Simplify VMware vSphere* 4 Networking with Intel® Ethernet 10 Gigabit Server Adapters

Today's Intel® Ethernet 10 Gigabit Server Adapters can greatly reduce the level of networking complexity in VMware vSphere* 4 environments, relative to Gigabit Ethernet implementations. Advances in both Intel Ethernet and vSphere allow dramatic simplification of the environment without compromising areas such as security and traffic segmentation.

February 2010
Version 1.0

OVERVIEW

Advances in Intel® Ethernet 10 Gigabit (10GbE) Server Adapters and VMware vSphere* 4 enable migration away from legacy Gigabit Ethernet (GbE) networking. As the industry move toward 10GbE becomes more mainstream, IT organizations are considering its use for initiatives such as LAN/SAN consolidation and unification. Many of these organizations have found success with the practical first step of converging multiple LAN connections using 10GbE. This transition lowers the number of physical connections required, thus reducing overall complexity and cost, while providing advanced security and traffic segmentation.

Networking virtualized hosts with GbE is typically based on relatively large numbers of dedicated physical connections, which are used both to segregate different types of traffic and to provide sufficient bandwidth. Using 10GbE connectivity can help overcome shortcomings of this approach, including the following:

- **Excessive complexity.** Large numbers of server adapters and the attendant cabling make administration less efficient and increase the likelihood of connection and configuration errors.
- **High equipment costs.** Using many GbE network adapters requires the purchase of the actual adapters, as well as the use of large numbers of motherboard slots, switch ports, cables, and so on.
- **Increased energy use.** Powering many individual GbE server adapters and switch ports directly increases power usage and raises server-room cooling requirements.

This paper provides network architects and decision makers with guidance for moving from GbE to 10GbE networking in their virtualized data centers. It begins with a review of the current state of the typical virtualized data center and discusses some of the challenges and opportunities associated with maximizing the value of virtualization. Next, it explores the factors that underlie the excessive complexity associated with GbE networking and describes how 10GbE can address that complexity. It then provides some best practices for achieving optimal networking results in virtual infrastructures. Finally, it addresses some concerns decision makers might have with regard to security and performance.

Table of Contents

Overview **1**

The Benefits of Virtualization Come with Limitations **2**

Today's Reality: Undue Complexity from Gigabit Ethernet Solutions **2**

The Solution: Refresh Infrastructure with 10 Gigabit Ethernet **3**

Best Practices to Achieve Near-Native Performance **4**

Best Practice 1: Use Virtual Distributed Switches to Maximum Effect 4

Best Practice 2: Streamline Configuration Using Port Groups . . .5

Best Practice 3: Use VLANs with VLAN Trunking .5

Best Practice 4: Use Dynamic Logical Segmentation Across Two 10GbE Ports 5

Best Practice 5: Proactively VMotion VMs Away from Network Hardware Failures 5

Customer Concerns: Security, Traffic Segmentation, and Bandwidth **6**

Conclusion **8**

The Benefits of Virtualization Come with Limitations

The original value proposition embraced by many organizations when they first considered virtualizing their data centers still holds. By consolidating servers, they sought to take optimal advantage of the growing headroom from increasingly powerful servers while also reducing infrastructure requirements. The key benefits these organizations sought to derive included the following:

- **Efficiency and simplicity**. Having fewer physical hosts in the data center creates a simpler topology that can be more effectively managed.
- **Lower capital costs.** Buying smaller numbers of servers and cutting the requirements for the supporting infrastructure, such as switches, racks, and cables, potentially delivers very large capital savings.
- **Reduced operating expenses.** Lower power consumption and cooling requirements combined with greater agility from simplified provisioning and novel usage models such as automated load balancing can cut day-to-day costs.

Unfortunately, the success achieved by many organizations in attaining these benefits is limited by the complexity that has arisen from networking virtualized servers with GbE.

Today's Reality: Undue Complexity from Gigabit Ethernet Solutions

The common practice when deploying virtualized hosts has been to segregate network functions onto dedicated GbE ports, adding additional ports as demand for bandwidth increases. These ports are often installed in pairs to provide network failover, doubling the number of ports required per host. As a result, the number of network ports has a tendency to become bloated, leading to excessive complexity.

This practice came about in large part because many administrators did not fully understand or have experience with server virtualization. The following factors also played a role:

- **Physical server network connection paradigms** were extended to virtual infrastructures, leading to the use of separate physical connections to segment traffic and provide required bandwidth.
- **Previous VMware versions** required a dedicated connection for virtual machines (VMs) and for each of multiple traffic types, such as VM traffic, service console connections, IP storage, VMware VMotion*, and so on.
- **Security procedures** have often led network administrators to physically segregate traffic onto separate ports since 802.1Q trunking to the host was not allowed and was reserved for switch-to-switch traffic.

In many cases, hosts must have as many as eight or more GbE network ports to satisfy these requirements. As shown in Figure 1, several port groups are configured to support the various networking functions and application groupings, and in turn each port group is supplied with one or more physical connections. Virtual LAN (VLAN) tags may be implemented on these port groups as well.

This topology raises the following issues:

- **Complexity and inefficiency.** Many physical ports and cables make the environment very complex, and the large number of server adapters consumes a great deal of power.
- **Difficult network management.** The presence of eight to 12 ports per server makes the environment difficult to manage and maintain, and multiple connections increase the likelihood of misconfiguration.
- **Increased risk of failure.** The presence of multiple physical devices and cable connections increases the points of potential failure and overall risk.
- **Bandwidth limitations.** Static bandwidth allocation and physical reconnections are required to add more bandwidth to the GbE network.

The practice of using large numbers of GbE connections has persisted even though 10GbE networking provides the ability to consolidate multiple functions onto a single network connection, greatly simplifying the network infrastructure required to support the host.

Part of this continuing adherence to a legacy approach is due to the outdated understandings of security and networking. For example, some administrators believe that dedicated VMotion connections must be physically separated because they mistrust VLAN security and question bandwidth allocation requirements. Others assume that discrete network connections are required to avoid interference between network functions.

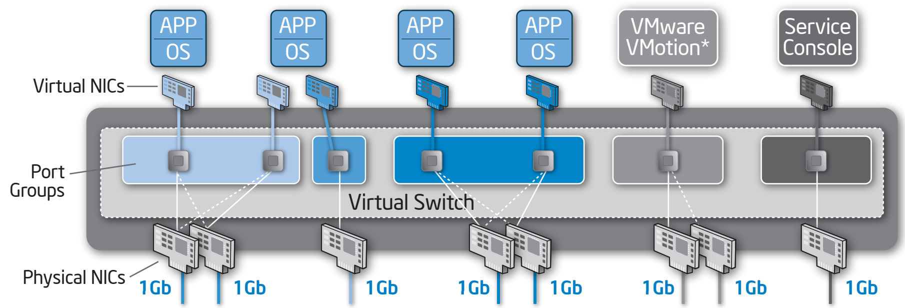

Figure 1. Virtual switch with multiple physical GbE server adapters.

Although these types of concerns may have been well founded in the past, modern switching equipment and server adapters have addressed them. Now is the time to take advantage of Intel Ethernet 10GbE Server Adapters.

The Solution: Refresh Infrastructure with 10 Gigabit Ethernet

The complexity issue and other limitations associated with GbE described above can be addressed by consolidating all types of traffic onto 10GbE connections.

With the advent of dynamic server consolidation and increasingly powerful servers such as those based on Intel's Nehalem architecture, more workloads and applications than ever before are being consolidated per physical host. As a result, the need is even greater for high-bandwidth 10GbE solutions. Moreover, features that provide high performance with multicore servers, optimizations for virtualization, and unified networking with Fibre Channel over Ethernet (FCoE) and iSCSI make 10GbE the clear connectivity medium of choice for the data center.

Moving from multiple GbE to fewer 10GbE connections will enable a flexible, dynamic, and scalable network infrastructure that reduces complexity and management overhead, and provides high availability and redundancy. Figure 2 shows an installation analogous to that in Figure 1, but using 10GbE connectivity and two GbE ports for the service console. This installation makes use of the VMware vNetwork Distributed Switch (vDS) feature (see Best Practice 1). Using vDS provides the same basic functions as standard virtual switches, but they exist across two or more clustered ESX

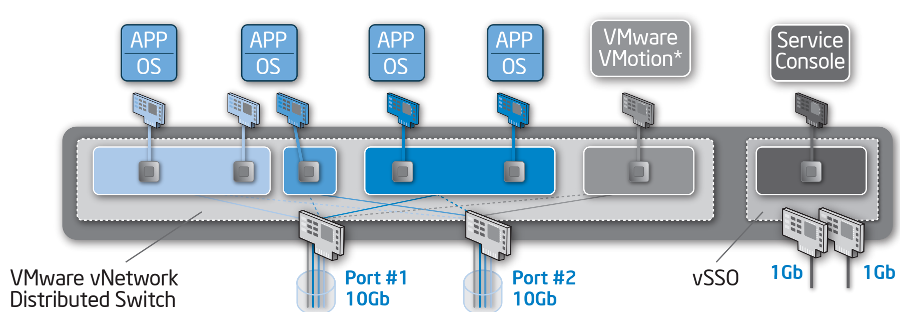

Figure 2. VMware vNetwork Distributed Switch with 10GbE server adapters for network traffic and GbE server adapters for service console traffic.

or ESXi hosts. Using a hybrid model of standard virtual switches and vDS allows for management of the vDS if the switch fails and keeps service console traffic on a physically separate network connection. Figure 3 takes the topology shown in Figure 2 one step further and shows the installation with only 10GbE connections.

The obvious advantage of the 10GbE solution is that it reduces the overall physical connection count, simplifying infrastructure management considerations and the cable plant. Moving to a smaller number of physical ports reduces the wiring complexity and the risk of driver incompatibility, which can help to enhance reliability. For additional reliability, customers may choose to use ports on separate physical server adapters. The new topology has the following characteristics:

- Two 10GbE ports for network traffic, using NIC teaming for aggregation or redundancy
- One to two GbE ports for a dedicated service console connection on a standard virtual switch (optional)
- SAN converged on 10GbE, using iSCSI
- Bandwidth allocation controlled by VMware ESX*

This approach increases operational agility and flexibility by allowing the bandwidth to the host and its associated VMs to be monitored and dynamically allocated as needed. In addition to reducing the number of GbE connections, the move from a dedicated host bus adapter to Intel Ethernet Server Adapters with support for iSCSI and FCoE can take advantage of 10GbE connections.

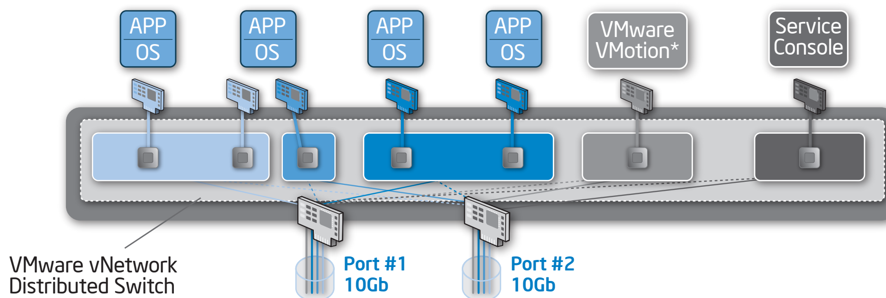

Figure 3. VMware vNetwork Distributed Switch with 10GbE server adapters for both network and service console traffic.

Best Practices to Achieve Near-Native Performance

The methods and techniques in this section provide guidance during network design that help engineers achieve high performance. The diagram in Figure 4 summarizes these best practices.

Best Practice 1: Use Virtual Distributed Switches to Maximum Effect

The VMware vDS feature manages traffic within the virtual network, providing robust functionality that simplifies the establishment of VLANs for the functional separation of network traffic. This virtual switch capability represents a substantial step forward from predecessor technologies, providing the following significant benefits to administrators:

- **Robust central management** removes the need to touch every physical host for many configuration tasks, reducing the chance of misconfiguration and improving the efficiency of administration.
- **Bandwidth aggregation** enables administrators to combine throughput from multiple physical server adapters into a single virtual connection

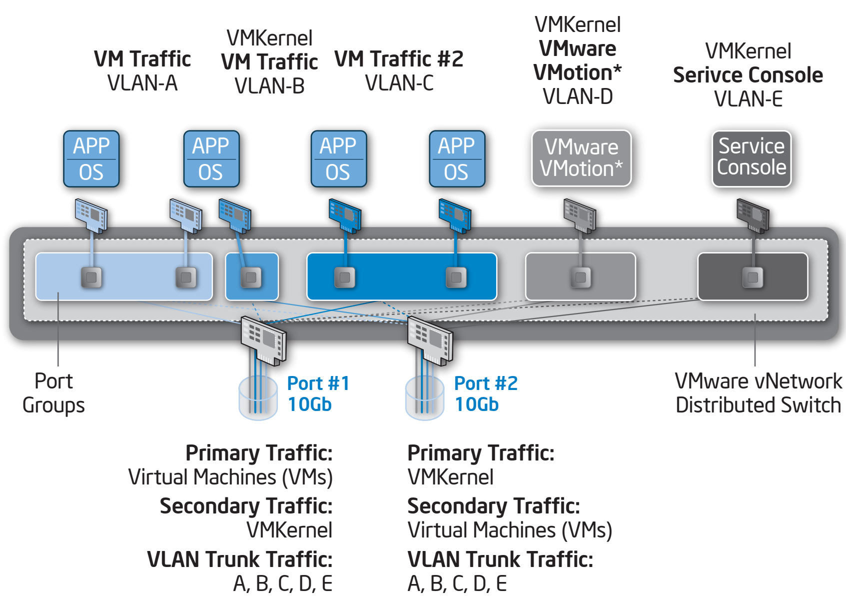

Figure 4. Summary of best practices for network design using VMware vSphere* 4 and Intel® Ethernet 10GbE Server Adapters.

- **Network failover** allows one physical server adapter to provide redundancy for another while dramatically reducing the number of physical switch ports required.
- **Network VMotion** allows the tracking of the VM's networking state (for example, counters and port statistics) as the VM moves from host to host.

Placing the service console network separate from the vDS can help avoid a race condition and provides additional levels of redundancy and security.

For a deeper discussion around best practices for the use of virtual switches, see the following resources:

- Ken Cline's blog entries on The Great vSwitch Debate[1]
- The VMware white paper, "What's New in VMware vSphere™ 4: Virtual Networking"[2]

Best Practice 2: Streamline Configuration Using Port Groups

Port groups provide the means to apply configuration policies to multiple virtual switch ports as a group. For example, a bi-directional traffic-shaping policy in a vDS can be applied to this logical grouping of virtual ports with a single action by the administrator. If changes need to be made in response to a security event or a change in business requirements, this capability enables faster response, and in normal day-to-day operations it enables greater efficiency by augmenting physical resources with more sophisticated virtual ones.

Best Practice 3: Use VLANs with VLAN Trunking

The use of VLANs allows network traffic to be segmented without dedicating physical ports to each segment, reducing the number of physical ports needed to isolate traffic types. As stated in VMware's paper "vSphere Hardening Guide: ESX and Virtual Networking"[3] (rev B: public draft January 2010): "In general, VMware believes that VLAN technology is mature enough that it can be considered a viable option for providing network isolation."

The following traffic should be configured on dedicated VLANs:

- **Service console** should be on a dedicated port group with its own dedicated VLAN ID; use port 2. An option to connect to a dedicated management network that is not part of a vDS can provide additional benefits but does add additional network connections and network complexity.
- **VMotion** should be on its own dedicated port group with its own dedicated VLAN ID; use port 2 in a two-port configuration.
- **IP-based storage traffic (iSCSI, NFS)** should be on its own port group in the vDS; use port 2 in a two-port configuration.
- **VM traffic** can use one or more VLANs, depending on the level of separation that is needed between the VMs. The VMs should not share the service console, VMotion, or IP-based storage traffic VLAN IDs and should use port 1 in a two-port configuration.

802.1Q VLAN trunking provides the ability to combine multiple VLANs onto a single wire. This capability makes the administration of VLANs more efficient because the grouping can be managed as a unit over one physical wire and broken into dedicated networks at the switch level instead of at each host.

Best Practice 4: Use Dynamic Logical Segmentation Across Two 10GbE Ports

If two physical 10GbE ports are used, as shown in Figure 3, place administrative, live migration, and other back-end traffic onto one physical connection and VM traffic onto the other. To provide redundancy between the two links, configure the network so that the traffic from each link fails over to the other if a link path failure with a NIC, cable, or switch occurs.

This approach provides redundancy, increases bandwidth because both ports are being utilized, and can provide additional security through physical isolation in a non-failed mode. While the use of two 10GbE ports helps to reduce solution cost, some organizations may prefer the use of four 10GbE ports to provide additional bandwidth, additional redundancy, or simply to interface with existing network infrastructure.

Best Practice 5: Proactively VMotion VMs Away from Network Hardware Failures

Install 10GbE ports in pairs so they can be configured in a redundant manner to enhance reliability. If two 10GbE ports are used, then run VM traffic primarily on port 1 and all other traffic on port 2. This design uses the bandwidth of both 10GbE ports and can be configured in a redundant manner for network failover.

Note that, in the event of a failover, all traffic would be travelling over the same wire. Because the duration of that status should be as short as possible, the host and management software should be reconfigured to migrate all VMs off the host with VMotion as quickly as possible, to maintain redundancy and help ensure reliability.

Using live migration in conjunction with network monitoring software helps to ensure uninterrupted operation of the VM(s) while minimizing the amount of time that they operate on a physical host where redundancy has been compromised.

To enhance redundancy further, a second option is to move to a four 10GbE port configuration that provides the two primary ports a dedicated backup or redundant port on separate adapters. The same practice of immediately moving VMs to a new host in case of failure (and lost redundancy) should be used. This configuration can provide greater bandwidth to the VMs and the VMKernel traffic types or physical separation of traffic types if required.

Customer Concerns: Security, Traffic Segmentation, and Bandwidth

Concerns reported by customers regarding consolidation of connections onto 10GbE include security and traffic segregation with dedicated bandwidth for critical networking functions.

When GbE server connections are consolidated, a way to isolate connections in the absence of dedicated physical connections is still necessary. This requirement reflects the need for security between different types and classes of traffic, as well as the ability to ensure adequate bandwidth for specific applications within the shared connection.

Security Considerations: Isolating Data among Traffic Streams

In our 10GbE model, VLANS provide some of the basic security features required in the installation. Security of VLANs has been debated, tested, and written about extensively. A review of the documentation suggests strongly that when properly implemented, VLANs provide a viable option for network isolation. For further inquiry on this subject, see the VLAN Security White Paper[4] from Cisco Systems or vSphere Hardening Guide: ESX and Virtual Networking[5] from VMware. In particular, note the following safeguards that help to protect data:

- **Logical partitioning protects individual traffic flows**. VMware vSphere can control the effects from any individual VM on the traffic flows of other VMs that share the same physical connection.
- **Internal and external traffic do not need to share a physical connection.** DMZs can be configured on different network adapters to isolate internal traffic from external traffic as security needs dictate.
- **Back-end services are handled by VMware ESX.** Administrative traffic and other back-end services are handled by a separate networking stack managed by the VMkernel, providing further isolation from VM traffic, even on the same physical connection.

Traffic Segmentation and Ensuring Bandwidth

10GbE provides enough bandwidth for multiple traffic types to coexist on a single port, and in fact, in many cases Quality of Service (QoS) requirements can be met simply by the availability of large amounts of bandwidth. The presence of sufficient bandwidth can also dramatically improve the speed of live VM migration using VMotion, removing potential bottlenecks for greater overall performance. Beyond the general availability of significant bandwidth, however, traffic segmentation can provide dedicated bandwidth for each class of network traffic.

While segmenting traffic flows onto discreet GbE connections is also a viable means of providing dedicated bandwidth to a specific traffic type or function, doing so has distinct shortcomings. For example, allocating two GbE connections to a VM traffic port group provides a potential of 2 Gbps of dedicated bandwidth, but if additional bandwidth is needed for sporadic traffic spikes from that port group, additional server adapters must be added (assuming additional PCI* slots are available). Additionally, the bandwidth allocated in the example cannot be used by any other traffic; it simply goes to waste.

On the other hand, handling the port group with bandwidth from a shared 10GbE server adapter allows additional bandwidth to be allocated for traffic spikes more seamlessly. Furthermore, multiple port groups can share the bandwidth headroom provided by the server adapter. The resource can be automatically and dynamically reallocated as various port groups are accessed by their associated VMs. Dedicated bandwidth is not needed for any one port group as long as the host network connection never reaches saturation. This method is very similar to how VMware shares processor resources, allowing VMs to burst processor utilization as needed due to the likelihood that many VMs won't burst at the same time.

The Intel Networking Test lab confirms the viability of replacing multiple GbE ports with a pair of 10GbE ports in the configuration mentioned in this paper. A number of VMs were installed on each of two hosts, and the process of migrating all of the VMs on one host to the other and back again was timed. This testing was done with no other work or traffic flows on the systems. As a point of comparison, network traffic flows were started to all of the VMs on each host, and the migration exercise was run again. The result was that the time required to migrate the VMs was only minimally affected.[1]

INTEL TECHNOLOGIES THAT ENABLE NEAR-NATIVE PERFORMANCE

In a virtualized environment, realizing the performance benefits of 10GbE requires the means to mitigate software overhead. To that end, features of Intel® Virtualization Technology improve networking performance.

Virtual Machine Device Queues (VMDq)

VMDq offloads data-packet sorting from the virtual switch in the virtual machine monitor (VMM) vmkernel onto the network adapter, which reduces processor overhead and improves overall efficiency. Data packets are sorted based on their MAC addresses, queued up based on destination virtual machine (VM), and then routed to the appropriate VMs by the VMM vmkernel virtual switch. A line rate throughput of 9.3 Gbps was seen on a virtualized server as compared to 4.2 Gbps without VMDq.[10] For more information, see Intel's VMDq Web page.[6]

Single Root I/O Virtualization (SR-IOV)

The latest Intel® Ethernet server controllers support SR-IOV, a standard created by the PCI* Special Interest Group (PCI-SIG*). Intel played a key role in defining this specification. Using SR-IOV functionality, Intel Ethernet Server Adapters can partition a physical port into multiple virtual I/O ports called Virtual Functions (VFs). Using this mechanism, each virtual I/O port or VF can behave as a single dedicated port directly assigned to a VM. For more information, see Intel's PCI-SIG SR-IOV Primer.[7]

Data Center Bridging (DCB)

DCB enables better traffic prioritization over a single interface, as well as an advanced means for shaping traffic on the network to decrease congestion. 10GbE with DCB can also help some low-latency workloads, avoiding the need to upgrade to a special purpose network to meet latency requirements, thereby reducing total cost of ownership. For more information, see Intel's DCB Web page.[8]

Intel® Virtualization Technology for Directed I/O (Intel® VT-d)

Intel VT-d enables the VMM to control how direct memory remapping (DMA) is processed by I/O devices, so that it can assign an I/O resource to a specific VM, giving unmodified guest OSs direct, exclusive access to that resource without the need for the VMM to emulate device drivers. For more information, see the Intel® Virtualization Technology for Directed I/O Architecture Specification.[9]

Conclusion

Using Intel Ethernet 10GbE Server Adapters in virtualized environments based on VMware vSphere allows enterprises to reduce the complexity and cost of network infrastructure. Advanced hardware and software technologies work together to ensure that security and performance requirements can be met without the large numbers of physical server connections required in GbE legacy networks. Additionally, more sophisticated management functionality is possible using 10GbE and VMware functionality features to replace methods that depend on physical separation with updated approaches that use logical separation.

Because migration to 10GbE can dramatically reduce the number of physical server adapters needed for a given configuration, organizations can realize a variety of cost advantages:

- Decreasing the number of physical server adapters, switch ports, and support infrastructure components required can potentially reduce capital costs.
- Providing a physically simpler environment where misconfigurations are more easily avoided can also help to decrease support costs, as well as enable greater automation through the extensive use of virtualization.
- Reducing the number of server adapters and switch ports being powered in data centers can help to create substantial energy savings.

In addition, the ability to dynamically allocate bandwidth on demand increases flexibility and enables organizations to get maximum value from their network hardware. Therefore, 10GbE is superior to GbE as a connectivity technology for virtualized networks, and as future generations of data centers continue to increase the demands placed on the infrastructure, the level of that superiority is expected to increase.

For more information about Intel® Ethernet 10GbE Server Adapters, see **www.intel.com/go/10GbE**

For more information about VMware vSphere* 4, see **www.vmware.com/products/vsphere**

For more information about virtual networking, see **www.intel.com/network/connectivity/solutions/virtualization.htm**

1 http://kensvirtualreality.wordpress.com/2009/03/29/the-great-vswitch-debate-part-1/

2 http://www.vmware.com/files/pdf/VMW_09Q1_WP_vSphereNetworking_P8_R1.pdf

3 http://communities.vmware.com/servlet/JiveServlet/downloadBody/11846-102-1-11684/vSphere%20Hardening%20Guide%20--%20%20vNetwork%20Rev%20B.pdf;jsessionid=849963B4EE5C82C85473272DCE9F0DF0

4 http://cio.cisco.com/warp/public/cc/pd/si/casi/ca6000/prodlit/vlnwp_wp.htm

5 http://communities.vmware.com/servlet/JiveServlet/downloadBody/11846-102-1-11684/vSphere%20Hardening%20Guide%20--%20%20vNetwork%20Rev%20B.pdf;jsessionid=849963B4EE5C82C85473272DCE9F0DF0

6 http://www.intel.com/network/connectivity/vtc_vmdq.htm

7 http://download.intel.com/design/network/applnots/321211.pdf

8 http://www.intel.com/technology/eedc/index.htm

9 http://download.intel.com/technology/computing/vptech/Intel(r)_VT_for_Direct_IO.pdf

10 Source: Internal testing by Intel.

Printed in USA 0210/BY/MESH/PDF Please Recycle 323336-002US