OpenStack NFV: Performance with OvS-DPDK for NFV and Connection Tracking

Bhanuprakash Bodireddy(bhanuprakash.bodireddy@intel.com), Sugesh Chandran(sugesh.chandran@intel.com), Antonio Fischetti(antonio.fischetti@intel.com)

OpenStack Summit, Boston, May’17
Introduction

- OvS-DPDK, compelling solution to achieve NFV goals
- Zero-Trust environments
- Connection Tracker and performance
- OvS-DPDK for NFV use cases
- Troubleshooting
- Hot Spot Analysis
Connection Tracker
Stateful vs Stateless Firewalls

Software Must Be Easier, Right?

White House Hacked In Cyber Attack That Used Spear-Phishing To Crack Unclassified Network

Experts: Google’s ‘Aurora’ hackers still at it years later

Attacks on Google and others three years ago have similarities to new attacks, including use of rare zero-day vulnerabilities.
Launch an instance with Security Groups
Verifying VM connectivity with SG enabled
Recirculation explained

Conntrack sets the connection bits in packet meta data
Recirculation explained

‘recirc’ flag is set and so the packets are sent back to OvS flow table
Recirculation explained

- Recirculated packet matches the ‘flow c’ due to connection bits
- ct(commit) - update the connection info in to CT Table
- output:2 - forward the packet on to output port 2
Recirculation explained

Once the connection is established, recirculated packets will match ‘Flow D’ and get forwarded on output port 2
**Compute Node 1 (egress)**

### Flow Entries

**Rule 1:** Match: `ct_state(-trk), recirc_id(0), in_port(6), eth(src=fa:16:3e:2a:05:ce), eth_type(0x0800), ipv4(src=10.0.0.3,proto=1,frag=no), …`,
   actions:`ct(zone=4), recirc(0x188)`

**Rule 2:** Match: `ct_state(-new+est-rel-rpl), recirc_id(0x188), in_port(6),
   eth(src=fa:16:3e:2a:05:ce,dst=fa:16:3e:9f:55:c1),
   eth_type(0x0800),ipv4(tos=0/0x3,frag=no), …`,
   actions:`tnl_push(tnl_port(5),header(size=50,type=4,eth(dst=90:e2:ba:6b:1f:61,sr=90:e2:ba:2b:49:41,dl_type=0x0800),
   ipv4(src=10.237.100.2,dst=10.237.100.3,proto=17,…),
   udp(src=0,dst=4789,csum=0x0),vxlan(flags=0x8000000,vni=0x4)),out_port(2))`

**Rule 3:** Match: `recirc_id(0), in_port(2), eth(src=90:e2:ba:2b:49:41,dst=90:e2:ba:6b:1f:61),
   eth_type(0x0800), ipv4(frag=no),
   actions:3`
Testing Stateful Firewall, Setup Details
**Key Performance Metrics**

**Connection**
- Transactions per second
- Connections per second
- Throughput
- TTFB (Time to first byte)
- TTLB (Time to last byte)

**Transaction**

**Concurrent connections**

**Transactions per second**

**Protocol latency**

**Desirable Connection**
1. TCP connection establishment (SYN, SYN-ACK, ACK)
2. layer 7 transaction (Request, Response)
3. TCP teardown (FIN, ACK).

**L4 performance with limited L7 transactions**
1. TCP connection establishment (SYN, SYN-ACK, ACK)
2. Partial/incomplete layer 7 transaction (only Request)
3. TCP teardown (FIN, ACK).

**Forceful connection tearowns**
1. TCP connection establishment (SYN, SYN-ACK, ACK)
2. Partial/incomplete layer 7 transaction (Request)
3. Forced TCP teardown (RST).
Waterfall view of response time

Results for **openstack.org** website retrieved using Webpagetest.org
HTTP client/server – Throughput test

**TCP connections**
- **TCP Connections Established**
  - conntrack enabled
  - disable conntrack
- **TCP SYN Sent**
  - conntrack enabled
  - disable conntrack
- **TCP SYN_SYN_ACK Received**
  - conntrack enabled
  - disable conntrack
- **TCP retries**
  - conntrack enabled
  - disable conntrack
- **TCP timeouts**
  - conntrack enabled
  - disable conntrack

**TCP Connect time (us)**
- **conn time(max)**
  - conntrack enabled
  - disable conntrack
- **conn time(avg)**
  - conntrack enabled
  - disable conntrack

**TCP Time to First Byte(us)**
- **TTFB(max)**
  - conntrack enabled
  - disable conntrack
- **TTFB(avg)**
  - conntrack enabled
  - disable conntrack

**TCP Time to Last Byte(us)**
- **TTLB(max)**
  - conntrack enabled
  - disable conntrack
- **TTLB(avg)**
  - conntrack enabled
  - disable conntrack

*The Higher the better*

*The Lower the better*
Analyzing Bottlenecks – Throughput test

- **μOp Issue**
  - Yes
  - **μOp Retired**
    - Yes
    - Retiring
  - No
    - **Bad Speculation**
      - No
      - BackEnd Bound
      - Yes
      - FrontEnd Bound

### Acceptable Values for HPC applications

<table>
<thead>
<tr>
<th>Category</th>
<th>The Higher the better</th>
<th>The Lower the better</th>
<th>OvS-DPDK contrack (http client/server throughput test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front-End Bound</td>
<td>5 – 10%</td>
<td></td>
<td>15.2%</td>
</tr>
<tr>
<td>Bad Speculation</td>
<td>1 – 5%</td>
<td></td>
<td>2.8%</td>
</tr>
<tr>
<td>Back-End Bound</td>
<td>20 – 40%</td>
<td></td>
<td>35.9%</td>
</tr>
<tr>
<td>Retiring</td>
<td>30 – 70%</td>
<td></td>
<td>46.1%</td>
</tr>
</tbody>
</table>
**Elapsed Time**: 120.011s

- **CPU Time**: 122.500s
  - Total Thread Count: 14
  - Paused Time: 0s

**Top Hotspots**

This section lists the most active functions in your application.

<table>
<thead>
<tr>
<th>Function</th>
<th>Module</th>
<th>CPU Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>csum_continue</td>
<td>ovs-vswitchd</td>
<td>42.831s</td>
</tr>
<tr>
<td>r40e_rev_pkts_vde</td>
<td>ovs-vswitchd</td>
<td>5.530s</td>
</tr>
<tr>
<td>clock_gettime</td>
<td>lib.so.6</td>
<td>3.500s</td>
</tr>
<tr>
<td>miniflow_extract</td>
<td>ovs-vswitchd</td>
<td>3.350s</td>
</tr>
<tr>
<td>tcp_conn_update</td>
<td>ovs-vswitchd</td>
<td>2.550s</td>
</tr>
</tbody>
</table>

**Effective Time by**

- **clone**
- **start_thread**
- **ovsthread_wrapper**
- **pmd_thread_main**
- **dp_netdev_process_rxq_port**
- **dp_netdev_input**
- **dp_netdev_input**
- **packet_batch_per_flow_execute**
- **dp_netdev_execute_actions**
- **odp_execute_actions**
- **dp_execute_cq**
- **conntrack_execute**
- **conn_key_extract**
- **extract_l4**
- **check_l4_tcp**
- **checksum_valid**

**Function Call Stack**

<table>
<thead>
<tr>
<th>Function</th>
<th>CPI Rate</th>
<th>Front-End Bound</th>
<th>Bad Speculation</th>
<th>Back-End Bound</th>
<th>Memory Bound</th>
<th>Core Bound</th>
<th>Retiring</th>
</tr>
</thead>
<tbody>
<tr>
<td>csum_continue</td>
<td>0.321%</td>
<td>8.8%</td>
<td>1.0%</td>
<td>42.3%</td>
<td>13.2%</td>
<td>0%</td>
<td>99.1%</td>
</tr>
<tr>
<td>miniflow_extract</td>
<td>0.312%</td>
<td>13.8%</td>
<td>0%</td>
<td>54.2%</td>
<td>0%</td>
<td>0%</td>
<td>99.5%</td>
</tr>
<tr>
<td>egress_lookup</td>
<td>0.702%</td>
<td>12.1%</td>
<td>13.8%</td>
<td>54.2%</td>
<td>0%</td>
<td>0%</td>
<td>99.3%</td>
</tr>
<tr>
<td>tcp_conn_update</td>
<td>0.360%</td>
<td>15.9%</td>
<td>5.2%</td>
<td>54.6%</td>
<td>0%</td>
<td>0%</td>
<td>99.4%</td>
</tr>
<tr>
<td>etcpovs_plts_vde</td>
<td>0.848%</td>
<td>6.6%</td>
<td>5.6%</td>
<td>55.3%</td>
<td>0%</td>
<td>0%</td>
<td>99.4%</td>
</tr>
</tbody>
</table>
HTTP client/server – 65k connections/sec

TCP Connections

- TCP Connections Established
- TCP SYN Sent
- TCP SYN_RECEIVED
- TCP retries
- TCP timeouts

The Higher the better

The Lower the better

TCP connect time(us)

- conn time(max)
- conn time(avg)

The Lower the better

TCP Time to First Byte(us)

- TTFB(max)
- TTFB(avg)

The Lower the better

TCP Time to Last Byte(us)

- TTLB(max)
- TTLB(avg)

The Lower the better
Datapath classifier

Exact match Cache (EMC)

miss

miss

Execute Action

Rx

Tx

Ofproto classifier

main thread:
  emc entries: 0 (0.00% full)
  emc hits: 0
  megaflow hits: 0
  avg. subtable lookups per hit: 0.00
  miss: 0
  lost: 0
  polling cycles: 572403 (100.00%)
  processing cycles: 0 (0.00%)

thread numa id 0 core id 3:
  emc entries: 8192 (100.00% full)
  emc hits: 173
  megaflow hits: 4960084
  avg. subtable lookups per hit: 1.57
  miss: 0
  lost: 0
  polling cycles: 3884380448 (39.72%)
  processing cycles: 5895049012 (60.28%)
  avg cycles per packet: 766.74 (9779429460/12754628)
  avg processing cycles per packet: 462.19 (5895049012/12754628)
HTTP client/server – 800k concurrent connections

**TCP connections**
- TCP Connections Established
- TCP SYN Sent
- TCP SYN_SYN_ACK Received
- TCP retries
- TCP timeouts

- **The Higher the better**
- **conntrack enabled**
- **disable conntrack**

**TCP connect time (us)**
- conn time (max)
- conn time (avg)

- **The Lower the better**
- **conntrack enabled**
- **disable conntrack**

**TCP Time to First Byte (us)**
- TTFB (max)
- TTFB (avg)

- **The Lower the better**
- **conntrack enabled**
- **disable conntrack**

**TCP Time to Last Byte (us)**
- TTLB (max)
- TTLB (avg)

- **The Lower the better**
- **conntrack enabled**
- **disable conntrack**
OVS-DPDK in real world telco deployments

- Not to showcase OVS-DPDK performance figures
- OVS-DPDK for Telco use cases.
- Debugging and performance fine-tuning of OVS-DPDK
- OVS-DPDK hotspot analysis
OvS-DPDK Troubleshooting
East-West traffic test
Debugging

```
sugeshch@silpixa00389816:/repo/ovs_master$ sudo ./utilities/ovs-ofctl dump-ports-desc br0 | more

OFPST_PORT_DESC reply (xid=0x2):
1(dpdk0): addr:90:e2:be:2b:49:40
   config: 0
   state: 0
   current: 10GB-FD
   speed: 10000 Mbps now, 0 Mbps max
   config: 0
   state: 0
   current: 10GB-FD
   speed: 10000 Mbps now, 0 Mbps max
3(dpdkvhostuser0): addr:00:00:00:00:00:00
   config: 0
   state: 0
   speed: 0 Mbps now, 0 Mbps max
4(dpdkvhostuser1): addr:00:00:00:00:00:00
   config: 0
   state: 0
   speed: 0 Mbps now, 0 Mbps max
5(dpdkvhostuser2): addr:00:00:00:00:00:00
   config: 0
   state: 0

--More--
```
Debugging
Debugging

List all openflow flow entries in switch tables.

```
sugeshch@silpixa00389816:~/.repo/ovs_master$ sudo ./utilities/ovs-ofctl dump-flows br0
NXST_FLOW reply (xid=0x4):
  cookie=0x0, duration=177.460s, table=0, n_packets=0, n_bytes=0, idle_age=177, in_port=1,dl_src=00:00:00:00:00:01 actions=output:3
  cookie=0x0, duration=177.412s, table=0, n_packets=38, n_bytes=7008, idle_age=12, in_port=3 actions=output:1
  cookie=0x0, duration=177.360s, table=0, n_packets=0, n_bytes=0, idle_age=177, in_port=4 actions=output:5
  cookie=0x0, duration=177.320s, table=0, n_packets=41, n_bytes=7659, idle_age=3, in_port=5 actions=output:4
  cookie=0x0, duration=177.272s, table=0, n_packets=0, n_bytes=0, idle_age=177, in_port=6 actions=output:7
  cookie=0x0, duration=177.228s, table=0, n_packets=40, n_bytes=7152, idle_age=0, in_port=7 actions=output:6
  cookie=0x0, duration=177.768s, table=0, n_packets=0, n_bytes=0, idle_age=177, in_port=8 actions=output:9
  cookie=0x0, duration=177.136s, table=0, n_packets=37, n_bytes=6394, idle_age=14, in_port=9 actions=output:8
  cookie=0x0, duration=177.068s, table=0, n_packets=0, n_bytes=0, idle_age=177, in_port=10 actions=output:2
  cookie=0x0, duration=177.024s, table=0, n_packets=0, n_bytes=0, idle_age=177, in_port=2 actions=output:10
sugeshch@silpixa00389816:~/.repo/ovs_master$
```
Debugging

List all openflow flow entries in switch tables.
Debugging

```
sugeshch@silpixa00389816:~/repo/ovs_master$ sudo ./utilities/ovs-ofctl dump-flows br0
NXST_FLOW reply (xid=0x4):
  cookie=0x0, duration=177.460s, table=0, n_packets=0, n_bytes=0, idle_age=177, in_port=1, dl_src=00:00:00:ff, dl_dst=00:00:00:00:00:01 actions=output:3
  cookie=0x0, duration=177.412s, table=0, n_packets=38, n_bytes=7008, idle_age=12, in_port=3 actions=output:1
  cookie=0x0, duration=177.360s, table=0, n_packets=0, n_bytes=0, idle_age=177, in_port=4 actions=output:5
  cookie=0x0, duration=177.320s, table=0, n_packets=41, n_bytes=7659, idle_age=3, in_port=5 actions=output:4
  cookie=0x0, duration=177.272s, table=0, n_packets=0, n_bytes=0, idle_age=177, in_port=6 actions=output:7
  cookie=0x0, duration=177.228s, table=0, n_packets=40, n_bytes=7152, idle_age=0, in_port=7 actions=output:6
  cookie=0x0, duration=177.768s, table=0, n_packets=0, n_bytes=0, idle_age=177, in_port=8 actions=output:8
  cookie=0x0, duration=177.136s, table=0, n_packets=37, n_bytes=6394, idle_age=14, in_port=9 actions=output:2
  cookie=0x0, duration=177.068s, table=0, n_packets=0, n_bytes=0, idle_age=177, in_port=10 actions=output:10
sugeshch@silpixa00389816:~/repo/ovs_master$
```

List all openflow flow entries in switch tables.
Datapath Debugging
Debugging
Debugging

```
netdev@ovs-netdev: hit:17543389635 missed:188
br0: 65534/1: (tap)
dpdk0 1/2: (apdk: configured rx queues=1, configured rxq descriptors=2048, configured tx queues=5, configured txq descriptors=2048, mtu=1500, requested rx queues=1, requested rxq descriptors=2048, requested tx queues=5, requested txq descriptors=2048, rx csum offload=true)
dpdk1 2/3: (dpdk: configured rx queues=1, configured rxq descriptors=2048, configured tx queues=5, configured txq descriptors=2048, mtu=1500, requested rx queues=1, requested rxq descriptors=2048, requested tx queues=5, requested txq descriptors=2048, rx_csum_offload=true)
dpdkvhostuser0 3/4: (dpdkvhostuser: configured rx queues=1, configured tx queues=1, mtu=1500, requested rx queues=1, requested tx queues=1)
dpdkvhostuser1 4/5: (dpdkvhostuser: configured rx queues=1, configured tx queues=1, mtu=1500, requested rx queues=1, requested tx queues=1)
dpdkvhostuser2 5/6: (dpdkvhostuser: configured rx queues=1, configured tx queues=1, mtu=1500, requested rx queues=1, requested tx queues=1)
dpdkvhostuser3 6/7: (dpdkvhostuser: configured rx queues=1, configured tx queues=1, mtu=1500, requested rx queues=1, requested tx queues=1)
dpdkvhostuser4 7/8: (dpdkvhostuser: configured rx queues=1, configured tx queues=1, mtu=1500, requested rx queues=1, requested tx queues=1)
```

---More---

Summary of datapath with configured port information (Openflow port no, Datapath port no and Type)
Summary of datapath with list of data-path-ports. --s shows the stats
Summary of datapath with list of data-path-ports. --s shows the stats

<table>
<thead>
<tr>
<th>Port</th>
<th>Description</th>
<th>RX Packets</th>
<th>RX Errors</th>
<th>RX Dropped</th>
<th>RX Overruns</th>
<th>RX Frame</th>
<th>TX Packets</th>
<th>TX Errors</th>
<th>TX Dropped</th>
<th>TX Aborted</th>
<th>TX Carrier</th>
<th>RX Bytes</th>
<th>TX Bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>ovs-netdev (tap)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>br0 (tap)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>dpdk0 (dpdk)</td>
<td>13798789152</td>
<td>0</td>
<td>1803077122</td>
<td>0</td>
<td>0</td>
<td>107</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Debugging

`ovs-appctl dpctl/dump-flows`

Display all active flow entries in a datapath
Debugging

`ovs-appctl dpctl/dump-flows`

Display all active flow entries in a datapath
Debugging

List all openflow flow entries in switch tables.
Debugging

• Flow configured wrongly.
  • set “dl_src = 00:00:00:00:00:0F”

```
    ovs-ofctl add-flow br0 in_port=1,dl_src=00:00:00:00:00:0F,dl_dst=00:00:00:00:00:01,action=output:3
```
ovs-appctl dpctl/dump-flows

Display all active flow entries in a datapath.
Debugging OVS-DPDK performance issues
OVS-DPDK Table Hierarchy

- **Exact Match Cache**
  - Logically, Single Table per datapath thread
  - Exact Match
  - 8192 entries / per thread

- **Datapath Classifiers**
  - Logically, Single Table per port
  - Wildcard Matches
  - 65536 entries
  - Each table is implemented as a priority list of subtables in order to implement wildcards

- **Ofproto Classifier**
  - Logically, Multiple (up to 255) Open Flow tables in pipeline per Open vSwitch bridge
  - Wildcard Matches
  - Each table is implemented as a priority list of subtables in order to implement wildcards

**rx**

**tx**

**EXECUTE ACTION**

**Datapath**
Debugging

Show performance stats for each pmd thread of the datapath.
Debugging

Description:

```
sugeshch@silpixa00389816:~/.repo/ovs_master/utilities$ sudo ./ovs-appctl dpif-netdev/pmd-stats-show | more
pmd thread numa id 0 core id 3:
- emc hits:819011745
- megaflow hits:329984399
  avg. subtable lookups per hit:1.00
  miss:48
  lost:0
  polling cycles:19459002510676 (97.46%)
  processing cycles:506349652804 (2.54%)
  avg cycles per packet: 13499.33 (19965252163480/1478980596)
  avg processing cycles per packet: 342.36 (506349652804/1478980596)
```

```
pmd thread numa id 0 core id 2:
- emc hits:62786086187
- megaflow hits:240496054
  avg. subtable lookups per hit:1.00
  miss:5
  lost:0
  polling cycles:1061900555584 (53.82%)
  processing cycles:9112873249736 (46.18%)
  avg cycles per packet: 311.88 (19731878805320/63267078300)
  avg processing cycles per packet: 144.04 (9112873249736/63267078300)
```

```
main thread:
- emc hits:0
--More--
```

**Summary:**

The command `ovs-appctl dpif-netdev/pmd-stats-show` shows performance stats for each pmd thread of the datapath.
Debugging

```
sugeshch@silpixa00389816:~-repo/ovs_master/utilities$ sudo ./ovs-appctl dpif-netdev/pmd-stats-show | more
pmd thread numa_id 0 core_id 3:
  emc hits:819011745
  megaflow hits:329984399
  avg. subtable lookups per hit:1.00
  miss:48
  lost:0
  polling cycles:19453002510576 (97.46%)
  processing cycles:506349652804 (2.54%)
  avg cycles per packet: 13499.33 (19965252163480/1478980596)
  avg processing cycles per packet: 342.36 (506349652804/1478980596)

pmd thread numa_id 0 core_id 2:
  emc hits:62786086187
  megaflow hits:240496054
  avg. subtable lookups per hit:1.00
  miss:5
  lost:0
  polling cycles:10619005555584 (53.82%)
  processing cycles:9112873249736 (46.18%)
  avg cycles per packet: 311.88 (19731878805320/63267078300)
  avg processing cycles per packet: 144.04 (9112873249736/63267078300)

main thread:
  emc hits:0
```

Shows performance stats for each pmd thread of the datapath.
Shows performance stats for each pmd thread of the datapath.
Better Performance

```
sugeshch@silpixa00389816:~/repo/ovs_master/utilities$ sudo ./ovs-appctl dpif-netdev/pmd-stats-show | more
pmd thread numa_id 0 core_id 3:
  emc hits:0
  megaflow hits:0
  avg. subtable lookups per hit:0.00
  miss:0
  lost:0
  polling cycles:72481927540 (100.00%)
  processing cycles:0 (0.00%)
pmd thread numa_id 0 core_id 2:
  emc hits:35667360
  megaflow hits:3825812
  avg. subtable lookups per hit:1.00
  miss:1
  lost:0
  polling cycles:58549367676 (80.61%)
  processing cycles:14085077144 (19.39%)
  avg cycles per packet: 1676.73 (72634444820/43318985)
  avg processing cycles per packet: 325.15 (14085077144/43318985)
main thread:
  emc hits:0
  megaflow hits:0
  avg. subtable lookups per hit:0.00
  miss:0
  lost:0
  polling cycles:742880 (100.00%)
```
Debugging

```
sugeshch@silpixa00389816:~/.repo/ova_master/utilities$ sudo ./ovs-appctl dpif-netdev/pmd-rxq-show
pmd thread numa_id 0 core_id 3:
  isolated : true
  port: dpdkvhostuser0 queue-id: 0
pmd thread numa_id 0 core_id 2:
  isolated : true
  port: dpdkvhostuser5 queue-id: 0
pmd thread numa_id 0 core_id 1:
  isolated : false
  port: dpdkvhostuser1 queue-id: 0
  port: dpdkvhostuser2 queue-id: 0
  port: dpdkvhostuser3 queue-id: 0
  port: dpdkvhostuser4 queue-id: 0
pmd thread numa_id 0 core_id 0:
  isolated : false
  port: dpdk0 queue-id: 0
  port: dpdk1 queue-id: 0
  port: dpdkvhostuser6 queue-id: 0
  port: dpdkvhostuser7 queue-id: 0
sugeshch@silpixa00389816:~/.repo/ova_master/utilities$
```

Shows dp ports/rxq assignment to PMD threads
Debugging

Affinitizing datapath Rx queues for uniform load distribution.

```
sugeshch@silpixa00389816:~/repo/ovs_master/utilities$ sudo ./ovs-appctl dpif-netdev/pmd-rxq-show
pmd thread numa_id 0 core_id 3:
  isolated : true
  port: dpdkvhostuser0 queue-id: 0
pmd thread numa_id 0 core_id 2:
  isolated : true
  port: dpdkvhostuser5 queue-id: 0
pmd thread numa_id 0 core_id 1:
  isolated : false
  port: dpdkvhostuser1 queue-id: 0
  port: dpdkvhostuser2 queue-id: 0
  port: dpdkvhostuser3 queue-id: 0
  port: dpdkvhostuser4 queue-id: 0
pmd thread numa_id 0 core_id 0:
  isolated : false
  port: dpdk0 queue-id: 0
  port: dpdk1 queue-id: 0
  port: dpdkvhostuser5 queue-id: 0
  port: dpdkvhostuser7 queue-id: 0
sugeshch@silpixa00389810:~/repo/ovs_master/utilities$ ovs-vsctl set interface dpdkvhostuser2 other_config:pmd-rxq-affinity="0:2"
```
Debugging

**ovs-appctl dpif-netdev/pmd-rxq-show**

```bash
sugeshch@silpixa00389816:~/repo/ovs_master/utilities$ sudo ./ovs-appctl dpif-netdev/pmd-rxq-show
pmd thread numa_id 0 core_id 3:
  isolated : true
  port: dpdkvhostuser0 queue-id: 0
  port: dpdkvhostuser3 queue-id: 0
  port: dpdkvhostuser6 queue-id: 0
pmd thread numa_id 0 core_id 2:
  isolated : true
  port: dpdkvhostuser4 queue-id: 0
  port: dpdkvhostuser5 queue-id: 0
  port: dpdkvhostuser7 queue-id: 0
pmd thread numa_id 0 core_id 1:
  isolated : false
  port: dpdk0 queue-id: 0
  port: dpdkvhostuser2 queue-id: 0
pmd thread numa_id 0 core_id 0:
  isolated : false
  port: dpdk1 queue-id: 0
  port: dpdkvhostuser1 queue-id: 0
```

Shows dp ports/rxq assignment to PMD threads
OVS-DPDK Hotspots
Cost of EMC

**EMC Thrashing**
- Logically, Single Table per datapath port
- Wildcard Matches
- 8192 entries / per thread

**Datapath Classifier**
- Logically, Single Table per datapath port
- Wildcard Matches
- 65536 entries

**Ofproto Classifier**
- Logically, Multiple (up to 255) Open Flow tables in pipeline per Open vSwitch bridge
- Wildcard Matches

Cost of lookup increasing
Classifier Cost

- **Exact Match Cache**
  - Logically Single Table per Open vSwitch
  - Exact Match
  - 8192 entries / per thread

- **Datapath Classifier**
  - **Megaflow Cost**
    - Wildcard Matches
    - 65536 entries

- **Ofproto Classifier**
  - Logically, Multiple (up to 255) Open Flow tables in pipeline per Open vSwitch bridge
  - Wildcard Matches

Cost of lookup increasing
Slow Vhost ports

- **rx**
  - **Slow Vhost**
  - **virtutal**
  - **physical**

**Execute Action**

- **tx**

**EMC Thrashing**
- Logically, Single Table per port
- Exact Match
- 8192 entries / per thread

**Exact Match Cache**

**Datapath Classifier**
- **Megaflow Cost**
- Wildcard Matches
- 65536 entries

**Ofproto Classifier**
- Logically, Multiple (up to 255) Open Flow tables in pipeline per Open vSwitch bridge
- Wildcard Matches

- **Cost of lookup increasing**

**Intel OvS Open vSwitch**
Packet recirculation

- **Exact Match Cache**
  - Logically Single Table per OVS Kernel
  - Exact Match
  - 8192 entries / per thread

- **Datapath Classifier**
  - **Megaflow Cost**
    - Wildcard Matches
    - 65536 entries

- **Ofproto Classifier**
  - Logically, Multiple (up to 255) Open Flow tables in pipeline per Open vSwitch bridge
  - Wildcard Matches

**Slow Vhost**

**Recirculation**

**EXECUTE ACTION**

**Cost of lookup increasing**
Summary

• OVS-DPDK can offer SR-IOV comparable performance

• OVS-DPDK offers
  • Fully fledged vswitch features.
  • Good Performance
  • Production Quality

• Understanding OVS commands helps to get better performance.

• Latest OVS-DPDK always offers better performance.
Legal Disclaimers

- No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

- Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

- This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest forecast, schedule, specifications and roadmaps.

- The products and services described may contain defects or errors known as errata which may cause deviations from published specifications. Current characterized errata are available on request.

- Copies of documents which have an order number and are referenced in this document may be obtained by calling 1-800-548-4725 or by visiting www.intel.com/design/literature.htm.

- Intel, the Intel logo, Xeon, VTune are trademarks of Intel Corporation in the U.S. and/or other countries.

- *Other names and brands may be claimed as the property of others

- Copyright © 2017 Intel Corporation. All rights reserved.