OvS for p2mp in NSM

About Open vSwitch in general
and for p2mp in NSM in particular

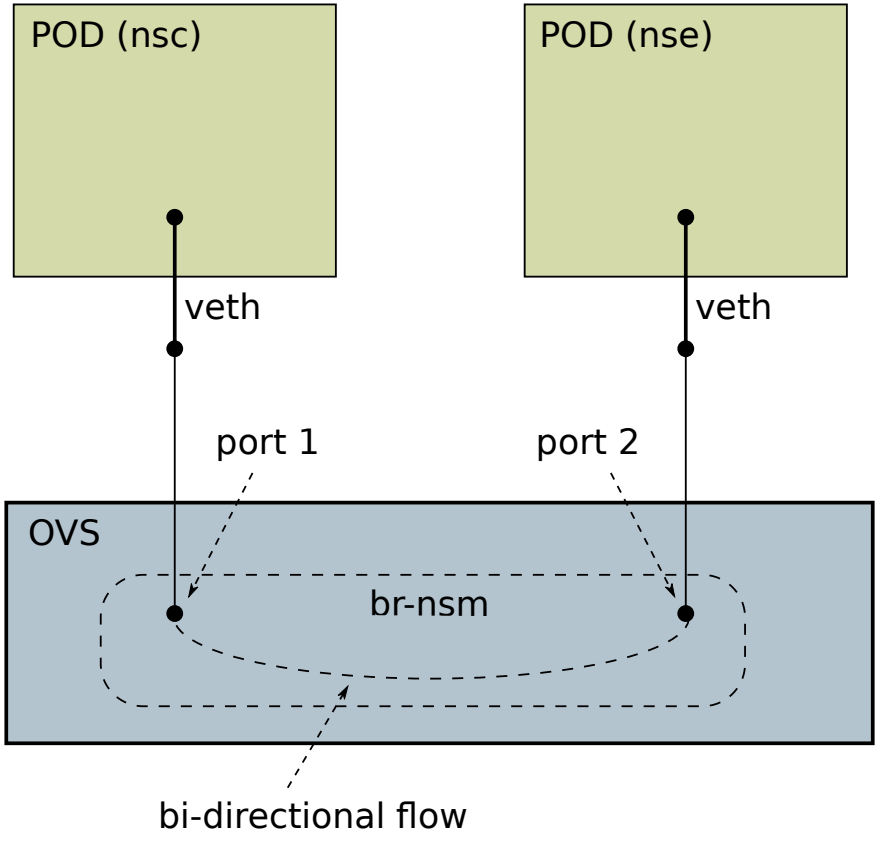
2022-06-25 lars.g.ekman@est.tech

Open vSwitch

- An OpenFlow enabled virtual switch on Linux with many bells and whistles
- Can run in the kernel or user-space (DPDK)
- Can utilize HW offload with a smart-nic
- In SDN many OpenFlow switches are handled by an “SDN controller”

NSM forwarder-ovs

- A working OvS is assumed on the host
- Only p2p (not L2)
- A bi-directional flow is setup with ovs-ofctl
- For remote connections a VXLAN tunnel is used (remote NSEs are **not** covered in this presentation)



Demo – local p2p

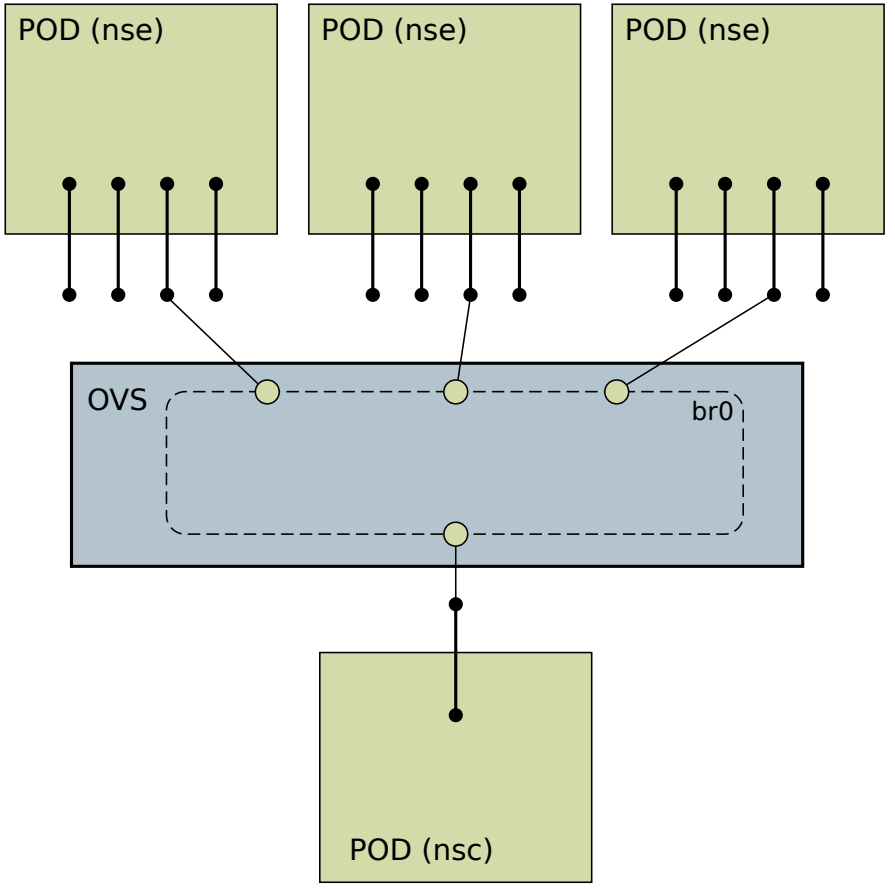
- Start cluster without K8s but with OvS
- PODs == network namespaces. Check veth’s
- Create a “nsm-br” OvS bridge (not L2)
- Attach the veth’s and test ping
- Create the bi-directional flow and test ping

NSM p2mp

- Only the data-plane is described in this presentation
- The BIG question is to choose between the vNSE Model and the IP Subnet Model
- There will be an explosion of bridges, one per NSC, dynamically created/deleted (now it’s only one per node)
- Implementations briefly discussed

NSM p2mp new NSC

- Create an OvS bridge
- Inject interfaces and connect all NSEs and the (one) NSC
- Setup flows (vNSE) or ECMP in the NSC (IP subnet)



IP Subnet Model – L2 switch

- Can be implemented with other forwarders than OvS
- NSCs must be aware of multiple NSEs (ECMP?)
- Is it really better than the “proxy”?

vNSE Model

- The function *may* only be possible with OvS. Does it matter?
- NSCs only see one NSE address (if any). Load-balancing between NSEs is done by OvS
- The same MAC+IP is shared among all NSEs. This causes problems (non-VIP IPs may not be needed)

Problems with same MAC+IP

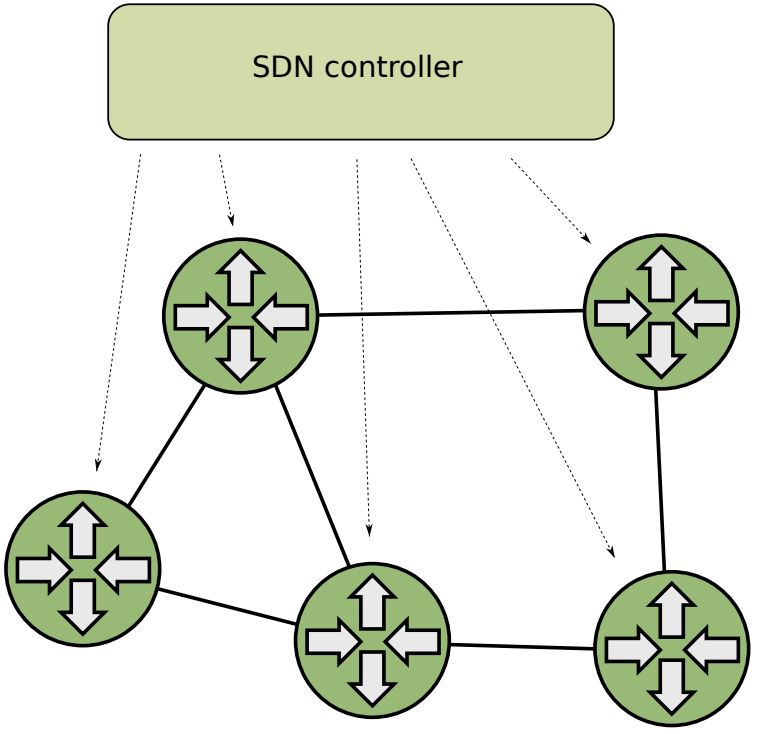
- How to configure/set them in NSM (control-plane)
- ARP (neighbor discovery) should be disabled
- IPv6 DaD *must* be disabled
- IPv6 generated link-local address should be disabled
- Since OvS doesn’t care about the MAC address we can use the same MAC everywhere (Linux cares though)

Demo – OvS load-balancing

- Start cluster. Show “PODs”
- Create a bridge (the bridge interface (br0) emulates the NSC)
- Attach POD (nse) interfaces (already injected) to the bridge. Check MACs
- Setup load-balancing (quite complicated)
- Add VIP addresses and route (differs from the NSE setup)
- Add return flows p2p POD→br0 (NSE→NSC)
- Test with mconnect

SDN

- In SDN many OpenFlow switches are managed by an SDN controller
- Packets without flow are sent to the controller who sets up flows



Demo - ovs-testcontroller

- Start cluster
- Create an ovs bridge and attach PODs
- Test ping (will not work)
- Start the ovs-testcontroller (tcp:127.0.0.1)
- Test ping (works!)
- Check flows

References

- Jan and Ed's investigation
- Xcluster ovl/ovs
- Nsm-test repo on github (where this presentation is stored)