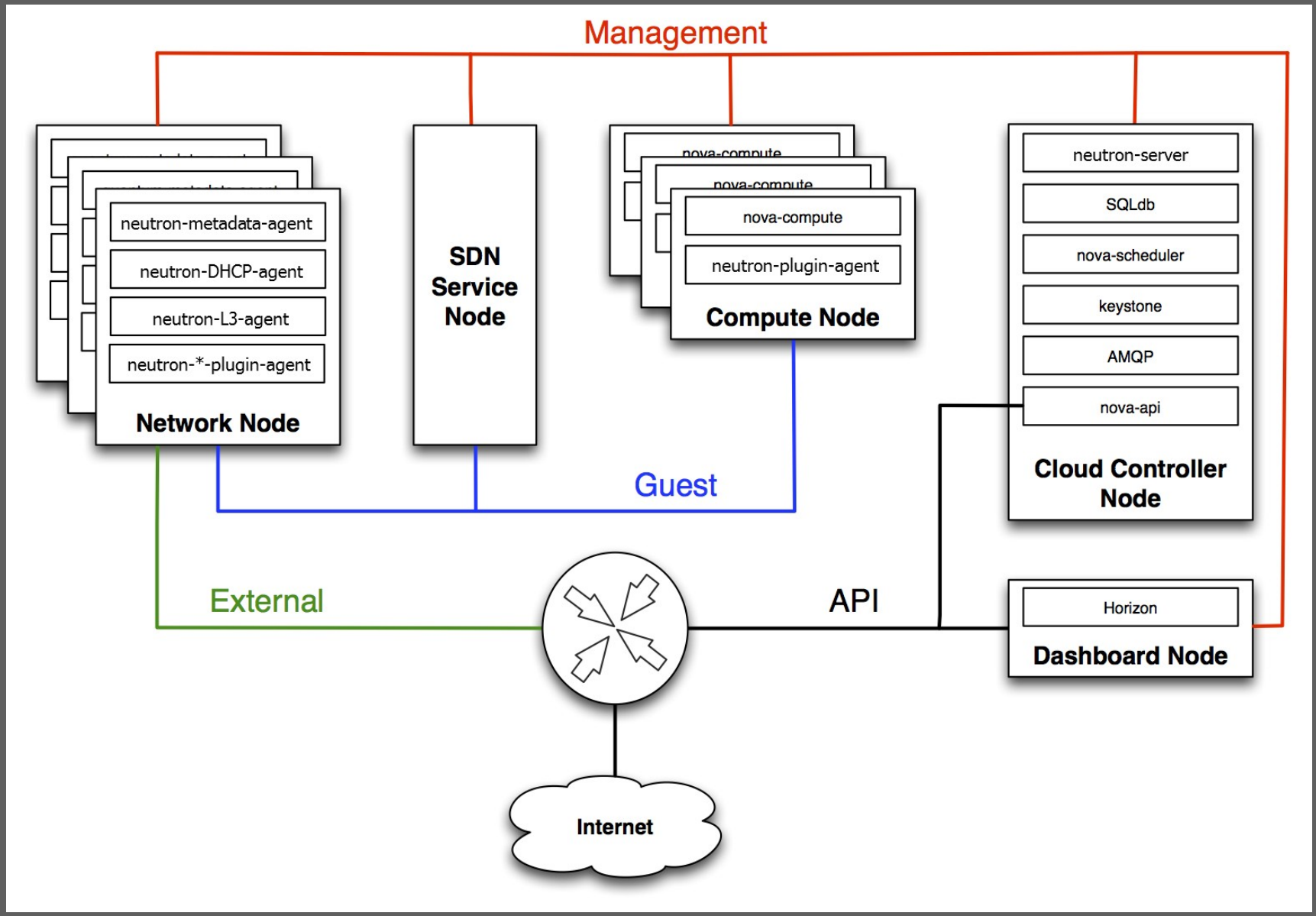
Neutron Performance Deep Dive

蒋趁心

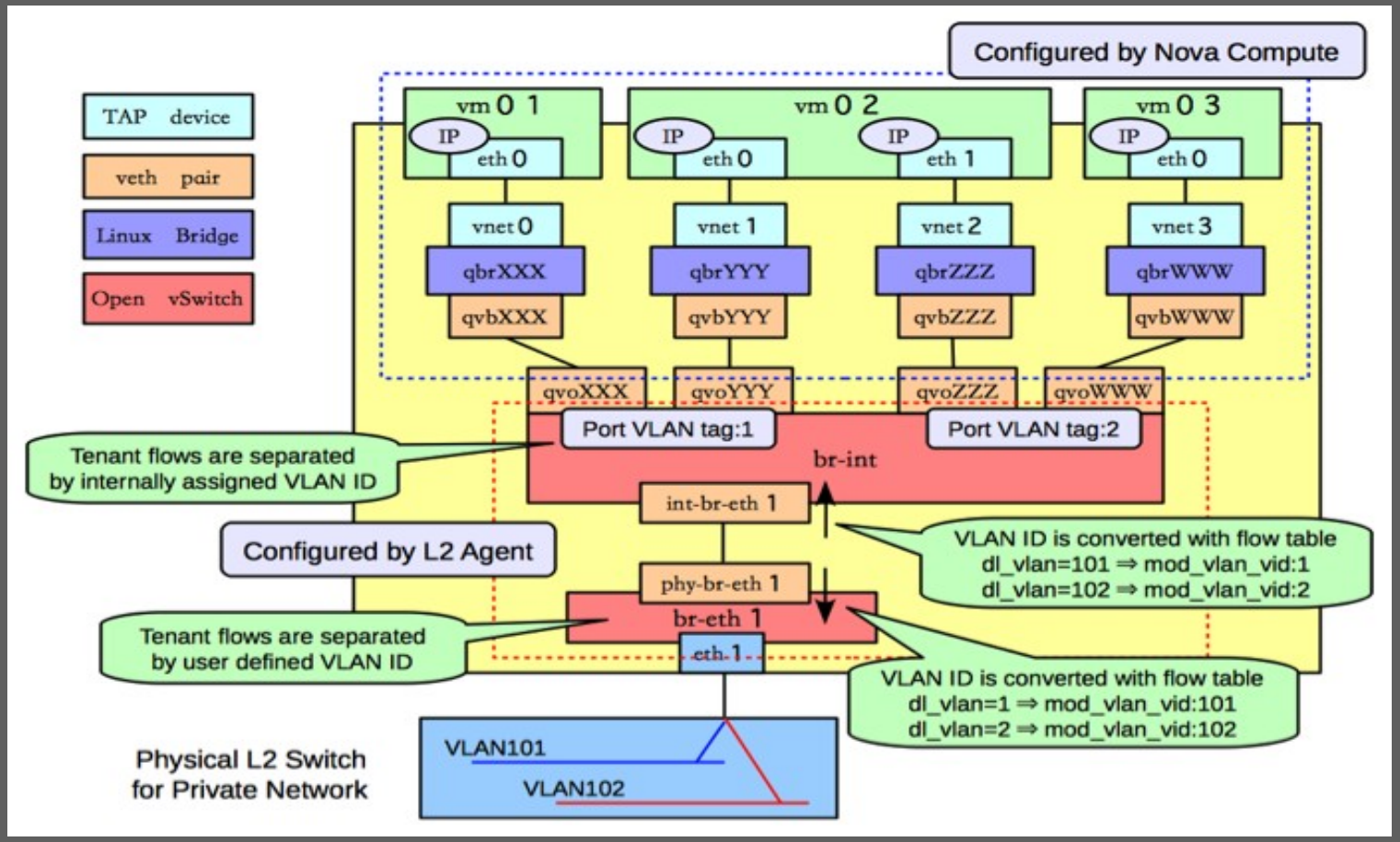
What is Neutron?

The Networking service, code-named neutron, provides an API that lets you define network connectivity and addressing in the cloud. The Networking service enables operators to leverage different networking technologies to power their cloud networking. The Networking service also provides an API to configure and manage a variety of network services ranging from L3 forwarding and NAT to load balancing, edge firewalls, and IPsec VPN.

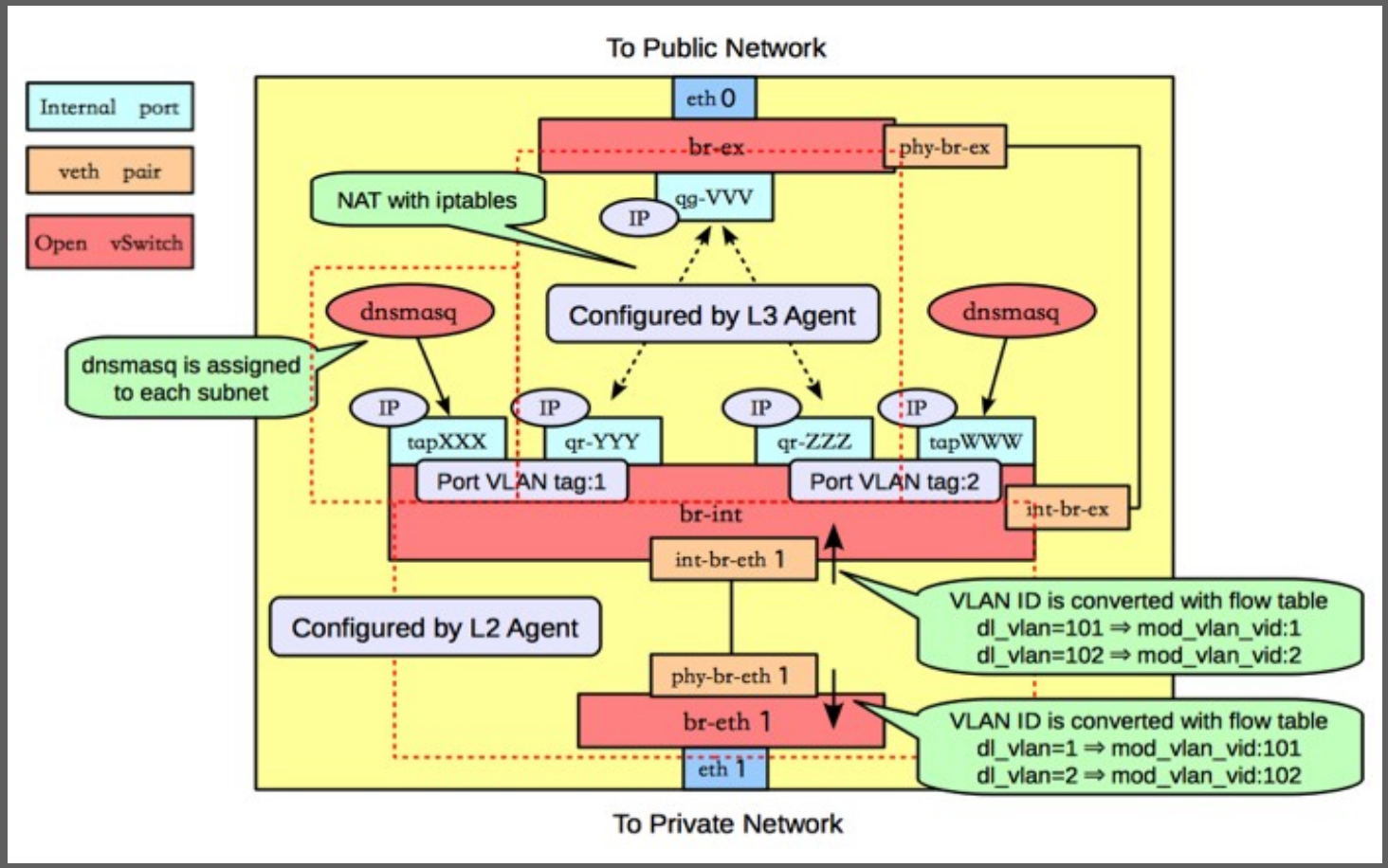
Neutron Architecture



Neutron ML2 OVS Impl



Neutron ML2 OVS Impl



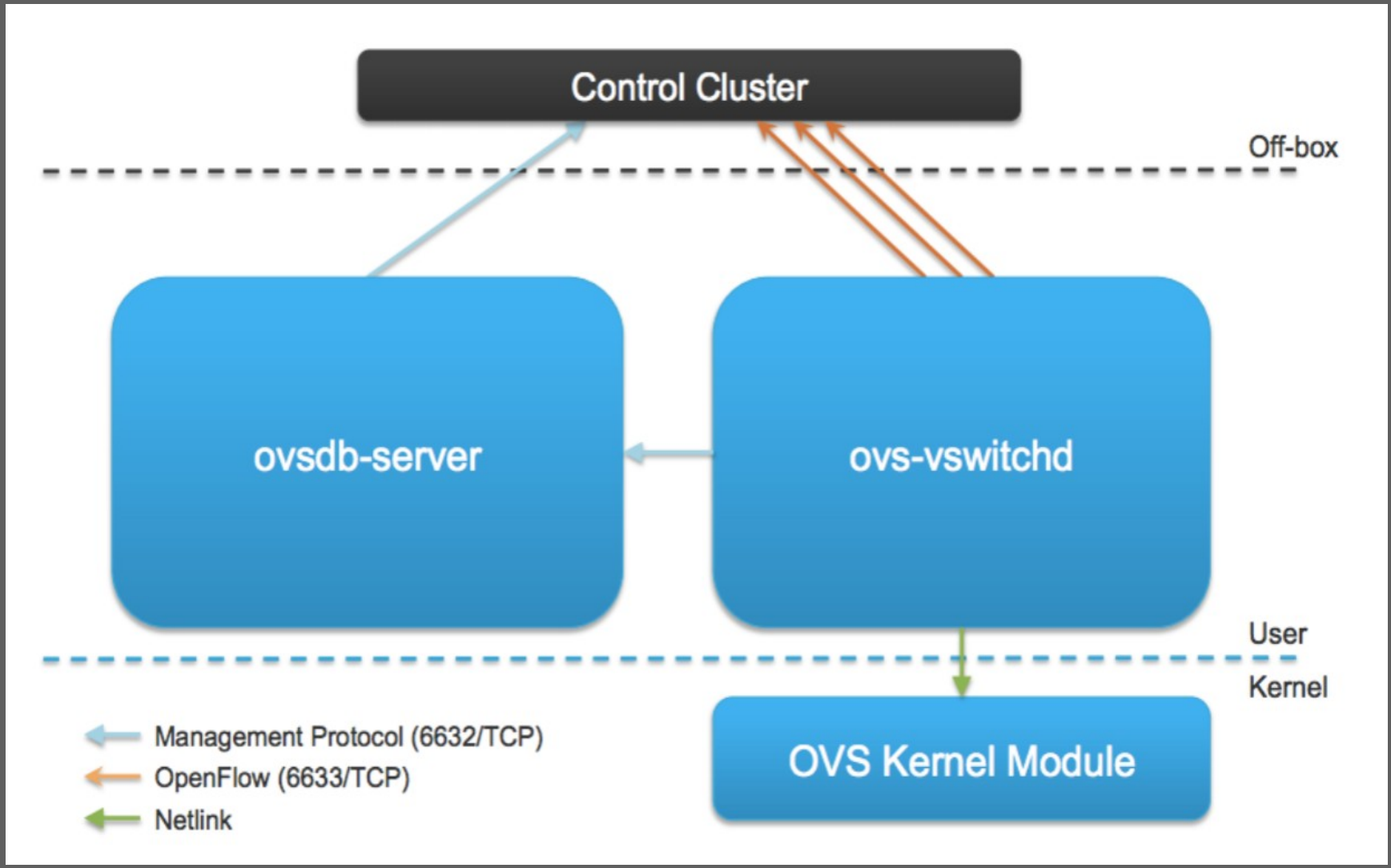
Performance tuning points

- What does performance mean?
  - Bandwidth
  - Small Packets
  - Flow setup rate
  - CPU usage
  - Latency
  - ...

OpenStack tuning points

- Guest OS
- VM NIC driver
- Tap device
- Linux Bridge
- Namespace
- IPTables
- Host OS
- NIC device
- Hardware Switch
- **Open vSwitch**

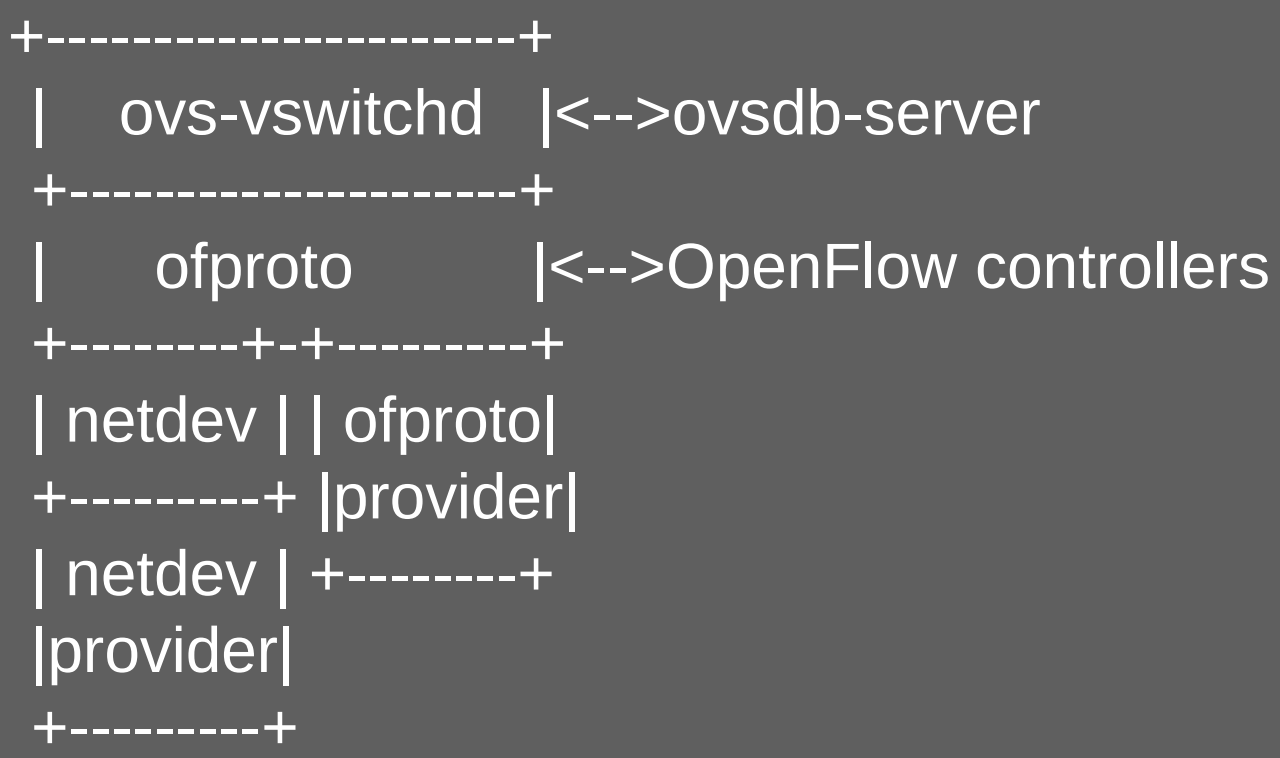
Open vSwitch Architecture



Open vSwitch Utilities

- Ovs-vsctl
  - Configures ovs-vswitchd, high level interface for ovsdb-server
  - ovs-vsctl set port port0 tag=4
- Ovsdb-tool
  - Command line tool for managing database file
  - ovsdb-tool show-log -m
- Ovs-ofctl
  - Speaks to OpenFlow module
  - ovs-ofctl dump-flows br-eth5
- Ovs-dpctl
  - Speaks to kernel module
  - ovs-dpctl show
  - ovs-dpctl dump-flows
-

Open vSwitch Architecture



OVS performance tuning points

- Userspace flow installation
- Flow matching
- Kernel space forwarding & packet header modification

OVS Evolution

- Userspace Multi-threading
- Mega Flows
- …
- Still not ideal

Solution – Fast Packet Process

- **DPDK**
  - DPDK is a set of libraries and drivers for fast packet processing. It was designed to run on any processors knowing Intel x86 has been the first CPU to be supported. Ports for other CPUs like IBM Power 8 are under progress. It runs mostly in Linux userland. A FreeBSD port is now available for a subset of DPDK features.
  - multicore framework
  - huge page memory
  - ring buffers
  - poll-mode drivers
- Netmap
- OpenOnload
- ...

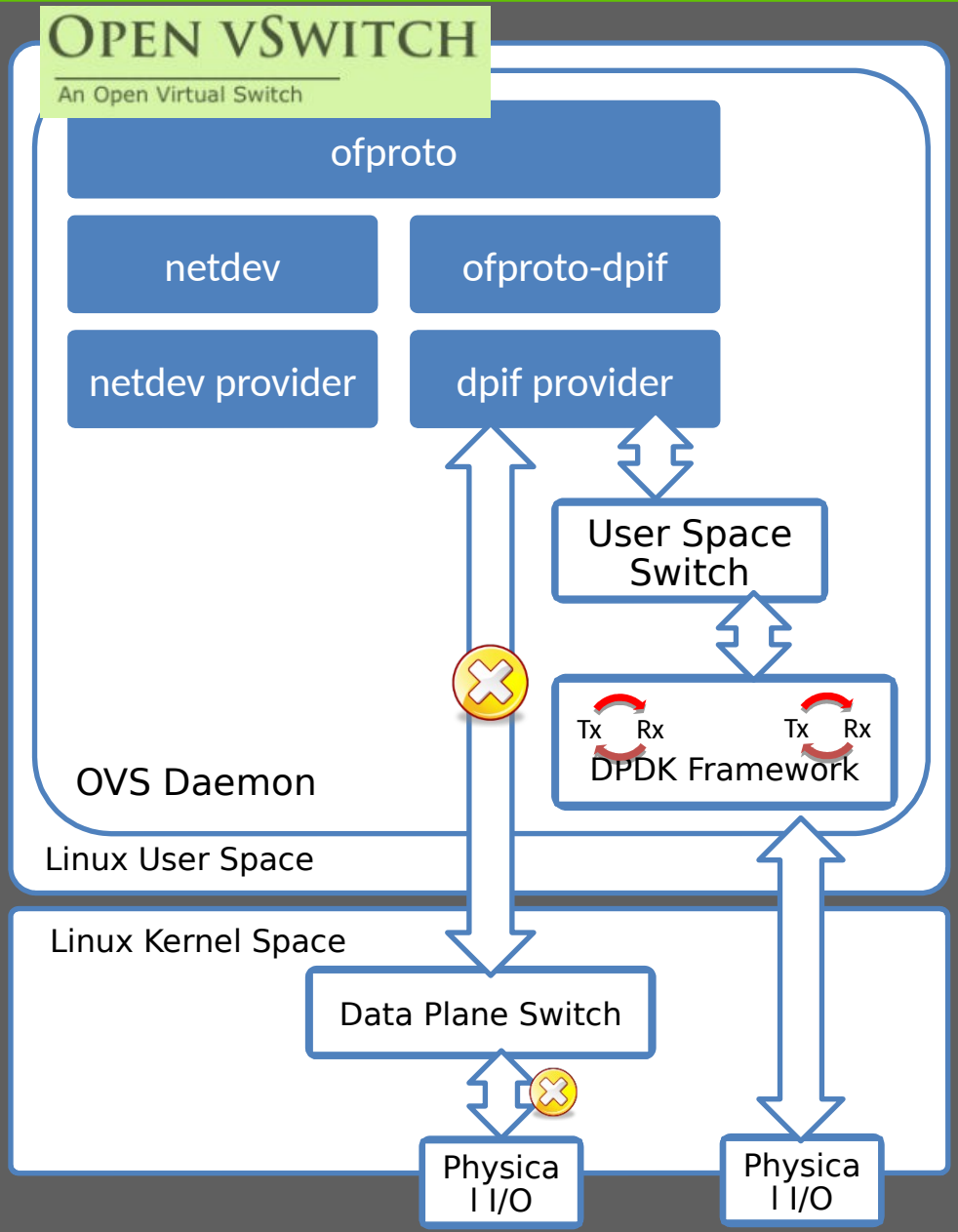
Why is DPDK fast?

- Parallel Processing
  - Multiqueue, MultiCore, 1 Loop percore
- Run to completion model or pipeline model(low overhead)
  - No time for scheduling
  -
- No Scheduler
  - Polling Mode Drivers for NICs
- Efficient Memory Access
  - Huge Pages instead of heap
  - Less pages needed so Less TLB(translation lookaside buffer) misses
  - NUMA allocation
- Buffer Management
  - Pre-allocated buffers reduced significant time
  - Multi-producer/Multi-consumer safe
  - Optimized(cache alignment, per core buffer caches, bulk alloc/free)

Why is DPDK fast?

- Queue Manager
  - Lockess. No spinlocks.
  - Bulk enqueue/dequeue
- Flow Classification
  - Based on Streaming SIMD extension
  - X86 instruction set extension with great performance
-

DPDK-OVS



- Trunked by Intel from OVS, integrated with DPDK
- Dead-ended by Intel in 2014
- DPDK integrated in main stream OVS 2.2 for experiment
- OVS with DPDK Will be released in 2.4
- Several times performance improvement

Solution – NIC offload

- **Intel 82599**
  - **40% performance improvement**
- Netronome FlowNIC

Intel 82599

Port-mapping Table

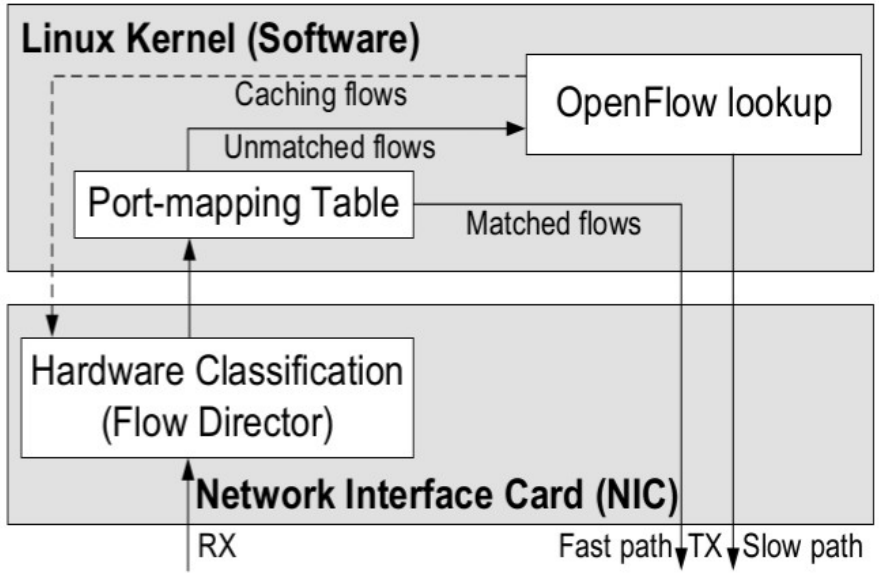
Input Port	Input Queue	Output Port	Output Queue
eth0	1	eth1	1
eth0	2	eth2	1

Map queue to output port

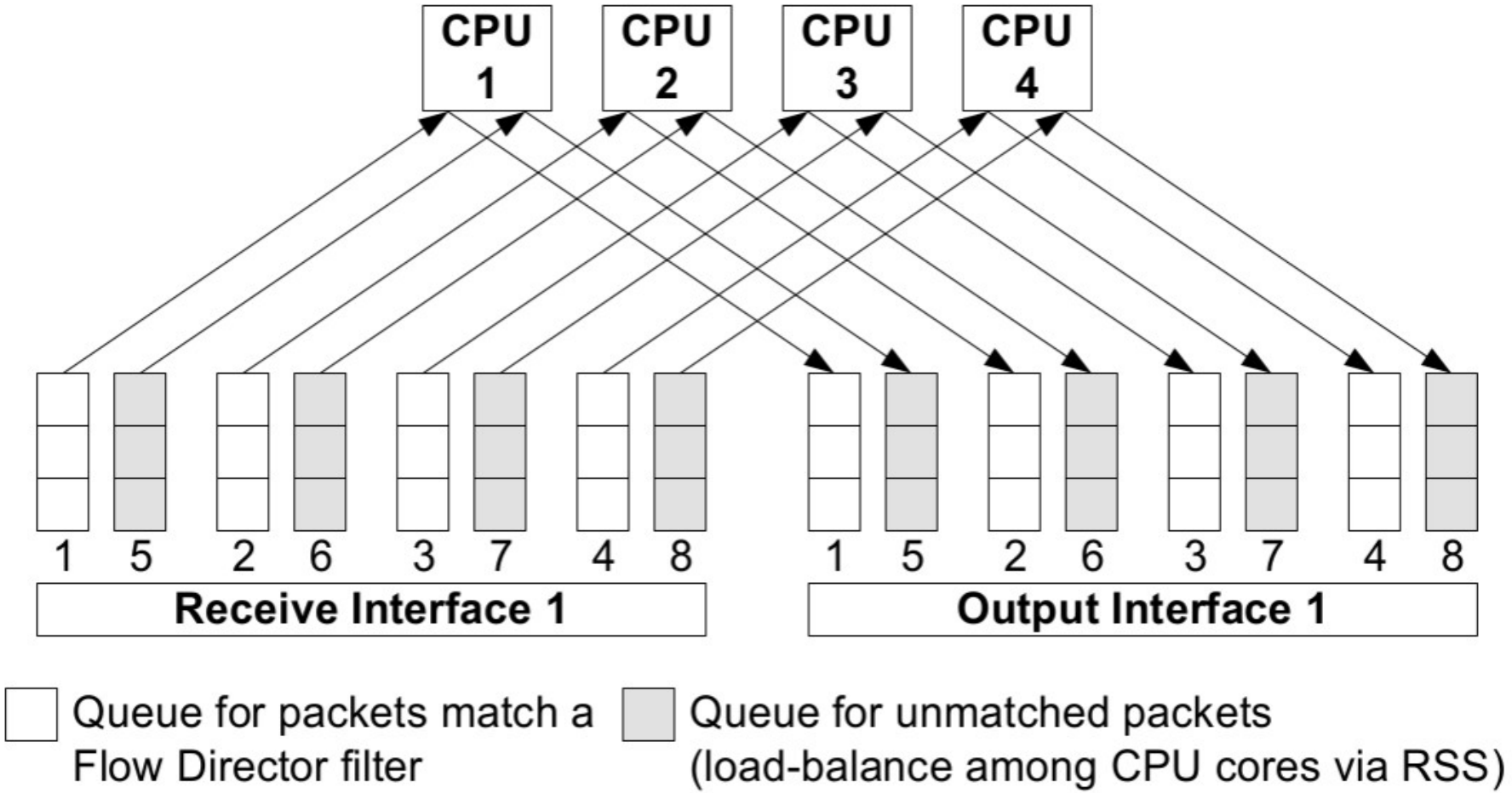
Offload lookup to NIC

Packet header					Queue
	VLAN	L2	IP	TCP	
Flow1 (src 1.1.1.1 dst 2.2.2.2)					1
Flow2 (src 2.2.2.1 dst 3.3.3.3)					2

Classifies packet to queue



Intel 82599


CPU 1
CPU 2
CPU 3
CPU 4
1 5 2 6 3 7 4 8
Receive Interface 1
1 5 2 6 3 7 4 8
Output Interface 1
Queue for packets match a Flow Director filter
Queue for unmatched packets (load-balance among CPU cores via RSS)

Solution – OVS offload

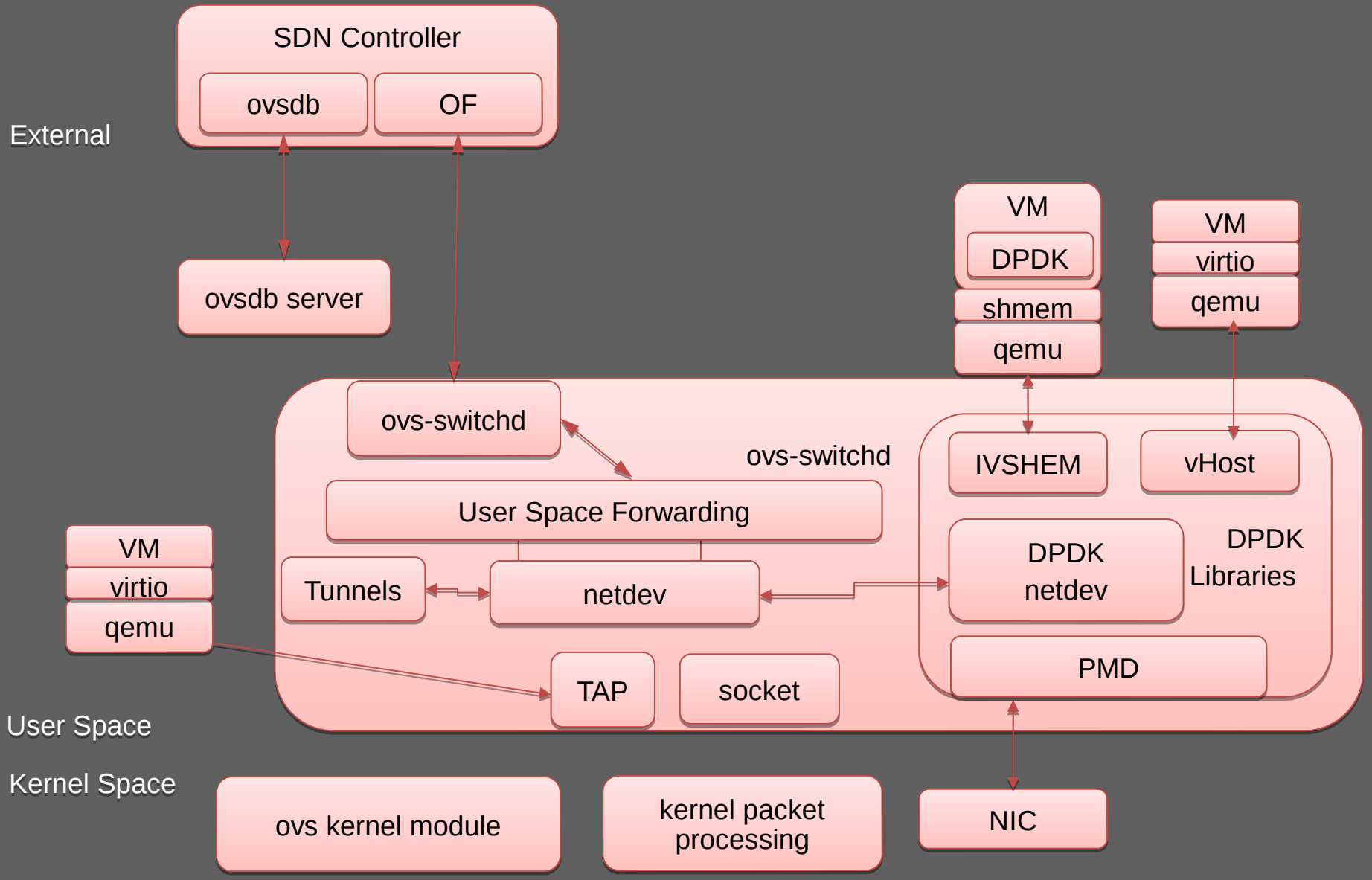
- Cisco Nexus
- Arista
- BigSwitch
- Centec
- ...

We are hiring

- admin@easystack.cn
- http://www.easystack.cn/en/jobs.php
- 欢迎应届毕业生、实习生

Q & A

DPDK


SDN Controller
ovsdb
OF
External
ovsdb server
VM
DPDK
shmem
qemu
VM
virtio
qemu
ovs-switchd
ovs-switchd
User Space Forwarding
IVSHEM
vHost
VM
virtio
qemu
Tunnels
netdev
DPDK
netdev
DPDK
Libraries
PMD
TAP
socket
User Space
Kernel Space
ovs kernel module
kernel packet
processing
NIC