Troubleshoot "KNI: Out of Memory" Errors on QvPC-DI Platforms

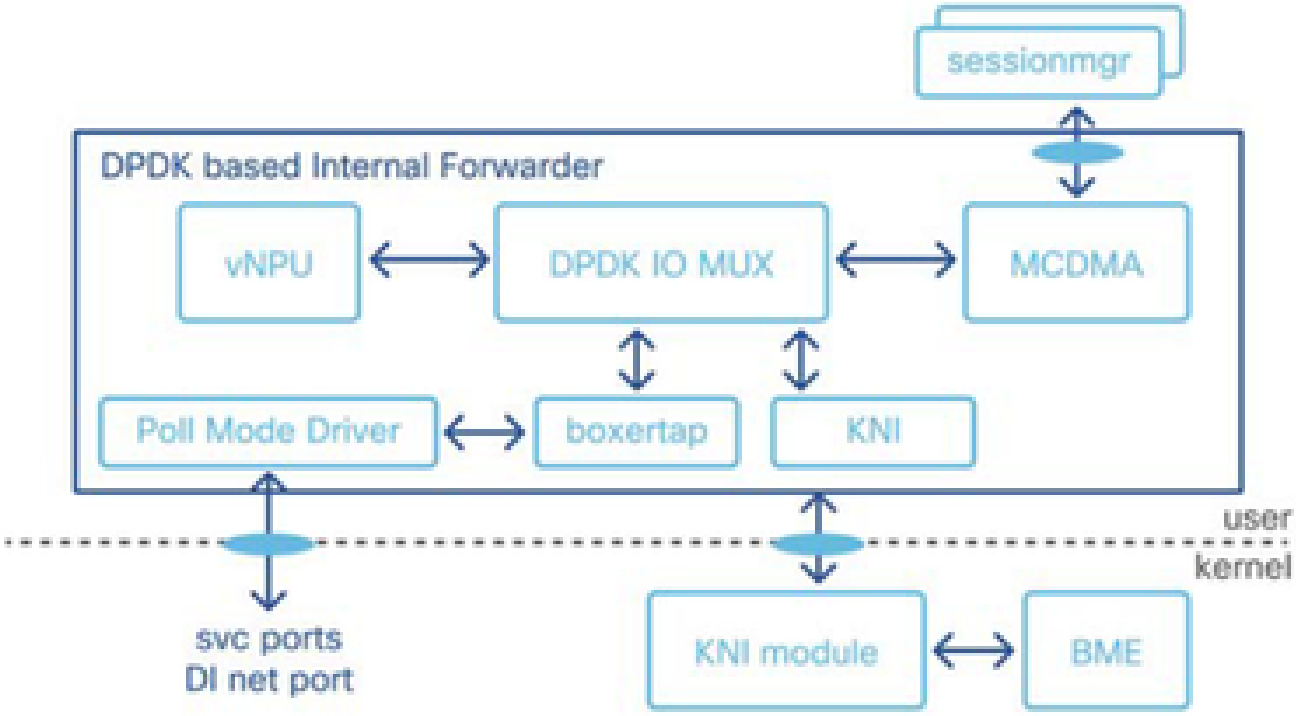
Contents

Introduction

This document describes how to determine if **StarOS KNI: Out of Memory** logs are caused by issues in the StarOS application or by hardware drivers.

Background

The Kernel Network Interface (KNI) module, within the DPDK Internal Forwarder (IFTASK) process, is a mechanism that allows user-space programs to receive packets directly from a network interface, bypassing the Linux networking and Linux IP stack completely.



KNI: Out of Memory logs rate-limiting warnings are produced when there is a resource contention issue affecting the KNI Module.

1. Memory buffers are not cleared at the bare-metal (hardware) level, causing an overrun of the buffer.
2. The KNI pools, from which the **iftask** allocates the message buffer for these packets, run out of space.
3. The virtual function queries for more packets, but the physical function responds that it does not have anything.
4. Once the **KNI: Out of Memory** condition occurs, the **iftask** goes into backup memory pool to allocate and process the packet further. If the backup pool also runs out of memory, the system drops the packets.
5. Because the **iftask** cannot read the burst of packets coming from the kernel, the **KNI: Out of Memory** log is produced on the StarOS.

Triggers for **KNI: Out of Memory** condition:

Potential triggers for the buffer overflow condition can vary, such as running SFTP or SCP applications or a very large file transfer between CF and SF cards.

Steps to Investigate

Step 1. Observe the Symptoms

Step 2. Check for DI-Network Health Degradation

Step 3. Check for Userspace KNI Drops

Step 4. Check the Hardware Drivers

Step 1. Observe the Symptoms

Correlate the timing of **KNI: Out of Memory** errors with other symptoms, such as packet losses or application layer degradations (**egtpc** path failures).

KNI: Out of Memory logs

- In the StarOS Syslogs, you can see logs lindicating that the kernal network interface is out of memory.

```
2023-Nov-16+09:18:03.205 [iftask 214701 error] [1/0/9602 <evlogd:0> evlgd_syslogd.c:236] [software inte
```

- If the backup memory is exhausted, you can see error messages indicating that the backup pool's memory is also exhausted.

```
RTE_LOG(ERR, KNI, "Out of memory from Backup pool, kni port %s, socket_id=%d, total=%d, iter=%d\n", kni
```

- In the IFTask logs, found in the tmp directory in the debug shell, you can observe the **KNI: Out of Memory** errors:

```
Wed Nov 15 17:20:30 2023 PID:7387 KNI: Out of memory, kni port cpbond0, socket_id=0, total=-759247296,
```

EGTPC path failures

- Spikes in gtpc path failures to various peers can occur with the cause **No response from peer** can occur during the time of the packet losses.

```
2023-10-23T00:14:33.813+00:00 Nodename evlogd: [local-60sec33.780] [egtpmgr 143137 info] [6/0/12364 <eg
```

Step 2. Check for DI-Network Health Degradation

Locate which connections are having the degradation. When seen on a sustained basis, higher drop or loss percentages in DI-network health outputs can indicate DI-network configuration or operational issues, traffic overload, or VM or Host issues.

show session recovery status verbose

- Use **show session recover status verbose** outputs to identify which virtual function card is serving as the Demux card.

```
******** show session recovery status verbose *******
Tuesday October 24 11:23:45 EDT 2023
Session Recovery Status:
  Overall Status        : Ready For Recovery
  Last Status Update    : 1 second ago

              ----sessmgr---  ----aaamgr----  demux
 cpu state    active standby  active standby  active  status
---- -------  ------ -------  ------ -------  ------  -------------------------
 3/0 Active   24     1        24     1        0       Good
 4/0 Active   24     1        24     1        0       Good
 5/0 Active   24     1        24     1        0       Good
 6/0 Active   0      0        0      0        10      Good (Demux)
 7/0 Active   24     1        24     1        0       Good
 8/0 Active   24     1        24     1        0       Good
 9/0 Active   24     1        24     1        0       Good
10/0 Active   24     1        24     1        0       Good
11/0 Active   24     1        24     1        0       Good
12/0 Standby  0      24       0      24       0       Good
```

show cloud monitor di-network detail

- Use "show cloud monitor di-network detail" outputs to identify which DI-network connections between virtual function cards have drops in heartbeats.

- Drops in heartbeats from CF and SF cards to SF Card 6 are shown. Outputs for CF and SF cards to other CF and SF cards show no heartbeat drops.

```
******** show cloud monitor di-network detail *******
Tuesday October 24 11:23:51 EDT 2023
Card 1 Heartbeat Results:
ToCard   Health     5Min-Loss     60Min-Loss
------   -------    ---------     ----------
…
   6      Good        0.00%         0.66%
…
Card 2 Heartbeat Results:
…
   6       Bad       14.67%         3.50%
…
Card 3 Heartbeat Results:
…
   6       Bad        5.35%         2.69%
…
Card 4 Heartbeat Results:
…
   6      Good        0.00%         0.00%
…
Card 5 Heartbeat Results:
…
   6       Bad       18.57%         3.90%
…
Card 6 Heartbeat Results:
…
   1      Good        0.00%         0.90%
   2       Bad       12.63%         3.31%
```

```
   3       Bad         2.90%         2.14%
   4      Good         0.00%         0.00%
   5       Bad        13.09%         3.30%
   7      Good         0.00%         0.00%
   8       Bad         2.91%         2.20%
   9      Good         0.00%         0.93%
  10       Bad        14.28%         3.38%
  11       Bad         3.67%         2.09%
  12      Good         0.00%         0.00%
…
Card 7 Heartbeat Results:
…
   6      Good         0.00%         0.00%
…
Card 8 Heartbeat Results:
…
   6       Bad         7.47%         2.85%
…
Card 9 Heartbeat Results:
…
   6       Bad         0.00%         1.07%
…
Card 10 Heartbeat Results:
…
   6       Bad        16.01%         3.73%
…
Card 11 Heartbeat Results:
…
   6       Bad         7.47%         2.71%
…
Card 12 Heartbeat Results:
…
   6      Good         0.00%         0.00%
```

show cloud monitor controlplane

- Use **show cloud monitor contolplane** outputs to identify which DI-network connections have degradation.

```
******** show cloud monitor controlplane *******
Tuesday October 24 11:24:22 EDT 2023

  Cards      15 Second Interval     5 Minute Interval    60 Minute Interval
 Src  Dst    Xmit   Recv  Miss%    Xmit   Recv  Miss%    Xmit   Recv  Miss%
 ---  ---  ------ ------ ------  ------ ------ ------  ------ ------ ------
…
  01   06      75     75   0.0%    1500   1500   0.0%   18000  17842   0.9%
…
  02   06      75     75   0.0%    1500   1265  15.7%   18000  17546   2.5%
…
  03   06      75     75   0.0%    1500   1396   6.9%   18000  17491   2.8%
…
  04   06      75     75   0.0%    1500   1500   0.0%   18000  18000   0.0%
…
  05   06      75     75   0.0%    1500   1267  15.5%   18000  17325   3.8%
…
  06   01      75     75   0.0%    1500   1500   0.0%   18000  17823   1.0%
```

```
  06   02       75      75   0.0%    1500   1301  13.3%   18000  17567   2.4%
  06   03       75      75   0.0%    1500   1419   5.4%   18000  17561   2.4%
  06   04       75      75   0.0%    1500   1500   0.0%   18000  18000   0.0%
  06   05       75      75   0.0%    1500   1294  13.7%   18000  17579   2.3%
  06   07       75      75   0.0%    1500   1500   0.0%   18000  18000   0.0%
  06   08       75      75   0.0%    1500   1417   5.5%   18000  17565   2.4%
  06   09       75      75   0.0%    1500   1500   0.0%   18000  17824   1.0%
  06   10       75      75   0.0%    1500   1296  13.6%   18000  17573   2.4%
  06   11       75      75   0.0%    1500   1422   5.2%   18000  17570   2.4%
  06   12       75      75   0.0%    1500   1500   0.0%   18000  18000   0.0%
…
  07   06       75      75   0.0%    1500   1500   0.0%   18000  18000   0.0%
…
  08   06       75      75   0.0%    1500   1426   4.9%   18000  17545   2.5%
…
  09   06       75      75   0.0%    1500   1500   0.0%   18000  17833   0.9%
…
  10   06       75      75   0.0%    1500   1278  14.8%   18000  17369   3.5%
…
  11   06       75      75   0.0%    1500   1408   6.1%   18000  17481   2.9%
…
  12   06       75      75   0.0%    1500   1500   0.0%   18000  18000   0.0%
```

show cloud monitor dataplane

- Use **show cloud monitor dataplane** poutputs to identify which DI-network connections have degradations and to identify any one-way degradations between virtual function cards.

```
******** show cloud monitor dataplane *******
Tuesday October 24 11:21:46 EDT 2023

  Cards      15 Second Interval     5 Minute Interval    60 Minute Interval
 Src  Dst    Miss    Hit    Pct    Miss    Hit    Pct    Miss    Hit    Pct
 ---  ---  ------ ------ ------  ------ ------ ------  ------ ------ ------
…
  06   01       0    150   0.0%       0   3000   0.0%       0  36000   0.0%
…
  06   02       0    150   0.0%       0   3000   0.0%       0  36000   0.0%
…
  06   03       0    150   0.0%       0   3000   0.0%       0  36000   0.0%
…
  06   04       0    150   0.0%       0   3000   0.0%       0  36000   0.0%
…
  06   05       1    149   0.7%       0   3001   0.0%       0  36000   0.0%
…
  01   06       0    150   0.0%       0   3000   0.0%       0  36000   0.0%
  02   06       0    150   0.0%     210   2790   7.0%    1015  34985   2.8%
  03   06      31    119  20.7%     540   2460  18.0%     995  35005   2.8%
  04   06      34    116  22.7%     554   2446  18.5%    1017  34983   2.8%
  05   06       0    150   0.0%     213   2787   7.1%     991  35009   2.8%
  07   06       0    150   0.0%       0   3000   0.0%     359  35641   1.0%
  08   06      29    121  19.3%     546   2454  18.2%    1009  34991   2.8%
  09   06       0    150   0.0%       0   3000   0.0%       0  36000   0.0%
  10   06       0    150   0.0%     208   2792   6.9%     992  35008   2.8%
  11   06      31    119  20.7%     548   2452  18.3%     993  35007   2.8%
  12   06      34    116  22.7%     547   2453  18.2%    1001  34999   2.8%
…
```

```
  06   07        0    150   0.0%       0    3000   0.0%       0   36000   0.0%
…
  06   08        0    150   0.0%       0    3000   0.0%       0   36000   0.0%
…
  06   09        0    150   0.0%       0    3000   0.0%       1   35999   0.0%
…
  06   10        0    150   0.0%       0    3000   0.0%       0   36000   0.0%
…
  06   11        0    150   0.0%       0    3000   0.0%       0   36000   0.0%
…
  06   12        0    150   0.0%       0    3000   0.0%       0   36000   0.0%
```

Step 3. Check for Userspace KNI Drops

show iftask stats

- Collect **show iftask stats** outputs multiple times to verify that KNI drops are not incrementing in the **IFTASK** userspace application level (StarOS).

```
******** show iftask stats *******
Tuesday October 24 11:22:06 EDT 2023
…
                          CARD 6 STATS
-------------------------------------------------------------------------------
Counters             SF6                  SF6_PPS
-------------------------------------------------------------------------------
svc_rx                           2587301598                       2203
svc_tx                           548969428                        295
di_rx                            2260147059                       2258
di_tx                            4072038717                       3966
__ALL_DROPS__                    0                                0
svc_tx_drops                     0                                0
di_rx_drops                      0                                0
di_tx_drops                      0                                0
sw_rss_enq_drops                 0                                0
kni_thread_drops                 0                                0
kni_drops                        0                                0
mcdma_drops                      0                                0
mux_deliver_hop_drops            0                                0
mux_deliver_drops                0                                0
mux_xmit_failure_drops           0                                0
mc_dma_thread_enq_drops          0                                0
sw_tx_egress_enq_drops           0                                0
cpeth0_drops                     0                                0
mcdma_summary_drops              0                                0
fragmentation_err                0                                0
reassembly_err                   0                                0
reassembly_ring_enq_err          0                                0
__DISCARDS__                     241984                           0
__BOND_DISCARDS__                55282718                         142
…
                             TOTAL STATS
-------------------------------------------------------------------------------
Counters             TOTAL                TOTAL_PPS
-------------------------------------------------------------------------------
svc_rx                           27964563261                      24791
svc_tx                           36109966153                      30168
```

```
di_rx                          74133486629                    51929
di_tx                          73958155063                    50897
__ALL_DROPS__                  0                              0
svc_tx_drops                   0                              0
di_rx_drops                    0                              0
di_tx_drops                    0                              0
sw_rss_enq_drops               0                              0
kni_thread_drops               0                              0
kni_drops                      0                              0
mcdma_drops                    0                              0
mux_deliver_hop_drops          0                              0
mux_deliver_drops              0                              0
mux_xmit_failure_drops         0                              0
mc_dma_thread_enq_drops        0                              0
sw_tx_egress_enq_drops         0                              0
cpeth0_drops                   0                              0
mcdma_summary_drops            0                              0
fragmentation_err              0                              0
reassembly_err                 0                              0
reassembly_ring_enq_err        0                              0
__DISCARDS__                   2324968                        0
__BOND_DISCARDS__              55635534                       149
-------------------------------------------------------------------------------------------
NDR is      100.0000
CONTINUE_TRAFFIC
-------------------------------------------------------------------------------------------
```

Step 4. Check the Hardware Drivers

With the application layer cleared from culpability, focus on underlying drivers at the hardware level to address the **KNI: Out of Memory** errors.

Because the bare-metal hardware driver allocates a certain amount of buffer for each virtual function, resource contention issues are commonly the result of a driver mismatch or defective drivers at the hardware level. The defective hardware driver that allocated the buffers that were needed for an application did not release the memory.

If third party (non-Cisco) virtualization software and/or hardware are in use, investigate the versions and drivers for potential compatibility mismatch or defect.

Summary

To determine if **KNI: Out of Memory** errors, are caused by application level processes or by underlying hardware drivers, check for evidence of DI-network degradation and userspace KNI drops. If DI-network degradation exist without a corresponding userspace KNI dedgradation, the cause can be concluded to be at the hardware level. **KNI: Out of Memory** errors with hardware level degredation indicate faulty hardware drivers.

An offload of the node and reload of the host computes upon which the affected application-level StarOS virtual function resides can temporarily clear the memory buffers on the underlying compute, resulting in a temporarily reduction in errors and packet losses. However, this is not a permanent solution! Packet losses and **KNI: Out of Memory** errors recur, when the buffer overflow condition recurs on the faulty hardware driver.