PFC

PFC (Priority Flow Control)

Priority Flow Control (PFC), which is defined in IEEE 802.1Qbb, provides hop-by-hop layer 2 congestion control. PFC extends the IEEE 802.3x Ethernet PAUSE, which allows pausing the traffic based on its priority (802.1p) rather than pausing all the traffic on a link. When a congestion happens on the ingress buffer of the downstream switch, it sends the PFC PAUSE frame to the upstream switch. The PFC PAUSE frame contains a 2-octet priority specifying which traffic of the priority class should be paused and a 2-octet time value specifying the pause time. When the upstream switch receives a PFC PAUSE frame, it stops transmitting based on the priorities and the period in the PFC frame. The upstream switch resumes transmitting after the time expires or receives a new PFC PAUSE frame with pause time is zero.

Topology:

Both Packet Generator (Tx) will inject 400G traffic to Packet Generator (Rx), with this situation the packet will be dropped by the Switch due to the traffic congestion.

To avoid AS9736-64D dropping packets, AS9736-64D will enable the PFC to achieve lossless packet forwarding.

Here's the Packet format.

The PCP (Priority Code Point) in the VLAN tag is "3".

Pre-configuration:
Add VLAN 10 to Ethernet0, Ethernet8, and Ethernet16.

```
admin@sonic:~$ show vlan brief
+-----------+-------------+----------+---------------+------------
+--------------+--------------+------------+
|   VLAN ID | IP Address   | Ports     | Port Tagging   | Proxy ARP   | DHCP
Helper   | DHCP Relay Configuration    |
|           |             |          |               |            | Address
|                          |
+===========+=============+==========+===============+============+=========
======+==============+=============+
|        10 |             | Ethernet0 | tagged         | disabled    |
| Source Interface:        |
|           |             | Ethernet8 | tagged         |            |
| Link Selection:          |
|           |             | Ethernet16| tagged         |            |
| Server Vrf:              |
|           |             |          |               |            |
| Server ID Override:      |
+-----------+-------------+----------+---------------+------------
+--------------+--------------+------------+
```

Procedure:
Step 1: Setting Buffer

```
admin@sonic:~$ sudo config qos reload
```

Step 2: Save the configuration

```
admin@sonic:~$ sudo config save -y
```

Step 3: Reboot the switch to activate the Buffer setting.

```
admin@sonic:~$ sudo reboot
```

Step 4: Checking the Buffer Pool setting

```
admin@sonic:~$ show buffer pool
Pool                   Type     Size              Xoff
---------------------  -------  ----------------  ----------------
egress_lossless_pool   egress   5,085,913 Bytes
egress_lossy_pool      egress   25,994,668 Bytes
ingress_lossless_pool  ingress  5,085,913 Bytes   25,429,567 Bytes
ingress_lossy_pool     ingress  25,994,668 Bytes
```

Step 5: Checking the Buffer Profile setting

```
admin@sonic:~$ show buffer profile
Profile                   Pool                   Size     Shared Mode    Shared
Size    Xoff           Xon             Xon-offset
------------------------  ---------------------  -------  -------------
-------------  -------------  ---------------  ------------
egress_lossless_profile   egress_lossless_pool   0 Bytes  dynamic        66%
egress_lossy_profile      egress_lossy_pool      0 Bytes  dynamic        50%
ingress_lossless_profile  ingress_lossless_pool  0 Bytes  dynamic        66%
971,176 Bytes  3,051,548 Bytes  18 KiB
ingress_lossy_profile     ingress_lossy_pool     0 Bytes  dynamic        66%
```

Step 6: Binding the Buffer profile on the Interface (including ingress and egress.)

```
admin@sonic:~$ sudo config interface buffer bind priority-group Ethernet0 3
ingress_lossless_profile
admin@sonic:~$ sudo config interface buffer bind priority-group Ethernet8 3
ingress_lossless_profile
admin@sonic:~$ sudo config interface buffer bind queue Ethernet16 3
egress_lossless_profile
```

Note:

```
admin@sonic:~$ sudo config interface buffer bind priority-group --help
Usage: config interface buffer [OPTIONS] {bind|unbind} {priority-group|queue}
                               {<interface_name>|all} {<pg>|<queue>} <profile>

   Set interface PG/queue buffer-profile configuration

Options:
-h, -?, --help Show this message and exit.
```

Here's the way to bind Buffer profile on all of the interfaces directly.

```
admin@sonic:~$ sudo config interface buffer bind priority-group all 3
ingress_lossless_profile
admin@sonic:~$ sudo config interface buffer bind queue all 3
egress_lossless_profile
```

The default setting for all interfaces is binding the lossy_profile.

```
admin@sonic:~$ show interfaces buffer priority-group Ethernet0
Interface    PG    Profile
-----------  ----  ---------------------
Ethernet0    0     ingress_lossy_profile
Ethernet0    1     ingress_lossy_profile
Ethernet0    2     ingress_lossy_profile
Ethernet0    3     ingress_lossy_profile
Ethernet0    4     ingress_lossy_profile
Ethernet0    5     ingress_lossy_profile
Ethernet0    6     ingress_lossy_profile
Ethernet0    7     ingress_lossy_profile

admin@sonic:~$ show interfaces buffer queue Ethernet0
```

```
Interface    Queue    Profile
-----------  -------  --------------------
Ethernet0    0        egress_lossy_profile
Ethernet0    1        egress_lossy_profile
Ethernet0    2        egress_lossy_profile
Ethernet0    3        egress_lossy_profile
Ethernet0    4        egress_lossy_profile
Ethernet0    5        egress_lossy_profile
Ethernet0    6        egress_lossy_profile
Ethernet0    7        egress_lossy_profile
```

Step 7: Check the interface that binds the Buffer Profile.

```
admin@sonic:~$ show interfaces buffer priority-group Ethernet0
Interface    PG    Profile
-----------  ----  -----------------------
Ethernet0    0     ingress_lossy_profile
Ethernet0    1     ingress_lossy_profile
Ethernet0    2     ingress_lossy_profile
Ethernet0    3     ingress_lossless_profile
Ethernet0    4     ingress_lossy_profile
Ethernet0    5     ingress_lossy_profile
Ethernet0    6     ingress_lossy_profile
Ethernet0    7     ingress_lossy_profile

admin@sonic:~$ show interfaces buffer queue Ethernet16
Interface    Queue    Profile
-----------  -------  ----------------------
Ethernet16   0        egress_lossy_profile
Ethernet16   1        egress_lossy_profile
Ethernet16   2        egress_lossy_profile
Ethernet16   3        egress_lossless_profile
Ethernet16   4        egress_lossy_profile
Ethernet16   5        egress_lossy_profile
Ethernet16   6        egress_lossy_profile
Ethernet16   7        egress_lossy_profile
```

Step 8: Enable the PFC for Priority 3 on the ingress Interface.

```
admin@sonic:~$ sudo config interface pfc priority Ethernet0 3 on
admin@sonic:~$ sudo config interface pfc priority Ethernet8 3 on
```

Step 9: Check the interface that enables the PFC.

```
admin@sonic:~$ show interfaces qos Ethernet0
 pfc-priority: 3

admin@sonic:~$ show interfaces qos Ethernet8
pfc-priority: 3
```

Result:

Packet Generator:

Note:

Column B and Column C belong to Packet Generator (Tx), and Column D belongs to Packet Generator (Rx).
Refer to the blue frame, there is no packet loss when triggering the traffic congestion.
92,130,119 + 94,781,831 =186,911,950

"Valid Frames Received" has two additional packets because the switch sends LLDP packets.

SONiC:

```
admin@sonic:~$ show pfc counter | grep -Ew "Ethernet0|Port (Rx|Tx)"
    Port Rx     PFC0    PFC1    PFC2    PFC3    PFC4    PFC5    PFC6    PFC7
  Ethernet0     0       0       0       0       0       0       0       0
    Port Tx     PFC0    PFC1    PFC2    PFC3    PFC4    PFC5    PFC6    PFC7
  Ethernet0     0       0       0       247,900 0       0       0       0

admin@sonic:~$ show pfc counter | grep -Ew "Ethernet8|Port (Rx|Tx)"
    Port Rx     PFC0    PFC1    PFC2    PFC3    PFC4    PFC5    PFC6    PFC7
  Ethernet8     0       0       0       0       0       0       0       0
    Port Tx     PFC0    PFC1    PFC2    PFC3    PFC4    PFC5    PFC6    PFC7
  Ethernet8     0       0       0       247,852 0       0       0       0
```

TC (Traffic Class) to PG (Priority Group) mapping table [Optional]
This mapping table has two purposes. One is to determine the priority-group (PG) for buffering. The other is to determine the priority in the PFC PAUSE frame. The behavior of the mapping varies by switch platform.

Broadcom Switch:
The TC is mapped to the priority-group (PG) based on the following rules:

If the packet is VLAN tagged, the packet will map the Dot1P-to-TC first (Please refer to this article.), then map the TC-to-PG.
If the packet is untagged, the packet will map the DSCP-to-TC first (Please refer to this article.) and then map the TC-to-PG.
The PG range is 0 to 7.
Default TC-to-PG mapping on Trident2plus, Trident3, Tomahawk, and Tomahawk2 switches.

PG   7
Default TC-to-PG mapping on Trident4, Tomahawk3, and Tomahawk4 switches.

PG   0   1   2   3   4   5   6   7
Intel Switch:
the packet will be mapped by QoS classification ("Dot1P-to-TC" or "DSCP-to-TC", Please refer to this article) and then map the TC-to-PG.
The PG range is 0, 1 to 5. The PG 0 is the default PG, and it is not able to change the buffer thresholds and be mapped by TC to PG mapping table.
Default TC-to-PG mapping

PG   0

Procedure:
Step 1: Create a profile for TC to PG.

```
admin@sonic:~$ sudo config qos tc-pg add tc-pg-prof --tc 0 --pg 0
```

Note:

```
admin@sonic:~$ sudo config qos tc-pg add --help
```

```
Usage: config qos tc-pg add [OPTIONS] <profile>

  Add tc-pg map profile.

Options:
  --tc TEXT Traffic-class(TC) value [required]
  --pg INTEGER RANGE Priority-group(PG) value [required]
  -h, -?, --help      Show this message and exit.
```

Step 2: Modify the existing TC to PG profile.

```
admin@sonic:~$ sudo config qos tc-pg update tc-pg-prof --tc 3 --pg 3
```

Note:

```
admin@sonic:~$ sudo config qos tc-pg update --help
Usage: config qos tc-pg update [OPTIONS] <profile>

  Update tc-pg map profile

Options:
  --tc TEXT Traffic-class(TC) value [required]
  --pg INTEGER RANGE Priority-group(PG) value
  --remove Delete the mapping entry
  -h, -?, --help      Show this message and exit.
```

Step 3: Check the profile of TC to PG profile

```
admin@sonic:~$ show qos tc-pg
tc-pg policy: tc-pg-prof
  TC PG
---- ----
   0    0
   3    3
```

Step 4: Binding the TC to PG mapping table to all the interfaces.

```
admin@sonic:~$ sudo config interface qos tc-pg bind all tc-pg-prof
```

Step 5: Check the interface that binds TC to PG mapping table.

```
admin@sonic:~$ show interfaces qos Ethernet0
 TC to PG: tc-pg-prof
```