



Standardizing Data Center Server- Network Edge Virtualization

October 2010

The following companies collaborated in the development of this white paper: Blade Network Technologies, Broadcom, Brocade, Citrix, Emulex, Extreme Networks, HP, IBM, Intel, Juniper Networks and QLogic.

In data centers worldwide, IT sprawl is reaching a breaking point, and managing the proliferating data center resources has become arduous. Virtualization is fueling this sprawl, spurring dramatic growth in the number of server instances and adding complexity to data center management.

As data centers pack in more servers to support virtualization and boost application performance, they also must add more network switches. With virtualization, a physical server runs multiple virtual servers, called virtual machines (VMs); each VM requires a connection to a virtual switch (vSwitch). The vSwitches are connected to the physical network to allow communication among the VMs and the external network.

An increasingly important part of the data center infrastructure is what is sometimes referred to as the 'virtual edge': where the physical and virtual realms overlap, and where the virtualized network extends into the virtual server. The virtual edge blurs the traditional demarcation between network and server equipment and administration.

The virtual edge architecture is composed of both server and network elements. It overlaps the server-network edge of the data center – i.e., the link between servers and the first tier of the network infrastructure. It is the virtual nature of the server-network edge that allows for the dynamic creation, deletion, and movement of network resources in support of the dynamic creation, deletion, and movement of virtualized servers.

Within modern data centers, each class of technology has typically been controlled by a different set of experts with different domain knowledge, each with their own complex, multi-step processes for deploying new infrastructure or implementing changes. Now, the advent of vSwitches – typically Layer 2 bridges within the server that allow VMs to talk to each other and optionally to the external network – has created the need for even higher-level coordination between server and network managers.

Still, despite best intentions, data center management remains largely a manual process requiring many complex and time-consuming steps, as shown in Figure 1.

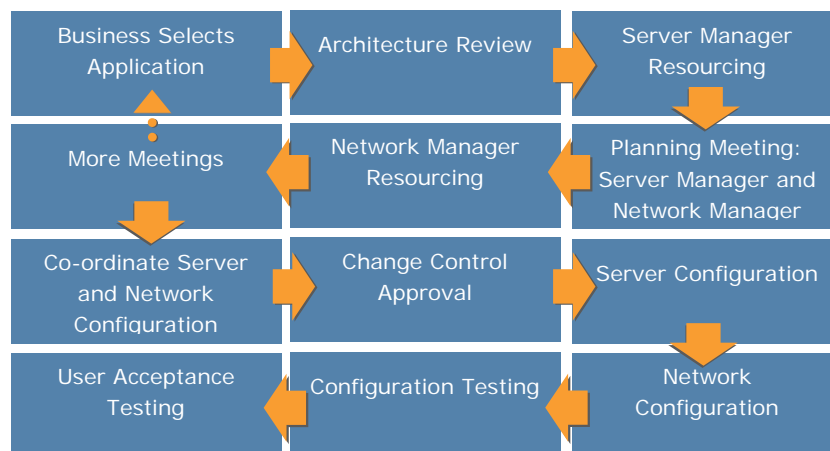


Figure 1: Data Center Management Before EVB

Manual provisioning typically takes a number of people – most often, valuable and busy experts – many steps and multiple weeks to accomplish, and the result is likely to be error-prone.

As an example, a server administrator who wants to deploy a new VM must first call a network administrator to get a switch provisioned. At least two experts are tied up in the switch provisioning process, which itself is complicated. If the vSwitch and edge switch are configured improperly, network connectivity can be interrupted or, worse, security vulnerabilities can be introduced. Deploying and configuring VMs, in this way, not only jeopardizes business operations, it also squanders considerable time and patience.

Automating and simplifying the interactions between the server and networking departments offers a number of potential benefits:

- Decreased overall operational expense (OpEx) in the data center due to streamlined and consistent processes between server admins and network admins.
- More reliable operation that is less prone to 'pilot error.'
- Greater efficiency of both people and equipment.
- Faster provisioning, with the ability to reuse pre-defined configurations.
- Optimal resource utilization and less application downtime.
- Increased flexibility, so the infrastructure can respond and adapt more easily to change.
- Improved ability for companies to take advantage of cloud computing and other advanced IT services while avoiding the coordination challenges.

If the automation is based on open standards, additional benefits accrue, such as the ability to choose vendors' solutions best able to meet individual companies' needs and budgets, as well as expertise that extends across multiple products and vendors.

This paper addresses the standardization of server-network edge virtualization, which provides the primary means for virtual server communication. We describe how support for Edge Virtual Bridging (EVB) technologies, including the new Virtual Ethernet Port Aggregator (VEPA) technologies now making their way through the IEEE standards bodies, can make it easier for businesses to achieve server-network edge virtualization in the data center. In addition to the benefits of automation and open standards, the EVB approach also helps preserve existing IT investments by allowing migration to new, modern systems without having to replace current equipment.

We believe that companies implementing an EVB approach to server-network edge virtualization can reduce IT costs, support cloud computing and mobility initiatives, and ultimately remain more competitive.

Limitations of current approaches

Industry analysts predict that by the middle of the decade, as much as three-quarters of data center server workloads will be virtualized. In addition to servers, data center networks and storage also achieve advanced initiatives, such as public and private cloud computing, by way of virtualization.

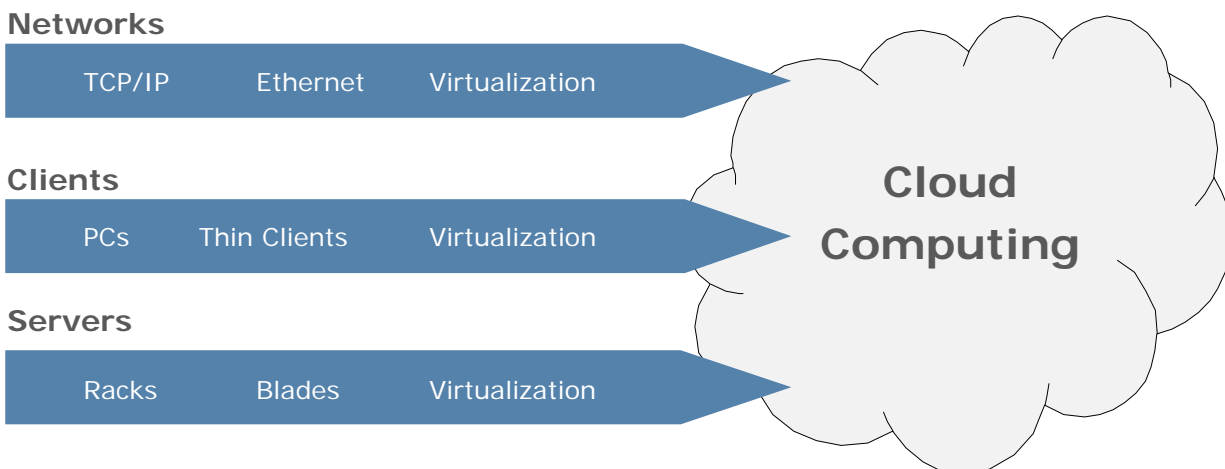


Figure 2: The Path to Cloud Computing

Data centers have historically been designed for physical environments. This means that companies today are applying yesterday's tools and capabilities to address current and future opportunities: trying to virtualize IT architectures that are not necessarily optimized for virtualization, especially in terms of the network.

In addition, a lack of process automation can severely complicate the migration to VMs. Reliance on manual configuration leads to inefficient process interaction between disciplines, or the dreaded 'silo impasse' syndrome. When the experts in charge of managing servers, networks, and storage are separated into their own isolated silos, they can't easily coordinate shared activities or support virtualization that spans two or more domains. Manual coordination drains experts' time, which leads to inefficiency, extra costs, and a general climate of frustration.

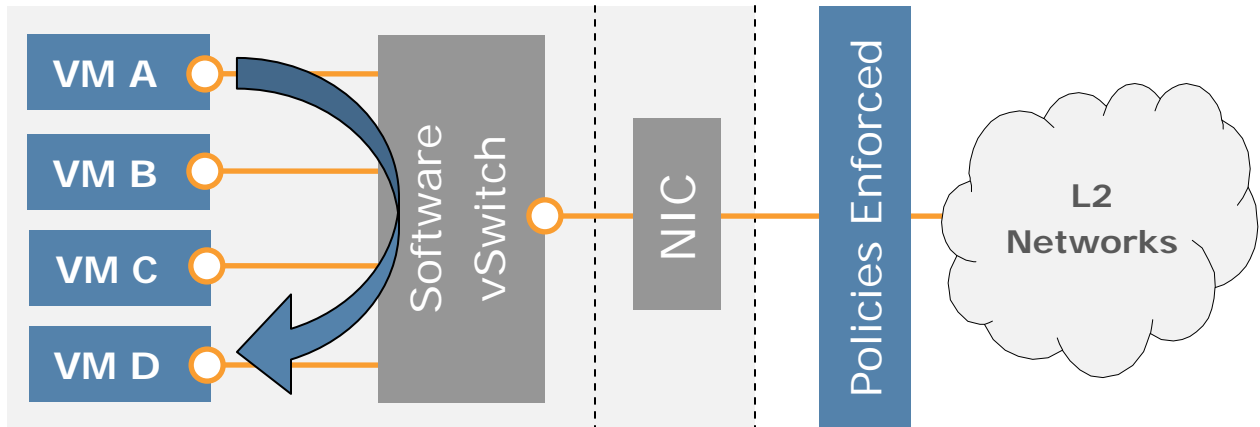
At the server-network edge of the data center, server I/O can be consolidated by combining multiple lower-bandwidth connections into a single, higher-bandwidth connection, or by converging different network types (LAN and SAN) to reduce the amount of physical infrastructure required at the server-network edge. To simplify the data center infrastructure, many data center administrators are deploying converged networking technologies such as iSCSI, Fibre Channel over Ethernet (FCoE), and Data Center Bridging (DCB).

Still, these converged networking technologies need adaptation before they can function in a virtualized environment. Virtual Ethernet Bridges (VEBs) provide one approach to this adaptation between VMs and their associated physical networks. VEBs are useful for both traditional and newer, converged network connections.

The VEB supports local bridging between multiple virtual end stations and, optionally, connectivity to the external bridging environment. VEBs expand the capabilities of hypervisors, but by definition, Layer 2 traffic between VMs connected to a VEB remains within the server. To allow consistent policy enforcement, fully visible VM-to-VM traffic, VEBs must provide Layer 2 access and traffic controls that are comparable to those available in physical network switches.

A VEB can be implemented within the server software either as a virtual Ethernet switch or as embedded hardware within a Network Interface Controller (NIC), as seen in Figure 3.

Software VEB (aka vSwitch)



Hardware VEB

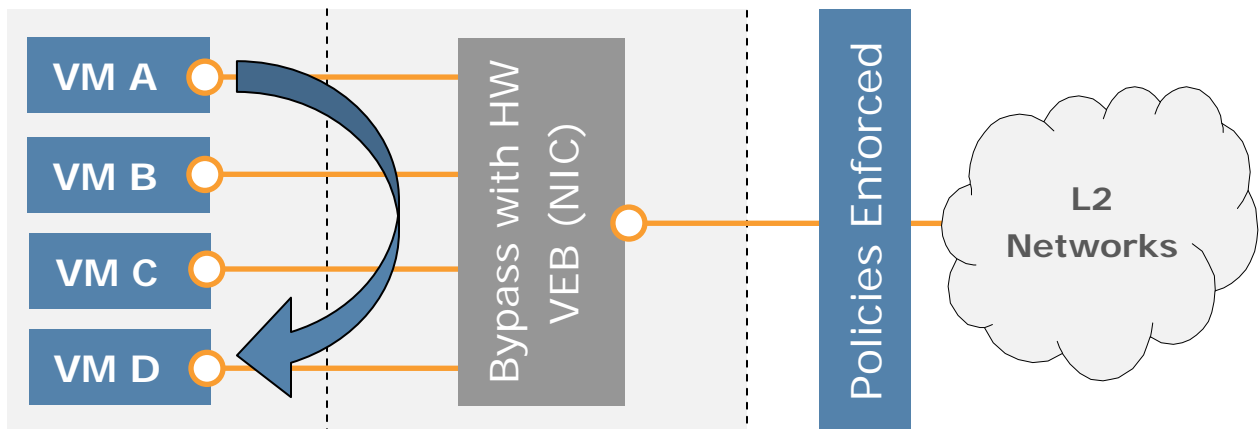


Figure 3: Software and Hardware VEBs

In its software configuration as a virtual switch, the VEB's advantages include:

- Deployment without an external switch.
- High-level memory-bandwidth performance for VM-to-VM traffic.
- Ability to work with existing operating systems, servers, adapters, and switches.
- Automated creation and migration of vSwitch state.
- Support for large numbers of VMs as well as a wide variety of Ethernet-based use cases.

In hardware configurations, VEBs also offer lower latency and high performance with very low CPU cycles, incurring no hypervisor overhead for most VM traffic.

Most of today's VEBs do not include a full suite of monitoring capabilities, network policy enforcement, or management scalability. This may change over time, as hardware-based VEBs increase hardware capabilities and leverage software functions, but they will likely remain less capable than the physical network switches to which they directly connect.

In addition to VEBs, some proprietary approaches also promise to ease the coordination between servers and networks in a virtualized environment. These proprietary approaches, however, require new hardware for their implementation, with changes to NIC and switch devices, as well as new software.

The requirements posed by proprietary environments impact the tools, diagnostics, and processes used to deploy and manage the virtualized data center environment. Businesses have to perform a forklift upgrade to migrate to these proprietary solutions, which adds significant data center IT costs rather than improving their capital expenditures (CapEx) or investment protection calculations.

EVB: a standard framework for server-network edge configuration

Edge Virtual Bridging (EVB) is the umbrella term for the range of new technologies and associated protocols that are being standardized to ease the coordination and management issues faced by virtualized servers at the server-network edge.

More specifically, the EVB environment encompasses VEBs, Virtual Ethernet Port Aggregator (VEPA) technologies, and protocols that help automate the coordination and configuration of network resources. VEBs and VEPAs in the EVB environment have unique visibility to the configuration of virtual NICs and thus can simplify deployment of the system.

The EVB standard, being created by the IEEE 802.1 Work Group as IEEE 802.1Qbg, is defining, among other things, how external bridges and VEBs can talk to each other to exchange configuration information. An industry coalition of leading vendors is developing important new EVB-related technologies that include both VEPA and the Virtual Station Interface Discovery Protocol (VDP).

A VEPA, which provides an alternative switching approach to VEB, sends all VM-generated traffic to an external switch. VEPA also moves the network demarcation back to the adjacent switch, giving the security responsibility for VM access to the physical network switch.

The VDP protocol is used to automatically associate and de-associate a VM to a state resident in the network. In doing so, VDP can automate the migration of network state ahead of a VM's migration across systems. VDP can be used in conjunction with both VEB- and VEPA-based solutions.

When VEPA mode is enabled, the external switch applies filtering and forwarding rules to the VMs and sees all traffic generated by the VMs. If necessary, traffic destined for VMs on the same physical server is returned to that server via a 'reflective relay' process commonly known as a

'hairpin turn' – aptly named because the traffic is capable of doing a 180-degree turn, but only when necessary.

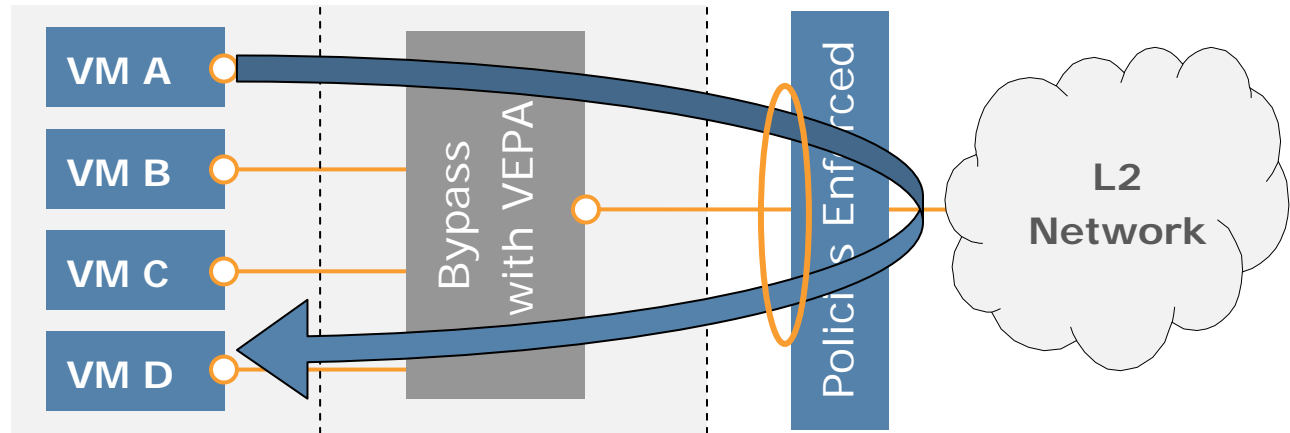


Figure 4: Hardware VEPA with Reflective Relay

VEPA: exploiting physical network capabilities for virtual systems

The IEEE 802.1 Work Group's Data Center Bridging (DCB) Task Group is well underway in the formal standardization process for the IEEE 802.1Qbg EVB standard. The project authorization request (PAR) for the EVB standard was approved in December 2009 by IEEE, and the first EVB standard draft has completed task group ballot. Additionally, a version of the IEEE 802.1Qbg EVB proposal was provided by a coalition of vendors, enabling early implementation. This process is similar to the coalition that created the Converged Enhanced Ethernet (CEE) proposal that has been implemented by many Ethernet switch vendors.

The IEEE 802.1 Work Group has also decided to standardize on reflective relay: the on-the-wire behavior needed to support VEPA in end stations. VEPA is seen as an appealing technology due to its minimal impact on data centers' existing NICs, bridges, standards, and frame formats, which allows easy retrofit into most existing data centers.

VEPA allows an external switch to have visibility into, and – when combined with VDP – policy control over, all of the VM traffic flows. The reflective relay used by VEPA can be enabled in many existing switches through a simple firmware upgrade that does not require hardware changes.

VEPA supports all types of network communications, including broadcast, multicast, and unicast delivery. For the most efficient delivery of broadcast and multicast, a VEPA replicates the traffic to each VM locally on the server.

A critical design goal for VEPA is to incorporate existing IEEE standards and, if necessary, specify minimal changes so that most existing NICs and virtual switches can implement it without a hardware upgrade. As a result, VEPA complements a VEB because it is an easy addition.

VEPA simplifies NIC implementations while allowing the external switch to be leveraged, so that VMs can take advantage of all the features and policies provided by access switches. VEPA enables a consistent level of network policy enforcement, leveraging the more complete policy-enforcement capabilities already included in the adjacent switch.

VEPA also provides visibility of traffic between VMs, so that network administrators can use the same network management tools designed for physical adjacent switches to monitor the VMs. With the automation that VEPA enables, server administrators do not have to do as much network configuration, reducing the time and complexity of their jobs.

Benefits of using VEPA at the server-network virtual edge include:

- A completely open, industry-standard architecture without proprietary attributes or frame formats – and with the flexibility and choice that comes with multi-vendor solutions.
- An architecture that achieves better bandwidth utilization (for both software- and hardware-based VEPAs), because it works without having to send all traffic all the way to the network core (or controlling bridge).
- Decreased hardware complexity for hardware-based VEPAs, with low overhead and latency, especially for small packet sizes.
- Easy implementation, typically as a simple software upgrade.
- A consistent switching and security model across different hypervisor technologies and across virtualized and non-virtualized workloads.
- Low-cost implementation along with investment protection, because the VEPA approach minimizes changes to NICs, software switches, and external switches.
- Ability to leverage mature, robust, and scalable implementations of switching and security in purpose-built hardware-based external networking equipment.
- Simplified management and lower OpEx for both networking and server operations.
- Dramatically lower cost of server virtualization by using an integrated solution compared to a collection of disparate applications from multiple vendors.

Implementation of VEPA as software or hardware

Server-side VEPA solutions can be implemented either in software or in conjunction with embedded hardware within a NIC.

A software VEPA can support a physical NIC as well as a logical uplink to a virtual NIC. Software-based VEPA solutions can be implemented as simple upgrades to existing software virtual switches in hypervisors. For example, the Red Hat enterprise Linux (RHEL) 6.0 kernel-based virtual machine (KVM) hypervisor provides integrated VEPA support.

A hardware implementation of a VEPA can be used by NICs that support the PCIe I/O virtualization (IOV) standards, including the Alternative Routing ID Interpretation (ARI) and single-root IOV (SR-IOV) specifications. NICs that support the SR-IOV standard can be upgraded easily to support the VEPA mode of operation.

Software and hardware VEPAs can also be combined. In this case, SR-IOV-attached virtual NICs (i.e., hardware VEPAs) provide VM interfaces delivering the highest levels of performance, while highly scalable software VEPAs are used to handle large number of virtual NICs with less-stringent performance requirements.

Multichannel technology

The IEEE 802.1 Work Group is also defining a way to allow the simultaneous operation of both VEB and VEPA modes in the server, in data centers where administrators choose VEB switches for their performance as well as VEPA switches for their network manageability. To address the coexistence of VEB and VEPA, the group added an optional Multichannel technology that allows the traffic on a physical network connection or port (e.g., NIC) to be logically separated into multiple channels, called S-channels, as if they were independent, parallel connections to the external network. Each logical channel can be assigned to any type of virtual switch (VEB or VEPA) or directly mapped to any VM or other services within the server.

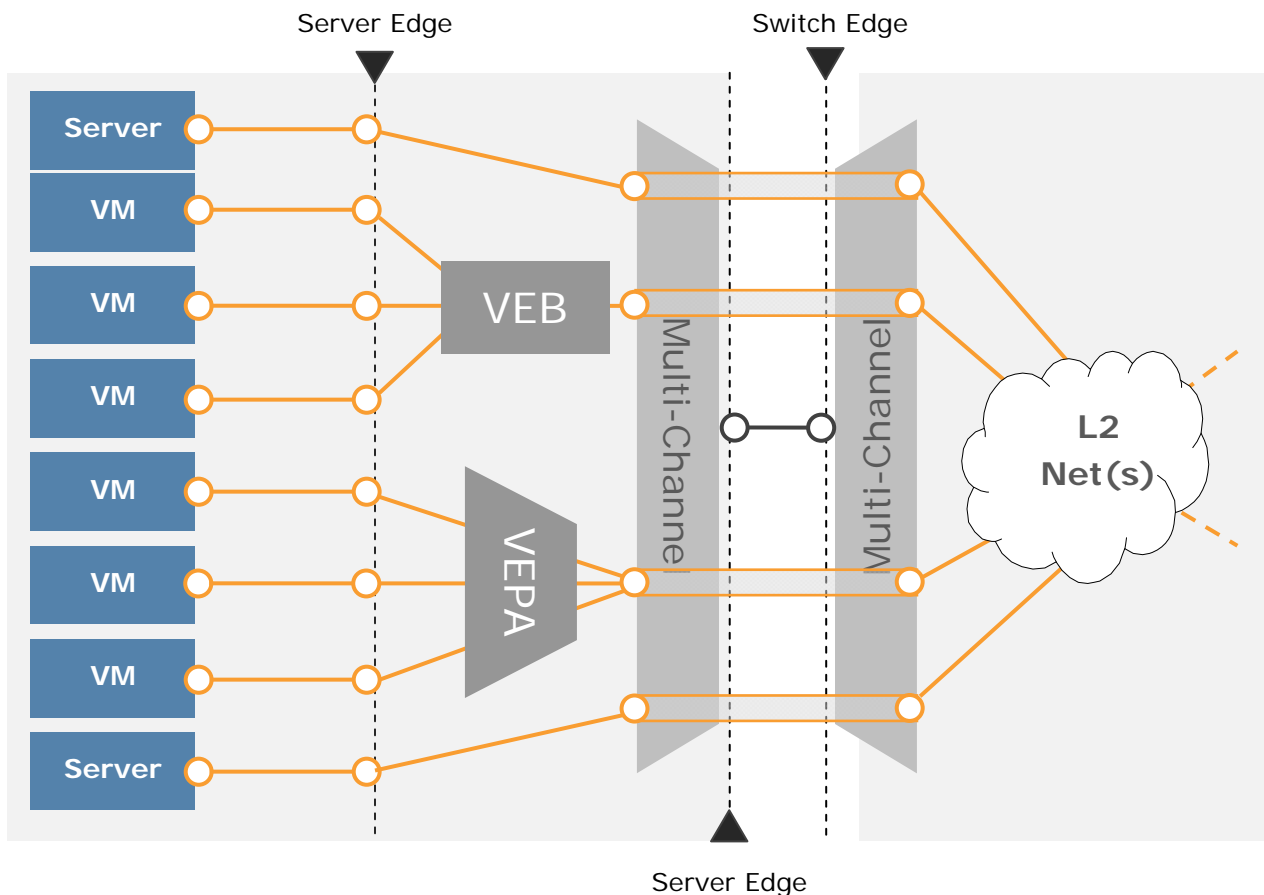


Figure 5: Multichannel Technology

With Multichannel technology, IT architects can better match application requirements to the design of their specific network infrastructure: VEB for increasing performance of VM-to-VM traffic; VEPA for management visibility into the VM-to-VM traffic; sharing physical NICs with direct-mapped VMs; and optimized support for individual applications such as traffic monitors, firewalls, and virus detection software.

Multichannel uses existing standard Service VLAN tags (S-Tags) that were standardized in IEEE 802.1ad (Provider Bridging), commonly referred to as QinQ. Multichannel technology uses the S-Tag and incorporates VLAN IDs in these tags to represent the logical channels of the physical network connection.

Automation of virtual edge provisioning: overcoming the 'silo impasse'

Automation of the provisioning processes is important for freeing the various technology disciplines and experts within the data center from the burden of lengthy interactions among IT staff. By providing the infrastructure and tools to automate the management of the virtual edge, EVB technologies change the nature of the management process and can help bridge the traditional silos. As a result, they enable automatic, dynamic, and rapid network resource provisioning among the various data center functions.

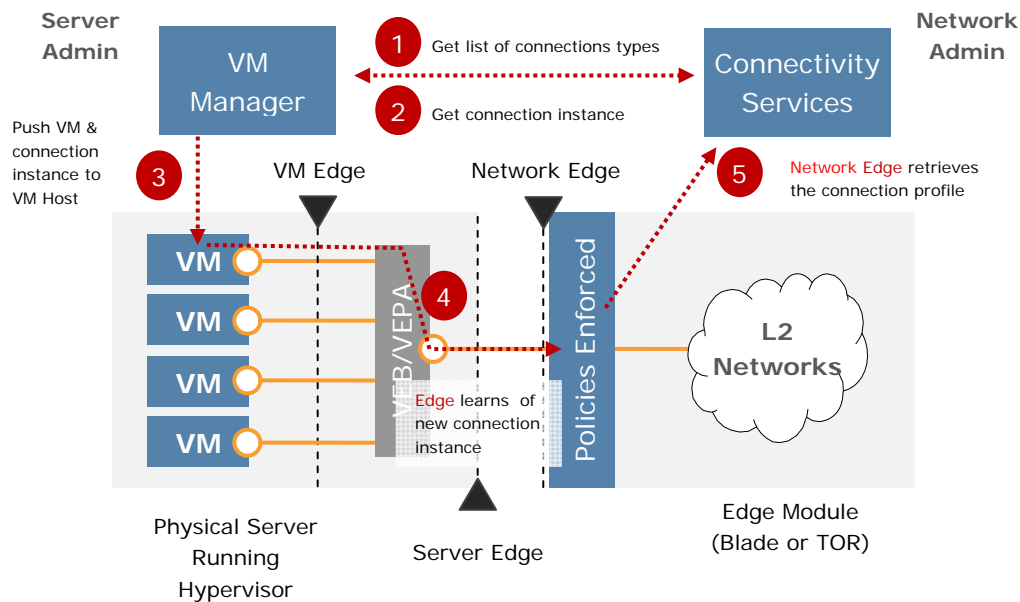


Figure 6: Virtual Network Provisioning with EVB

One key to successful automation is maintaining a consistent platform of features and allowing clear delineation of roles and responsibilities. The VEPA and VDP environment supports successful automation by allowing network functionality to be applied consistently to all VM traffic, whether it goes through a VEPA or a VEB.

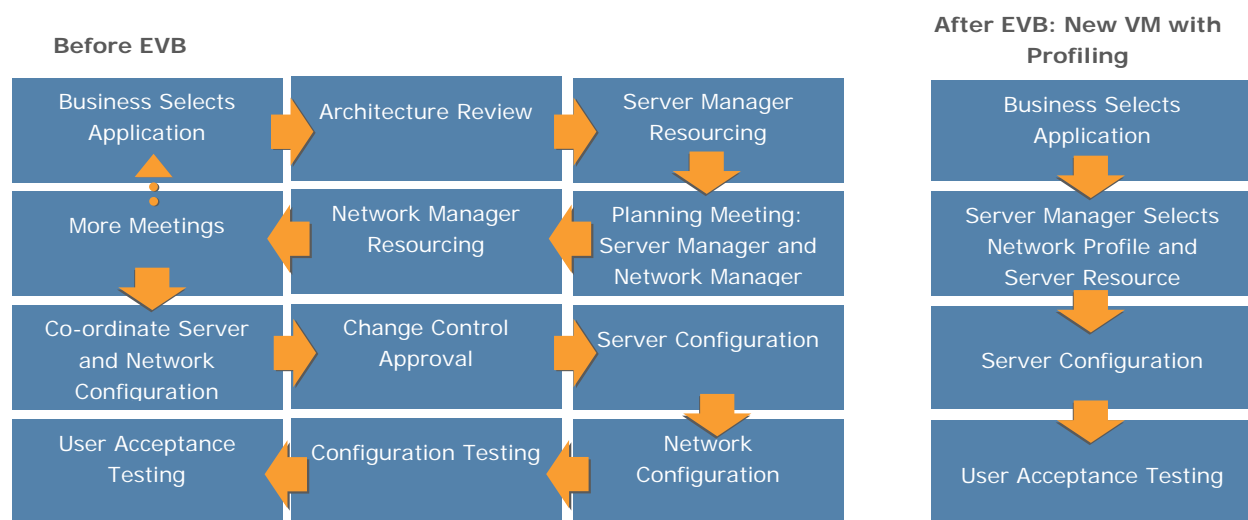


Figure 7: Data Center Management Before and After VEB

Protocols for discovery and auto-configuration

As part of the IEEE 802.1Qbg standards efforts, a number of protocols have been defined for Virtual Station Interface (VSI) discovery and auto-configuration of VEPA, VEB, and Multichannel. These discovery and configuration protocols allow the data center network to discover each physical NIC's virtualization capabilities and automate the configuration of network state associated with each virtual NIC – and they enable VEPA and VEB to provide the full benefits of network visibility and management of the virtualized server environment.

These protocols include:

- **Edge Discovery and Configuration Protocol (EDCP).** EDCP informs an adjacent switch of the presence of a VEPA (or VEB), which can be attached to a physical port or a logical S-channel (when using Multichannel).
- **Virtual Station Interface Discovery and Configuration Protocol (VDP).** VDP supports the dynamic creation of VMs by signaling the state, identity, and resource requirements for each VM attached to each VEPA or VEB. VDP enables the association and de-association between a VM (aka VSI Instance in the proposed IEEE 802.1Qbg specification) and network resident state (aka VSI Type). It simplifies and automates Virtual Server (VS) network configuration by essentially enabling the movement of the network resident state (i.e., a VSI Type) when the VM (i.e., VSI Instance) moves. VDP also can be used to automate the configuration of network state and the association of network state to a VSI Instance, and it is the VDP protocol that supports profiling for each VM.
- **S-channel Discovery and Configuration Protocol (CDCP).** CDCP is used to create and delete S-channels (Multichannel).

Today, IEEE 802.1 control plane discovery operations are performed over unacknowledged protocols such as Link Layer Discovery Protocol (LLDP), which carries configuration information for EDCP, CDCP, and Data Center Bridging Capability Exchange Protocol (DCBX).

For VDP and potentially other upper-layer protocols, the Edge Control Protocol (ECP) is used to send acknowledgement frames. The ECP acknowledgements signal the sender that the receiver is able to receive an additional ECP data unit, providing both flow control and frame retransmission. ECP enables the sender to transmit discovery operations more frequently than would be the case with timer-based approaches, allowing support of thousands of VMs with very fast response times.

Summary

Businesses must wring the most from their data center IT resources, optimizing performance and support for mobility, cloud computing, collaborative technologies, and other services that can boost agility and competitiveness. At the same time, businesses are more cost-conscious than ever before. IT needs to streamline processes in the data center to reduce OpEx and simplify operations and management. Automation and coordination of roles and responsibilities support that goal.

The path forward for businesses worldwide will pass through virtualization on the way to cloud environments, in which data centers embody a true convergence of servers, networks, and storage. One important step on this path is to automate virtualization of the server-network edge of the data center, where servers and the first network tier meet.

The EVB standards from the IEEE 802.1 Work Group offer an excellent solution. Advantages of the open, industry-standard EVB architecture include:

- Easy integration into existing data center environments, avoiding the cost and hassle of equipment retrofit and forklift upgrades.
- Lower IT costs, both CapEx and OpEx.
- The full benefits of network visibility and management in a virtualized server environment.
- Flexibility in choosing from a number of vendors' offerings (across the key components of the stack) to build a solution that meets a company's individual needs.
- Less-expensive deployment compared to proprietary solutions.
- Faster path to virtualization and the potential of cloud computing, which are important capabilities for both short- and long-term business competitiveness.
- The easiest, most cost-effective path to next-generation server-network edge virtualization in the data center. The result will be greater visibility and control in an increasingly virtualized world.

GLOSSARY OF ACRONYMS & ABBREVIATIONS

ARI	Alternative Routing ID Interpretation
CapEx	Capital Expenditures
CDCP	S-channel Discovery and Configuration Protocol
CEE	Converged Enhanced Ethernet
DCB	Data Center Bridging
DCBX	Data Center Bridging Capability Exchange Protocol
ECP	Edge Control Protocol
EDCP	Edge Discovery and Configuration Protocol
EVB	Edge Virtual Bridging
FCoE	Fibre Channel over Ethernet
I/O	Input/Output
IOV	I/O Virtualization
iSCSI	Internet Small Computer System Interface
LAN	Local Area Network
LLDP	Link Layer Discovery Protocol
NIC	Network Interface Controller
OpEx	Operating Expenditures
PCIe	Peripheral Component Interconnect Express
QinQ	IEEE 802.1ad (an amendment to IEEE 802.1Q)
RHEL	Red Hat Enterprise Linux
SAN	Storage Area Network
SR-IOV	Single-Root I/O Virtualization
S-tag	Service VLAN tag
TOR	Top of Rack switch
VEB	Virtual Ethernet Bridge
VEPA	Virtual Ethernet Port Aggregator
VDP	Virtual Station Interface Discovery and Configuration Protocol
VLAN	Virtual Local Area Network
VM	Virtual Machine
VS	Virtual Server
VSI	Virtual Station Interface
vSwitch	Virtual Switch

