DPDK PMD for AF_XDP

Zhang Qi, Li Xiaoyun
Intel
DPDK UserSpace - Dublin 2018
Agenda

- Introduction to AF_XDP
- DPDK PMD for AF_XDP
- Future Work
Introduction to AF_XDP

- An upgraded version of AF_PACKET
  - Use XDP program to trigger Rx path for selected queue
- XDP programs can redirect frames to a memory buffer in user-space by eBPF.
- DMA transfers use user space memory (Zero Copy)
Introduction to AF_XDP -- Benefits

- Performance improvement
  - Zero copy between user space and kernel space
  - Achieve 3-20X times improvement comparing to AF_PACKET
- Connect the XDP pass-through to user-space directly
  - An eBPF program that processes packets can be forwarded to an application in a very efficient way
- For DPDK
  - No change to DPDK apps, kernel driver handles hardware
  - Provide a new option for users
**Introduction to AF_XDP**

- **AF_XDP socket**

```c
sfd = socket(PF_XDP, SOCK_RAW, 0);
buffs = calloc(num_buffs, FRAME_SIZE);
setsockopt(sfd, SOL_XDP, XDP_MEM_REG, buffs);
setsockopt(sfd, SOL_XDP, XDP_{RX|TX|FILL|COMPLETE}_RING, ring_size);
mmap(..., sfd, .......); /* map kernel rings */
bind(sfd, "/dev/eth0", queue_id,....);
for (;;) {
    read_process_send_messages(sfd);
};
```
Introduction to AF_XDP

• Datapath
DPDK PMD for AF_XDP

- Enable AF_XDP socket

<table>
<thead>
<tr>
<th>User Space</th>
<th>NETDEV queue 0</th>
<th>NETDEV queue 1</th>
<th>NETDEV eth0</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPDK q0</td>
<td>XSK A</td>
<td>XSK B</td>
<td>XSK C</td>
</tr>
<tr>
<td>DPDK q1</td>
<td>XSK D</td>
<td>XSK E</td>
<td>XSK A</td>
</tr>
<tr>
<td>DPDK q0</td>
<td>XSK D</td>
<td>XSK E</td>
<td>XSK A</td>
</tr>
<tr>
<td>DPDK q0</td>
<td>XSK D</td>
<td>XSK E</td>
<td>XSK A</td>
</tr>
<tr>
<td>DPDK q1</td>
<td>XSK D</td>
<td>XSK E</td>
<td>XSK A</td>
</tr>
</tbody>
</table>

- eBPF

.NETDEV eth0

<table>
<thead>
<tr>
<th>ID</th>
<th>KEY</th>
<th>XSK</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>XSK A</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>XSK B</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>XSK C</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>XSK D</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>XSK E</td>
</tr>
</tbody>
</table>
DPDK PMD for AF_XDP

- Load eBPF program
- Link ebpf program to specific network interface
- Expose the xdp socket redirect map id and entries number to user

```bash
./build/app/testpmd -c 0xc -n 4 --vdev eth_af_xdp,
iface=en0,queue=1,xsk_map_id=56,xsk_map_key_start=3,
xsk_map_key_count=2 --i --rxq=2 --txq=2
```
DPDK PMD for AF_XDP

• Mapping between rte_mempool and xsk umem
DPDK PMD for AF_XDP

AF_XDP Experimental Setup

- V2 patch set of AF_XDP published on April 27, 2018 + ZC patches in house
  - Very few optimizations performed
- Broadwell E5 2660 @ 2.0 GHz
- IXIA load generator blasting at full 40 Gbit/s
- Intel XL710 card (40G, i40e driver)

http://lists.openwall.net/netdev/2018/08/28/62
AF_XDP performance

- PMD performance is aligned with this

http://lists.openwall.net/netdev/2018/08/28/62
Future Work

• Refine zero copy enabling in PMD
• Usertools for AF_XDP in DPDK
• Follow up kernel AF_XDP’s changes
• Performance optimization
Q & A


Zhang Qi
qi.z.zhang@intel.com

Li Xiaoyun
xiaoyun.li@intel.com