CERN – Switch OVN Sync

cristian.contescu@switch.ch

16th December 2024

Agenda

- Deployment Overview
  - General setup.
  - Network stack.
  - Load-balancer implementation.
- Motivation for OVN
  - Design choices and Tradeoffs.
- Work in Progress
  - Geneve tunnel vs VLAN interfaces.
    - Performance and some bugs.
  - Some other tenant network issues.
- Outlook

Deployment Overview - Overall

- Use Kubernetes for deploying Openstack control plane.
- All Openstack services are deployed using Openstack-Helm.
- FluxCD making sure declared configuration is provisioned on target environment.
- Customizations needed for OVN deployment are deployed utilizing Helm or Kustomize.

Deployment Overview – Network Stack

- We currently run Openstack Antelope (2023.1):
  - Neutron: 22.2.2
  - OVN: 24.03.2
  - Open vSwitch: 3.3.0
- Provider network: VLAN.
- Tenant networks: Geneve tunnels and VLANs.
- External access to tenant networks:
  - IPv4: distributed floating IPs on provider network
  - IPv6:
    - Public IP ranges for tenant networks
    - Highly available virtual routers (OVN default: prio-based chassis assignment)
    - Route advertisements via ovn-bgp-agent

Deployment Overview – Load-balancers

- Running Octavia (12.0.1) with ovn-octavia-provider (4.0.2)
  - One internal customer using OVN LBs already.
  - Will support amphora as well (for L7 LBs, not a priority now).
- OVN LBs have less functionalities but are quicker to set up and update: big plus for our internal customer.
- Compared to Amphorae, requires much less resources and Octavia's configuration is very straight-forward.

Motivation for OVN

- Distributed floating IPs (traffic flows directly to/from compute nodes):
  + Better performance.
  - Only works for IPv4.
  - OVN load-balancers break the VLAN tenant networks (unusable for this use-case).
- Resource optimised:
  - No longer need dedicated DHCP agent.
  - No dedicated network nodes required.
- Optimised ACLs: Security groups are handled by OVN.
- Direct Routing: E/W traffic is always routed directly between the hypervisors.

WIP: What could possibly go wrong(TM)?

- Nothing, we're just here to spread the word! 
- Using solutions like `keepalived` to float IPs between VMs on the same tenant network seems to have an inconsistent behavior.
- Geneve tunnel low performance (maybe MTU related?) - ~ 40% of the line speed.
- OVN Load-balancer hair-pinning issue:
  - Member of a load-balancer tries to reach the external IP (floating IP attached to the VIP) of the LB -> works only ~ 66% of the time (if 3 backends)
  - Work around in place: allowing public IP of the LB in the security groups

WIP: What could possibly go wrong(TM)?

- Loadbalancer listener '`allowed_cidrs`' option is not (yet) implemented for OVN LBs.
- Before Antelope: issues with the metadata service when Octavia loadbalancer health-monitors were added.
- Important to note: if using VLAN tenant network, make sure you configure unique `external-ids:ovn-chassis-mac-mappings` in OVS for each hypervisor [1], otherwise centralized traffic (i.e.: IPv6) will not flow correctly.

[1] https://www.ovn.org/support/dist-docs/ovn-controller.8.html

Outlook – Things we'd like to see.

- The basic network functionality seems covered well by OVN, we'd like to see some of the more advanced use-cases covered as well (i.e.: HA with keepalived not breaking N/S traffic).
- Improved troubleshooting support:
  - Currently very bare bones (from an operator point of view).
  - Requires a lot of effort to provide the correct input.
  - If not familiar with the results, hard to spot if it is a real issue, or if input was wrong.

Summary

- The benefits of OVN
  - Better performance (when combined with distributed floating IPs).
  - Easy to setup Layer 4 load-balancers.

- The draw-backs
  - Not yet widely adopted, so harder to get (community) support.
  - Difficult to troubleshoot in case of bugs.

- Some considered alternatives to our current deployment
  - Addressing poor performance:
    - OVN in conjunction with OVS-DPDK
      + would allow hardware offload.
      - does not (yet) support IPv6 BGP advertisement out of the box (and we use this feature).
  - Disable distributed floating IPs solves some issues, at the expense of performance.

Questions?