VMready
Implementation of a Profile Based Virtual Switching Framework

Vijoy Pandey and Jay Kidambi

DC-CAVES 2010
The evolving switching primitive
The evolving switching primitive
The evolving switching primitive
The evolving switching primitive
The evolving switching primitive
The evolving switching primitive
Physical interfaces are the old switching unit
- Not aware of server-side virtualization technologies
- Configuration per physical interface only
  - No flexibility for unique VM needs
- No flexibility to handle VM migrations

Virtual interfaces are the new switching unit
- Fully aware of server-side virtualization technologies
- Configuration per virtual interface
- Automated Network mobility
  - Configuration follows Virtual Machines in real-time during migration

VMready: Virtualization Aware Networking
Big Picture: What is needed

Virtual Interface Identification

Virtual Interface Data Path Switching

Virtual Interface Tracking & Policy Enforcement

External Edge Switch

Network Cloud (L2 or L3)

Policy dB

Hypervisor

IOV CNA

VM Migration
 vmready uses MAC based identification of VMs today
  — MAC assignment pools for the different vendors
 vmready uses Q-in-Q for NIC-partition (vNIC) identification
A: Secure Identification

- MAC based identification of VMs is potentially **insecure**
  - Prone to spoofing attacks
- DCBX-based vNIC identification is **secure**
  - Agent running on both ends of the wire guarantees spoof-free operation
A: Secure Identification for VMs

- Closed loop secure VM verification
  - Switch verifies VM & Port attach from hypervisor and vCenter
- VM spoof protection across data center
  - VM cannot be spoofed at any hypervisor in DC
  - Visibility into which VM is connected to which BLADE Switch & Port
Number of entries: 3
* indicates VMware ESX Service Console Interface
+ indicates VMware ESX/ESXi VMKernel or Management Interface
0.0.0.0 indicates IP address not yet available

>> Virtual Machine# /cfg/virt/vmg
Enter group number: (1-32) 1

[VM group 1 Menu]
  Vlan  - Set the group's vlan (only for groups with no VM profile)
  vmap  - Set VMAP for this group
  tag   - Enable vlan tagging on all VM group ports
  addvm - Add a virtual entity to the group
  remvm - Remove a virtual entity from the group
  validate - Sets secure mode for all VMs in this group
  addprof - Add a VM profile to the group
  remprof - Delete any VM profile associated with the group
  addport - Add ports to the group
  remport - Remove ports from the group
  addtrunk - Add trunk to the group
  remtrunk - Remove trunk from the group
  addkey - Add LACP trunk to the group
  remkey - Remove LACP trunk from the group
  stg   - Assign VM group vlan to a Spanning Tree Group
  del   - Delete group
  cur   - Display current group configuration

>> VM group 1# addvm 00:50:585:8b:48:a8

>> VM group 1# apply
.

Apply complete; don't forget to "save" updated configuration.

>> VM group 1# 2000-08-09 15:06:03:Unauthenticated MAC [00:58:56:8b:48:a8] from port 1 switchId [172.31.1]
Disabling port 1 because verification failed on the port

Aug  9 6:58:41 172.31.213.1 NOTICE server: link down on port INT1

Aug  9 7:00:22 172.31.213.1 NOTICE vm: Virtual Machine with IP address 172.31.213.2 came online

Aug  9 7:02:37 172.31.213.1 NOTICE mgmt: Port MGT1 ENABLED and MGT2 DISABLED because Management Module
vCenter Visibility of BLADE Switch & Ports

172.31.46.6 VMware ESX, 4.0.0, 164009

Hardware
- Processors
- Memory
- Storage
  - Networking
    - Storage Adapters
    - Network Adapters
    - Advanced Adapters
- Advanced Settings

Software
- Licensed Features
- Time Configuration
- DNS and Routing
- Power Management
- Virtual Machine Startup/Shutdown
- Virtual Machine Swapfile Location
- Security Profile
- System Resource Allocation
- Advanced Settings

View: Virtual Switch, Distributed Virtual Switch

Networking

Virtual Switch: vSwitch0
- VMkernel
  vmk0: 172.31.46.7
- Virtual Machine Port Group
  - vM Network
  - 3 virtual machine(s)
    vm2
    vMA
    ubuntu
- Service Console Port
  - Service Console
    vsww1: 172.31.46.6
- Virtual Machine Port Group
  - BNT_Test
    VLAN ID: 2
- Virtual Machine Port Group
  - BNT_Default
    1 virtual machine(s)
    vm1

Properties
- Version: 1
- Timeout: 0
- Time to live: 83
- Samples: 723929
- Device ID: BLADE
- Address: 172.31.213.1
- Port ID: INT2
- Software Version: Unreleased: FW_VIEW:
- Hardware Platform: BNT 1/10Gb Uplink
- IP Prefix: 0.0.0.0
- IP Prefix Length: 0
- VLAN: 0
- Full Duplex: false
- MTU: 0
- System Name
- System Oid
- Management Address: 0.0.0.0
- Location
Virtual Ethernet Bridge (VEB) mode of operation today
- Relies on a soft switch in the hypervisor for intra-server data path
- External edge switch provides inter-server data path
- Configures soft switch to provide consistency with the Edge Switch
B: Data Path Switching (vNICs)

- Isolation is key property
  - DCBX to exchange vNIC parameters, S-Tags for ID
  - Two models: switch mode and I/O extender mode of operation
  - S-Tags either
    - carried to Network Cloud (Host I/O extender mode) or
    - stripped on ingress (Switch mode)
C: Stack Switch Tracking & Policy Enforcement

- Policy-based configuration on Edge Switch
  - Profiles created on switch or distributed switch
  - Virtual Interfaces (live or pending) assigned to Port Group
  - Profile attached to Port Group
- Migrate events cause Migrated VM ID to be attached to the same (BLUE) Port Group
  - Post-verification with VM Manager
- Drawback
  - Post-creation / post-move association of profile
  - Window small in the normal case since RARP sent before VM traffic
C : DC-wide Tracking & Policy Enforcement

- Policy-based configuration on Policy DB Manager
  - Profiles created on BHM
  - On creation, switch queries BHM for Profile
  - On Migrate, destination switch queries BHM for Profile
  - BHM performs post-verification with VM Manager
- Drawback
  - Post-move / creation association of profile
  - Window small in the normal case since RARP sent before VM traffic
Futures : ABC of Qbg

VM Migration

A. Virtual Interface Identification
B. Virtual Interface Data Path Switching
C. Virtual Interface Tracking & Policy Enforcement

Network Cloud (L2 or L3)

External Edge Switch

Hypervisor

IOV CNA

VM1
VM2
VM3

BLADE Network Technologies
VMready Changes for Qbg & Value Proposition

- **VEPA Mode**
  - All VM Traffic handled by external switch for larger context
  - Hardware-based Policy enforcement & traffic switching

- **VDP**
  - Pre-move or Pre-creation policy enforcement

---

**Virtualization & Migration**

**VM Migration**

**External Edge Switch**

**Hypervisor**

**IOV CNA**

**VM1**

**VM2**

**VM3**

---

**Network Cloud** (L2 or L3)

**Policy dB**

**BLADE Harmony Manager**
Network Manager & Profile Database

**Integration for VSI Types**

**VM Manager**
Firmware Architecture

- Hypervisor-vendor-specific Plugins
  - AAA
  - Configuration
  - CLI
  - XML based
  - SNMP v1/v2/v3
  - BBI
- EVB Control Stack
- Layer 2 Protocols & Layer 2 High Availability
- Layer 3 Protocols (IPv4, IPv6) & Layer 3 High Availability
- Protocol Application
- Virtual Port Switching Abstraction Layer
- v-port Identification Plugins
  - FC | FCoE v-port
  - Host I/O Extension
  - EVB v-port
  - IOV vNIC v-port
- Switching Silicon Abstraction
- OS and Hardware Abstraction Layer
  - Linux
  - Switch Silicon Device Drivers
  - Peripheral Drivers
- Hardware