Hardware Offload of QoS

Simon Horman — Open vSwitch and OVN 2021 Fall Conference
Agenda

- Packet-per-second rate-limiting
- Modeling OpenFlow Metering with Policer Action
Packet-Per-Second Rate-Limiting
Open vSwitch Rate Limiting

- Oldest QoS feature in OVS?
- Limit applied to packets received by an OVS port
  - bits/s and burst
- Over-limit packets are dropped
- Configured via OVSDB

```
 ovs-vsctl set interface $dev ingress_policing_rate=1000
 ovs-vsctl set interface $dev ingress_policing_burst=100
```
TC Policer

- For kernel datapath TC police action is used to implement OVS rate limiting
- Also provides hardware offload facility

```
tc filter add dev $dev ingress protocol all \matchall \action police rate 1Mbit burst 125Kb
```
Packet-per-Second Rate-Limiting

PPS

- Add mechanism to apply rate-limit based on packets/s
- Now a policer action instance may enforce rate limit as either:
  - Bits/s (long-standing behaviour)
  - Packets/s (new behaviour)
- Modes may be combined using multiple actions

```
tc filter add dev $dev ingress protocol all \  
match all \ 
action police pkts_rate 2000 pkts_burst 400 \ 
action police rate 1Mbit burst 125Kb
```
OVS and PPS

- Configurable via OVSDB
- May be combined with bit/s rate limit on same port

```bash
ovs-vsctl set interface $dev ingress_policing_kpkts_rate=2000
ovs-vsctl set interface $dev ingress_policing_kpkts_burst=400
ovs-vsctl set interface $dev ingress_policing_rate=1000
ovs-vsctl set interface $dev ingress_policing_burst=100
```
Kernel Hardware Offload of PPS

- Offload via Flow Rule API
- Support dependent on offload driver
- Currently only supported by NFP driver with flower

```c
struct flow_action_entry {
    ...
    struct { /* FLOW_ACTION_POLICE */
        u32    index;
        u32    burst;
        u64    rate_bytes_ps;
        u64    burst_pkt;
        u64    rate_pkt_ps;
        u32    mtu;
    } police;
    ...
};
```
Status of PPS

Packet-per-second rate limiting is fully upstream

- Kernel
  - v5.13
  - TC policer action
  - Flow rule
  - NFP driver

- OVS
  - v2.16
Open Flow Meters meet the TC Police Action
Open Flow Meters

• n-band

• Each band has a target rate
  • bytes/s or packets/s
• Drop or DSCP mark packets

• Are independent of flows
  • May be used by zero or more flows via meter action

• Are independent of ports
  • May be used in flows attached to different ports
ovs-ofctl -O openflow13 add-meter br0 \ 
    meter=99, kbps, burst, band=type=drop, rate=1000, burst_size=100
for dev in $DEV1 $DEV2; do
    ovs-ofctl -O openflow13 add-flow br0 \ 
        in_port=$dev, ip, action=meter:99, output=$DEV3 \ 
    ovs-ofctl -O openflow13 add-flow br0 \ 
        in_port=$dev, ipv6, action=meter:99, output=$DEV3 \ 
done
Meters and Police Action

Open Flow Meters
• n-band

• Each band has a target rate
  • bits/s or packets/s
• Drop or DSCP mark packets

• Are independent of flows
  • May be used by zero or more flows via meter action
• Are independent of ports
  • May be used in flows attached to different ports

TC Police Action
• 1-band
  • multiple bands via multiple actions

• Each instance has a target rate
  • bits/s or packets/s
• May drop packets
  • Or mark with assistance from other actions

• May be independent of filters
  • May be used by zero or more flows using index of action instance
• May be independent of netdevs
  • May be used by filters attached to different ports

Wow, that’s a pretty good match!
TC Action Independent of Filters

tc action add action police index 99 rate 1mbit burst 100k
for dev in $DEV1 $DEV2; do
    tc filter add dev $dev ingress protocol ip \
        flower ip_proto tcp action police index 99
    tc filter add dev $dev ingress protocol ipv6 \
        flower ip_proto tcp action police index 99
done
Hardware Offload Prototype

Offload via Flow Rule API

- New tc_setup_type
  - TC_SETUP_ACT
  - First class citizen along with U32, Flower, Block, ...

- Hook into cls_api and act_api for action:
  - Creation/Modification
  - Deletion
  - Statistics
Status of Meters via TC

Upstreaming in progress

- Kernel
  - Posted
  - Targeting inclusion in v5.17
  - cls_api and act_api
  - Flow rule
  - NFP driver

- OVS
  - Targeting v2.17
THANKS