New Approach to OVS Datapath Performance

Founder of CloudNetEngine
Jun Xiao
Agenda

• VM virtual network datapath evolvement
• Technical deep dive on a new OVS datapath
• Performance comparisons
• Q & A
VM virtual network datapath evolvement

- VNIC emulation
- VNIC paravirtualization
- VNIC/PNIC Multiple queues/load balance
- VNIC offloading and PNIC H/W acceleration

- Overlay
- Overlay awareness offloading

- Stateful actions, i.e. conntrack

- Very high packet rate processing

OVS Kernel DP

OVS-DPDK DP
Why a new approach to OVS datapath performance?

- VNIC emulation
- VNIC paravirtualization
- VNIC/PNIC Multiple queues/load balance
- VNIC offloading and PNIC H/W acceleration

- Overlay
- Overlay awareness offloading

- Stateful actions, i.e. conntrack

- Very high packet rate processing

---

OVS Kernel DP

OVS-DPDK DP

CPU efficiency is very important!
A new approach to OVS datapath performance

VNIC emulation
VNIC paravirtualization
VNIC/PNIC Multiple queues/load balance
VNIC offloading and PNIC H/W acceleration

Overlay
Overlay awareness offloading

Stateful actions, i.e. conntrack

Very high packet rate processing

...
Technical deep dive on CloudNetEngine virtual switch
Design principles

• Datapath needs to be as reliable as possible
• High performance for all typical workloads
  – Throughputs on both BPS and PPS wise
  – CPU efficiency is very critical
• Easy of integration for various virtual networking solutions
• Easy of maintenance
CloudNetEngine virtual switch architecture

- Openflow APIs
- OVSDB APIs
- ovs-vswitchd* (dpif-netlink)
- ovsdb-server

- ODP APIs
- Adv APIs
- CDP APIs

- Scheduler
- Adaptive poll
- Mem mgmt
- Timer mgmt
- Classifier*
- Flow cache*
- FWD engine*
- Multi queue
- Overlay
- Security group
- QoS
- Offloading
- Sniffer

- DPDK*

- Net chain

CDP (CloudNetEngine Data Path)
Performance practices

• Packet handle layout, lazy metadata reset.
• Improve instruction per cycles.
• Load balancing rxq processing.
• Inline filtering for packet monitoring.
• CPU efficiency:
  – Hybrid polling + RX interrupt
  – Packet group metadata
  – Zero copy
  – S/W H/W Offloading depending on system runtime configuration
Extensibility

• A lightweight and efficient framework (net chain) to plugin new features
  – It’s RCU protected so that updating a net chain won’t have any performance penalty on the datapath
  – A net chain can use packet group metadata to very quickly decide whether the net chain is applicable to the input packet vector or not
Performance comparisons
Performance test configuration

Host H/W
CPU:
- Xeon E5-2620 v3 2.40GHz
- 6 physical cores, 12 logical cores
NIC:
- 82599ES 10-Gigabit
MEM:
- 16G

Host S/W
- Ubuntu 16.04 x86_64 + KVM
- Qemu 2.5.1
- 1G size hugepages are used

All QEMU instances set cpu affinity

Guest H/W
- 4 vCPUs/ 2vNICs/ 4G memory for NFV tests
- 1 vCPUs/ 1vNICs/ 1G memory for non-NFV tests
- for NFV tests, virtio mrg_rxbuff=off, all other offload flags are enabled
- for non-NFV tests, virtio all offload flags are enabled
- vNICs use default queues

Guest S/W
- buildroot kernel 4.4.3 x86_64
- testpmd io mode forward for NFV test
- iperf 3.1.1 for TCP test
- netperf 2.7.0 for TCP_RR test

Virtual Switches Under Test

Native OVS
- OVS 2.6
- kernel module bundled with Linux kernel 4.4.0

OVS-DPDK
- OVS 2.6
- DPDK v16.11

CNE vSwitch
- CNE vSwitch 1.0
NFV Test Topology

- Bi-directional traffic
- Each direction with 250 concurrent flows
- 0.5% PDR
MPPS (Higher is better)

NFV 64 Bytes 0.5% PDR test Throughput
TCP Single Host Test Topology
TCP/VXLAN Two Hosts Test Topology
TCP/VXLAN Two Hosts Test Throughput

TCP/VXLAN Two Hosts Test CPU Usage
TCP_RR/VXLAN Two Hosts Test Throughput

TCP_RR/VXLAN Two Hosts Test CPU Usage
Demo: CNE vSwitch integration with OVN/OpenStack
Q & A

www.cloudnetengine.com

info@cloudnetengine.com

Twitter: @cloudnetengine