Balance-TCP Bond Mode Performance Improvement

Vishal Deep Ajmera
Nitin Katiyar
Pradeep Venkatesan
Anju Thomas

Ericsson

AGENDA

- OVS Bond Modes – Pros & Cons
- Limitations with current design
- Proposed optimization
- Test topology & Test results
- New CLI
- Summary

BOND MODES

- Balance-SLB (Source Load Balance)
  - Uses source mac address and VLAN hash to identify member link.
  - Better packet throughput as no additional recirculation required.
  - Poor load balancing when using overlay tunnels like VxLAN.

- Balance-TCP
  - Uses 5-tuple hash to identify member link.
  - Better distribution and load balancing.
  - Low packet throughput due to additional recirculation of packet.

CURRENT DESIGN

ERICSSON

- Uses hash() & recirc() actions
- Recirculation of packets reduces the packet throughput.
- Post recirculation flows (a.k.a. pr-rules) occupy EMC cache entries.
- Unique recirculation id for each bond port.

EXAMPLE DPCLS FLOWS

- With 8 IP-UDP flows (with random UDP source port):

recirc_id(0),in_port(7),packet_type(ns=0,id=0),eth(src=02:00:00:02:14:01,dst=0c:c4:7a:58:f0:2b),eth_type(0x0800),ipv4(frag=no),
actions:hash(hash_l4(0)),recirc(0x1)

Post-recirculation flows (pr-rules):

recirc_id(0x1),dp_hash(0xf8e02b7e/0xff),in_port(7),packet_type(ns=0,id=0),eth_type(0x0800),ipv4(frag=no), actions:2

recirc_id(0x1),dp_hash(0xb236c260/0xff),in_port(7),packet_type(ns=0,id=0),eth_type(0x0800),ipv4(frag=no), actions:1

recirc_id(0x1),dp_hash(0x7d89eb18/0xff),in_port(7),packet_type(ns=0,id=0),eth_type(0x0800),ipv4(frag=no), actions:1

recirc_id(0x1),dp_hash(0xa78d75df/0xff),in_port(7),packet_type(ns=0,id=0),eth_type(0x0800),ipv4(frag=no), actions:2

recirc_id(0x1),dp_hash(0xb58d846f/0xff),in_port(7),packet_type(ns=0,id=0),eth_type(0x0800),ipv4(frag=no), actions:2

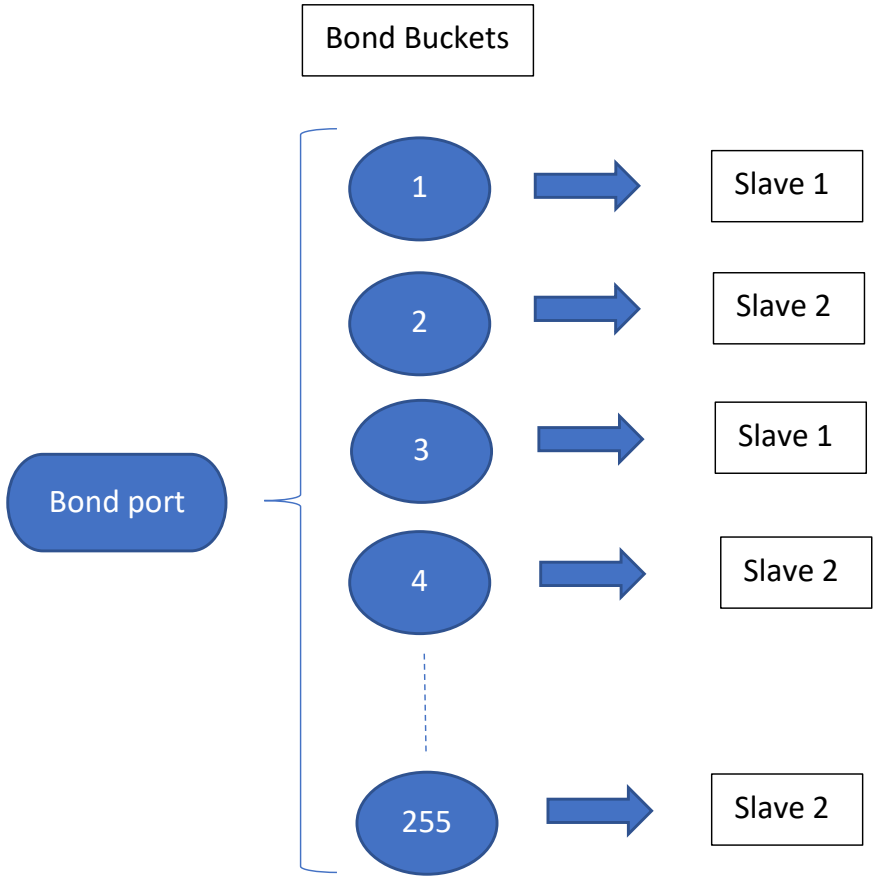
recirc_id(0x1),dp_hash(0x24534406/0xff),in_port(7),packet_type(ns=0,id=0),eth_type(0x0800),ipv4(frag=no), actions:1

recirc_id(0x1),dp_hash(0x3cf32550/0xff),in_port(7),packet_type(ns=0,id=0),eth_type(0x0800),ipv4(frag=no), actions:1

- Up to 255 unique pr-rules matching each possible hash (8-bits) for bond-port.

PROPOSED DESIGN

- Bond buckets
  - Pre-create 255 buckets (equivalent to 255 pr-rules).
  - Bucket indexed using RSS hash (if available) or 5-tuple hash.
  - Statistics maintained at bucket level for load balancing.
- Bond-id used to identify bond port.
- Each bond port has its own set of bond buckets.
- Each PMD maintains bond cache with mapping of bond buckets to slaves.



NEW DATAPATH ACTION

- **lb_output(bond, <bond-id>)**
  - Replaces hash() and recirc() actions for balance-tcp mode.
  - Bond id is same as recirculation id

- Example datapath flow for packets from vm-port to bond-port :

in_port(7),packet_type(ns=0,id=0),eth(src=02:00:00:02:14:01,dst=0c:c4:7a:58:f0:2b),
eth_type(0x0800),ipv4(frag=no), actions: lb_output(bond,1)

- Only one dpcls flow entry irrespective of hash.

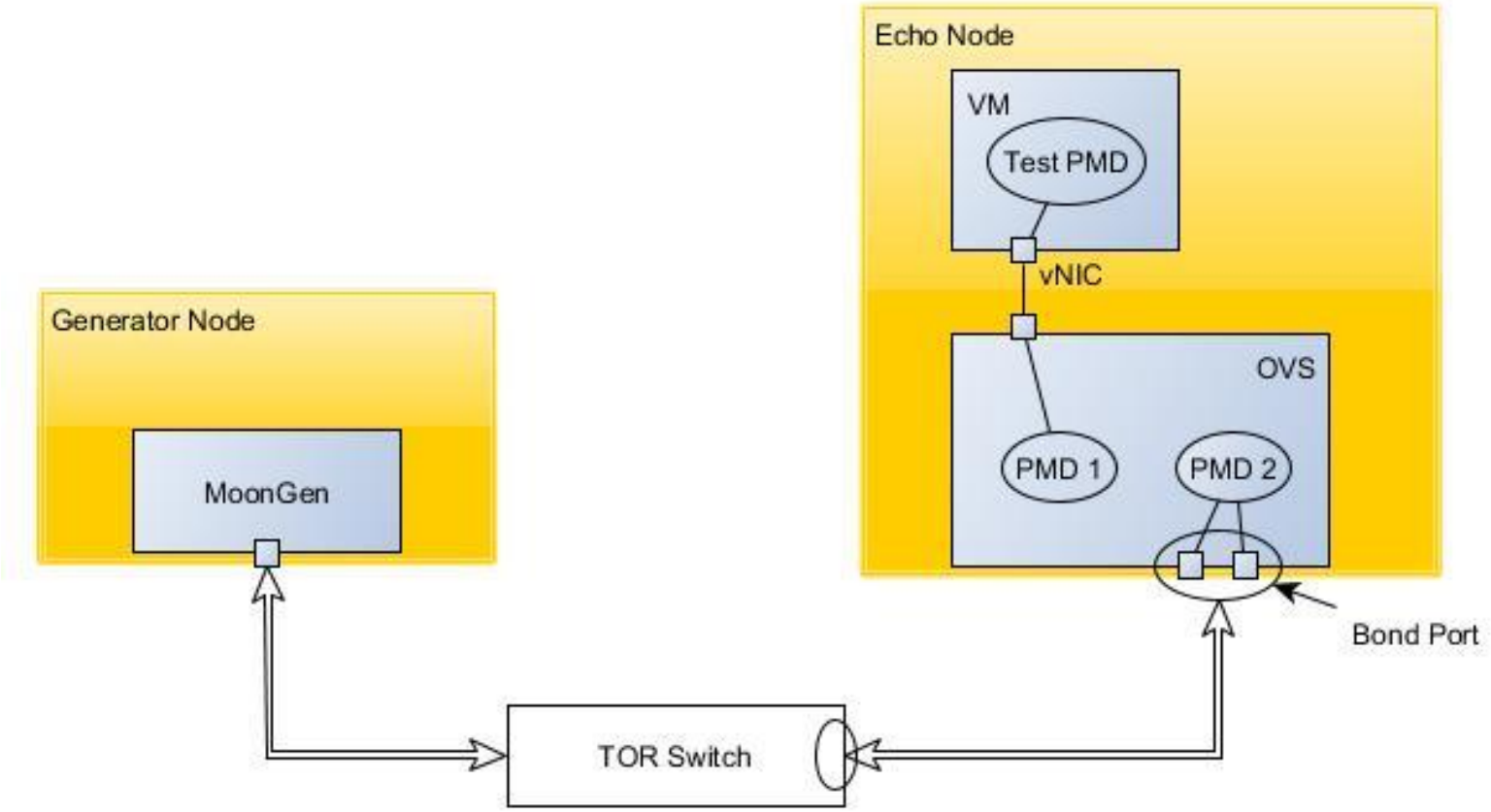
LOAD BALANCING

- Each bucket maintain packets and bytes count.
- Core load balancing logic (ofproto) remains intact.
- For redistribution modify buckets to use different slave id.
- Change in bucket mapping needs PMD to refresh bond cache.

bond-id 1:
bucket 0 – slave 2
bucket 1 – slave 1
bucket 2 – slave 2
bucket 3 – slave 1

bond-id 1:
bucket 0 – slave 1
bucket 1 – slave 1
bucket 2 – slave 2
bucket 3 – slave 2

TEST TOPOLOGY



TEST PARAMETERS

- Phy -> VM -> Phy test (PVP)
- Loss profile < 10 ppm
- Packet size 64 bytes, UDP
- Multiple unique packet streams
- 2 PMDs

TEST RESULTS

# of packet streams	OVS Master (Mpps)	OVS Master + Optimization (Mpps)	% Improvement
1	4.47	5.73	28%
10	4.17	5.35	28%
1000	3.41	5.25	53%
10000	2.53	4.57	80%
100000	2.33	4.27	83%
500000	2.33	4.27	83%

NEW CLI

ERICSSON

- To enable/disable:
  - *ovs-vsctl set port <bond port> other_config:lb-output-action=<true|false>*
  - By-default it is false.

- To dump bond cache in datapath:
  - *ovs-appctl dpif-netdev/dp-bond-show [dp]*

- To check if bond port is using new action:
  - *ovs-appctl bond/show*

SUMMARY

ERICSSON

- Balance-TCP bond mode provides better load distribution due to 5-tuple hash.
- Use of bond buckets eliminates recirculation of packets for bond member selection.
- Supported in ‘netdev’ datapath only. Kernel datapath will continue to use existing actions.
- Patch-set in the mailing list for review.

  https://mail.openvswitch.org/pipermail/ovs-dev/2019-September/362758.html