Revisiting the Open vSwitch Dataplane Ten Years Later

William Tu[*], Yi-Hung Wei[*], Gianni Antichi[+], Ben Pfaff[*]

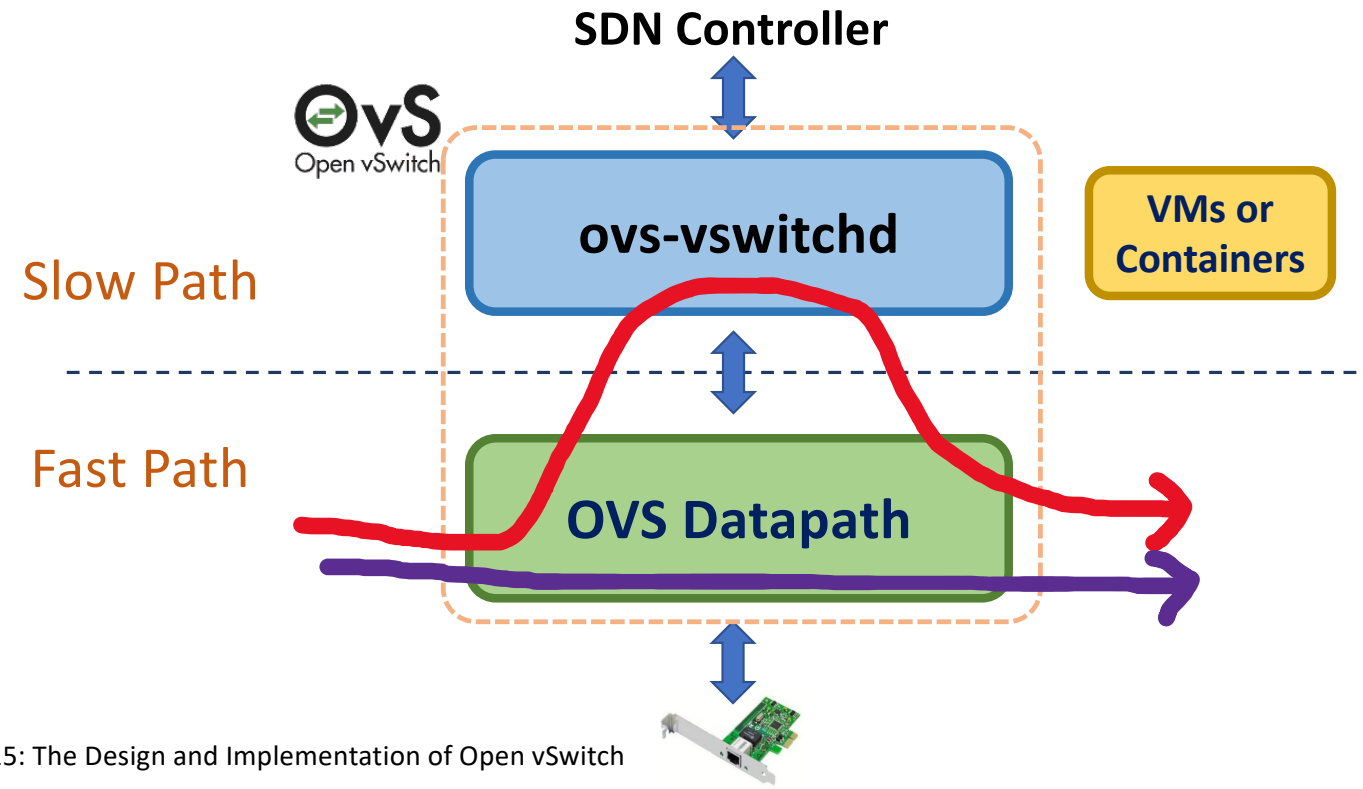
VMware, US[*]

Queen Mary University of London, UK[+]

ACM SIGCOMM 2021

What is Open vSwitch (OVS)?

An open-source software switch runs on multiple platforms



DigitalOcean

openstack.

vmware NSX

OPENSHIFT

ANTREA

And many research projects!

Classic OVS design (A Split User/Kernel)

Performance

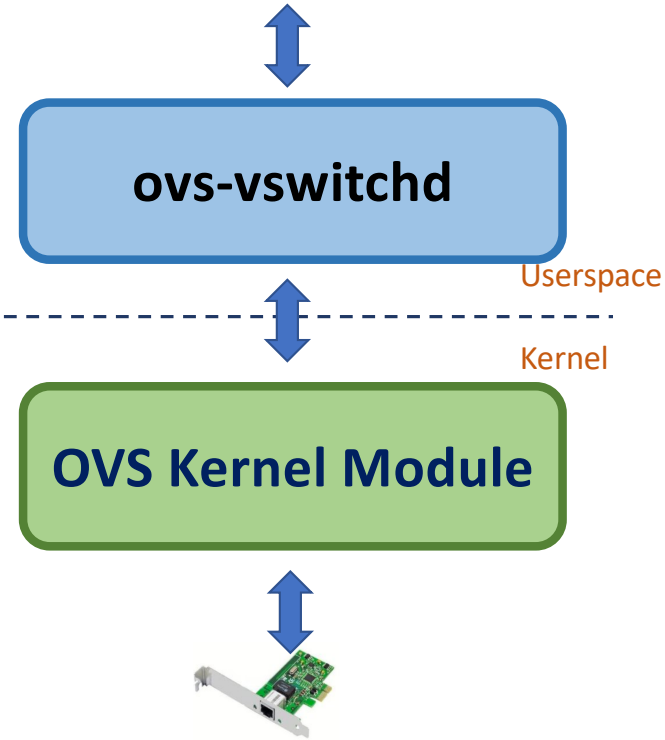
Before DPDK/netmap, most performant way to do packet I/O at that time.

Maintainability

Developer needs permission from Linux kernel community and OVS community.

Operability

OVS upgrade or bug fixes need to be on both components and rebooting is required.



OVS eBPF Datapath

Performance

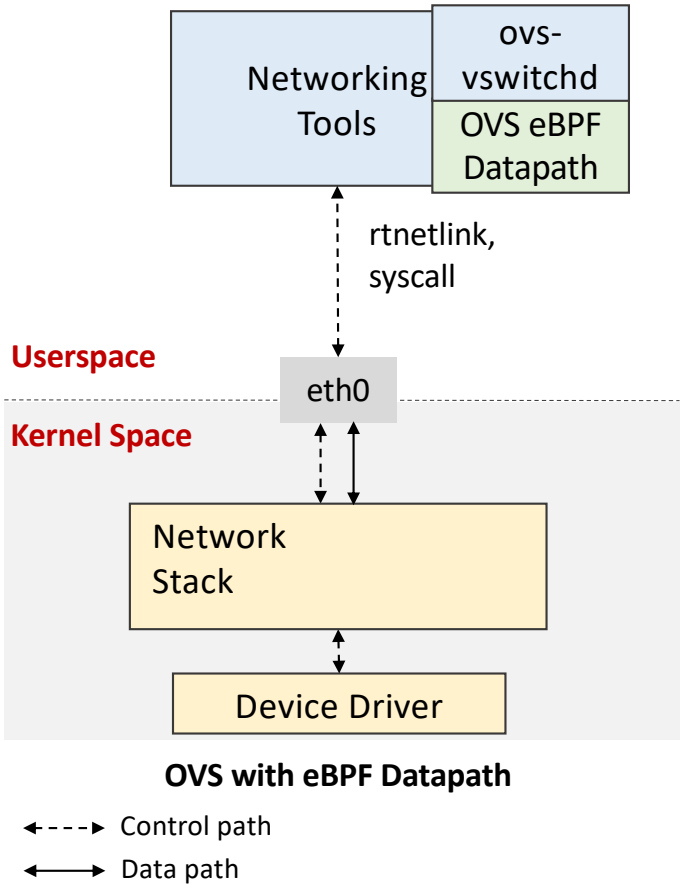
10-20% performance degradation, depending on the implementations.

Maintainability

Developer only needs to work with OVS and OVS community to add new features.

Operability

OVS upgrade or bug fixes done by replacing the new eBPF datapath. Tools remains working.



OVS-DPDK: Userspace Datapath

Performance

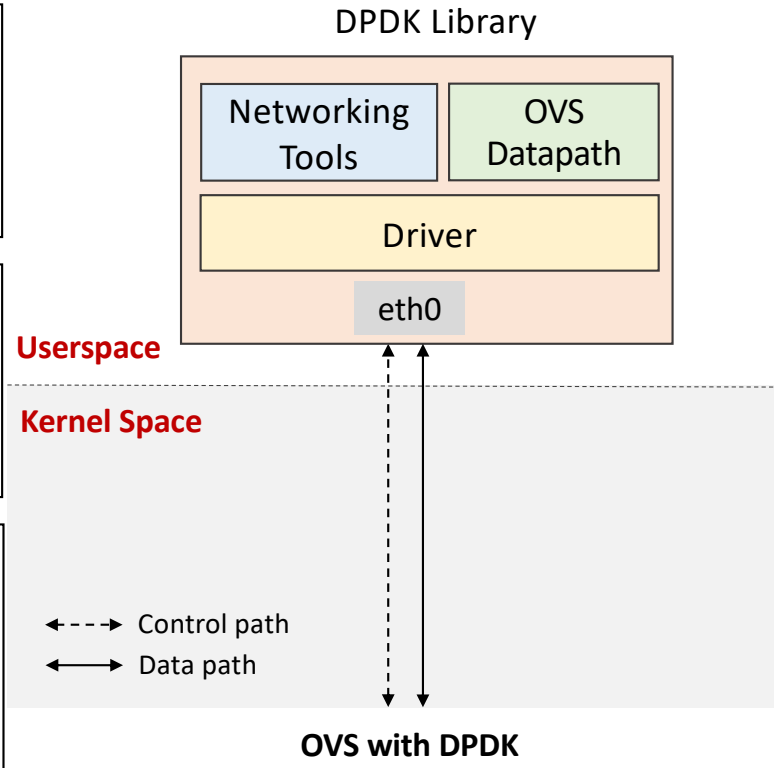
Closed to line-rate performance, ideal for high performance network appliance use cases.

Maintainability

Developer only needs to work with OVS and OVS community to add new features.

Operability

Manage two separate network configs: DPDK and Linux Kernel. Existing networking tools might not work on DPDK.



OVS Userspace Datapath with AF_XDP

Performance

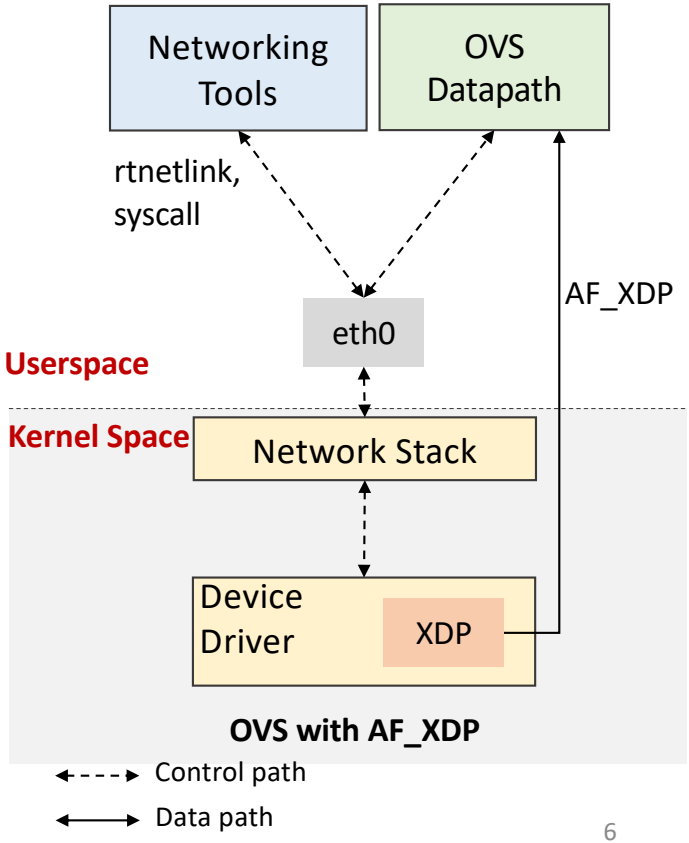
Fast packet processing in XDP, or DPDK-like performance when by-pass kernel w. AF_XDP

Maintainability

Developer only needs to work with OVS and OVS community to add new features.

Operability

Easier upgrade, bug fixes, and deployment. Networking tools remains working.



Implementation and Evaluation

- Naively apply AF_XDP socket has some performance overhead
  - See section 3 about optimizations

- Performance Evaluation
  - A. VMware NSX production platform with Intel NIC
  - B. Forwarding/Transaction rate using packet generator with Mellanox NIC

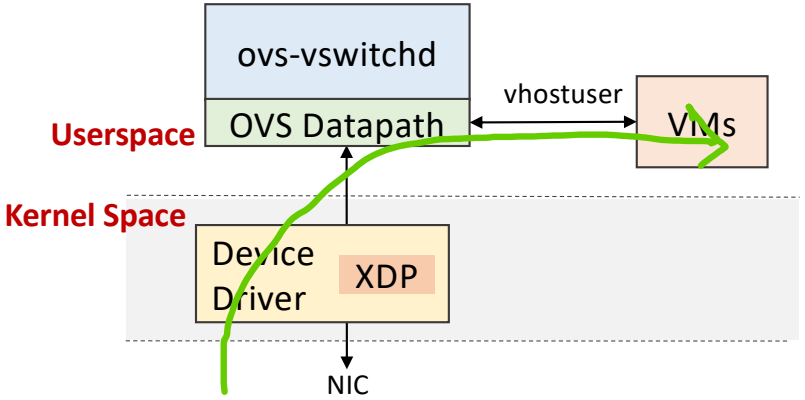
OVS Entity on NSX Platform	Count
Geneve Tunnel	291
Virtual ports (VMs per host)	15
OpenFlow Rules	103,302
OpenFlow Tables	40
Matching fields among all rules	31

Artifacts: https://github.com/williamtu/sigcomm21-ovs-artifacts

Evaluation: VM Performance

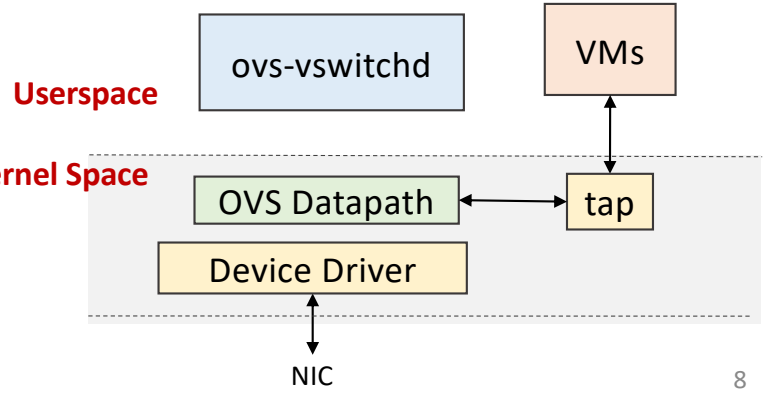
DPDK, AF_XDP

Zero-copy NIC to userspace datapath and to VMs.



Kernel

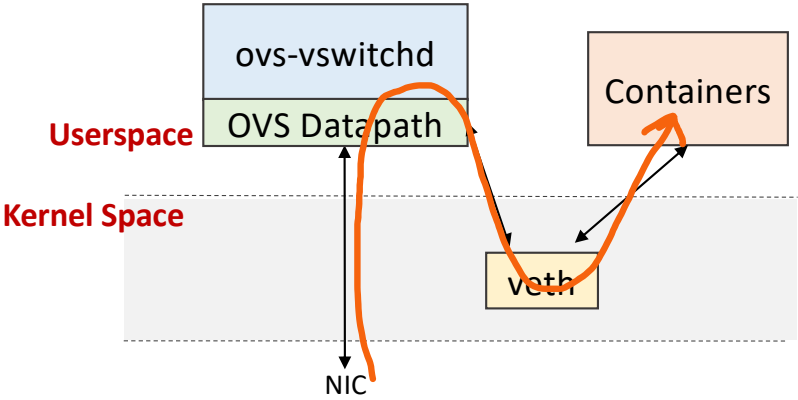
Packet copying from kernel to userspace using tap device.



Evaluation: Container Performance

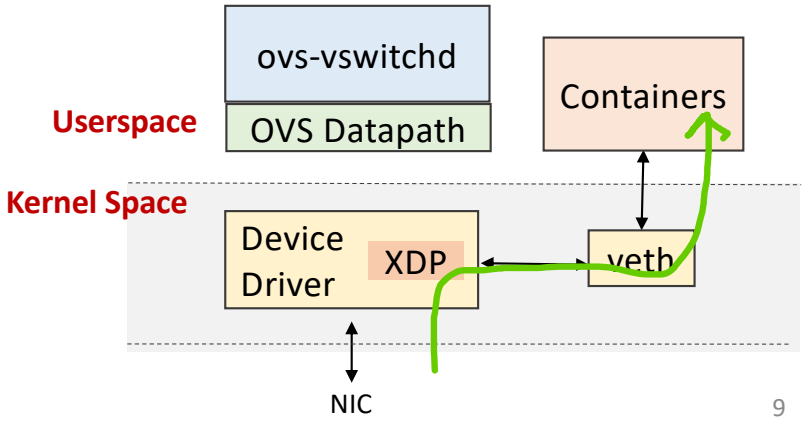
DPDK

High latency due to packet traversing kernel and userspace.



Kernel, AF_XDP

Better performance due to keep packets in kernel using kernel module or XDP program!



Lesson Learned and Conclusion

AF_XDP is not always the fastest option

AF_XDP is not yet available for Windows

Easier patching, development and troubleshooting

Easier packaging and validation using CI/CD

A better cross-platform design

Try it out!

$git clone https://github.com/openvswitch/ovs.git

Questions?