primary concern in today's highly virtualized data centers and in the implementation of private or public cloud computing environments. In the traditional data center, internal security has generally been implemented by deploying dedicated physical security appliances at the Aggregation layer of a 3-tier or 2-tier network. This reduces the number of physical devices required and allows firewalls to filter traffic flowing from one access

VLAN to another. This approach has been successful in relatively static non-virtualized environments that require infrequent changes to the location and configuration of both servers and physical security appliances. This traditional model does not address inspection of inter-VM traffic within a single physical server.

With the advent of server virtualization and the dynamic migration of workloads within and between data centers, there is a growing need to make the workload's complete security environment as easily provisioned and migrated as the VMs themselves. In addition to being dynamic and virtualization-aware, the security solution needs to be both scalable and automated to the degree possible.

For enterprise data centers and Private Cloud Networking, the prevalent traffic isolation solution has been to make extensive uses of VLANs to isolate VMs performing different workloads or different aspects of the workload (e.g., web, application, and database tiers). In addition, firewalls, Intrusion Prevention Systems and other security appliances are generally required to filter and monitor inter-VM and inter-VLAN traffic in order to provide an additional layer of security for critical workloads and data resources.

In multi-tenant environments, it is highly desirable to be able to secure traffic within the tenant network as well as firewalling traffic at the tenant edge. The problem is most significant in highly virtualized IaaS data centers where a physical server may host VMs from multiple clients. In order to meet the demand for highly dynamic provisioning of resources IaaS service providers will focus on maximizing the use of virtual security appliances rather than physical devices. Traffic isolation in multi-tenant environments will be increasingly based on network virtualization based on either overlays or OpenFlow or possibly a combination of these techniques.

One approach for securing highly virtualized server environments is to use virtual security appliances on the same servers as the virtualized applications. Virtual appliances can be dynamically provisioned and migrated along with application VMs. Some virtual security appliances can support multiple security functions in a single VM. A virtual security appliance integrated with the hypervisor vNICs can provide security services for all the VMs on a host, inspecting both inter-VM traffic and traffic from external sources. Where the virtual security appliance also supports routing functionality, it can also inspect inter-VLAN traffic on the same host. When the VMs and the virtual security appliances are on separate VLANs and on separate hosts, traffic between them is typically switched at the Layer 3 tier of the physical network (typically at the aggregation layer). This means that a significant volume of security traffic may have to make a rather inefficient round trip through the physical network even if the application VM and the virtual security appliance are in the same POD or even on the same physical server (i.e., where the virtual security appliance doesn't support routing).

A second approach, more applicable in enterprise data centers because it does not involve virtual appliances on the servers, is to deploy a virtualized physical security appliance that can support a large number of instances of virtual security devices, such as firewalls, IDS/IPS, WAF, etc. Potentially. these instances could be implemented as VMs running on the security device's hypervisor. This type of integrated security device can also include its own physical Layer 2 and Layer 3 switching functionality, which allows the device to be installed in line between the access and aggregation layers of the physical data center LAN. The VLANs used by the virtualized servers are trunked to the virtualized security appliance via the hypervisor vSwitches and the physical access switches. There are a number of benefits of the integrated virtualized security appliance including:

- Specialized or dedicated hardware support for a number of security functions
- Ability to flexibly serialize different security services (firewall, IPS, etc) without having to change switch configurations or install additional physical security appliances
- Support for dynamic changes to security configurations for traffic among VLANs
- Ability to switch inter-POD security traffic without involving the aggregation layer switches

With the advent of DVSs and Layer 2 network virtualization using overlays, network partitioning can be based on virtual Ethernet overlay networks rather than simple VLANs. This vastly increases the number of virtual networks that can co-exist in the enterprise or multi-tenant IaaS data center and provides support for overlapping IP addresses among multiple tenants. Also, because a virtual network can span Layer 3 boundaries, VMs on the same physical server can communicate with each other across subnet boundaries via the DVS without involving Layer 3 switching in upstream physical devices. This can optimize securing local communication between co-resident VMs running different applications on separate subnets or VMs accessing the security services provided by co-resident virtual appliances on separate subnets. The overlay tunnels eliminate the need for inline security services and makes it possible to direct traffic to security services provided by virtual or physical security devices anywhere in the network.

As noted earlier, another potential approach to network virtualization is based on OpenFlow. The OpenFlow network can potentially be partitioned into multiple virtual networks based on certain characteristics of the 12-tuple used to differentiate flows. Each of the OpenFlow virtual networks can have its own independent OpenFlow controller, providing isolation of virtual networks at the control plane as well as the data plane. OpenFlow also provides a high degree of flexibility where the controller can direct flows to either physical security devices or virtual security appliances. It is also possible that the OpenFlow controller itself would provide some of the security services required.

Summary of Third Generation Data Center LAN Technologies

The data center LAN is still in the throes of rather dramatic technology developments, summarized in **Table 6**. As shown in the table, a number of standards have been completed in the last year or so, creating the expectation that more products supporting these standards will be announced in the near future.

Table 6: Status of Data Center Technology Evolution

Technology Development	Status
Two-tier networks with Layer 2 connectivity extending VLANs across the data center.	On-going deployment
Standardized edge virtualization automating Layer 2 configuration for VM creation and mobility. Possible changing role for the hypervisor vSwitch as a port aggregator (VEPA) for EVB, potentially eliminating the vSwitch tier.	The 802.1Qbg standard is in place and some implementations are available.
Reduced role for blade switches to eliminate switch tier proliferation.	On-going with proprietary fabric extenders. Work on the IEEE802.1BR standard is in progress

Table 6: Status of Data Center Technology Evolution

Technology Development	Status
Multi-chassis LAG and switch virtualization technology to address STP issues and provide active-active redundant server connectivity.	On-going deployment
Multi-core servers with notably more VMs per server and 10 GbE connectivity to the LAN.	Adoption stage.
40 GbE and 100 GbE uplinks and core switches.	A standard has been in place for some time: 40 GbE is becoming widely available on access and core switches 100 GbE is becoming available. But adopted primarily by service providers due to economic considerations
TRILL enabling new Layer 2 data center LAN topologies; e.g., fully meshed, fat tree with equal cost multi-path forwarding	The TRILL standard RFC 6325 has also been approved. Enhancement being proposed to IETF. Pre-standard switch implementations of TRILL with proprietary extensions are available. No standard TRILL yet.
SPB enabling new Layer 2 data center LAN topologies; e.g., fully meshed, fat tree with equal cost multi-path forwarding	SPB (IEEE 802.1aq) has been finalized and switch products are available.
SPB Network Virtualization	Layer 2 virtualization covered in IEEE 802.1aq. Products are available. Layer 3 virtualization is the subject of an Internet draft and implemented by Avaya in its SPB switches
VXLAN Network Virtualization	A draft was recently submitted to the IETF. Pre-standard implementations are available in vSwitches and some access switches
NVGRE and STT Network Virtualization	Drafts were recently submitted to the IETF. STT is implemented by Nicira
SDN	Vendors are beginning to offer SDN solutions based on OpenFlow. ONF standards are limited to OpenFlow
OpenFlow	OF V1.0 hybrid switches and controllers are available from multiple vendors OF V1.3 spec has been released
DCB delivering lossless Ethernet for 10 GbE and higher speed Ethernet	Standards are in place. Switches with DCB are available.
10 GbE FCoE approach to fabric unification	FCoE standard is in place and products are available
10 GbE iSCSI approach to fabric unification	Early implementations
Management tools that integrate, coordinate, and automate provisioning and configuration of server, storage and network resource pools	These are proprietary and have varying levels of maturity.

About the Webtorials® Editorial/Analyst Division

The Webtorials® Editorial/Analyst Division, a joint venture of industry veterans Steven Taylor and Jim Metzler, is devoted to performing in-depth analysis and research in focused areas such as Metro Ethernet and MPLS, as well as in areas that cross the traditional functional boundaries of IT, such as Unified Communications and Application Delivery. The Editorial/Analyst Division's focus is on providing actionable insight through custom research with a forward looking viewpoint. Through reports that examine industry dynamics from both a demand and a supply perspective, the firm educates the marketplace both on emerging trends and the role that IT products, services and processes play in responding to those trends.

For more information and for additional Webtorials® Editorial/Analyst Division products, please contact Jim Metzler at jim@webtorials.com or Steven Taylor at taylor@webtorials.com.

Published by Webtorials Editorial/Analyst Division
www.Webtorials.com

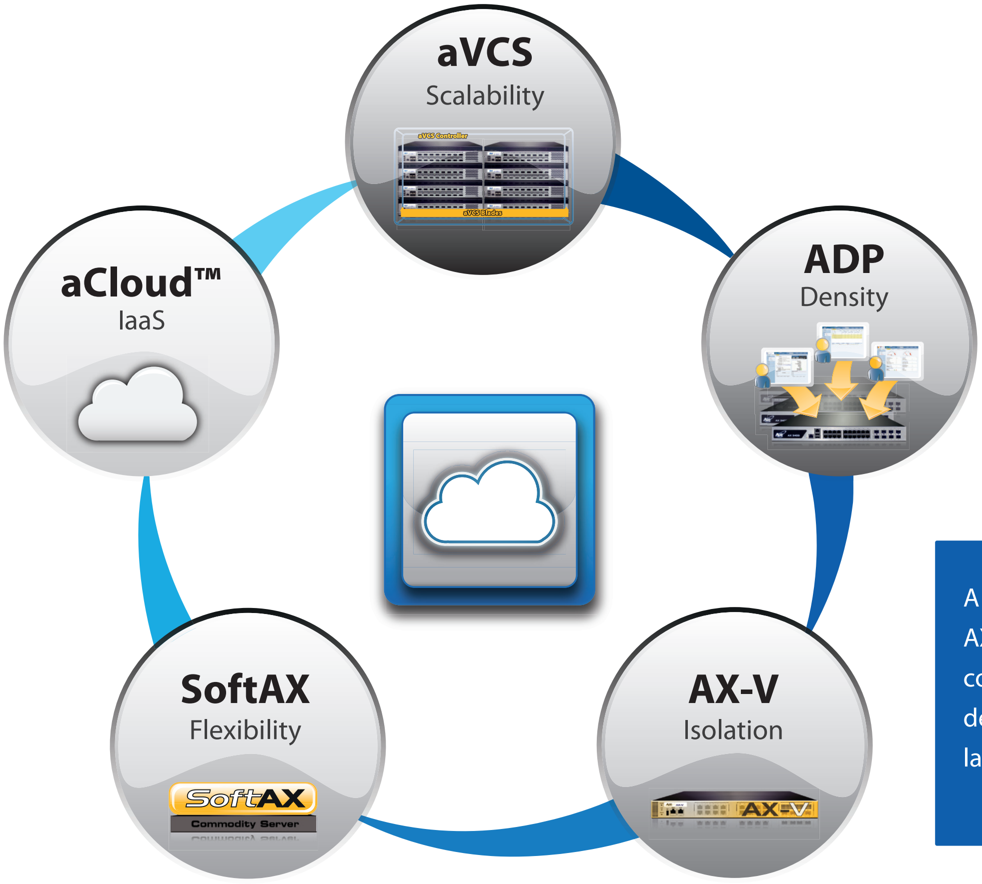
Division Cofounders:
Jim Metzler
jim@webtorials.com
Steven Taylor
taylor@webtorials.com

Professional Opinions Disclaimer
All information presented and opinions expressed in this publication represent the current opinions of the author(s) based on professional judgment and best available information at the time of the presentation. Consequently, the information is subject to change, and no liability for advice presented is assumed. Ultimate responsibility for choice of appropriate solutions remains with the reader.

For editorial and sponsorship information, contact Jim Metzler or Steven Taylor. The Webtorials Editorial/Analyst Division is an analyst and consulting joint venture of Steven Taylor and Jim Metzler.

Customer Driven Innovation

Cost Effective Cloud Networking | Virtualization as the Enabler



A10 also offers a powerful choice of AX Series ADC form factors with comprehensive management options, delivering flexibility and efficiency for large scale deployments.

AX Series Virtualization Products & Solutions

Based on A10's award-winning AX Series Application Delivery Controllers (ADC) and Advanced Core Operating System (ACOS™) architecture, enterprises and service providers will have the flexibility to choose the following scale-as-you-grow virtualization options.

SoftAX™

- SoftADC: AX virtual machine (VM) on commodity hardware
- Rapidly scale with commodity hardware
- Reduce hardware costs and upload to compatible cloud providers

AX-V Appliance

- SoftADC: AX virtual machine (VM) on AX Series hardware
- SoftAX flexibility with AX hardware performance and reliability
- Guaranteed performance, certifications, support and optimized hardware

AX Virtual Chassis System (aVCS™)

- Cluster multiple AX devices to operate as a unified single device
- Scale while maintaining single IP management
- Reduce cost and simplify management while adding devices as you grow

Application Delivery Partitions (ADPs)

- Divide the AX platform resources for individual applications
- Enables quality multi-tenancy with granular resource allocation
- Reduce the number of appliances to host multiple applications

The Application Fluent Data Center Fabric

Introduction

The rise of virtualization and cloud computing requires the selection of a best-of-breed data center switching solution as part of an enterprise's overall data center strategy. And at the heart of this strategy is the need to deliver a high quality user experience with new virtualized applications, including video, on new devices such as smart phones and tablets. However, the traditional 3-layer networks designed for a client/server communication model cannot meet the requirements of these new applications and devices, nor can it address the new requirements of virtualized servers and desktops.

Application Fluency for the Data Center

Resilient Architecture

- Simplified 10 & 40 GigE network with low latency and ready for 100 GigE
- Multi-path data center network extends between data center sites and to public cloud
- Supports definition of virtual data centers
- Ready for storage convergence with lossless Ethernet

Automatic Controls

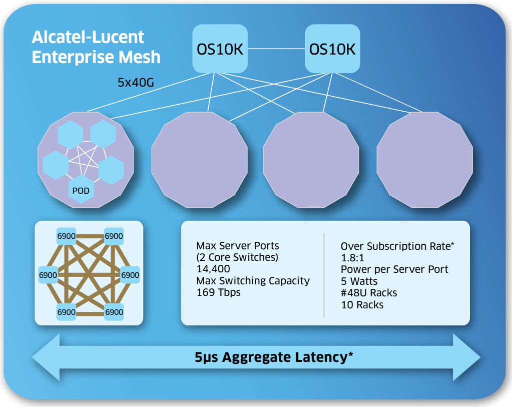
- Application profiles ensure that the network is aware of application provisioning, security and QoS requirements
- The network will automatically sense virtual machine location and movement
- The network will automatically adjust to VM motion within and between data center sites

Streamlined Operations

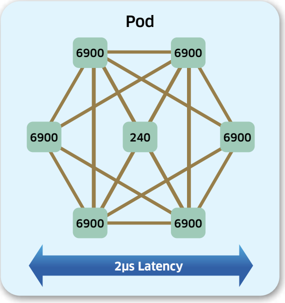
- Applications are automatically provisioned
- Core switches automatically configure top of rack switches
- Converged management for data center network and virtual machine mobility
- Low power consumption

The Alcatel-Lucent Mesh

Alcatel-Lucent provides a unique Application Fluent approach to maximize the benefit from virtualization technologies for servers, the desktop, as well as the network. Alcatel-Lucent's application fluent data center fabric can scale from several hundred to over 14,000 server facing ports while keeping aggregate latency at 5ms, and can automatically adapt to virtual machine movement no matter which server virtualization platform is used.



The Alcatel-Lucent Virtual Network Profile (vNP), embedded in the Alcatel-Lucent Mesh, includes the critical information the fabric needs to understand each application, including provisioning requirements, security profiles, and expected quality of service levels. With this knowledge, the network can manage applications as services, including automatically discovering the location of each virtual machine, modifying the network configuration to follow virtual machine moves and providing an integrated view on visibility on VM movement and current location from a network perspective.



Application fluency in the corporate data center includes its transformation into a multi-site private cloud by extending layer 2 connectivity between data center sites and allowing for seamless delivery of public cloud-based services on the corporate network.

The Alcatel-Lucent Mesh enables enterprises to provide a high quality user experience with mission critical, real-time applications, and to improve agility in deploying new applications while significantly reducing data center costs.

Open Ecosystems and Market Success

Alcatel-Lucent Enterprise is committed to open standards, allowing enterprises to select best-of-breed suppliers for their complete data center solution: servers, storage, data center fabric, and data center interconnect.

- Winner: Best of Interop 2011 for Data Center Switching and Storage
- Data center ecosystem partners include Emulex, NetApp, VMware, Citrix, and QLogic
- Participant in IEEE sponsored Shortest Path Bridging interoperability test with Avaya, Huawei, Solana and Spirent
- Over 20 million Ethernet ports shipped

For More Information

Alcatel-Lucent Data Center Switching Solution
Alcatel-Lucent Application Fluent Networks
Alcatel-Lucent Enterprise

Assuming Server to Server Traffic 70% within a Pod, 20% between Pods and 10% Via Core

Alcatel·Lucent
Enterprise

www.alcatel-lucent.com Alcatel, Lucent, Alcatel-Lucent and the Alcatel-Lucent logo are trademarks of Alcatel-Lucent. All other trademarks are the property of their respective owners. The information presented is subject to change without notice. Alcatel-Lucent assumes no responsibility for inaccuracies contained herein.

Visibility. Control. Optimize SaaS, BYOD, and Social Media

How to Lower Networking Costs and Safely Improve Performance

So many of the dominant trends in applications and networking are driven from outside the organization, including software-as-a-service (SaaS), bring-your-own-device (BYOD), Internet streaming video, and social networking. These technologies of an Internet connected world are fundamentally changing how we live and work every day. Yet, Network Administrators struggle to see and control these traffic streams from the Internet.

As businesses have opened their networks to SaaS applications, users are quickly starting using business bandwidth to access recreational websites and download BYOD updates, applications, and upload photos, videos and backups. This has created overburdened networks and slows the response of both cloud-based and internally delivered applications.

But with Visibility and Control from Blue Coat, Network Administrators can see all traffic on their networks and apply policies that can separate and control application traffic, and ensure internal and SaaS application performance.

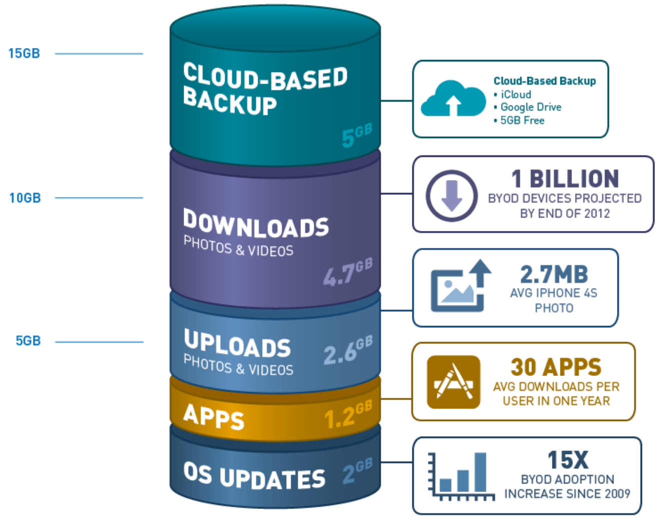
First: Visibility of all traffic on all ports – Understand what is on your network

Blue Coat PacketShaper leverages Blue Coat WebPulse™, an Internet Intelligence Service powered by a global community of 75 million users, the Cloud Service is able to deliver real-time categorization of Internet applications and web traffic.

WebPulse is based on sound analysis-system design principles:

- Massive input: WebPulse analyzes up to 1 billion web requests per day.
- In-depth analysis: 16 layers of analysis support over 80 categories in 55 languages.
- Granular policy: Up to 4 categories can be applied to each web request for multi-dimensional ratings.
- Speed: Automated systems process inputs – in most cases, in real time.
- Results: This collective intelligence allows WebPulse to categorize new Internet applications and websites quickly to PacketShaper without software updates/upgrades.

BYOD BANDWIDTH CONSUMPTION - JUNE 2011 TO JUNE 2012



The graphic details the impact of BYOD and Recreational video traffic can have on a network if left unchecked.

Second: Optimize Performance

SaaS, BYOD, Video and Social Media present challenges to network capacity and user patience. Blue Coat WAN Optimization helps overcome these challenges.

Chatty protocols and multi-megabyte files can hurt SaaS performance. Video requirements destroy capacity plans. Blue Coat's asymmetric, on-demand video caching and live stream splitting boost video capacity up to 500x – whether it's corporate or recreational video. For SaaS, our CloudCaching Engine improves performance by 3-93x, dramatically raising productivity for SaaS users at branch locations.

And now Blue Coat ProxySG/MACH5 technology secures SaaS applications as it accelerates their performance. ProxySG/MACH5 connects directly to the Blue Coat Cloud Service, enforcing SaaS user policies and leveraging WebPulse to scan and filter cloud traffic. Branch users can access applications like SAP, Salesforce, and RightNow without the burden of bandwidth slowdowns or risk of malware threats.

On The Road To The Cloud?

With Converged Infrastructure Management and Network Automation, CA Technologies' allows you to transform your IT management functionality…reduce complexity and proactively optimize infrastructure while reducing costs…for a superior customer experience.

The Cloud Challenge…Increasingly CIO'-s and CEO'-s are looking to the IT organization to help deliver differentiation to the marketplace through innovation. As well, some organizations are looking to the Cloud to help them become more agile. Today "Cloud is synonymous with "Agility" but can you ensure your business services and guarantee application performance and availability in the cloud? How can you be proactive and optimize your infrastructure for lower costs while still delivering the highest quality user experience?

Cloud-Enable Your Network…CA Technologies Converged Infrastructure Management delivers ease of use and simple deployment while getting you up and running quickly with prescriptive OOTB capabilities- the benefit of IT organizations that say "It works as advertised." As well as functionality that can go deeper for dedicated IT teams showing them visibility into the infrastructure they specifically manage.

Access a single user interface for actionable performance, availability, flow capacity and application response information for all Layer 2 and Layer 3 technologies.

CA Technologies Converged Infrastructure Management delivers up to 25X Faster Problem Resolution While Reducing Total Cost by as Much as 50%. It helps you deliver a superior, differentiated customer experience – quickly and economically while -

Speeding proactive triage and remediation with less effort

- Analytics translate disparate data into intelligent views for up to 25x faster problem resolution

Meeting massive scalability demands cost-effectively

- Monitoring leading nationwide voice and video network with only two management servers

Shifting operations costs to innovation

- Converged infrastructure management reduces total costs by as much as 50%

Improving revenue streams

- Generate differentiated new sources of revenue and onboard new clients faster

The Cloud and Network Automation…CA Technologies Network Automation enables cloud-readiness all across your network, making your operation more efficient, more cost-effective and safer. Automation allows your workers to be more productive, improves your compliance and security issues, diminishes the risk of failure and ensures safe and immediate disaster recovery.

Automated dashboard for data collection and analysis to improve remediation options like manual time and level of effort.

Just some of the ways Network Automation helps enable Cloud is:

- Tasking over manual, error-prone processes of provisioning network devices.
- Detecting network changes and addressing their impact with troubleshooting and notifying in real time when issues are detected.
- Knowing and showing who is on the network, where and when at any given time, as well as archiving historical configurations.
- Updating network configuration changes on a wide number of devices from a central location automatically.
- Obtaining a current inventory of all components on the network and detecting policy and compliances failures in real time.
- Backing up all network configuration son a near real time basis, allowing restoration to take place in a matter of minutes.

Whether you are looking for ease-of-use, enterprise scalability or automation on your journey to the cloud, CA Technologies will help you deliver the innovation and agility that today's business services demand.

Visit us at http://www.ca.com/converge or http://www.ca.com/us/it-automation.aspx

Simplify and Accelerate Private Cloud Deployments with Cisco’s Virtual Networking Portfolio

Cisco and a Multi-Vendor Ecosystem Provide Cloud-ready Network Solutions

ROLE OF THE NETWORK PLATFORM IN CLOUD

Access to Critical Data, Services, Resources and People

- Core fabric connects resources within the data center and data centers to each other
- Pervasive connectivity links users and devices to resources and each other
- Network provides identity- and context-based access to data, services, resources and people

Granular Control of Risk, Performance and Cost

- Manages and enforces policies to help ensure security, control, reliability, and compliance
- Manages and enforces SLAs and consistent QoS within and between clouds, enabling hybrid models and workload portability
- Meters resources and utilization to provide transparency for cost and performance

Robustness and Resilience

- Supports self-healing, automatic redirection of workload and transparent rollover
- Provides scalability, enabling on-demand, elastic computing power through dynamic configuration

Innovation in Cloud-specific Services

- Context-aware services understand identity, location, proximity, presence, and device
- Resource-aware services discover, allocate, and pre-position services and resources
- Comprehensive insight accesses and reports on all data that flows in the cloud

The Power of Cloud for the Enterprise

Business and IT executives are confronted daily by conflicting and exaggerated claims of how cloud will transform their industries, but the lure of transformative efficiency and agility is hard to ignore. Understanding the objectives and obstacles to cloud, as well as the solutions to overcome those obstacles is the key to achieving cloud-readiness.

Defining Cloud

In the simplest terms, cloud is IT delivered as a service over the network. Going a level deeper, cloud is a model in which IT resources and services are abstracted from the underlying infrastructure and provided on demand and at scale in a multi-tenant environment.

- *On demand* means that resources can be provisioned immediately when needed, released when no longer required, and billed only when used.
- *At scale* means the service provides the experience of infinite resource availability to meet whatever demands are made on it.
- *Multi-tenant environment* means that the resources are provided to many consumers - for example, business units -from a single physical infrastructure.

Note that the physical location of resources (on or off premises) is not a part of this statement. From the perspective here, that aspect has more to do with the way the cloud is sourced than with what the cloud does.

CISCO VIRTUAL NETWORK PORTFOLIO

Routing and Switching

- Cisco Nexus 1000V virtual switch
- Cisco Cloud Services Router (CSR) 1000V

Security and VPN

- Cisco Virtual Security Gateway for Nexus 1000V (included in Nexus 1000V Advanced Edition)
- Cisco Adaptive Security Appliance (ASA) 1000V Cloud Firewall

WAN Optimization

- Cisco Virtual Wide Area Application Services (vWAAS)

Network Analysis and Monitoring

- Cisco Prime Virtual Network Analysis Module (NAM)

Application Delivery Controllers

- Citrix NetScaler VPX virtual application delivery controller

Virtual Services Deployment Platform

- Cisco Nexus 1100 Series Virtual Services Appliance

Cloud Orchestration and Management

- Cisco Intelligent Automation for Cloud
- Cisco Virtual Network Management Center (VNMC)

To learn more about Cisco’s complete virtual networking portfolio: http://cisco.com/go/1000v

Barriers to Adoption

While most enterprises have recognized the potential benefits of cloud, practical concerns and perceived challenges have hampered the widespread adoption of cloud technologies and services. Many of these barriers can be understood as questions of trust: Can the cloud be trusted to deliver the same capabilities at the same service levels in the same controlled way as traditional IT?

- **Security**: Can the same security available to applications be applied in the cloud?
- **Compliance**: Can applications in the cloud meet the same regulatory compliance requirements?
- **Reliability and quality of service (QoS)**: Can the same service-level agreements (SLAs) for reliability and QoS be met in the cloud, especially given the multi-tenant use of the underlying IT infrastructure?
- **Control**: Can application owners still have the same amount of control over their applications and the infrastructure supporting them in the cloud?
- **Fear of vendor lock-in**: Will use of a particular vendor for cloud services or infrastructure prevent use of a different one in the future, or will the enterprise’s data and applications be tightly locked into a particular model?

These concerns represent questions of technology and governance, but do not address any potential organizational friction that might arise from adopting cloud. For example, who will manage which part of the cloud or who will determine which applications to migrate to the cloud. Cisco believes that all these concerns can be met with the right technology, architecture, and approach.

Practical Solutions for Cloud-ready Virtual Networks and Infrastructure

The Cisco Virtualized Multi-Tenant Data Center (VMDC) architecture provides an end-to-end architecture and design for a complete private cloud providing IaaS capabilities. VMDC consists of several components of a cloud design, from the IT infrastructure building blocks to all the components that complete the solution, including orchestration for automation and configuration management. The building blocks are based on stacks of integrated infrastructure components that can be combined and scaled: Vblock™ Infrastructure Packages from the VCE coalition developed in partnership with EMC and VMware and the Secure Multi-Tenancy (SMT) stack developed in partnership with NetApp and VMware. Workload management and infrastructure automation is achieved using BMC Cloud Lifecycle Management (CLM). Clouds built on VMDC can also be interconnected or connected to service provider clouds with Cisco DCI technologies. This solution is built on a service delivery framework that can be used to host other services besides IaaS on the same infrastructure: for example, a virtual desktop infrastructure VDI).

These solutions for building private clouds are also being used by service providers to build cloud infrastructures on which to provide public, hybrid, and virtual private clouds to their enterprise customers. With service providers and enterprises, Cisco is developing an ecosystem of cloud providers, builders, and consumers. This ecosystem will be able to take advantage of common approaches to cloud technology, management, interconnection, and operation.

Where to Begin Your Cloud Journey

Cisco is working with its broad ecosystem of partners to assist some of the world's leading institutions in their initial cloud deployments. Cisco will have a central role in the unique journeys of enterprises, small and medium-sized businesses (SMBs), public-sector organizations, and service providers as they move to cloud.

When the topic of cloud comes up, the conversation often focuses on the newest technologies and the latest service provider offerings. However, Cisco believes that every conversation needs to begin with an understanding of the expected business outcomes. Is the goal lower total cost of ownership (TCO) or greater agility and innovation, or some blend of the two? The journey to cloud has many paths; starting the journey without a clear understanding of the destination can lead to disappointing results.

Enterprises should start the journey to cloud by answering some basic questions:

- What is the expected impact of cloud on my business?
- Which applications can and should I move to the cloud?
- What cloud deployment model is best suited for each of my applications?
- How do I maintain security and policy compliance in the cloud?
- How do I transition my organization to best take advantage of cloud?

The answers to these questions will fundamentally shape your cloud strategy. We are helping customers define and implement a pragmatic approach to cloud. We deliver solutions that address our customers' unique business architecture and needs, align with regulatory constraints, and are optimized according to the customer's individual preferences for performance, cost, and risk.

For More Information

As you begin your own journey to the cloud, we invite you to discuss the right approach for your organization with your Cisco account manager, channel partners, and other IT advisors. For additional information about cloud, please visit: http://www.cisco.com/go/cloud.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at **www.cisco.com/go/offices.**

Application Performance for Business Efficiency

The unique way to guarantee business application performance over the WAN, increase IT productivity and save on IT costs.

82% *

of organizations suffer application performance problems.

63% *

of organizations don't know the number of apps using the network.

72% *

of organizations use very occasionally their network to its full data transmission capacity.

Business and IT performance are tightly coupled…

Losing 5 minutes per day for poor application performance means 1% of productivity drop which can turn down profitability by 10%.

*****Ipanema Killer Apps survey 2012*

IT departments are witnessing change at a pace never seen before

Transformation is occurring as CIOs seek to access the benefits offered by Unified Communications, cloud computing, internet-based applications and consolidation, amongst many other strategic projects.

These initiatives are aimed at increasing enterprise's business efficiency. While they simplify the way IT is delivered to users, they increase the complexity and the criticality of corporate networking as applications and users rely more than ever on the continuous, reliable and consistent flow of data traffic.

In order to protect the business and the significant investments made in transformative applications such as Unified Communications and SaaS the network must be more intelligent, more responsive and more transparent. Ipanema's revolutionary self-learning, self-managing and self-optimizing Autonomic Networking System™ (ANS) automatically manages all its tightly integrated features to guarantee the application performance your business requires over the global network:

- Global Application Visibility
- Per connection QoS and Control
- WAN Optimization
- Dynamic WAN Selection
- SLA-based Network Rightsizing

Business efficiency requires guaranteed application performance

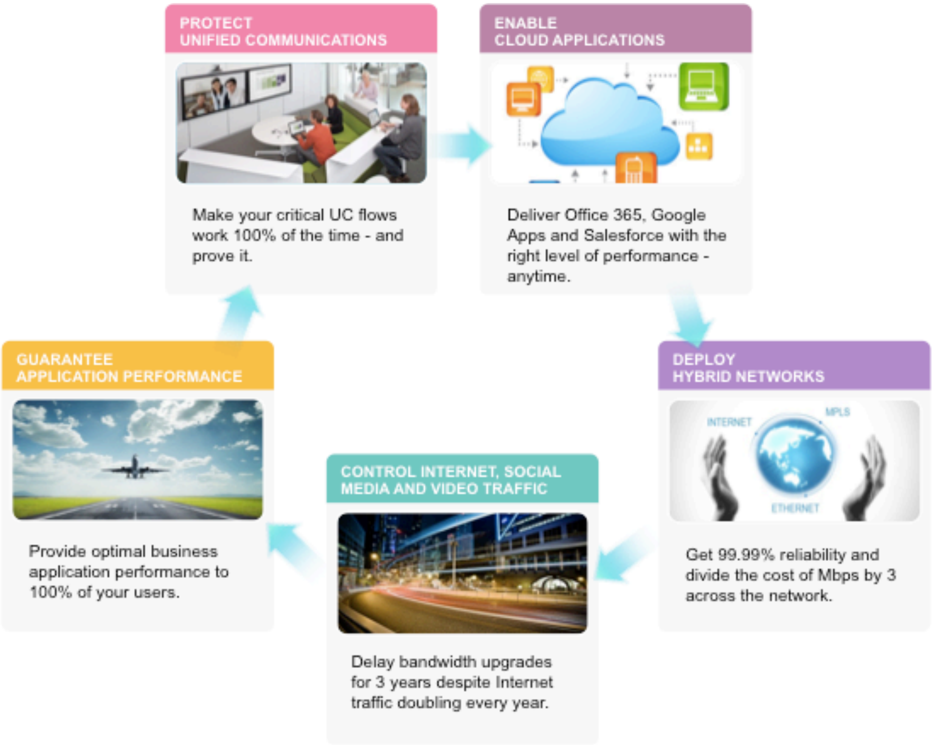
- Know which applications make use of your network…
- Guarantee the application performance you deliver to users…
- Manage cloud applications, Unified Communications and Internet growth at the same time…
- Do more with a smaller budget in a changing business environment, and to prove it…

With Ipanema, control all your IT transformations!

For $3/employee/month, you guarantee the performance of your business applications… and can save 10 times more!

Ipanema's global and integrated approach allows enterprises to align the application performance to their business requirements. With an average TCO of $3/employee/month, Ipanema directly saves x10 times more and protects investments that cost x100 times more:

- **Application performance assurance**: Companies invest an average of $300/employee/month to implement the applications that support their business. At a mere 1% of this cost, Ipanema can ensure they perform according their application SLAs in every circumstance, maximizing the users' productivity and customers' satisfaction. While they can be seen as "soft money", business efficiency and investment protection are real value to the enterprise.
- **Optimized IT efficiency**: Ipanema proactively prevents most of the application delivery performances problems that load the service desk. It automates change management and shortens the analysis of the remaining performance issues. Global KPIs simplify the implementation of WAN Governance and allow better decision making. This provides a very conservative direct saving of $15/employee/month.
- **Maximized network efficiency**: Ipanema's QoS & Control allows to at least doubling the actual capacity of networks, deferring upgrades for several years and saving an average of $15/employee/month. Moreover, Ipanema enables hybrid networks to get access to large and inexpensive Internet resources without compromising the business, typically reducing the cost per Mbps by a factor of 3 to 5.



PROTECT UNIFIED COMMUNICATIONS

Make your critical UC flows work 100% of the time - and prove it.

ENABLE CLOUD APPLICATIONS

Deliver Office 365, Google Apps and Salesforce with the right level of performance - anytime.

DEPLOY HYBRID NETWORKS

Get 99.99% reliability and divide the cost of Mbps by 3 across the network.

CONTROL INTERNET, SOCIAL MEDIA AND VIDEO TRAFFIC

Delay bandwidth upgrades for 3 years despite Internet traffic doubling every year.

GUARANTEE APPLICATION PERFORMANCE

Provide optimal business application performance to 100% of your users.

What our customers say about us:

Do more with less

"Whilst data volume across the Global WAN has increased by 53%, network bandwidth upgrades have only grown by 6.3%. With Ipanema in place we have saved $987k this year alone."

Guarantee Unified Communications and increase network capacity

"Ipanema is protecting the performance our Unified Communication and Digital Signage applications, improving our efficiency as well as our customers' satisfaction. Moreover, we have been able to multiply our available capacity by 8 while preserving our budget at the same time."

Reduce costs in a cloud environment

"With Ipanema, we guaranteed the success of our cloud messaging and collaboration deployment in a hybrid network environment, while dividing per 3 the transfer cost of each gigabyte over our global network."

NEC
Take control of your network.
GRAND PRIZE
WINNER
best of
INTEROP
Awards 2012
Grand Prize
PRESENTED BY: InformationWeek reports

Enabling the cloud:

Award-winning NEC ProgrammableFlow® Open Software Defined Networking…

…delivering automated, efficient, and agile networks for the cloud

NEC's ProgrammableFlow network suite was the first commercially available SDN solution to leverage the OpenFlow protocol—enabling network-wide virtualization, allowing customers to easily deploy, control, monitor, and manage multi-tenant network infrastructure in a cloud environment. This architecture delivers better utilization of all IT assets, and helps provide ongoing investment protection as customers add functionality or upgrade their networks. NEC's approach simplifies network administration and provides a programmable interface for unifying the deployment and management of network services with the rest of IT infrastructure.

Specific functions customers prize include:

- **Drag and drop network design:** The GUI interface to the ProgrammableFlow Controller includes the familiar CLI found on most routers and switches today, so with minimal training a network admin can easily point and click to design an entire network from the single pane provided by the PF6800. This can radically reduce network programming and design time and errors caused previously by human intervention.
- **VM mobility:** With the ability to readily direct traffic throughout the data center—or throughout multiple data centers, it is possible to better manage all of the resources in a data center. For example, in NEC's own data centers in Japan, where they have recently implemented the ProgrammableFlow Fabric, it has enabled them to spread traffic between East and West Japan, offloading servers in East Japan that were nearing capacity, and postponing purchase of new servers, for a substantial saving. VM Mobility also enabled Nippon Express to complete a data center consolidation move that normally would have taken 2 months down to 10 days.
- **Bandwidth monitoring and traffic flow visualization:** This feature of the PF6800 provides performance monitoring of network flows and centralized management of network traffic, reducing bottlenecks and enabling smooth, streamlined network operations with substantially improved network admin productivity.
- **Secure, multi-tenant networks:** Secure, multi-tenant networks from the PF6800 enables customers like Genesis Hosting to expand their service offering with new sources of revenue potential. Genesis also reports software engineering investments were reduced by 100 hours each month with the advancements provided by ProgrammableFlow multi-tenancy.
- **Automation and administration of business policy to network management:** With network services aligned with business policy, automation such as prioritizing classes of applications or specific applications over other enterprise activity during peak loads is now possible with the ProgrammableFlow Network Suite, with multiple paths provided automatically. These capabilities offer significant value, particularly to enterprises engaged in heavy transaction loads.
- **Load balancing:** Traditional networking protocols often lead to performance-reducing bottlenecks. ProgrammableFlow uses path selection algorithms to analyze traffic flow across the network, check all available paths, and customize traffic flows to maintain performance and fully utilize network capacity. This increases the utilization of the network and improves application performance.

Backed by a 100-year history
of technology innovation,
NEC helps customers improve
performance and solve their
toughest IT challenges.

To learn more about how NEC can help you optimize your network for the cloud, visit necam.com/pflow or call your NEC Account Manager today.

Empowered by Innovation NEC

® ProgrammableFlow is a registered trademark of NEC Corporation radware

Expand Your Cloud Offering with Advanced Cloud ADC Solutions

Challenges in the Cloud Provider Business

The broad adoption of cloud based services by enterprise organizations and the multiple entrants into the cloud and hosting business challenges cloud providers to differentiate their service offerings and attract customers. Cloud providers face multiple challenges in establishing their business.

radware Research by TechValidate

Enabling Cost-Effective Cloud Operations

“ Radware ADC enables us to support public/private cloud environments. ”

Source: Engineer, Large Enterprise Telecommunications Services Company

www.techvalidate.com/product-research/radware-application-delivery TVID: 085-477-B07

The first challenge is the infrastructure availability challenge. In an effort to provide uptime assurance at the base service level, or as a value added service offering, cloud providers must provide continuous availability of customer resources. One threat impacting the business availability is general connectivity: infrastructure outages and disruption events in which providers are dependent on external utilities and their running equipment. Failure to these can have significant adverse affect on the providers' business. Furthermore, part of the scalability value proposition of a cloud provider is the ability to scale-out application infrastructures – without load balancers, application scale-out is virtually impossible.

radware Research by TechValidate

Better SLA with Radware ADC

“ As a cloud provider, we use multiple Radware ADC services for our hosted applications to ensure our customers receive better SLA. ”

Source: Senior IT Manager, Medium Enterprise Computer Services Company

www.techvalidate.com/product-research/radware-application-delivery TVID: C6F-218-D63

Above all, cloud providers are pressed to build solutions with minimal capital expenditure, maintain low operational costs and rapidly meet spikes in customer demand. Flexible procurement models by vendors and platforms that are easily scalable and centrally managed support the overall operational constraints faced by cloud providers.

Radware Solutions for Cloud Service Providers

Radware offers a set of fully integrated infrastructure availability and security solutions to meet the demands of cloud providers worldwide. Radware's solutions are comprised of the following components as illustrated in the figure below:

- **Radware ADC-VX™** – highly scalable ADC virtualization and consolidation solution offering high speed global and local load balancing, application acceleration and SSL offloading that supports dynamic availability requirements of cloud customers. ADC-VX can host multiple fully isolated, fully featured vADC instances.

- **Radware Alteon VA®** – flexible virtual ADC instance running atop most commercial, general purpose x86 server hypervisors.

- **Radware VADI®** – comprehensive virtual application delivery infrastructure solution including Alteon VA and ADC-VX-based virtual ADCs (vADC) and vDirect, an ADC service automation plug-in that simplifies ADC service deployment in cloud environments.

Radware's solutions enable cloud providers and hosts to offer more reliable and scalable infrastructure services to their customers. Resilience and scalability are key attributes of a cloud service as enterprises are contemplating the extent of cloud service adoption.

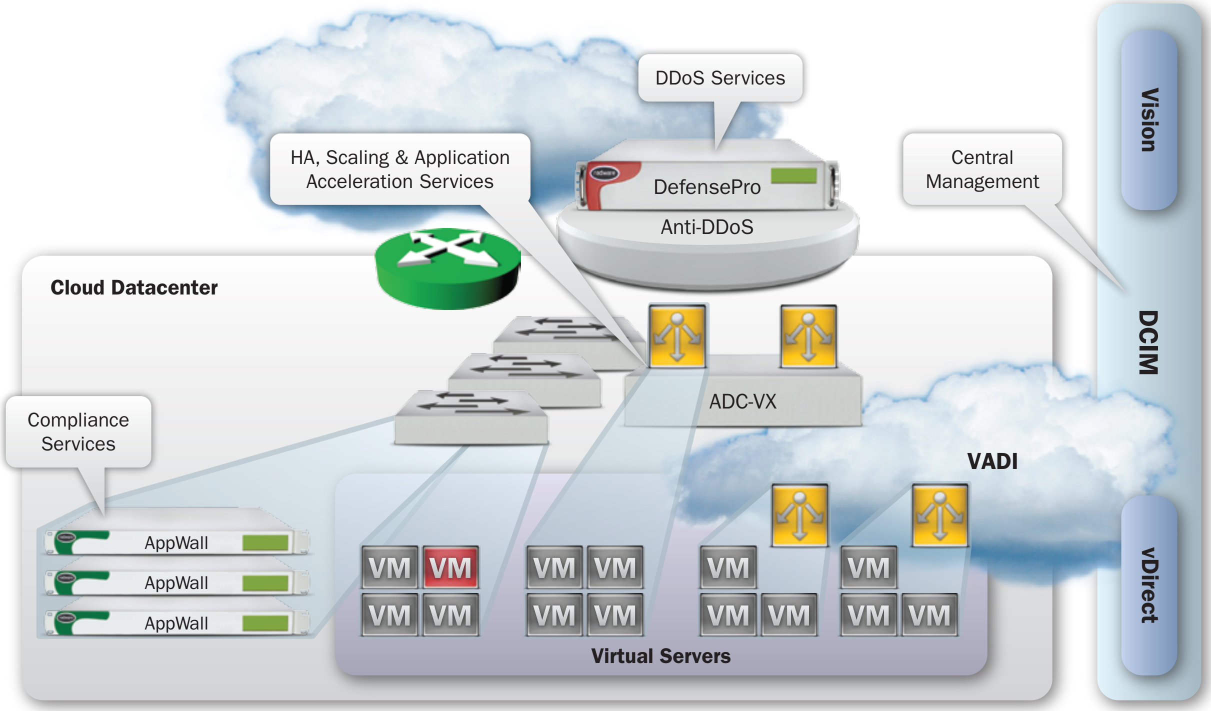


Figure 1 - Radware Service Architecture for Cloud

Benefits of Radware Solutions for Cloud Service Providers

1. Offer increased level of availability to cloud customers through highly available deployments of load balancing and application delivery services. High availability can be offered across any hardware form factor and location.
2. Seamlessly offer scale-out services to cloud customers inside cloud datacenters and across cloud datacenters by leveraging advanced health monitoring and KPI based global server load balancing.
3. Host a large scale of diverse services over a shared, purpose-built ADC infrastructure while fully isolating ADC instances associated with the different services.
4. Easily integrate application delivery and load balancing services into existing cloud service orchestration frameworks, home grown management tools and applications.
5. Simplify operations with a single management system controlling the entire set of Radware products in the cloud datacenter.
6. Cloud providers can offer additional value-add services such as application acceleration and application performance monitoring to their customers. All this while easily bundling the services into service packages and increasing customer confidence of rolling out applications in the cloud.

Summary

Radware application delivery and security solutions for cloud and hosting providers offer exceptional capabilities that greatly enhance the resilience, scalability and breadth of services offered by cloud and hosting providers. The value of the Radware is derived from 3 main benefits: (1) ability to enhance stability and scalability of cloud provider infrastructure (2) capability to help cloud providers build value added network services and offer these to their customers and (3) enabling these capabilities with minimal integration efforts and enhanced control.

Radware works with cloud providers globally addressing the key application delivery requirements presented in a cloud infrastructure through innovative cloud specific solutions.

For more information please visit http://www.radware.com