Edge Virtual Bridging with VEB and VEPA
(Summary)

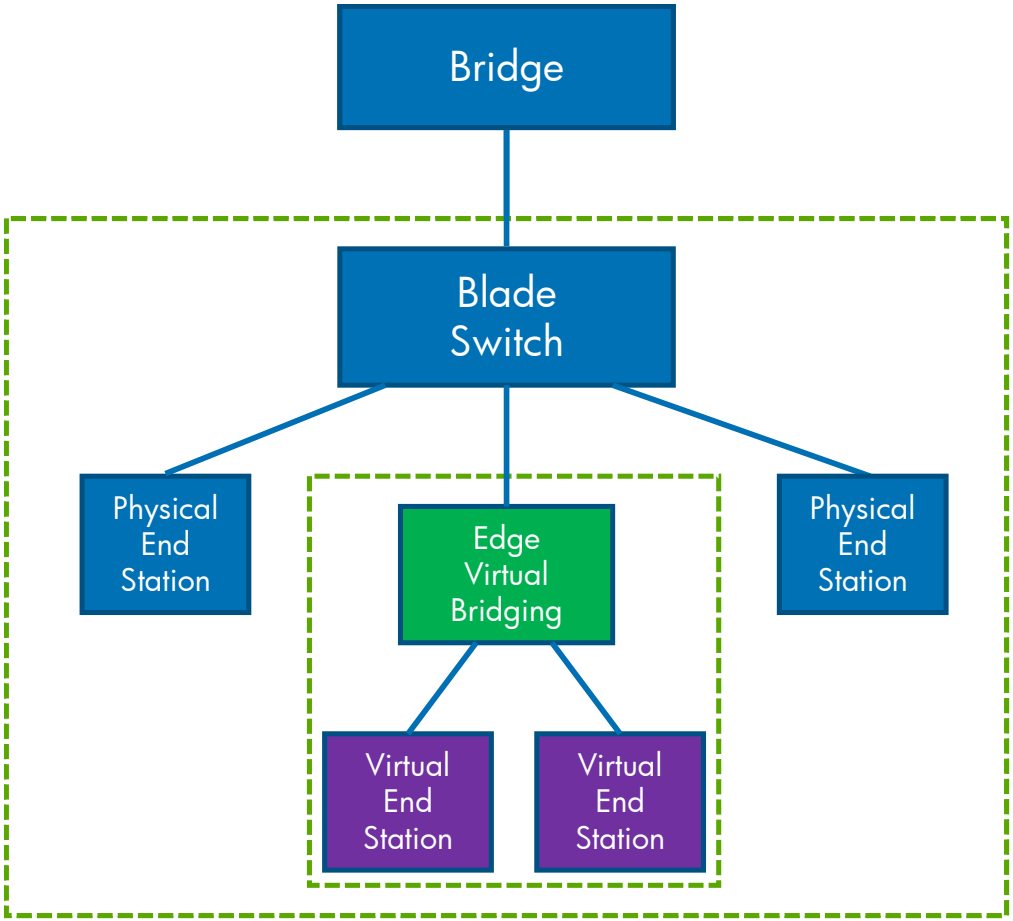
May 2009

Chuck Hudson (HP) **c.hudson@hp.com**
Paul Congdon (HP) **ptcongdon@ucdavis.edu**

Edge Virtual Bridging
An Edge-centric Definition

Edge Virtual Bridging (EVB) is an environment where physical end stations contain multiple virtual end stations that participate in the Ethernet network environment.

Note: EVB environments are unique in that vNIC configuration information is available that is not normally available to an 802.1Q bridge.



Edge Virtual Bridging
focused on…

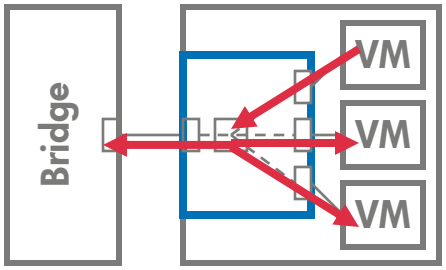
- Virtual Machine Environments (Virtual Switch)
  - VMware ESX Server
  - Microsoft HyperV
  - Citrix XEN
  - Linux KVM (linux-kvm.org)
  - Others

- NICs with multiple vNICs that share a single link
  - PCI Single-root or Multiple-root IO Virtualization (SR-IOV, MR-IOV)
  - Other multi-vNIC technologies

Challenges at the Edge

- Visibility & Control
  - System admins own the physical end stations
  - Lack of network admin control can mean inadequate:
    - Control of network access
    - Visibility of networking traffic
    - Support for debugging network issues
- Limited Embedded Capability
  - NICs have cost & complexity constraints (no TCAMs, no learning)
  - End-stations and bridges evolve independently

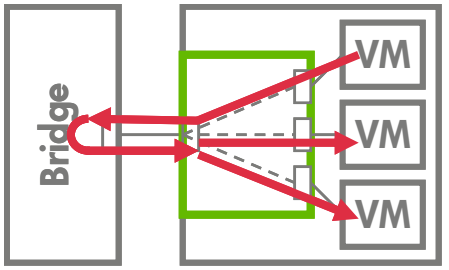
VEB & VEPA



Virtual Ethernet Bridge (VEB)

uses MAC+VID to steer frames

- Emulates 802.1 Bridge
- Loop-free, No STP
- Address Table:
  - No learning required, vNICs register MAC addresses
  - Local packet replication using address table
- Configured by hypervisor
- Requires settings for vPorts



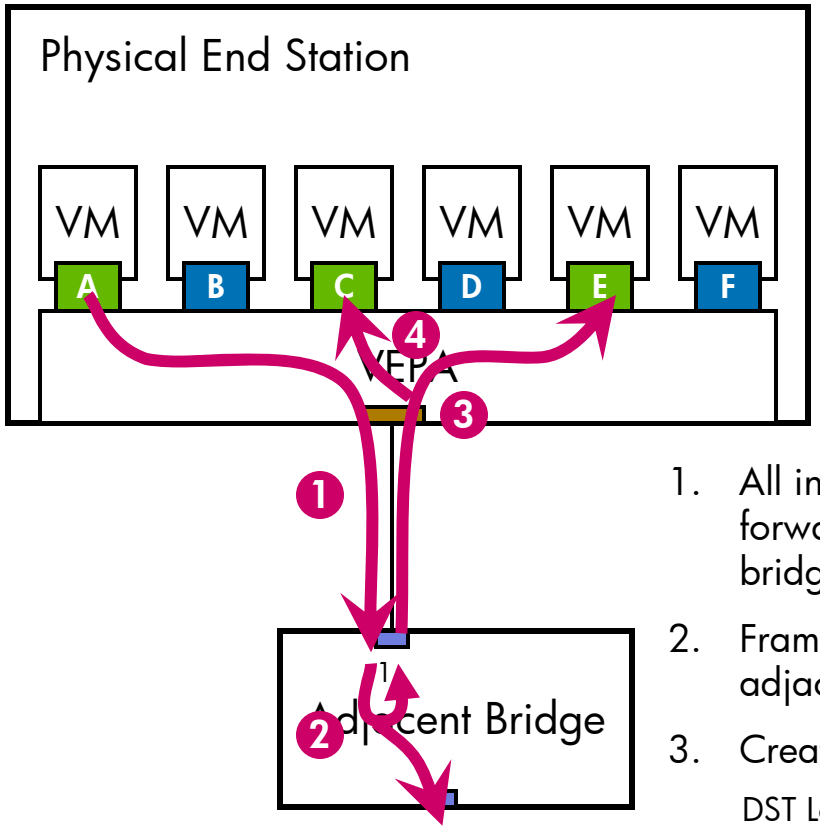
Tag-less VEPA

uses MAC+VID to steer frames

- Steers frames via adjacent bridge
- Loop-free, No STP
- Address Table:
  - No learning required, vNICs register MAC addresses
  - Local packet replication using address table
- Configured by hypervisor
- Requires the same settings for vPorts

Basic VEPA Operation
Multicast

SRC = A; DST = MulticastC



1. All ingress frames forwarded to adjacent bridge
2. Frame forwarded by adjacent bridge.
3. Create delivery mask

   DST Lookup = 101010
   SRC Lookup = 100000
   Delivery Mask = 001010

4. Deliver Frame Copies

VEPA Address Table

DST MAC	VLAN	Copy To (ABCDEF)
A	1	100000
B	2	010000
C	1	001000
D	2	000100
E	1	000010
F	2	000001
Bcast	1	101010
Bcast	2	010101
MulticastC	1	101010
Unk Mcast	1	100010
Unk Mcast	2	010101
Unk Ucast	1	000000
Unk Ucast	2	000000

Promiscuous Ports at the edge…

- Promiscuous ports are not common at the edge in a virtualization environment… However,
  - Simultaneous operation of a VEB and VEPA provides both performance and flexibility
  - A small number of inline virtual appliances may be useful
  - The VN-Tag alternative believes it is necessary
- Approaches for handling promiscuous ports
  - Use a VEB
  - Use hypervisor security APIs instead
  - Have the VEPA learn (not really practical)
  - Use VLANs to isolate promiscuous ports

Benefits of VEB/VEPA Solution

- Simple extension to VEB
  - Similar port configuration
  - Similar address table
  - Minor changes to frame forwarding behavior
- Solves nearly all of the issues with VEBs
- Allows easy migration between VEB and VEPA modes
  - Could allow simultaneous operation of VEB and VEPA
- Requires minimal 802.1 standards effort
  - Configuration of hair-pin mode
- Easiest to implement
  - Can be implemented in many existing switches with a firmware update
  - Simple extension to existing vSwitches/VEBs

Issues with VN-Tag Approach

- Using multiple layers of VN-Tag network concentrators…
  - Significantly limits the network cross-sectional bandwidth
  - Increases congestion
  - Often increases the number of links traversed
- Constrains innovations in distributed computing
  - Blocks advantages of locality in distributed systems
    - Distributed storage solutions, nearby caching servers, etc.
  - Blocks benefits of increased end-station capabilities over time
- VN-Tags increases hardware complexity to end stations
  - Significantly different than already-required VEB
  - New forwarding and frame replication mechanisms
- VN-Tags require significant new standards efforts
  - New tag format
  - Management of remote frame replication
- VN-Tags will not work with any switch not specifically designed for it
  - Adds significant cost and complexity to controlling bridge
  - Constrains other bridges to be remote line cards for controlling bridge

Proposed

- IEEE 802.1 standardization of
  - Switch port operation when in 'hairpin' mode
  - Configuration of 'hairpin' mode
    - LLDP/DCBX capabilities exchange & configuration
    - Managed object definition
    - Possible configuration of hairpin for specific VIDs
- Industry Standardization of EVB management
  - Coordinated configuration of vPort settings for both VEB/VEPA
  - Mechanism & standards forum is still TBD
  - Probably not 802.1