InformationWeek
:: analytics
Analytics.InformationWeek.com

September 2010
$199

Strategy Session

Presented in conjunction with

Network Computing

Networking in a Virtualized World

Virtualization is rapidly evolving into a core element of next-generation data centers. This expanded role places new strains on the network. This report explore the technical issues exposed by virtualized infrastructure and looks at standards, technologies and best practices that can make your network ready to support virtualization.

By Kurt Marko

Report ID: S1820910

TABLE OF CONTENTS

ABOUT US | ***InformationWeek Analytics'*** experienced analysts arm business technology decision-makers with real-world perspective based on a combination of qualitative and quantitative research, business and technology assessment and planning tools, and technology adoption best practices gleaned from experience.

If you'd like to contact us, write to managing director **Art Wittmann** at *awittmann@techweb.com,* executive editor **Lorna Garey** at *lgarey@techweb.com* and research managing editor **Heather Vallis** at *hvallis@techweb.com.* Find all of our reports at *www.analytics.informationweek.com.*

Kurt Marko
Network Computing

Kurt Marko is a technology writer and IT industry veteran. After graduating from Stanford University with a BS and MS in Electrical Engineering, Kurt spent several years as a semiconductor device physicist, working on process design, modeling and testing. He then joined AT&T Bell Laboratories as a memory chip designer and CAD and simulation developer.

Moving to Hewlett-Packard, Kurt started in the laser printer R&D lab doing electrophotography development, for which he earned a patent. His love of computers eventually led him to join HP's nascent technical IT group. He spent 15 years as an IT engineer and was a lead architect for several enterprisewide infrastructure projects at HP, including its Windows domain infrastructure, remote access service, Exchange e-mail infrastructure and managed Web services.

For the past five years, Kurt has been a frequent contributor to IT and consumer technology publications.

Executive Summary

Early server virtualization implementations have demonstrated significant cost savings while greatly enhancing deployment flexibility and adaptability. As virtualization software has matured, simple server consolidation, with modest 10-to-1 consolidation ratios, are giving way to more demanding strategies involving wholesale rearchitecting of the server environment. Next-generation server hardware will support scores of VMs, and most software is designed with virtualization in mind—meaning no enterprise application is off limits.

This prospect has significant network implications: It imposes unprecedented demands on bandwidth and switching capacity, greatly increases topological complexity, and complicates management and support by erasing the lines between physical and virtual, network and server, and blurring the distinction between data and storage networks.

The good news is that standards activity and innovative technologies promise to usher in a new era of virtualization-optimized networks. For instance, one issue is that as virtual machines migrate from server to server and VLAN to VLAN, the appropriate policies and QoS settings may not migrate with them. Standards such as 802.1Qbh Bridge Port Extension are being developed to address this issue. At the same time, vendors are developing technologies such as port profiles that append the correct policy settings to VMs even as they move among servers or VLANs.

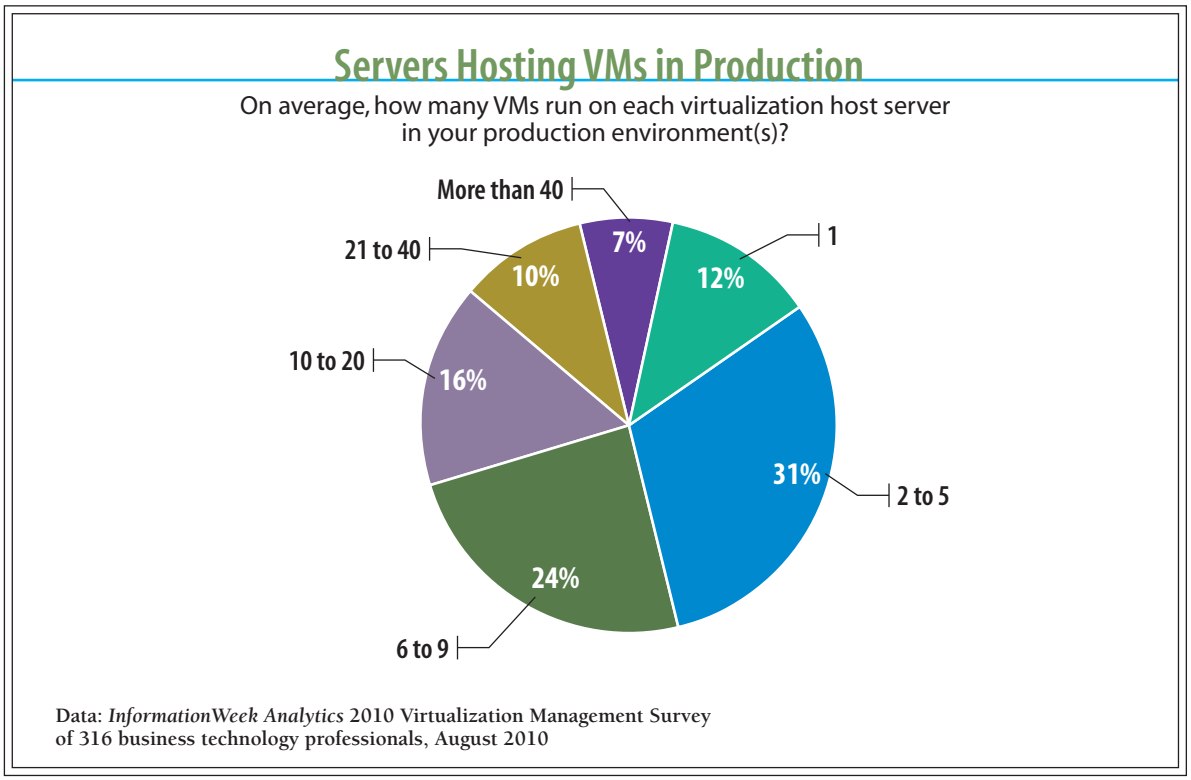
This report outlines the networking complications that arise in highly virtualized environments and examines different approaches to addressing these complications.

Server Virtualization vs. the Network

Server virtualization has been a boon for IT. Virtual machines are easier to deploy than traditional instances of an operating system or application. Organizations can also cut costs by running multiple virtual machines on a single server. For instance, in a 2010 InformationWeek Analytics survey on virtualization, 24% of respondents run between six and nine VMs per server in their production environments, and another 16% run between 10 and 20 VMs (see Figure 1, below). When asked what percentage of production servers will be virtualized by the end of next year, 20% of respondents said at least one-quarter of their servers would be virtualized. Another 19% said half or more (see Figure 2, next page).

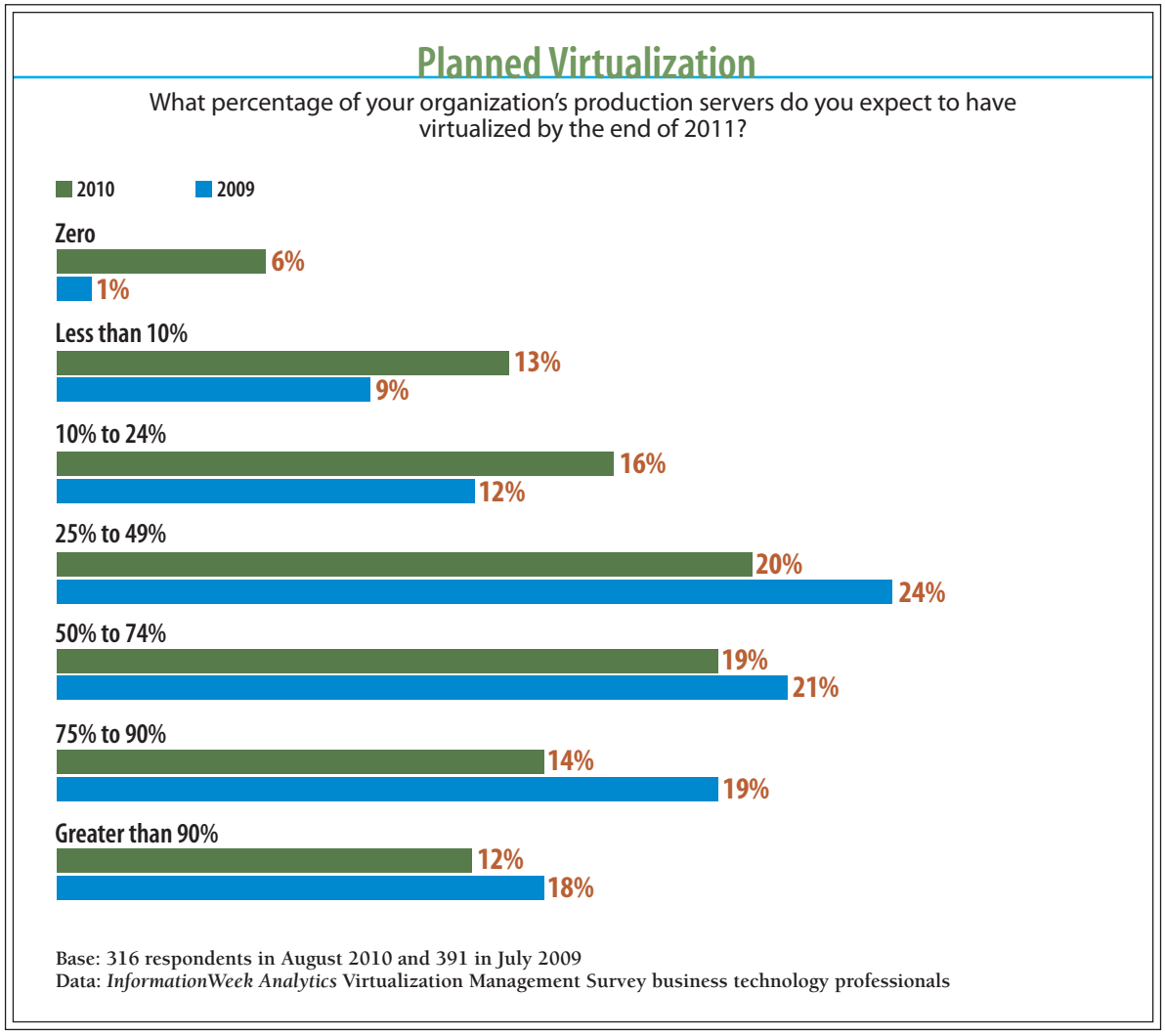
But there are downsides to virtualization, both for server and network administrators. For instance, IT is placing greater emphasis on management tools that can address VM sprawl in 2010 than they did in 2009 (see Figure 3, page 7). On the network side, while virtualization

Figure 1



reduces the number of physical servers, it snowballs the number of virtual devices and network devices; from a switching perspective, there's little difference between a virtual network port and physical one. Additionally, the proliferation of VMs, each handling richer data sets, is exploding network traffic, both at the core and edge. Ten or 20 virtual machines sharing the same physical network port, with each running applications passing increasing amounts of data—whether it's bigger e-mail attachments or streaming Web video—adds up to potential network bottlenecks and management headaches. Additionally, new hypervisor releases

Figure 2



continue to improve their network stacks, pushing ever closer to being able to saturate even 10 Gbps links.

This paradox of server simplicity vs. network complexity is analogous to what would happen if thousands of commuters gave up their individual cars for shared minivans, with each passenger going to a different destination. While this would reduce the number of vehicles on the freeway, it doesn't reduce the number of *trips*—the driver still must crisscross town dropping off passengers at their offices. It also greatly complicates the routing calculus. Instead of each commuter finding the quickest path between home and office, the van driver must optimize the pickup and delivery schedule to minimize drive time and distance.

These broad-scale issues break down into five problems. First, increased network complexity affects performance. Aside from merely increasing the number of network devices, virtualization adds tiers to the switching fabric, increasing latency, power consumption and management complexity.

Figure 3

Importance of VM Management Functions

Please rank VM management functionality in order of importance to your organization from 1 to 9, where 1 is "most important" and 9 is "least important."

	2010	2009
Interoperability with existing enterprise management tools	1	1
Application performance monitoring	2	2
Lifecycle management tools; control of virtual machine sprawl	3	5
Automated physical-to-virtual server conversion tools	4	4
Regulatory/compliance/audit capabilities	5	7
Virtual security management	6	3
Storage management/SAN visibility	7	6
Live VM migration	8	N/A
Cross-vendor capability/heterogeneous VM management	9	8

Base: 316 respondents in August 2010 and 391 in July 2009
Data: *InformationWeek Analytics* Virtualization Management Survey business technology professionals

Most data centers use a three-tier architecture of edge switches, aggregation switches and top-of-rack (TOR) or end-of-row (EOR) switches. Hypervisors add an additional layer with a software switch (also known as a virtual switch) to manage intra-hypervisor traffic. Some servers may use intelligent NICs with hardware port virtualization, while blade chassis often have switch modules, each of which adds another switching tier (see Figure 4, page 9).

Second, the consolidation of virtual machines on physical servers affects switching scalability and performance. As dual-processor servers with six-, eight- and even 10-core CPUs become common, consolidation ratios will climb. Currently, a hypervisor virtual switch with a workload of 10 to 15 VMs per system extracts a modest overhead of about 10% to 15%, but that figure that will undoubtedly increase when handling scores of VMs.

Additionally, because virtual switches operate at Layer 2, their proliferation implicitly changes the LAN topology to one that is larger and flatter. This resurrects all those problems—like broadcast traffic overhead and address table explosion—that LAN designers battled in the primeval days before edge switches. VM consolidation ratios also limit the effectiveness of VLAN partitioning. Because virtual switches operate at Layer 2, and look more like bridges to the external network, and because VLAN tags are also port-based, any application-specific VLANs are visible to all VMs on a system. Thus, if more VMs share the same Ethernet port, and different applications on those VMs are members of different VLANs, then each port could conceivably need access to many more VLANs, which greatly limits their effectiveness.

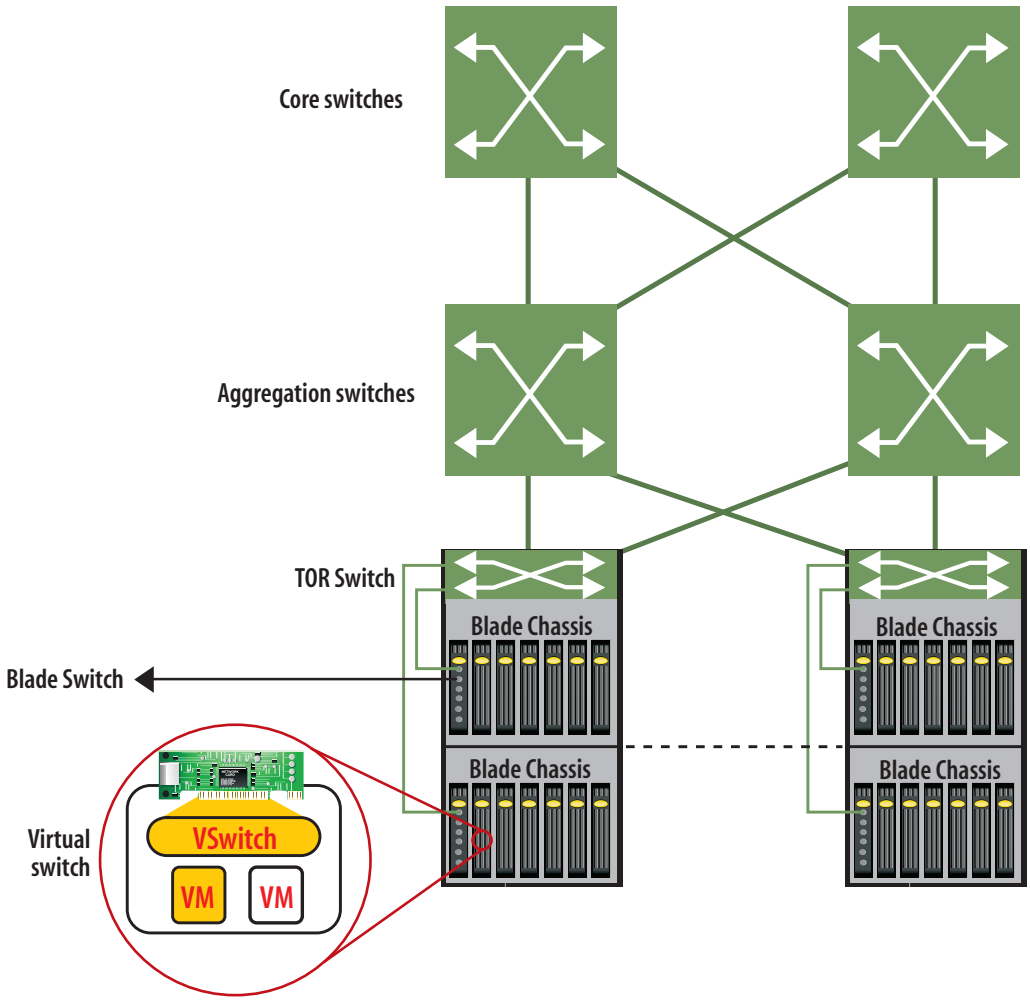
The third challenge is that software switching complicates management and security. Network monitoring, management, traffic reporting and security tools use standard protocols operating on physical ports, but as more traffic is switched within the hypervisor, these tools lose visibility into a significant amount of network activity. Some vendors make their monitoring and analysis software available on VMs to regain visibility, but these are proprietary solutions that typically support only one or two hypervisor vendors, and usually come with additional license costs.

Virtualization also exacerbates administrative silos that divide IT departments. For example, the server group might manage the hypervisor, VMs and virtual switches; network managers handle the switch fabric and policies; and the storage team handles the SAN and associated Fibre Channel (FC) fabric. Such divisions may seem sensible on an organizational chart, but without close coordination among these groups, problems can arise. For instance, server and storage

Figure 4

Switch Tiers Add Up

High-density, virtualized servers can add two additional switch tiers to the typical configuration of network core, aggregation and edge switches (either end-of-row or top-of-rack). Blade chassis often rely upon an internal switching module. When paired with a virtual switch in each hypervisor, this creates five switching tiers.



Source: Extreme Networks, Network Computing

admins with little networking training or experience often make configuration changes with wider network ramifications.

Fourth, the ability to seamlessly and transparently move VMs from one physical server to another complicates management and security. Such dynamic movement of application workloads becomes a headache when keeping network policies aligned with applications. Network managers may bind different application servers to specific VLANs, or assign application network flows different QoS priorities and security ACLs. But if the VM moves to another server, on a completely different LAN, how does the network management system know to migrate the policies accordingly? Today, it doesn't, unless the networking vendor has integrated its switching products with the hypervisor. The integration can be switch-to-hypervisor, such as Cisco (via its Nexus 1000v) and Arista do with VMware APIs. The integration can also pass through a management station, which Extreme Networks offers.

Network standards and technology have evolved under the assumption that policies like VLAN assignments, QoS and ACLs are set at the physical interface (Layer 2) or port (Layer 3) layers. While binding these policies to specific VMs is problematic enough, keeping them in sync as VMs move between servers is even more difficult.

Fifth, virtualization exacerbates demands for shared storage, due to the inherent need to decouple OS images, applications and data from the underlying server hardware. The solution has traditionally been a separate, dedicated SAN, which to most people still means Fibre Channel. Yet SANs are expensive and complex to manage, adding an entirely new network protocol, switching fabric and (often) management team.

As the number of VMs and their storage demands increase, scaling and managing the SAN becomes as challenging and complex as scaling the data network. Wouldn't it be better to solve this set of problems once instead of twice?

Addressing the Problems

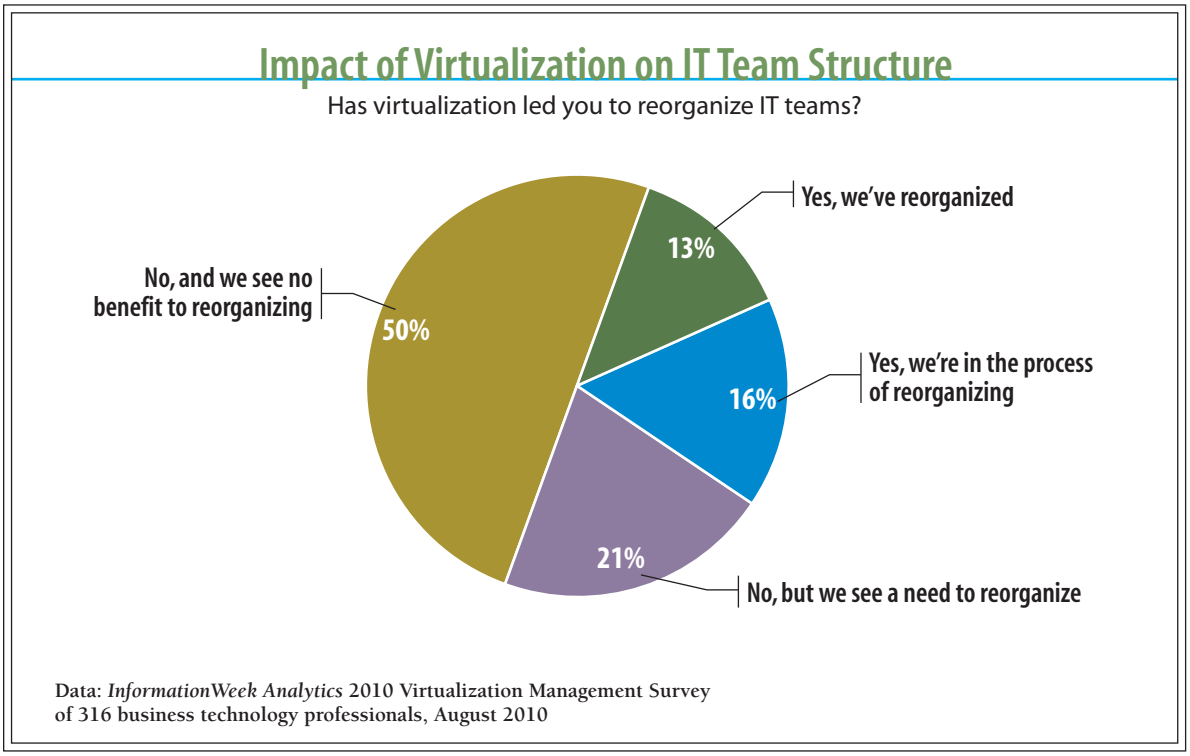
This litany of problems hasn't gone unnoticed by the industry and is being addressed by a host of new products and standards efforts designed to improve the performance, scalability and manageability of virtualized server networks.

For example, one effort is to pull switching out of the hypervisor. Virtual switches were something of an ad hoc solution to the problem of switching multiple, independent operating systems, each with distinct IP and MAC addresses, and sharing the same physical interface—a situation the Ethernet protocol didn't anticipate. While hypervisor switching certainly works, alternatives are coming.

One alternative is Virtual Ethernet Bridging (VEB), which moves the bridging function onto an intelligent NIC that supports so-called single root I/O virtualization (SR-IOV), a standard proposed by the PCI-SIG. SR-IOV essentially carves up an intelligent NIC into multiple virtual NICs—one for each VM. It does this by providing independent memory space, interrupts and DMA streams for each VM. This allows bridging to occur in hardware rather than in the hypervisor.

This seemed like a promising solution when Intel proposed it several years ago, but it's failed

Figure 5



to gain market momentum due to poor interoperability between NICs and scalability concerns as the number of VMs per server grows. Aside from lackluster industry adoption, the problem is that each embedded bridge is yet another device to manage (no different from a software switch), and that management function is not integrated with the overall network management system. Due to implementation differences (that is, extended functions not part of the standard), different NICs may have different bridging capabilities, and these often don't interoperate.

Another solution involves ditching the virtual switch entirely and moving the switching function back out to edge devices where every good network engineer believes it belongs. This is where two new IEEE standards projects come in, with work proceeding on two parallel and largely complementary paths. Both are amendments to the base IEEE 802.1Q VLAN tagging standard.

The first and more mature project is 802.1Qbg Edge Virtual Bridging (EVB). It's sometimes known as virtual Ethernet port aggregation (VEPA), after HP's technology submission, although HP's VEPA actually includes additional, proprietary features. EVB is designed to allow multiple VMs to share a common port while obtaining services from an external bridge (that is, an edge switch acting as a reflective relay). Normally, Ethernet frames are not forwarded back out of the same interface they came in on. This action, called hairpinning, causes a loop in the network at the port. EVB provides a standard way to solve the hairpinning problem. As a relatively simple protocol extension, EVB is attractive because it can be implemented on existing hardware with a software upgrade to the switch and hypervisor.

Unfortunately, EVB is only a baby step because it doesn't solve the policy management problem and can burden switches with more broadcast and multicast traffic. These thornier issues are being tackled by the 802.1Qbh Bridge Port Extension project. Whereas Qbg doesn't modify the underlying Ethernet packet, the Qbh port extension standard adds a tag, much like standard VLAN tags, allowing network flows to be mapped to specific VMs and followed as those VMs move about the network.

The first drafts of each standard were just released this past winter, and given the warring technical debates between the two primary proponents, Cisco and HP, don't expect a resolution anytime soon. While both vendors would augment the protocol with a form of VM tags, they differ in details, with HP proposing a modification to the MAC security tag (SecTAG) and Cisco

offering a new so-called port extension tag. Ultimately, the issues will get settled through the standards process. The result will be interoperable network switches and NICs that allow VM-generated traffic to be managed just like other network flows.

In the Meantime

The slow pace of standards development means the elimination of the virtual switch tier is several years away. However, there's still plenty of technology in the toolbox to simplify edge and storage networks.

10 Gigabit Ethernet switch ports and NIC interfaces have cracked the $500 mark, and with LAN-on-motherboard single-chip solutions rolling out this year, further steep reductions are inevitable. With bandwidth to spare over copper, 10 GbE makes the ideal edge network transport. In a dense environment of blades or 1U pizza boxes, a top-of-rack (TOR) topology makes much more sense than routing a rope of cables to row-end switches.

Yet, even here, cable management inside a rack can be problematic, particularly when using blades. With a dozen or more server modules per chassis, each rack might have 50 or more servers—that's 100-plus Cat6 cables to a TOR switch for a fully redundant topology. The shear bulk and complexity of routing all that copper isn't feasible. It's the primary reason many blade designs have turned from pass-through modules, which essentially route each server's LAN connection from the chassis backplane to a glorified patch panel, to integrated switching modules that aggregate traffic from a single chassis. Yet, as we've mentioned, this adds another tier to the switching fabric.

There are two solutions for maintaining a flatter physical network. One entails using new high-density cable standards, specifically MRJ21, which aggregate six RJ45 cables into a single harness and is then routed from a blade pass-through module to a TOR switch. The other approach uses a fabric extender, such as Cisco's UCS 2100, the blade equivalent of the Nexus 2000 Fabric Extender, which essentially extends the Nexus 5000 TOR switch's backplane into the UCS blade chassis itself. Both of these products are relatively new; while they work suitably within specific vendor ecosystems, your choice is bound by the solution embraced by your incumbent switch vendor.

LAN Plus SAN

The days of running parallel data and storage networks are ending for all but the most demanding storage I/O workloads. 10 GbE has achieved rough price-per-bandwidth parity with GigE, with ample bandwidth for networking functions. It should be the default storage network for next-generation virtual server farms.

Depending on the specific I/O workload, backhaul from a TOR switch to the aggregation tier can be achieved by trunking 10 Gbps links or moving to the recently adopted 40 GbE standard. This year's Interop was 40 GbE's coming-out party, with Extreme Networks and Force10 announcing edge switch modules, with delivery slated for later this year. Admittedly, 40 GbE will be expensive at about $1,000 per port. Even so, it's at a rough price-per-gigabit parity with 10 GbE switching modules, albeit 40 GbE requires Cat6A cabling, which is roughly twice the cost per foot over Cat6.

10 GbE provides the raw bandwidth, but there are several ways to use it for shared storage. Building a block-level SAN on Ethernet can be achieved using either iSCSI or Fibre Channel over Ethernet (FCoE). While both use Ethernet as a transport, they differ in that iSCSI operates at L3 (using IP to encapsulate SCSI commands) while FCoE is pure L2—a detail that matters at the margins but is largely irrelevant for many workloads.

FCoE doesn't entirely replicate the network reliability provided by native Fibre Channel and is being augmented by the IEEE's Data Center Bridging (DCB) Task Group. DCB actually incorporates several standards tracks with the goal of addressing deficiencies in the native Ethernet protocol and replicating FC and Infiniband features like guaranteed reliability, congestion notification, flow control and traffic prioritization. The choice between iSCSI and FCoE largely comes down to one's existing infrastructure. For greenfield deployments, iSCSI is a perfectly serviceable solution that's more mature and less expensive than FCoE. Those with significant investments in FC arrays will find the transition to FCoE far more appealing, as it can piggyback on legacy equipment.

Some virtual environments may be able to avoid the SAN entirely because recent hypervisor releases (from VMware, Citrix Xen and Microsoft Hyper-V, among others) now support booting VM images via NFS from a NAS array.

For virtualized applications that don't require block I/O, there are many good reasons to use NFS

instead—it's simpler to install, manage and scale; it's cheaper (although probably not significantly so over iSCSI); and, because it's file-based, it's easier to back up.

Virtual ADC

The ability of the latest VM management platforms to automatically create new machines in response to application load or hardware failure improves both performance and reliability. Thus, VM live migration and on-demand provisioning should be a part of any next-generation environment. Yet, vagrant VMs pose a problem not only for network management and security tools, but also for traditional load balancers (also known as application delivery controllers). Here's one case where a problem created by virtualization can also be solved by it.

The traditional way to balance application load across many servers, or automatically redirect traffic to a redundant DR site in case of failure, entailed using hardware load balancers in conjunction with off-site data replication. A new breed of virtual network load balancing and WAN optimizing appliances mimics the functionality of their hardware predecessors. These software appliances are often a better fit for virtualized environments because they seamlessly integrate with the underlying VM management platform, enabling automatic re-configuration in response to VM movement, and can scale with increasing loads without a forklift upgrade.

Silo Busters

Tensions between IT's typical administrative silos, with different groups responsible for network, server and storage management, are exacerbated by the issues we've discussed, but here the solution isn't purely technical. IT managers must start by identifying clear responsibility for each management task, like server and storage provisioning or network configuration, with an eye toward consolidating shared roles in a single team. According to the InformationWeek survey, 29% of respondents have reorganized or plan to reorganize their IT teams because of virtualization. Another 21% see the need to reorganize but haven't done so yet (see Figure 5, page 11).

Virtual switches won't be going away anytime soon, but the configuration and management of these virtual network devices shouldn't reside with the server team merely by virtue of their ownership of the underlying VM management platform. Until the technology allows virtual port management to be pulled into a comprehensive management tool, it means the network

and server teams will have to share authority for the VM platform. Vendors are addressing these management issues through software, as evidenced by the Cisco-VMware partnership that integrates VMware's vSphere with Cisco's Nexus 1000V software switch.

The network team must then consider extending its architecture to facilitate monitoring, security and troubleshooting on software switches—for example, by using virtual span ports (promiscuous-mode port mirroring).

Virtual Infrastructure, Real Impact

Enterprise virtualization is evolving from simple server consolidation into the foundation for a new cloud-like data center infrastructure. This transformation will necessarily place new demands on the network, requiring increased capacity, a redesigned switch topology, new management software and converged data and storage networks.

New technology and standards are emerging to address many of the issues raised by virtualization's impact on the network, but organizations must also take a strategic approach to virtualization to ensure that benefits in one sector of the IT shop don't turn into problems in another.

While the journey toward a highly virtualized infrastructure will be long, and at times arduous, the result will bring the enterprise to new levels of performance, reliability, agility and efficiency.