



ons

EUROPE

OPEN NETWORKING //
Integrate, Automate, Accelerate

September 25 - 27, 2018
Amsterdam, The Netherlands

A Brief Guide to Virtual Switching

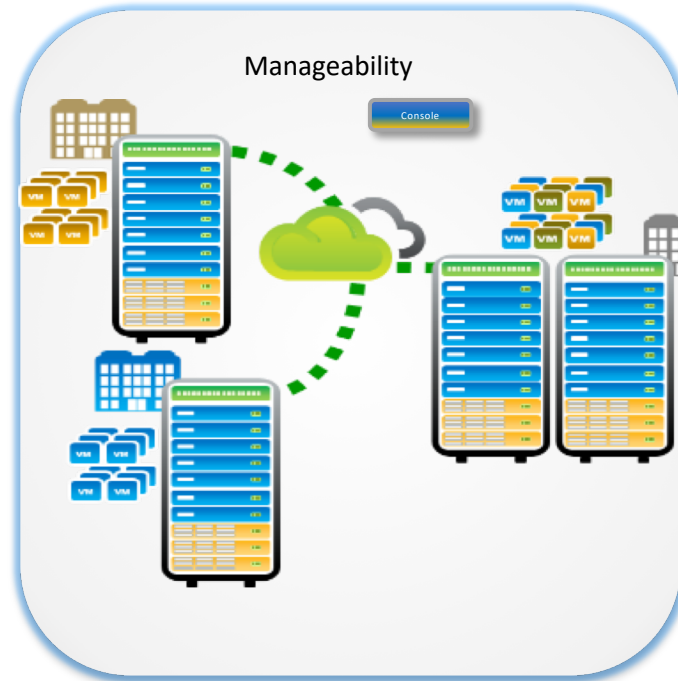
Franck Baudin (Red Hat)
Billy O'Mahony (Intel)



vSwitching: Different Use Cases

Enterprise Data Center

- Larger Packet Mix for Endpoint Use
- 10/25G connectivity
- Software switching
- Out-of-box platform software
- SDN controllers Priority
- Mainstream hardware features
- Live Migration Typical



Telco Network Infrastructure

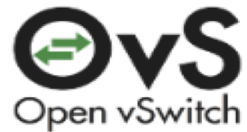
- Smaller Packets in Network Switching
- 10/25G connectivity and greater
- Software augmented with hardware
- Custom platform software
- Network Functions Virtualization
- Low jitter/latency
- Lower Downtime





ons
EUROPE
OPEN NETWORKING //
Integrate, Automate, Accelerate

A Brief History of some Open vSwitches



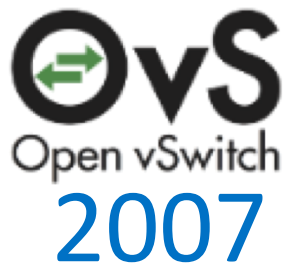
2015



2016



2013



2019+

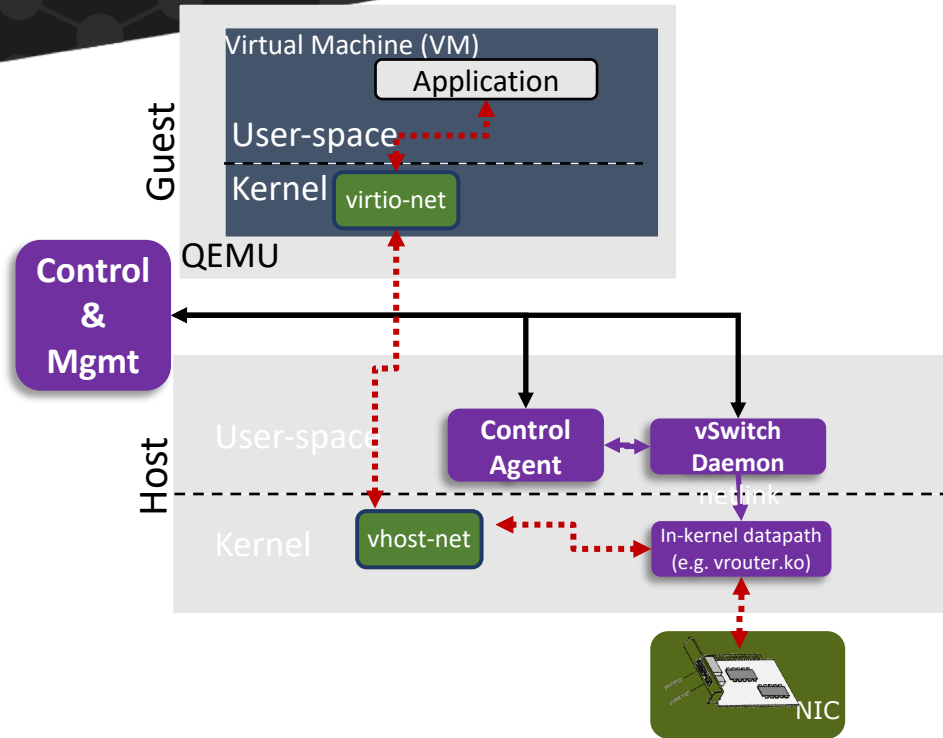
Offload/Acceleratio

Intel is leading the innovation with industry partners

*Other names and brands may be claimed as the property of others

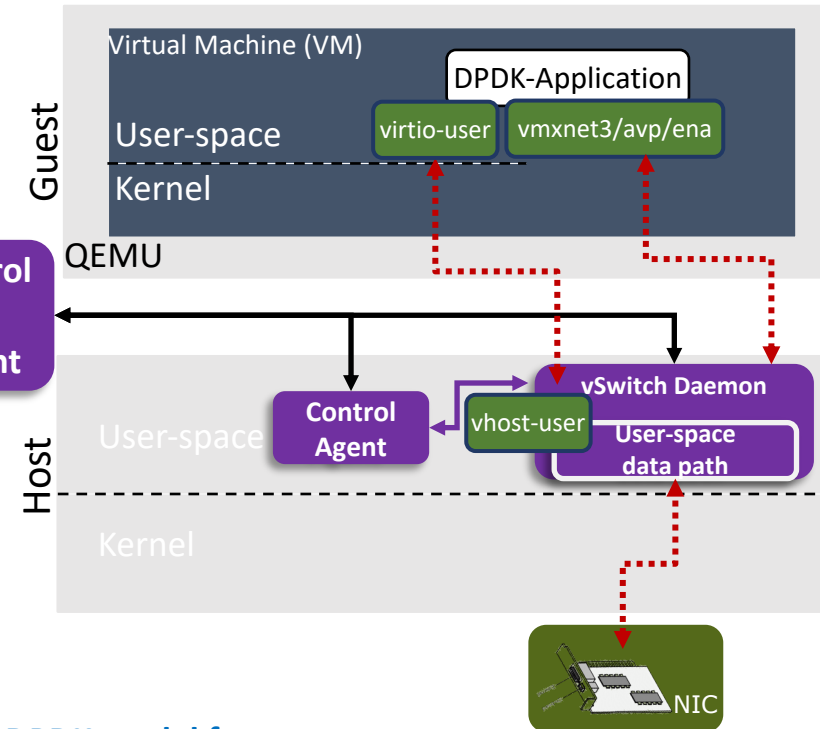


Kernel v. User-space Datapath



Kernel model slow:

- Packet needs to be copied across kernel/user-space boundary in both host and guest
- Kernel code responsible for packet I/O, tunneling, etc.



DPDK model fast:

- NIC DMA directly into user-space.
- DPDK memory allocation with hugepages reduces TLB misses.
- DPDK in host & guest copies directly from host user-space to guest.



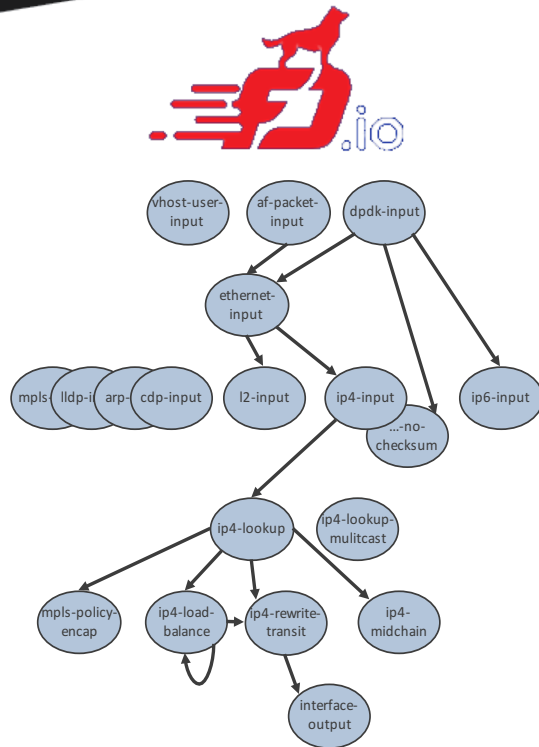
Dataplanes Compared

	OVS-DPDK	vRouter-dpdk	FD.io vPP
Forwarding Tables	Openflow tables "NORMAL-ACTION" (L2 LEARNING)	Interface table Nexthop table htable (flow table) entries (routing table/FIB) Bridge table (for each interface)	Interface: Physical & Virtual, BONDING Layer 2 : VLAN Q, Q-in-q Layer 3 : Unicast/MultiCast FIB Lookup Layer 4: Sessions/Endpoint TM: Classifiers, POLICING, NAT ...
Tunnel / Overlay	VXLAN, VLAN, GRE , MPLS, GENEVE, STT	MPLSoGRE/MPLSoUDP, VXLAN ...	IPSEC, SFC, SRv6, GRE, VXLAN GENEVE, L2TP, PPP, GTPU ...
Control Plane	Openflow, OVN, Neutron ML2, ...	XMPP (XML), VROUTER agent INTEGRATES WITH nEUTRON, ODL, etc.	VPP-API (LOCAL IPC) with control Planes apps for NETCONF/YANG, ODL, neutron ML2,



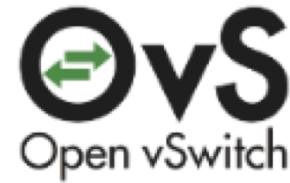
ons
EUROPE
OPEN NETWORKING //
Integrate, Automate, Accelerate

Dataplanes Compared



Each node is applied to all relevant packets before next node

- Leverages instruction locality cache
- No flow caching – performance scales with #flows
- Graph is easily extensible



Cost of lookup increasing

Ofproto Classifier

- Open Flow Tables
- Wildcard
- Ordered entries - tuples can overlap
- Has next table.

Datapath Classifier

- Wildcard matches
- Multiple *disjoint* match-tuple
- Flows expire

Exact Match Cache

- Exact match
- 8,192 entries
- N-way cache

Open Flow tables traversal is squashed to single datapath rule:

- Processing an existing flow is fast
- New flows must be processed by Ofproto



ons
EUROPE
OPEN NETWORKING //
Integrate, Automate, Accelerate

OpenVSwitch In Red Hat Portfolio and communities

RED HAT®
VIRTUALIZATION

oVirt



RED HAT®
OPENSTACK®
PLATFORM



RED HAT®
OPENS SHIFT®



OVS is distributed by layered products / communities, **not** by the OS distribution

RED HAT®
ENTERPRISE LINUX®



CentOS®



**Other names and brands may be claimed as the property of others*



ons
EUROPE
OPEN NETWORKING //
Integrate, Automate, Accelerate

Red Hat OpenStack supported SDN/Neutron plugins



JUNIPER[®]
NETWORKS

NEC



Some are based on OVS, some on tungsten fabric, some on FD.io, some on others.

<https://access.redhat.com/ecosystem/search/#/category/Software>

**Other names and brands may be claimed as the property of others*



ons
EUROPE
OPEN NETWORKING //
Integrate, Automate, Accelerate

OpenStack/NFV Datapath Performances

Low Range
Kernel OVS

No Tuning, default
deployment
Up to 50 Kpps

Mid Range
OVS-DPDK

Up to 4 Mpps per NUMA socket*
*Lack of NUMA Awareness
[OpenStack Blueprint NUMA aware vswitch \(Rocky\)](#)

High Range
SR-IOV
OVS HW Offload
TC-flower w/ SR-IOV**

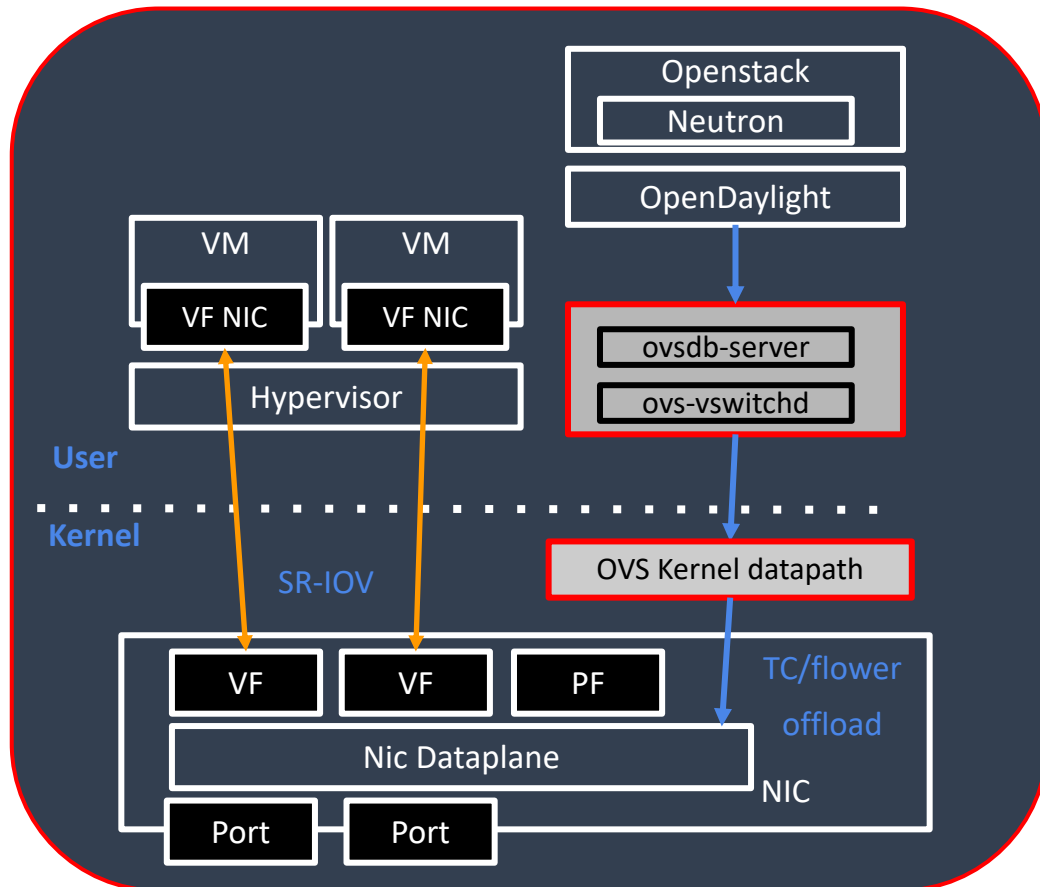
20+ Mpps per core (line rate?)
OVS 2.9, OpenStack Queens
**RHOSP 13/RHEL 7.5 (Tech Preview)

Throughput Measured in Packets per second with 64 Byte packet size (VSPerf PVP)

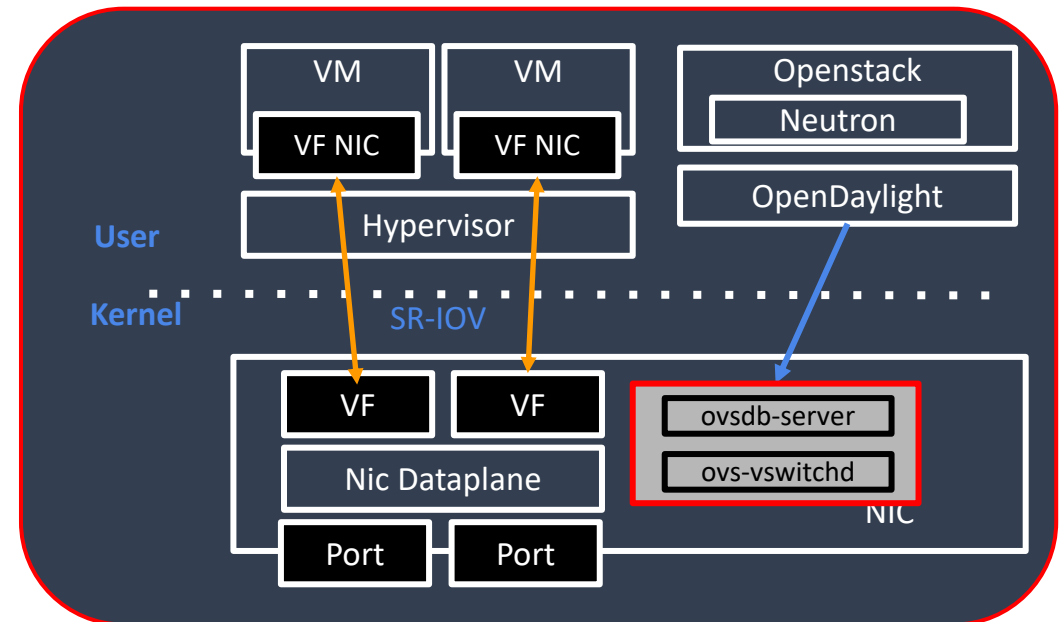


OVS OFFLOAD product roadmap

Kernel offload: RHOSP13 (TP*)



Complete offload: RHOSP15 (TP*)



Other offload

Two offload already
in-progress!

What's next?

VF specific driver =>
virtio in HW (vDPA)



ons
EUROPE
OPEN NETWORKING //
Integrate, Automate, Accelerate

VSPerf Overview

used by Red Hat CI

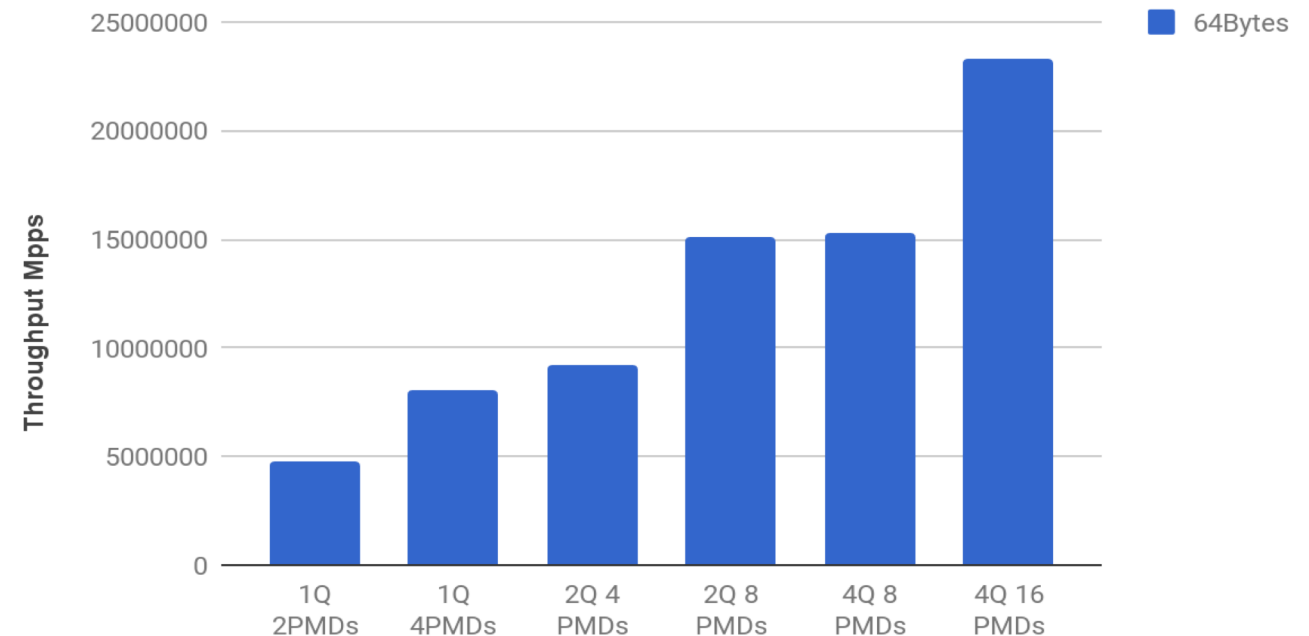
Performance measured PVP

- PVP: Physical -> Virtual -> Physical
- Traffic pattern until 2020 at least
- Platform calibration: testpmd PP and PVP
- Supported vSwitches: testpmd, OVS (kernel and DPDK), FD.io

Field reality

- Multi-queue works in lab only
- Realistic configuration (OpenFlow pipeline)
- Real traffic pattern
- All items above are WIP

64Bytes OVS-DPDK Zero Loss Multi-queue



	1Q 2 PMDs	1Q 4PMDs	2Q 4 PMDs	2Q 8 PMDs	4Q 8 PMDs	4Q 16 PMDs
64Bytes	4791550	8043502	9204914	15108820	15244256	23257998



ons
EUROPE
OPEN NETWORKING //
Integrate, Automate, Accelerate

Time to make a choice?

Production's path



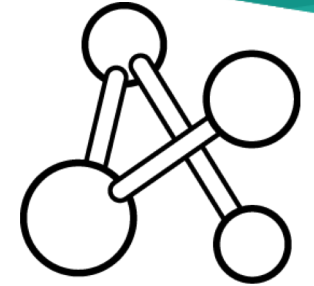
End user focused

Realistic expectations

Troubleshooting materials

Distributions/SDN commercial support

Innovation's path



Developers focused

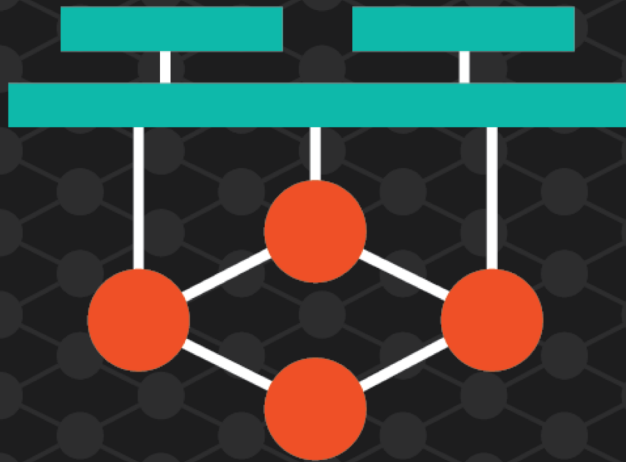
Supported SDNs / Nth. Bound API

Containers, Source-Routing, ...

10x performances (in some lab...)

“one size fits all” or modular stack allowing multiple vSwitches/vRouters?

September 25 - 27, 2018
Amsterdam, The Netherlands



ons

EUROPE

OPEN NETWORKING //
Integrate, Automate, Accelerate

Thank you!