Data Center Networks: Virtual Bridging



Raj Jain
Washington University in Saint Louis
Saint Louis, MO 63130
Jain@cse.wustl.edu

These slides and audio/video recordings of this class lecture are at:
http://www.cse.wustl.edu/~jain/cse5700-25/

Student Questions

©2025 Raj Jain



Overview

1. Virtual Bridges to connect virtual machines
2. IEEE Virtual Edge Bridging Standard
3. Single Root I/O Virtualization (SR-IOV)
4. Aggregating Bridges and Links: VSS and vPC
5. Bridges with a massive number of ports: VBE

Student Questions

©2025 Raj Jain

Virtualization

"Virtualization means that Applications can use a resource without any concern for where it resides, what the technical interface is, how it has been implemented, which platform it uses, and how much of it is available."

-Rick F. Van der Lans
in Data Virtualization for Business Intelligence Systems

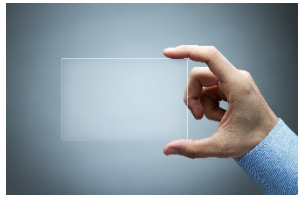
Student Questions

What is Virtual?

- If you can see it and it is there ⇒ Its real



- If you can't see it but it is there ⇒ It's transparent



- If you can see it and it is not there ⇒ It's virtual



- If you can not see it and it is not there ⇒ It's gone



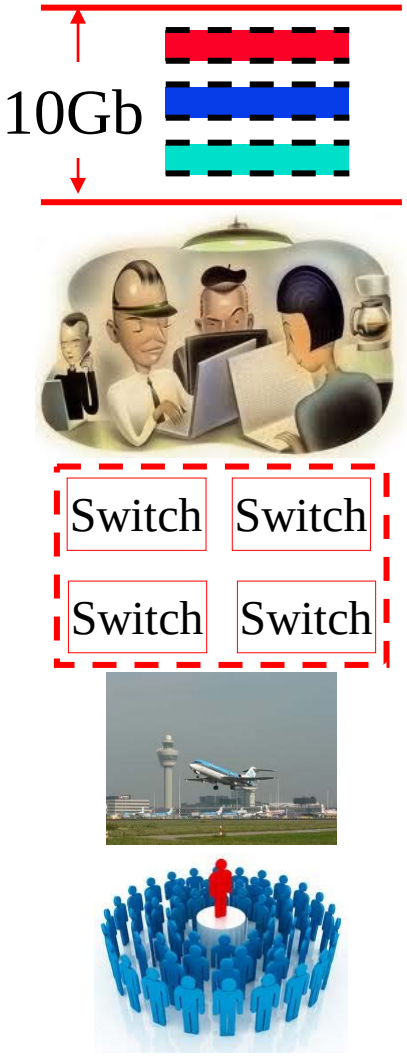
Ref: Cisco Live, "What you make possible," https://www.alcatron.net/Cisco%20Live%202013%20Melbourne/Cisco%20Live%20Content/Data%20Centre%20And%20Virtualisation/BRKVIR-2931%20%20End-to-End%20Data%20Centre%20Virtualisation.pdf

©2025 Raj Jain

Student Questions

5 Reasons to Virtualize

1. Sharing: Break up an extensive resource
   Large Capacity or high-speed
   E.g., Servers
2. Isolation: Protection from other tenants
   E.g., Virtual Private Network
3. Aggregating: Combine many resources into one, e.g., storage
4. Dynamics: Fast allocation, Change/Mobility, load balancing, e.g., virtual machines
5. Ease of Management ⇒ Easy distribution, deployment, testing



Student Questions

Advantages of Virtualization

- Minimize hardware costs (CapEx)
  Multiple virtual servers on one physical hardware
- Easily move VMs to other data centers
  - Provide disaster recovery. Hardware maintenance.
  - Follow the sun (active users) or follow the moon (cheap power)
- Consolidate idle workloads. Usage is bursty and asynchronous.
  Increase device utilization
- Conserve power
  Free up unused physical resources
- Easier automation (Lower OpEx)
  Simplified provisioning/administration of hardware and software
- Scalability and Flexibility: Multiple operating systems



Ref: http://en.wikipedia.org/wiki/Platform_virtualization
Ref: K. Hess, A. Newman, "Practical Virtualization Solutions: Virtualization from the Trenches," Prentice Hall, 2009, ISBN:0137142978

©2025 Raj Jain

Student Questions

Virtualization in Computing

- **Storage**:
  - Virtual Memory ⇒ L1, L2, L3, ... ⇒ Recursive
  - Virtual CDs, Virtual Disks (RAID), Cloud storage
- **Computing**:
  - Virtual Desktop ⇒ Virtual Server ⇒ Virtual Datacenter

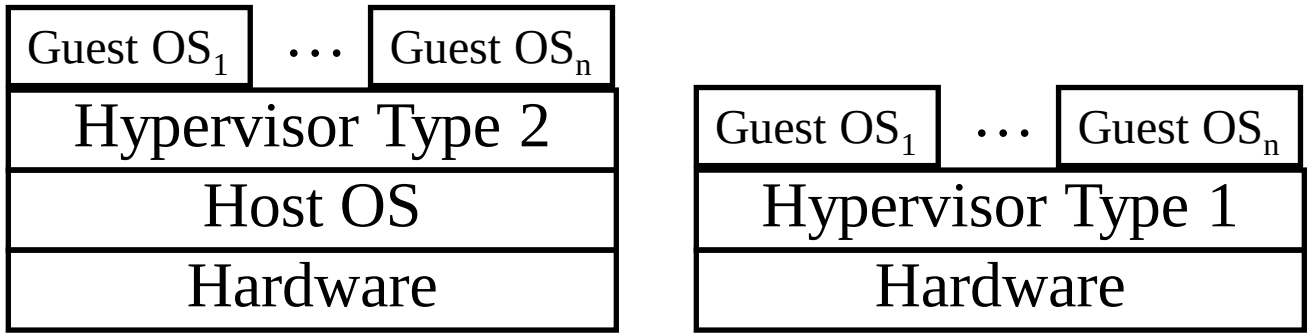
    Thin Client ⇒ VMs ⇒ Cloud
- **Networking**: Plumbing of computing
  - Virtual Channels, Virtual LANs, Virtual Private Networks





Student Questions

Server Virtualization Concepts



- Host OS: Runs on the bare metal
- Guest OS: Runs on the host OS,
  e.g., Windows 7 on Win 10
- Hypervisor: Software to support multiple virtual machines
  - Type 1: Runs on bare metal, e.g., Xen, VMware ESXi
  - Type 2: Runs on a host OS, e.g., MS Virtual PC
  - Type 0: Both 1 and 2,
    e.g., Linux Kernel-based Virtual Machine (KVM)

Ref: http://en.wikipedia.org/wiki/Hypervisor

Student Questions

©2025 Raj Jain

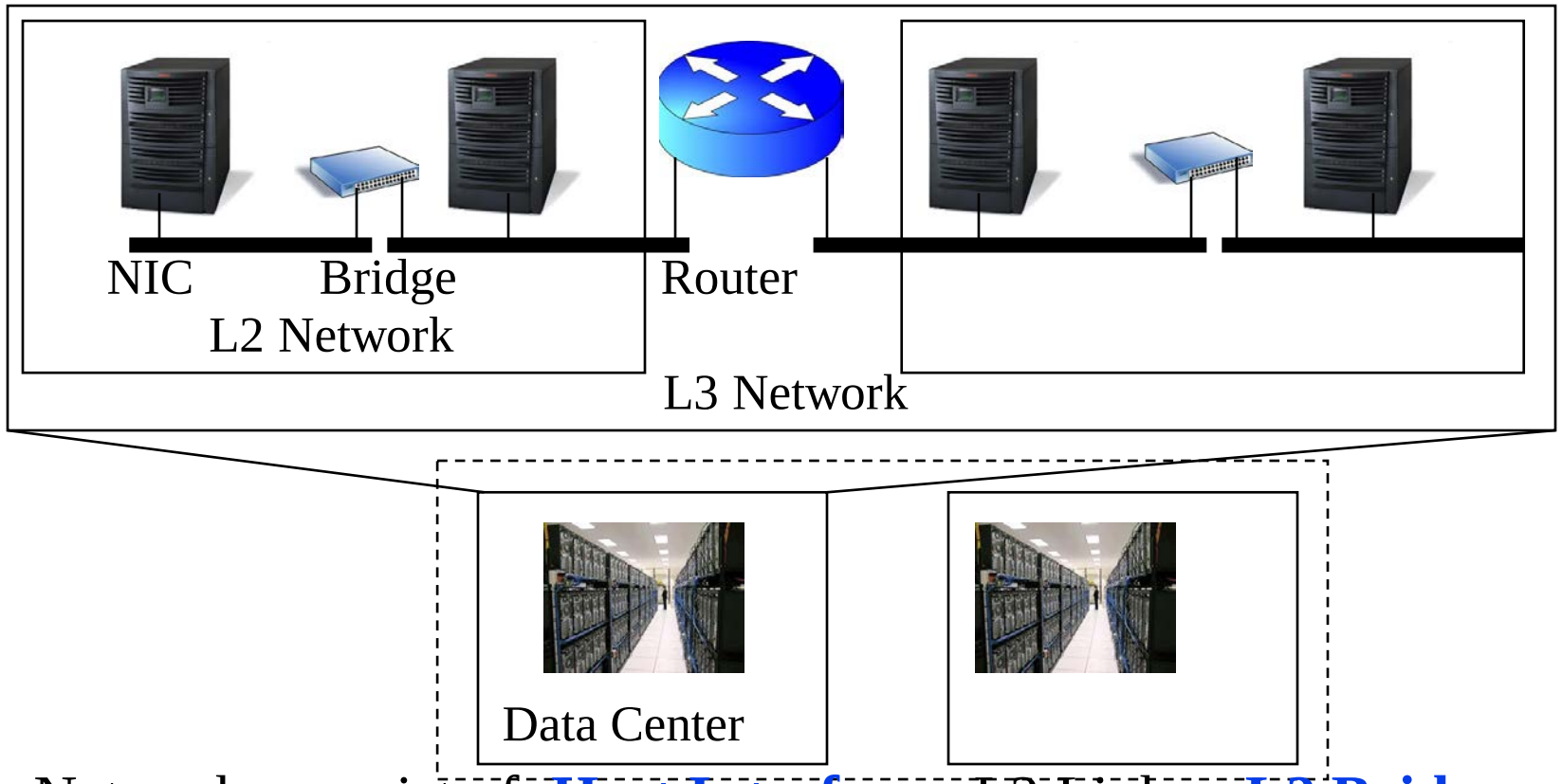
Network Virtualization

1. Network virtualization allows tenants to form an overlay network in a multi-tenant network such that tenants can control:
   1. Connectivity layer: The tenant network can be L2 while the provider is L3 and vice versa
   2. Addresses: MAC addresses and IP addresses
   3. Network Partitions: VLANs and Subnets
   4. Node Location: Move nodes freely
2. Network virtualization allows providers to serve a large number of tenants without worrying about:
   1. Internal addresses used in client networks
   2. Number of client nodes
   3. Location of individual client nodes
   4. Number and values of client partitions (VLANs and Subnets)
3. Network could be a single physical interface, a single physical machine, a data center, a metro, … or the global Internet.
4. . The provider could be a system owner, an enterprise, a cloud provider, or a carrier.

Student Questions

©2025 Raj Jain

Levels of Network Virtualization



- Networks consist of: **Host Interface** - L2 Links - **L2 Bridges** - **L2 Networks** - L3 Links - L3 Routers - L3 Networks – **Data Centers** – **Global Internet.**
- Each of these needs to be virtualized

Student Questions

©2025 Raj Jain

Network Virtualization Techniques

Entity	Partitioning	Aggregation/Extension/Interconnection**
NIC	SR-IOV	MR-IOV
Switch	VEB, VEPA	VSS, VBE, DVS, FEX
L2 Link	VLANs	LACP, Virtual PortChannels
L2 Network using L2	VLAN	PB (Q-in-Q), PBB (MAC-in-MAC), PBB-TE, Access-EPL, EVPL, EVP-Tree, EVPLAN
L2 Network using L3	NVO3, VXLAN, NVGRE, STT	MPLS, VPLS, A-VPLS, H-VPLS, PWoMPLS, PWoGRE, OTV, TRILL, LISP, L2TPv3 , EVPN, PBB-EVPN
Router	VDCs, VRF	VRRP, HSRP
L3 Network using L1		GMPLS, SONET
L3 Network using L3*	MPLS, GRE, PW, IPSec	MPLS, T-MPLS, MPLS-TP, GRE, PW, IPSec
Application	ADCs	Load Balancers

*All L2/L3 technologies for L2 Network partitioning and aggregation can also be used for L3 network partitioning and aggregation, respectively, by simply putting L3 packets in L2 payloads.

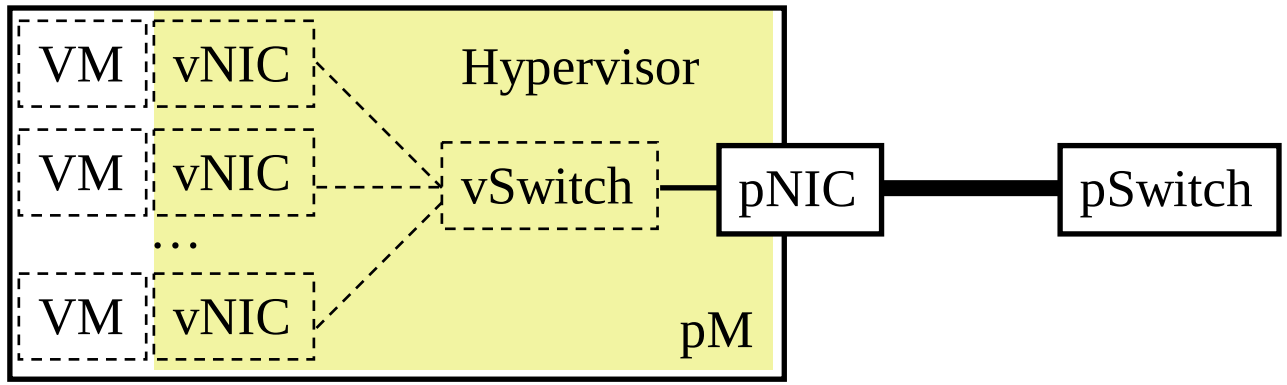
**The aggregation technologies can also be seen as partitioning technologies from the provider point of view.

Student Questions

©2025 Raj Jain

vSwitch

- **Problem**: Multiple VMs on a server need to use one physical network interface card (pNIC)
- **Solution**: Hypervisor creates multiple vNICs connected via a virtual switch (vSwitch)
- pNIC is controlled by the hypervisor and not by any individual VM
- **Notation**: From now on, prefixes p and v refer to physical and virtual, respectively. For VMs only, we use upper case V.

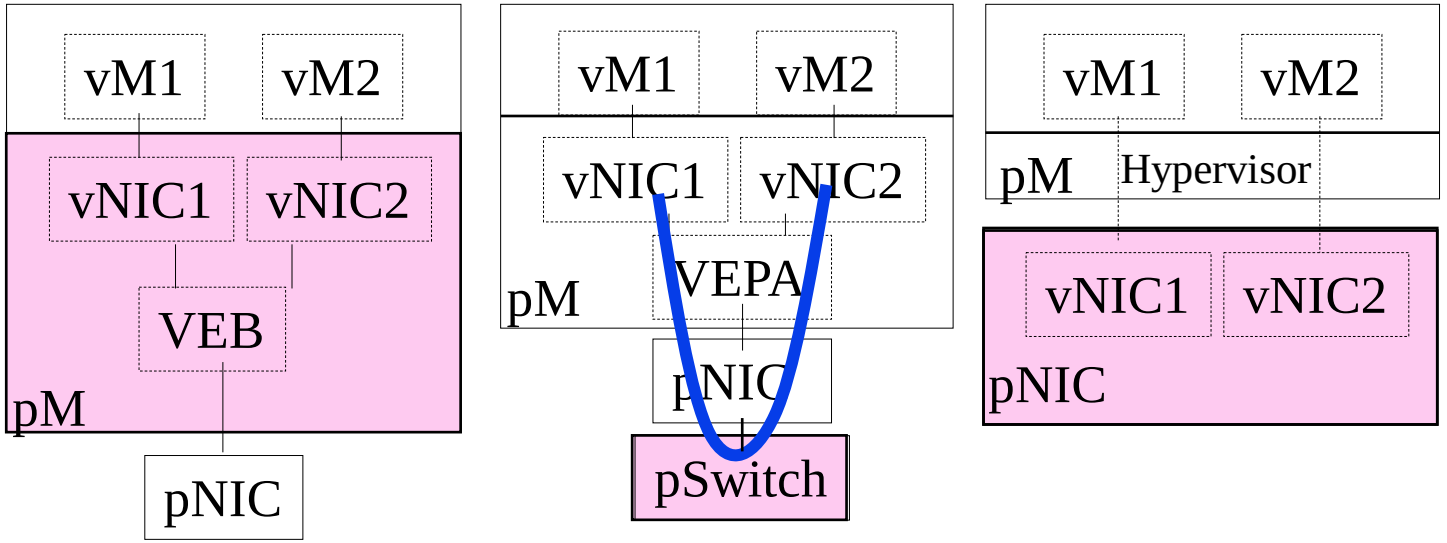


Ref: G. Santana, "Datacenter Virtualization Fundamentals," Cisco Press, 2014, ISBN: 1587143240

©2025 Raj Jain

Student Questions

Virtual Bridging



Where should most of the tenant isolation take place?

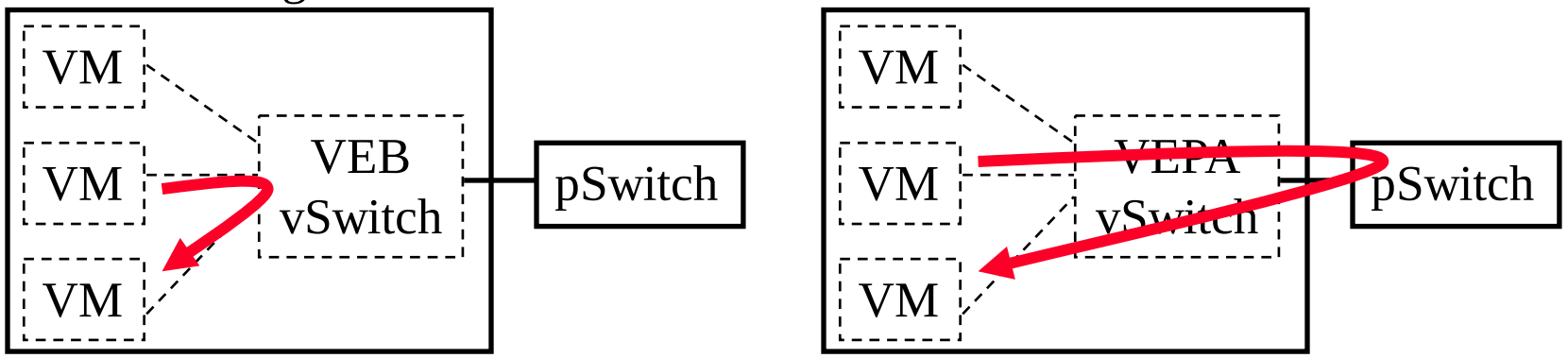
1. VM vendors: S/W NICs in Hypervisor w Virtual Edge Bridge (**VEB**): **802.1Qbg** (overhead, not ext manageable, not all features)
2. Switch Vendors: Switch provides virtual channels for inter-VM Communications using a virtual Ethernet port aggregator (**VEPA**): **802.1Qbg** (s/w upgrade)
3. NIC Vendors: NIC provides virtual ports using Single-Route I/O virtualization (**SR-IOV**) on PCI bus

Student Questions

©2025 Raj Jain

Virtual Edge Bridge

- IEEE 802.1Qbg-2012 standard for vSwitch
- Two modes for vSwitches to handle *local* VM-to-VM traffic:
  - **Virtual Edge Bridge (VEB):** Switch internally.
  - **Virtual Ethernet Port Aggregator (VEPA):** Switch externally
- VEB
  - could be in a hypervisor or network interface card
  - may learn or may be configured with the MAC addresses
  - VEB may participate in spanning tree or may be configured\
  - Advantage: No need for the external switch in some cases



Student Questions

©2025 Raj Jain

Virtual Ethernet Port Aggregator (VEPA)

- VEPA relays all traffic to an external bridge
- External bridge forwards the traffic. Called "*Hairpin Mode*."
  Returns local VM traffic to VEPA
  Note: Legacy bridges do not allow traffic to be sent back to the incoming port within the same VLAN
- **VEPA Advantages**:
  - Visibility: External bridge can see VM to VM traffic.
  - Policy Enforcement: Better. E.g., firewall
  - Performance: Simpler vSwitch ⇒ Less load on CPU
  - Management: Easier
- Both VEB and VEPA can be implemented on the same NIC in the same server and can be cascaded.

Student Questions

PCIe

- Peripheral Component Interconnect (PCI)
  Used in computers for I/O – storage, video, network cards
- Designed by PCI Special Interest Group (PCI-SIG)
- **PCI Express (PCIe)**: Serial point-to-point interconnect with multiple lanes, four pins per lane. X1=1 Lane, x32=32 lanes 2 GB/s/lane.
- **Root complex** is the head of connection to the CPU
- **Physical Function (PF)**: Ethernet, Fibre Channel, Video, …
- A PCIe card can provide multiple **virtual functions (VFs)** of the same type as PF, e.g., one 10Gbps pNIC = 2× 5Gbps vNICs

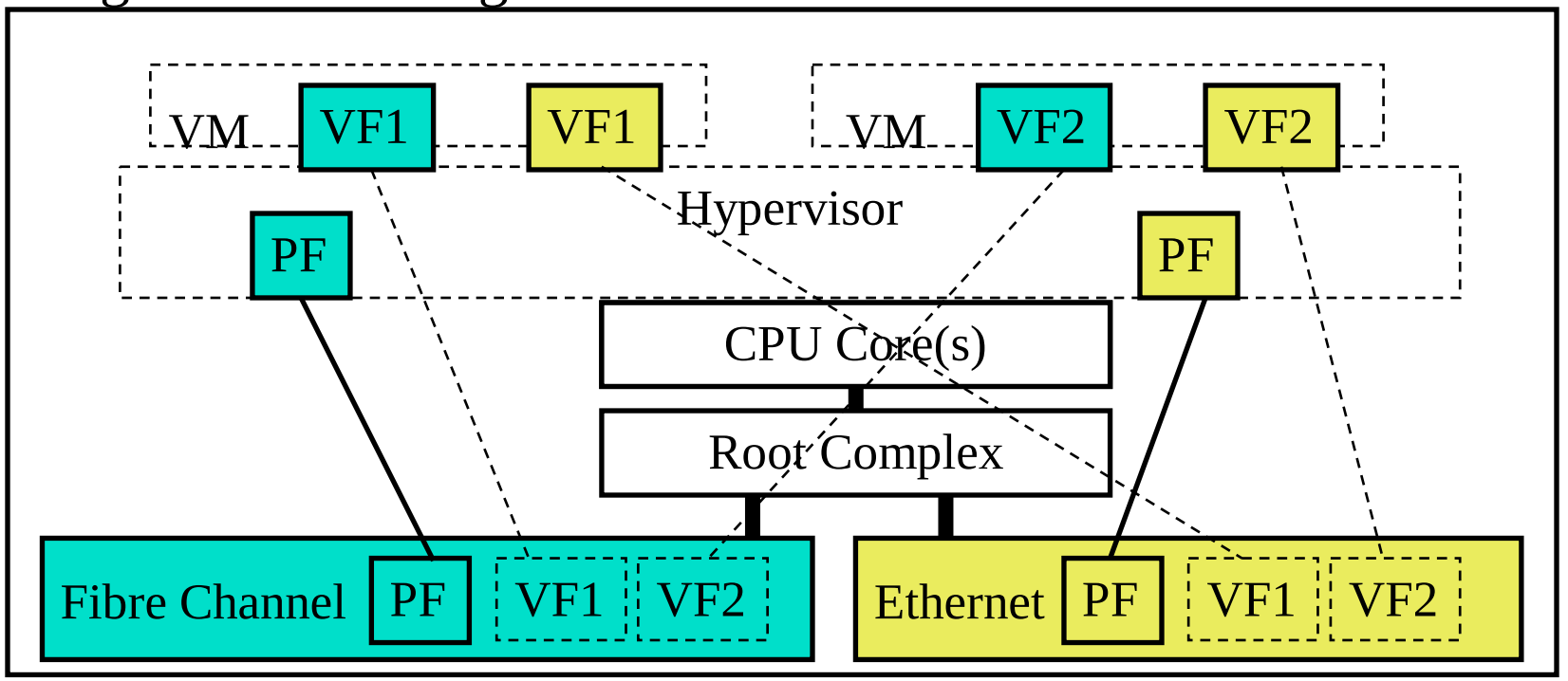
Student Questions

Ref: R. Emerick, "PCI Express IO Virtualization Overview," SNIA Education, 2012, http://www.snia.org/sites/default/files/RonEmerick_PCI_Express_IO_Virtualization.pdf (Excellent)

©2025 Raj Jain

Single Root I/O Virtualization (SR-IOV)

- After configuration by a hypervisor, VFs allow direct VM access without hypervisor overhead
- Single Root ⇒ Single hardware domain ⇒ In one Server

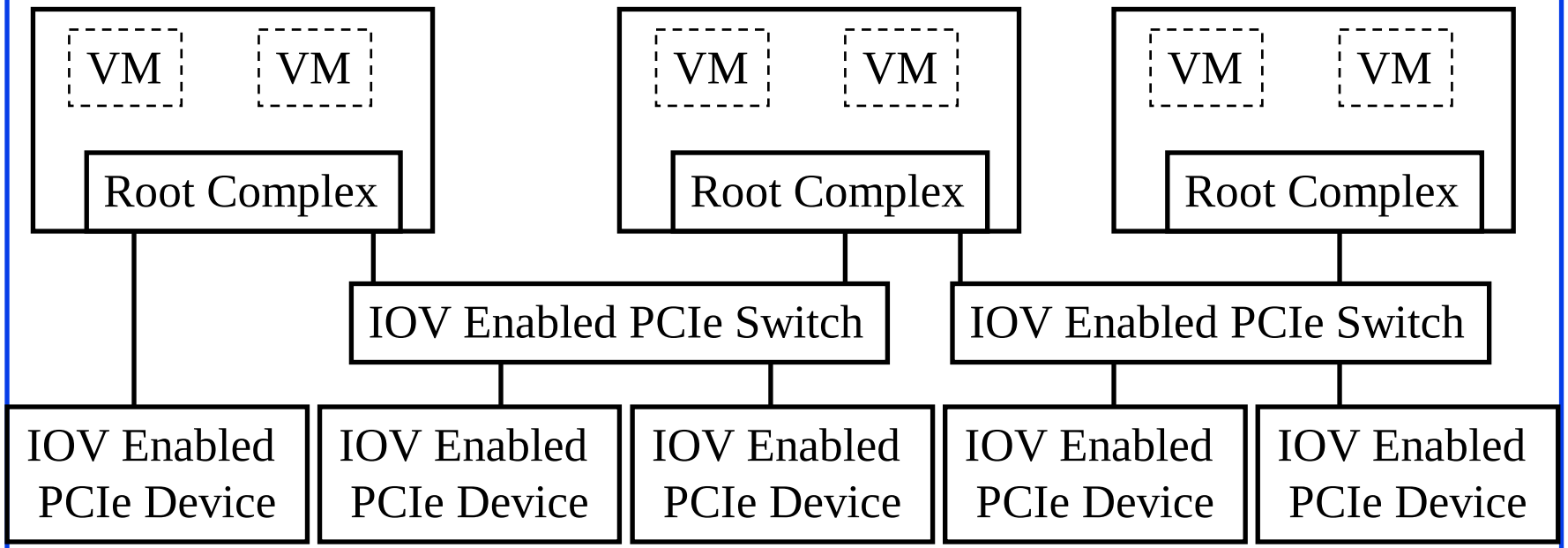


Student Questions

©2025 Raj Jain

Multi-Root IOV

- Multiple external PCIe devices accessible via a switch
  - Move the PCIe adapter out of the server into a switching fabric
  - Allows adapters to serve many physical servers
  - Used with rack-mounted or blade servers
- Fewer adapters → Less cooling. No adapters → Thinner servers



Student Questions

Combining Bridges

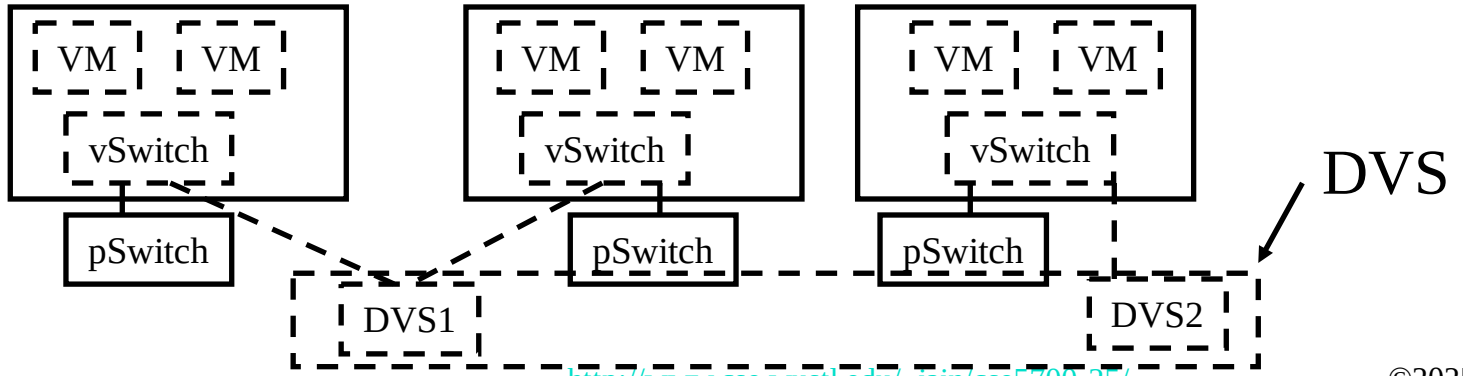
- **Problem**:
  - The number of VMs is growing very fast
  - Need switches with a massive number of ports
  - Easy to manage one bridge than 100 10-port bridges
  - How to make massive switches with ~1000 ports?
- **Solutions**: Multiple pSwitches to form a single switch
  1. Distributed Virtual Switch (DVS)
  2. Virtual Switching System (VSS)
  3. Virtual PortChannels (vPC)
  4. Fabric Extension (FEX)
  5. Virtual Bridge Port Extension (VBE)

Student Questions

©2025 Raj Jain

Distributed Virtual Switch (DVS)

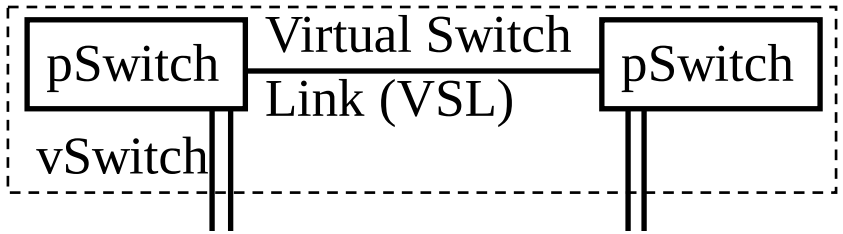
- VMware idea to solve the scalability issue
- A centralized DVS controller manages vSwitches on many physical hosts
- DVS decouples the control and data plane of the switch so that each VM has a virtual data plane managed by a centralized control plane
- Appears like a single distributed virtual switch
- Allows simultaneous creation of port groups on multiple pMs
- Provides an API so that other networking vendors can manage vSwitches and vNICs



Student Questions

Virtual Switch System (VSS)

- Allows two physical switches to appear as one
- Although VSS is a Cisco proprietary name, several vendors implement similar technologies. E.g., Virtual Switch Bonding by Enterasys.
- Implemented in Firmware → No degradation in performance
- Only one control plane is active.
  Data-plane capacity is doubled.
- Both switches are kept in sync to enable inter-chassis stateful switchover and non-stop forwarding in case of failure.

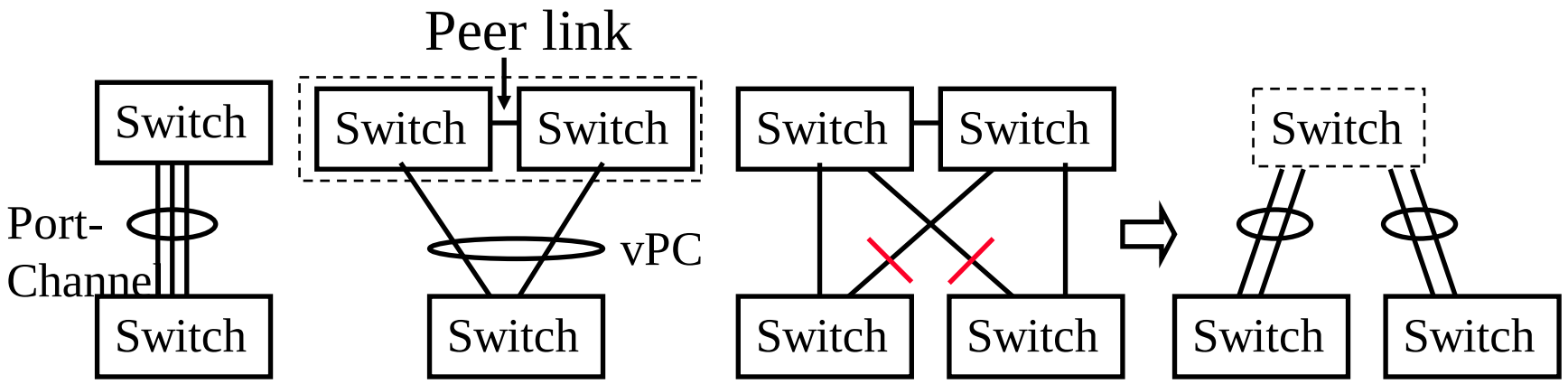


Student Questions

©2025 Raj Jain

Virtual PortChannel (vPC)

- **PortChannel**: Cisco name for aggregated link
- **Virtual PortChannel**: A link formed by aggregating links to multiple physical switches acting as a virtual switch
- The combined switch is called "**vPC Domain**"
- Each member of the vPC domain is called "**vPC peer**".
- vPC peer link is used to synchronize state and to forward traffic between the peers. No address learning on the peer link.
- All learned address tables are kept synchronized among peers. One peer learns an address ⇒ Sends it to everyone else.



Student Questions

©2025 Raj Jain

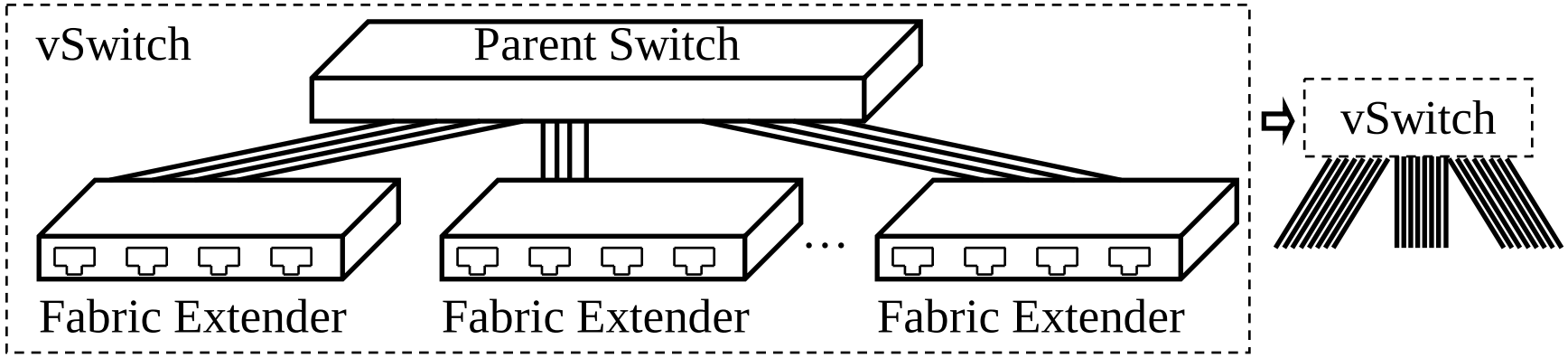
Virtual Port Channel (vPC)

- Allows aggregation of links going to different switches → STP does not block links → All capacity used
- Unlike VSS, it maintains two independent control planes
- Independent control plane → In-service upgrade
  Software in one of the two switches can be upgraded without service interruption.
- Falls back to STP → Used only in small domains
- vPC is Cisco proprietary. But other vendors have similar technologies. E.g., Split Multi-link Trunking (SMLT) by Nortel or "Multi-Chassis Link Aggregation (MC-LAG)" by Alcatel-Lucent.
- Standardized in RFC 7275 as "Multi-Chassis LACP"

Student Questions

Fabric Extenders

- Fabric Extenders (FEX) consists of ports that are managed by a remote parent switch
- 12 Fabric extenders, each with 48 host ports, connected to a parent switch via four 16 Gbps interfaces to a parent switch provide a virtual switch with 576 host ports
  ⇒ **Chassis Virtualization**
- All software updates/management, forwarding/control plane is managed centrally by the parent switch.
- A FEX can have an active and a standby parent.



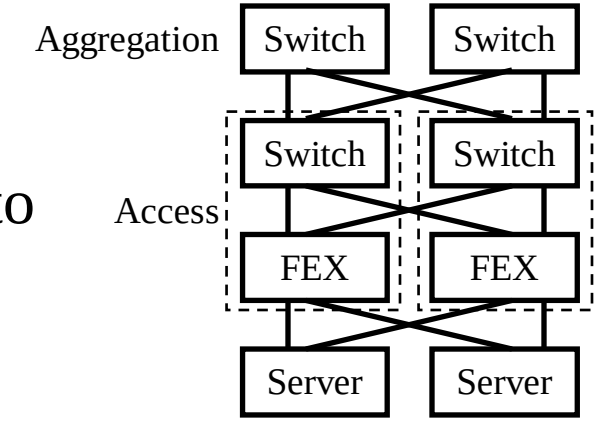
Ref: P. Beck, et al., "IBM and Cisco: Together for a World Class Data Center," IBM Red Book, 2013, 654 pp., ISBN: 0-7384-3842-1, http://www.redbooks.ibm.com/redbooks/pdfs/sg248105.pdf

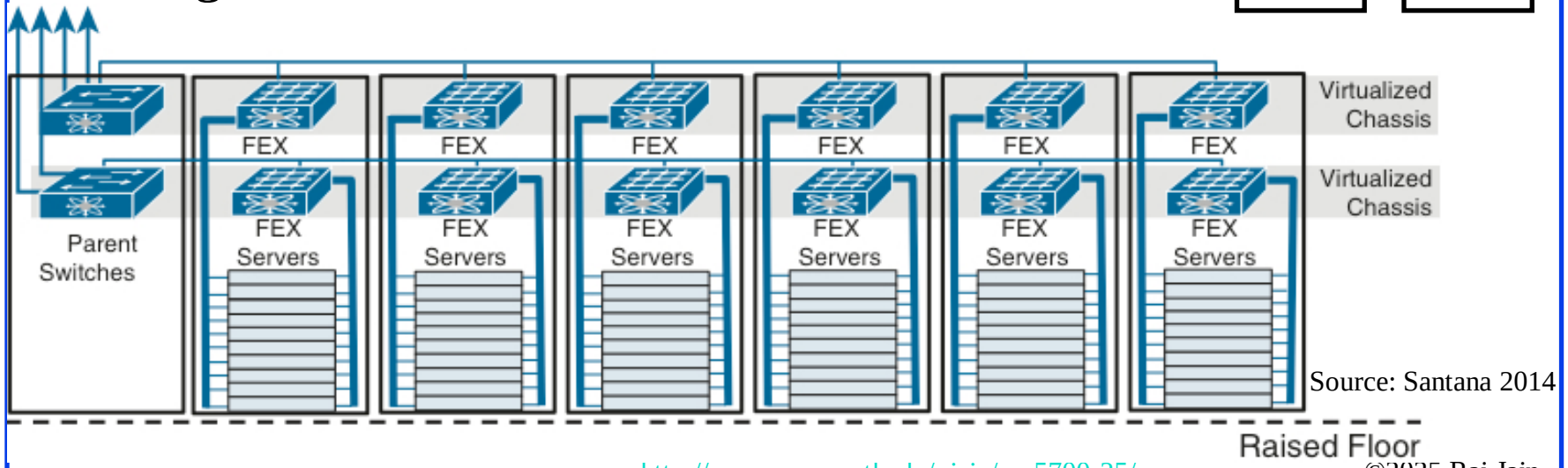
©2025 Raj Jain

Student Questions

FEX Topology Example

- All hosts are dual-homed to FEX
  ⇒ Two FEX per rack
- Both FEX are dual-homed to two parents
  ⇒ Two virtual access switches
- Virtual Access switches are dual-homed to aggregation switches.
- Using vPCs, all links can be active.





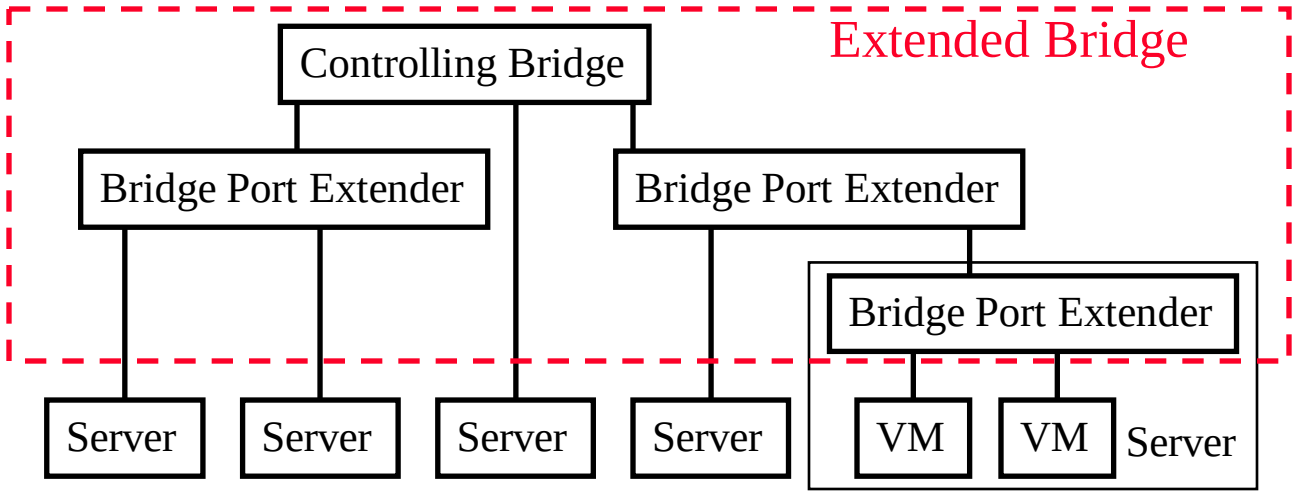
Source: Santana 2014

Student Questions

©2025 Raj Jain

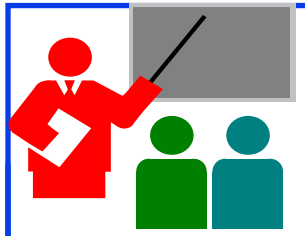
Virtual Bridge Port Extension (VBE)

- IEEE 802.1BR-2012 standard for fabric extender functions
- Specifies how to form an extended bridge consisting of a controlling bridge and Bridge Port Extenders
- Extenders can be cascaded.
- Some extenders may be in a vSwitch in a server hypervisor.
- The controlling bridge relays all traffic
  ⇒ The extended bridge acts as a bridge.



Student Questions

©2025 Raj Jain



Summary

1. Network virtualization includes the virtualization of NICs, Bridges, Routers, and L2 networks.
2. Virtual Edge Bridge (VEB) vSwitches switch internally while Virtual Ethernet Port Aggregator (VEPA) vSwitches switch externally.
3. SR-IOV technology allows multiple virtual NICs via PCI and avoids the need for internal vSwitch.
4. VSS allows multiple switches to appear as one logical switch vPortChannels allow links to multiple switches to appear as one.
5. Fabric Extension and Virtual Bridge Extension (VBE) allow creating switches with a large number of ports using port extenders (which may be vSwitches)

Student Questions

©2025 Raj Jain

Reading List

- G. Santana, "Datacenter Virtualization Fundamentals," Cisco Press, 2014, ISBN: 1587143240 (Safari Book)
- H. Shah, "Management Standards for Edge Virtual Bridging (EVB) and Network Port Profiles," Nov 2010, http://www.ieee802.org/1/files/public/docs2011/bg-shah-dmtf-evbportprofile-overview-0311.pdf

Student Questions

©2025 Raj Jain

References

- Intel, "PCI-SIG SR-IOV Primer," Jan 2011, https://www.intel.com/content/dam/doc/application-note/pci-sig-sr-iov-primer-sr-iov-technology-paper.pdf
- P. Beck, et al., "IBM and Cisco: Together for a World Class Data Center," IBM Red Book, 2013, 654 pp., ISBN: 0-7384-3842-1, http://www.redbooks.ibm.com/redbooks/pdfs/sg248105.pdf
- R. Emerick, "PCI Express IO Virtualization Overview," SNIA Education, 2012, https://www.snia.org/sites/default/education/tutorials/2012/fall/networking/RonEmerick_PCI%20Express_%20IO_Virtualization_Overview-r2.pdf (Excellent)
- R. Sharma et al., "VSI Discovery and Configuration," Jan 2010, http://www.ieee802.org/1/files/public/docs2010/bg-sharma-evb-VSI-discovery-0110-v01.pdf

Student Questions

©2025 Raj Jain

Wikipedia Links

- http://en.wikipedia.org/wiki/Address_Resolution_Protocol
- http://en.wikipedia.org/wiki/EtherChannel
- http://en.wikipedia.org/wiki/IEEE_802.1aq
- http://en.wikipedia.org/wiki/Link_aggregation
- http://en.wikipedia.org/wiki/MC-LAG
- http://en.wikipedia.org/wiki/Network_virtualization
- http://en.wikipedia.org/wiki/PCI_Express
- http://en.wikipedia.org/wiki/Port_Aggregation_Protocol
- http://en.wikipedia.org/wiki/Root_complex
- http://en.wikipedia.org/wiki/Virtual_Routing_and_Forwarding

Student Questions

©2025 Raj Jain

Acronyms

- A-VPLS Advanced Virtual Private LAN Service
- Access-EPL Access Ethernet Private Line
- Access-EVPL Access Ethernet Virtual Private Line
- ADC Application Delivery Controllers
- API Application Programming Interface
- CDCP S-Channel Discovery and Configuration Protocol
- CPU Central Processing Unit
- DMTF Distributed Management Task Force
- DVS Distributed Virtual Switching
- ECP Edge Control Protocol
- EPL Ethernet Private Line
- EVB Edge Virtual Bridging
- EVP-Tree Ethernet Virtual Private Tree
- EVPL Ethernet Virtual Private Line
- EVPLAN Ethernet Virtual Private Local Area Network
- EVPN Ethernet Virtual Private Network
- FEX Fabric Extender

Student Questions

Acronyms (Cont)

GB	Giga Byte
GMPLS	Generalized Multi-Protocol Label Switching
GRE	Generic Routing Encapsulation
H-VPLS	Hierarchical Virtual Private LAN Service
HSRP	Hot Standby Router Protocol
IBM	International Business Machines
IO	Input/Output
IOV	Input/Output Virtualization
IP	Internet Protocol
IPoMPLSoE	IP over MPLS over Ethernet
IPSec	Internet Protocol Security
L2TPv3	Layer 2 Tunneling Protocol Version 3
LAG	Link Aggregation
LISP	Locator ID Split Protocol
MAC	Media Access Control

Student Questions

©2025 Raj Jain

Acronyms (Cont)

- MPLS-TP Multiprotocol Label Switching Transport
- MPLS Multi-Protocol Label Switching
- MR-IOV Multi-Root I/O Virtualization
- NIC Network Interface Card
- NVGRE Network Virtualization using GRE
- NVO3 Network Virtualization Over L3
- OTV Overlay Transport Virtualization
- PB Provider Bridge
- PBB-EVPN Provider Backbone Bridging with Ethernet VPN
- PBB-TE Provider Backbone Bridge with Traffic Engineering
- PBB Provider Backbone Bridge
- PCI-SIG Peripheral Component Interconnect Special Interest Group
- PCI Peripheral Component Interconnect
- PCIe Peripheral Component Interconnect Express
- PF Physical Function

Student Questions

©2025 Raj Jain

Acronyms (Cont)

- pM Physical Machine
- pNIC Physical Network Interface Card
- pSwitch Physical Switch
- PW Pseudo Wire
- PWoGRE Pseudo Wire Over Generic Routing Encapsulation
- PWoMPLS Pseudo Wire over Multi-Protocol Label Switching
- SMLT Split Multi-link Trunking
- SNIA Storage Networking Industry Association
- SR-IOV Single Root I/O Virtualization
- STP Spanning Tree Protocol
- STT Stateless Transport Tunneling
- TP Transport Profile
- T-MPLS Transport Multiprotocol Label Switching
- TRILL Transparent Interconnection of Lots of Link
- VBE Virtual Bridge Extension

Student Questions

©2025 Raj Jain

Acronyms (Cont)

- VDC Virtual Device Context
- VEB Virtual Edge Bridge
- VEM Virtual Ethernet Module
- VEPA Virtual Ethernet Port Aggregator
- VF Virtual Function
- VIP Virtual IP
- VLAN Virtual Local Area Network
- VM Virtual Machine
- vNIC Virtual Network Interface Card
- vPC Virtual PathChannel
- VPLS Virtual Private LAN Service
- VPN Virtual Private Network
- vPort Virtual Port
- VRF Virtual Routing and Forwarding

Student Questions

©2025 Raj Jain

Acronyms (Cont)

- VRRP Virtual Routing Redundancy Protocol
- VSI Virtual Station Interface
- VSL Virtual Switch Link
- VSS Virtual Switch System
- VXLAN Virtual eXtensible Local Area Network

Student Questions

Scan This to Download These Slides



Raj Jain
http://rajjain.com

http://www.cse.wustl.edu/~jain/cse5700-25/m_07avb.htm

Student Questions

©2025 Raj Jain

Related Modules



CSE567M: Computer Systems Analysis (Spring 2013),
https://www.youtube.com/playlist?list=PLjGG94etKypJEKjNAa1n_1X0bWWNyZcof

CSE473S: Introduction to Computer Networks (Fall 2011),



https://www.youtube.com/playlist?list=PLjGG94etKypJWOSPMh8Azcgy5e_10TiDw



Wireless and Mobile Networking (Spring 2016),
https://www.youtube.com/playlist?list=PLjGG94etKypKeb0nzyN9tSs_HCd5c4wXF

CSE571S: Network Security (Fall 2011),



https://www.youtube.com/playlist?list=PLjGG94etKypKvzfVtutHcPFJXumyyg93u



Video Podcasts of Prof. Raj Jain's Lectures,
https://www.youtube.com/user/ProfRajJain/playlists

Student Questions

©2025 Raj Jain