NFVI = Openstack => OVS:
quelles performances?

L3 est-il oublié?

Vincent JARDIN - CTO
ETSI NFV Model: Openstack the defacto VIM
ETSI NFV Model
Neutron: APIs and Python agents to configure hypervisor’s networking
Requirements of Virtual Infrastructure
OpenStack: analysis of a Linux KVM NFVI case

1. Driver Level - pNICs
2. L2 Virtual Switch or L3 Virtual Routing
3. Virtual NICs (vNIC - virtio) - Host to Guest
4. Virtual Machine (KVM/Qemu)
5. VM-to-VM Communication (ETSI Phase2)
NFVI survival kit: learn OVS debugging

- ovs-appctl
- ovs-vsctl
- ovs-dpctl
- ovs-ofctl

+ what if OpenDayLight is being used?
Neutron: Linux’s consumer of pNICs to vNICs technologies

pNICs = 10G/40G PCI Ethernet ports  
vNICs = virtio

Neutron leverages:

- L2
  - OVS
  - Linux Bridge

- L3
  - netns / Multi Tenants
  - iptables
    - Firewall -> Security Groups
    - NAT -> Openstack’s floating IP addresses
  - L3 routes

- VPN
  - Kernel IPsec (XFRM)

- Overlay / tunneling
  - OVS vPort or Netdevices
  - VXLAN, GRE
  - (GENEVE, MPLS)
Test 1: Linux Open vSwitch and Linux VM

14 Gbps

Limited Bandwidth To Linux Based Virtual Machines

IP Forwarding

Linux Based Virtual Machine

Open vSwitch

Hypervisor

X86 server with 12x10G Ports
6WIND Virtual Accelerator in OpenStack Compute Node

- 240Gbps 6WIND Virtual Accelerator throughput on 12 cores of Xeon E5-2697 v2 @ 2.70GHz

- 1 core provides a 20Gbps Virtual Accelerator bandwidth

- Examples on a dual socket / 24 cores server
  - 120Gbps North-South traffic delivered to standard VMs or VNFs with 12 cores remaining for VMs
  - 40Gbps North-South traffic with 20 cores remaining for VMs
  - 40Gbps North-South traffic and 160Gbps East-West traffic for service chaining
Compute Nodes Neutron Diagram
Test 2:
6WIND Virtual Accelerator + Linux VM

8X Bandwidth Increase

14 Gbps

118 Gbps

Linux Based Virtual Machine

IP Forwarding

Hypervisor

Open vSwitch

Virtual Accelerator

X86 server with 12x10G Ports
L2 : OVS interconnect

- **6WINDGate OVS L2 switching performance**
  - 6.8 Mfps per core
  - Up to 67.8 Mfps using 10 cores (20 threads)

- Performance scales linearly with the number of cores configured to run the fast path.

- Performance is independent of frame size.

- **This benchmark is simplified, other overhead have to be included:**
  - Host to VM (virtio)
  - VM to host (virtio)
  - OVS over VXLAN
**L3 processing**

- **6WINDGate IP forwarding performance**
  - 9.57 Mpps per core
  - Up to 226.83 Mpps with 22 cores

- Performance scales linearly with the number of cores configured to run the fast path.

- Performance is independent of frame size.
L2 or L3 interconnect?

L2

- All VMs share the same broadcast L2 overlay network
- VM’s VLAN over OpenStack’s VXLAN
- How to distribute MAC addresses, how to manage L2 overlayed topologies?
  - Proxy ARPs
  - ISIS?
  - EVPN (BGP extension)?
  - PVSTP?
- Live migration => MAC tables of switches to refresh (GARP is not reliable)

L3

- L3 interconnects: all IP traffic from VMs are routed through its VRF such as VPNs (RFC2547)
- No broadcast domain
- But L2 interconnect can still be supported:
  - Each tenant can configure its own L2 over IP (VXLAN, NVGRE, MPLS over GRE, etc.)
- VRF aware L3 routing is mature
- Live migration => route updates
OpenStack Neutron: L2 today, L3 tomorrow?

Benefits of Neutron without L2

- No OVS, no Linux Bridge
  - less networking bugs and less provisioning bugs

- no VLAN

- Less processing layers => Less CPU cycles per packets => Increase packet rates

- Simplify the design of PCI Ethernet boards (HW offloads)

Why L2 is leading, but not L3 with OpenStack?

- No L3 alternative to OVS
  - Quagga: EVPN not available
  - LISP: No proof points with Neutron

- Like OVS, any solutions have to be
  - 1st – democratized into Linux kernel, Linux networking community
  - 2nd – integrated use with OpenStack / Neutron
  - 3rd – described for their usage through standard bodies (ETSI, IETF, ONF, etc.)
Conclusion:
ETSI NFV does not assume L2 or L3 interconnect
SDN for VM2VM can be built based on L3 over L3 (aka VPNs)
L3 shall restore performance + simplicity