Named Pools

In This Section

This section provides information to configure Named Pools QoS policies using the command line interface.

Topics in this section include:

Overview

The named buffer pool feature allows for the creation of named buffer pools at the MDA and port level. Named pools allow for a customized buffer allocation mode for ingress and egress queues that goes beyond the default pool behavior.

Named pools are defined within a named pool policy. The policy contains a q1-pools context which is used to define port allocation weights and named pools for buffer pools on Q1 based IOMs (all IOMs that are currently supported). The policy may be applied at either the port or MDA level at which time the pools defined within the policy are created on the port or MDA. When the policy is applied at the MDA level, MDA named pools are created. MDA named pools will typically be used when either a pool cannot be created per port or when the buffering needs of queues mapped to the pool are not affected by sharing the pool with queues from other ports. MDA named pools allow buffers to be efficiently shared between queues on different ports mapped to the same pool. However, MDA named pools do present the possibility that very active queues on one port could deplete buffers in the pool offering the possibility that queues on other ports experiencing buffer starvation. Port named pools are created when the policy is applied at the port level and allow for a more surgical application of the buffer space allocated for a physical port. MDA pool names do not need to be unique. If a name overlaps exists, the port pool will be used. The same pool name may be created on multiple ports on the same MDA.

The named pool policy is applied at the MDA ingress and egress level and at the ingress and egress port level. Each MDA within the system is associated with a forwarding plane traffic manager that has support for a maximum of 57 buffer pools. The following circumstances affect the number of named pools that can be created per MDA (these circumstances may be different between ingress and egress for the MDA):

- The forwarding plane can be associated with multiple MDAs (each MDA has its own named pools).
- A single system level pool for system created queues is allocated.
- Each system must have default pools for queues that are not explicitly mapped or are incorrectly mapped to a named pool.
- Default pools for most IOM types (separate for ingress and egress).
- Access pool.
- Network pool.
- The number of named per-port pools is dependant on the number of ports the MDA supports which is variable per MDA type.
- Per-port named pools cannot be used by ingress network queues, but pools defined in a named pool policy defined on an ingress all network port are still created.
  → Ingress network queues use the default network pool or MDA named pools.
  → Ingress port buffer space allocated to network mode ports is included in the buffers made available to ingress MDA named pools.

→ Ingress port buffer space on channelized ports associated with network bandwidth is included in the buffers made available to ingress MDA named pools

→ Ingress port named pools are only allocated buffers when the port is associated with some access mode bandwidth

- Per-port named pools on ports aggregated into a LAG are still created per physical port

- Default, named MDA and named per-port pools are allocated regardless of queue provisioning activity associated with the pool

If the named pool policy is applied to an MDA or port that cannot create every pool defined in the policy, the policy application attempt will fail. Any pre-existing named pool policy on the MDA or port will not be affected by the failed named pool policy association attempt.

When buffer pools are being created or deleted, individual queues may need to be moved to or from the default pools. When a queue is being moved, the traffic destined to the queue is first moved temporarily to a 'fail-over' queue. Then the queue is allowed to drain. Once the queue is drained, the statistics for the queue are copied. The queue is then returned to the free queue list. A new queue is then created associated with the appropriate buffer pool, the saved stats are loaded to the queue and then the traffic is moved from the fail-over queue to the new queue. While the traffic is being moved between the old queue to the fail-over queue and then to the new queue, some out of order forwarding may be experienced. Also, any traffic forwarded through the fail-over queue will not be accounted for in billing or accounting statistics. A similar action is performed for queues that have the associated pool name added, changed or removed. Please note this only applies to where fail-over queues are currently supported.

The first step in allowing named pools to be created for an MDA is to enable 'named-pool-mode' at the IOM level (config card slot-number named-pool-mode). Named pool mode may be enabled and disabled at anytime. When MDAs are currently provisioned on the IOM, the IOM is reset to allow all existing pools to be deleted and the new default, named MDA and named port pools to be created and sized. If MDAs are not currently provisioned (as when the system is booting up), the IOM is not reset. When named pool mode is enabled, the system changes the way that default pools are created. The system no longer creates default pools per port, instead, a set of per forwarding plane level pools are created that are used by all queues that are not explicitly mapped to a named pool.

After the IOM has been placed into named pool mode, a named pool policy must be associated with the ingress and egress contexts of the MDA or individual ports on the MDA for named pools to be created. There are no named pools that exist by default.

Each time the default pool reserve, aggregate MDA pool limit or individual pool sizes is changed, buffer pool allocation must be re-evaluated.

Pools may be deleted from the named pool policy at anytime. Queues associated with removed or non-existent pools are mapped to one of the default pools based on whether the queue is access or ingress. The queue is flagged as 'pool-orphaned' until either the pool comes into existence, or the pool name association is changed on the pool.

An ingress or egress port managed buffer space is derived from the port's active bandwidth. Based on this bandwidth value compared to the other port's bandwidth value, the available buffer space is given to each port to manage. It may be desirable to artificially increase or decrease this bandwidth value to compensate for how many buffers are actually needed on each port. If one port has very few queues associated with it and another has many queues associated, the commands in the port's "modify-buffer-allocation-rate" CLI context may be used to move one port's bandwidth up, and another port's bandwidth down. As provisioning levels change between ports, the rate modification commands may be used to adapt the buffer allocations per port.

Buffer allocation rate modification is supported for both standard and named pool mode buffer allocation methods.

The system allocates buffers based on the following criteria:

- "named-pool-mode" setting on the IOM.
- Amount of path bandwidth on channelized ports.
- Existence of queues provisioned on the port or channel.
- Current speed of each port.
- Each ports "ing-percentage-of-rate" and "egr-percentage-of-rate" command setting.
- The port-allocation-weights setting for default, MDA and port.
- The ports division between network and access bandwidth.
- Each individual named pool's network-allocation-weight and access-allocation-weight.

System reserved named pool names (cannot be used when configuring a named pool) are: **default**, **SAP Shared** and **MC Path Mgmt**.

Named Pool Mode for IOM3-XP Card

IOM3-XP has one forwarding complex for MDA1 and MDA2. The total available buffer space is divided in the ingress and egress direction. The total ingress buffer space is shared by both MDAs in the ingress direction and the total egress buffer space is shared by both MDAs in the egress direction.

Each MDA can use a different named pool policy.

Network ingress queues can only use the MDA1 named pools. If named pools are configured for MDA2 they will not be used by network ingress queues. Network ingress queues configured to use MDA2 named pools will be considered pool orphaned. To check for orphan queues, use the command **show mda** *mda* **qos ingress orphaned-queues**. The same restriction applies for SAP shared queues using named pools.

Each named pool policy can have a maximum of 57 named pools configured.

The total number of named pools that can be configured per IOM3 is 245. The named pool usage per card can be checked with the **tools dump system-resources** *slot-number* | match Pools command.

```
A:ALA-48>tools>dump# system-resources 1 | match Pools
        Ingress Q1 Named Pools |         57|          0|        57
         Egress Q1 Named Pools |         57|          0|        57
        Ingress Q1 Named Pools |         57|          0|        57
         Egress Q1 Named Pools |         57|          0|        57
A:ALA-48>tools>dump#
```

Basic Configuration

A basic named pool QoS policy must conform to the following:

- Default values can be modified but parameters cannot be deleted.

Create a Named Pool QoS Policy

To create a new named pool policy, the following must be define.

- A named pool policy ID value. The system will not dynamically assign a value.

Named pool Configuration Procedure

Step 1. Configure the named pool policy.

CLI Syntax:
```
config# configure qos named-pool-policy 3pools create
     q1-pools
         pool p1 create
         exit
         pool p2 create
         exit
         pool p3 create
         exit
     exit all
```

Step 2. Apply the named pool policy on ingress and/or egress MDA and/or port.

Since the named pool mode is not yet active on the card, all queues are drawing buffer from the default pool.

Queue to named pool association.

Configure the queues to get buffers from a named pool.

Configure the named pool policy.

CLI Syntax:
```
config# network queue
     configure qos
         copy network-queue default 15
         network-queue 15
         queue 1 pool p1
         exit all
```

Step 3. Configure the above queue profile to be used by the respective applications or port.

Configure the named pool policy.

CLI Syntax: `config# configure card 1 mda 2 network ingress queue-policy 15`

Step 4. Turn on named pool on the card. The card will reboot and the named pool mode will be active on the card.

CLI Syntax: `configure card 1 named-pool-mode now`

To check the pools after named pool mode was enabled, use "show pools mda". There are no port pools unless a port has configured a named pool policy. Only named pools that have active queues associated will be shown with a non-zero size; this means that the named pool was created. If a named pool is configured but has no active queue associated, the size of the pool is zero. The named pool would be instantiated but not created. An example is shown below.

Example:

```
A:SR7-10# show pools 1/2
===============================================================================


===============================================================================
Type    Id        App.    Pool Name                   Actual ResvCBS  PoolSize
                                                       Admin ResvCBS
-------------------------------------------------------------------------------
MDA     1/2       Acc-Ing default                     3072            10240
                                                       Sum
MDA     1/2       Acc-Ing MC Path Mgmt                0               0
                                                       50%
MDA     1/2       Acc-Egr default                     12288           40960
                                                       Sum
MDA     1/2       Net-Ing default                     20480           40960
                                                       Sum
MDA     1/2       Net-Egr default                     81920           163840
                                                       Sum
MDA     1/2       Ingress p1                          12288           28672
                          Policy: 3pools               30%
MDA     1/2       Ingress p2                          0               0
                          Policy: 3pools               30%
MDA     1/2       Ingress p3                          0               0
                          Policy: 3pools               30%
===============================================================================
A:SR7-10#
```

Allocation Steps

Whether one or multiple MDAs share the same buffer space, the buffer space is portioned out on a per port basis. Each port gets an amount of buffering which is its fair-share based on the port's bandwidth compared to the overall active bandwidth. This is identical to current behavior.

This mechanism takes the buffer space available and divides it into a portion for each port based on the ports active bandwidth relative to the amount of active bandwidth for all ports associated with the buffer space. The number of ports sharing the same buffer space depends on the type of IOM the pools are being created on and the type of MDAs populated on the IOM. An active port is considered to be any port that has an active queue associated. Once a queue is created for the port, the system will allocate the appropriate amount of buffer space to the port. This process is independently performed for both ingress and egress.

Normally, the amount of active bandwidth is considered as opposed to total potential bandwidth for the port when determining the ports fair share. If a port is channelized and not all bandwidth is allocated, only the bandwidth represented by the configured channels with queues configured is counted towards the bandwidth represented by the port. Also, if a port may operate at variable speeds (as in some Ethernet ports), only the current speed is considered. Based on the above, the number of buffers managed by a port may change due to queue creation and deletion, channel creation and deletion and port speed variance on the local port or other ports sharing the same buffer space.

After the active bandwidth is calculated for the port, the result may be modified through the use of the 'ing-percentage-of-rate' and 'egr-percent-of-rate' commands. The default value of each is 100% which allows the system to use all of the ports active bandwidth when deciding the relative amount of buffer space to allocate to the port. When the value is explicitly modified, the active bandwidth on the port is changed according to the specified percentage. If a value of 50% is given, the ports active bandwidth will be multiplied by .5; if a value of 150% is given, the active bandwidth will be multiplied by 1.5. This capability is independent of named pool mode. The ports rate percentage parameters may be modified at any time.

When named pool mode is configured on the buffer space, the ingress and egress chunk of buffering assigned to a port is now split into 3 smaller chunks for default pools, mda named pools and port named pools. The way the buffering is split into the 3 smaller chunks is based on the 'port-allocation-weights' given in the named-pool-policy. The weights may come from either the MDA level applied named pool policy or the local port applied named pool policy. If a named pool policy is assigned on both locations, the defined 'port-allocation-weights' from the port associated policy will apply. Any of the weights may be set to '0', indicating that none of the buffers allocated to the port should be given to the pool category. If only MDA named pools are created, the port weight should be set to '0'; if only port named pools are created, the MDA weight should be set to '0'. Setting the default weight to '0' should be done with care as any queues without named pool definitions or queues with non-existent pool names use the default pools. The weights are summed and then each individual weight is divided by the sum. The result is

multiplied by the buffer space managed by port to determine the amount of buffer space given to each category.

The portion of buffering set aside for default pools is applied to the default network and access pools based solely on the amount of bandwidth in network and access modes on the physical port. Combining the default pool buffering chunks from all ports result in the aggregate size of the default pools. For Ethernet ports, all the bandwidth is either network or access (there is never both). For channelized ports, the amount given to network and access default pools is based on the ratio between access and network bandwidth and not based on weights defined within the pools. The default pool configuration is not changed when named pool mode is enabled.

The portion of buffering managed by the port relative to the ports MDA weight parameter is separated into two sub-portions based on the ratio between network and access bandwidth on port. If the port is all network or access bandwidth, that opposite sub-portion will be empty. The sub-portion associated with network bandwidth on the port and the sub-portion associated with access bandwidth on the port are individually summed with the network and access MDA named pool bandwidth set aside by the other ports on the MDA. The outcome is a total amount of access and a total amount of network bandwidth that will be given to MDA named pools based on each pools network and access weights.

The portion of buffering assigned to port named pools based on the port weight is also separated into two sub-portions in the same manner as step 4 above. The network and access bandwidth defines the ratio of bandwidth population between the network and access sub-portions. If the port is all network or access bandwidth, that opposite sub-portion will be empty. Each sub-portion is divided between the port named pools based on each pools network and access defined weights.

On the ingress side, network queues may only be associated with the default ingress network pool or one of the MDA named pools. Because ingress network queues may not use ingress port based pools, the port's network sub-portion is added to the ports MDA named pool network sub-portion to be distributed to the named MDA pools based on each pools network weight.