Chapter 13

Shaping Input and Output Traffic on Ethernet IQ2 Interfaces

Gigabit Ethernet Intelligent Queuing 2 (IQ2) 4-port and 8-port Type 2 Physical Interface Cards (PICs) are oversubscribed, which means the amount of traffic coming to the PIC can be more than the maximum bandwidth from the PIC to the Flexible PIC Concentrator (FPC).

The 10-Gigabit Ethernet IQ2 PIC (xe-) is unlike other Gigabit Ethernet IQ2 PICs in that this PIC does not have oversubscription. The bandwidth from the PIC to the FPC is sufficient to transmit the full line rate. However, the 10-Gigabit Ethernet IQ2 PIC has the same hardware architecture as other Gigabit Ethernet IQ2 PICs and supports all the same class-of-service (CoS) features. For more information, see the PIC guide for your routing platform.

To handle oversubscribed traffic, you can configure input shaping and scheduling based on Layer 2, MPLS, and Layer 3 packet fields. Gigabit Ethernet IQ2 PICs also support simple filters, accounting, and policing. This chapter discusses input and output shaping and scheduling. For information about simple filters, see “Example: Configuring a Simple Filter” on page 72 and the *JUNOS Policy Framework Configuration Guide*. For information about accounting and policing, see the *JUNOS Network Interfaces Configuration Guide*.

NOTE: The class-of-service (CoS) functionality supported on Gigabit Ethernet IQ2 PICs is not available across aggregated Ethernet links. However, if you configure a CoS scheduler map on the link bundle, the configuration is honored by the individual links within that bundle.

Therefore, CoS on a per-link level behaves as configured, but CoS does not behave as configured across the aggregated links. For example, if you configure a shaping transmit rate of 100 megabits per second (Mbps) (by including the transmit-rate exact statement in a scheduler) on an aggregated Ethernet bundle with three ports, each port is provisioned with a 33.33 Mbps shaping transmit rate.

By default, transmission scheduling is not enabled on logical interfaces. Logical interfaces without shaping configured share a default scheduler. This scheduler has a committed information rate (CIR) that equals 0. (The CIR is the guaranteed rate.) The default scheduler has a peak information rate (PIR) that equals the physical interface shaping rate. The default operation can be changed by configuring the software.

To configure input and output shaping and scheduling, include the following statements at the [edit class-of-service] and [edit interfaces] hierarchy levels of the configuration:

```
[edit class-of-service]
traffic-control-profiles profile-name {
    delay-buffer-rate (percent percentage | rate);
    guaranteed-rate (percent percentage | rate);
    scheduler-map map-name;
    shaping-rate (percent percentage | rate);
}
interfaces interface-name {
    input-scheduler-map map-name;
    input-shaping-rate rate;
    scheduler-map map-name; # Output scheduler map
    shaping-rate rate; # Output shaping rate
    unit logical-unit-number {
      input-scheduler-map map-name;
      input-shaping-rate (percent percentage | rate);
      input-traffic-control-profile profile-name shared-instance instance-name;
      output-traffic-control-profile profile-name shared-instance instance-name;
    }
}

[edit interfaces interface-name]
per-unit-scheduler;
shared-scheduler;
```

This chapter discusses the following topics:

Differences Between Gigabit Ethernet IQ and Gigabit Ethernet IQ2 PICs

The following CoS differences arise because Gigabit Ethernet IQ PICs and Gigabit Ethernet IQ2 PICs use different hardware:

- Gigabit Ethernet IQ2 PICs support a transmission rate within a queue, but do not support an exact rate within a queue. You can apply a weight to a queue, but you cannot put an upper limit on the queue transmission rate that is less than the logical interface can support. Consequently, the exact statement at the [edit class-of-service schedulers *scheduler-name* transmit-rate (*rate* | *percent*)] hierarchy level is not supported for Gigabit Ethernet IQ2 interfaces.

- Gigabit Ethernet IQ2 PICs support only one queue in the scheduler map with high or strict-high priority. If more than one queue is configured with high or strict-high priority, the one that appears first in the configuration is implemented as strict-high priority. This queue receives unlimited transmission bandwidth. The remaining queues are implemented as low priority, which means they might be starved.

- To ensure that protocol control traffic (such as OSPF, BGP, and RIP) are not dropped at the oversubscribed ingress direction, the software puts control protocol packets into a separate control scheduler. There is one control scheduler per port. These control schedulers are implemented as strict-high priority, so they transmit traffic until they are empty.

- On Gigabit Ethernet IQ2 PICs, you can configure a single traffic-control profile to contain both a PIR (the shaping-rate statement) and a CIR (the guaranteed-rate statement). On Gigabit Ethernet IQ PICs, these statements are mutually exclusive.

- Gigabit Ethernet IQ2 PICs support only two fill levels in the RED drop profile. The recommended definition of the RED drop profile is as follows:

```
class-of-service {
    drop-profiles {
        drop-iq2-example1 {
            fill-level 20 drop-probability 0;
            fill-level 100 drop-probability 80;
        }
    }
}
```

  This configuration defines a drop profile with a linear drop probability curve when the fill level is between 20 and 100 percent, and a maximum drop probability of 80 percent.

  You can configure more than two fill levels in the drop profile, but the software only uses the points (min_fill_level, 0) and (max_fill_level, max_probability) and ignores other fill levels. The drop probability at the minimum fill level is set to 0 percent even if you configure a non-zero drop probability value at the minimum fill level. The following example shows a sample configuration and the software implementation:

Configuration

```
class-of-service {
    drop-profiles {
        drop-iq2-example2 {
            fill-level 30 drop-probability 10;
            fill-level 40 drop-probability 20;
            fill-level 100 drop-probability 80;
        }
    }
}
```

Implementation

```
class-of-service {
    drop-profiles {
        drop-iq2-example2-implementation {
            fill-level 30 drop-probability 0;
            fill-level 100 drop-probability 80;
        }
    }
}
```

If you configure more than two fill levels, a system log message warns you that the software supports only two fill levels and displays the drop profile that is implemented.

Though the interpolate statement is supported in the definition of a RED drop profile, we do not recommend using it. The following example shows a sample configuration and the software implementation:

Configuration

```
class-of-service {
    drop-profiles {
        drop-iq2-example3 {
            interpolate {
                fill-level [ 30 50 80 ];
                drop-probability [ 10 20 40 ];
            }
        }
    }
}
```

Implementation When you use the interpolate statement and the maximum fill level is not 100 percent, the software adds the point (100, 100). Therefore, the drop-iq2-example3 drop profile is implemented as:

```
class-of-service {
    drop-profiles {
        drop-iq2-example3-implementation {
            fill-level 2 drop-probability 0;
            fill-level 100 drop-probability 100;
        }
    }
}
```

The implemented minimum fill level is not 30 percent as configured, but 2 percent because of the 64-point interpolation.

Configuring a Shared Scheduler and Shaper

Shared scheduling and shaping allows you to allocate separate pools of shared resources to subsets of logical interfaces belonging to the same physical port. You configure this by creating a traffic-control profile, which specifies a shaping rate and references a scheduler map. To share this set of shaping and scheduling resources, you apply an instance of the traffic-control profile to a subset of logical interfaces. You can apply a separate instance of the same (or a different) traffic-control profile to another subset of logical interfaces, thereby allocating separate pools of shared resources.

To configure a traffic-control profile, perform the following steps:

1. Include the shaping-rate statement at the [edit class-of-service traffic-control-profiles *profile-name*] hierarchy level:

   ```
   [edit class-of-service traffic-control-profiles profile-name]
   shaping-rate (percent percentage | rate);
   ```

   You can configure the shaping rate as a percentage from 1 through 100 or as an absolute rate from 1000 through 160,000,000,000 bits per second. The shaping rate corresponds to a peak information rate (PIR). For more information, see "Oversubscribing Interface Bandwidth" on page 153.

2. Include the scheduler-map statement at the [edit class-of-service traffic-control-profiles *profile-name*] hierarchy level:

   ```
   [edit class-of-service traffic-control-profiles profile-name]
   scheduler-map map-name;
   ```

   For information about configuring schedulers and scheduler maps, see "Configuring a Scheduler" on page 119 and "Configuring the Scheduler Map" on page 139. Gigabit Ethernet IQ2 interfaces support up to eight forwarding classes and queues. For more information, see "Configuring Up to Eight Forwarding Classes" on page 98.

3. Include the delay-buffer-rate statement at the [edit class-of-service traffic-control-profiles *profile-name*] hierarchy level:

   ```
   [edit class-of-service traffic-control-profiles profile-name]
   delay-buffer-rate (percent percentage | rate);
   ```

   You can configure the delay-buffer rate as a percentage from 1 through 100 or as an absolute rate from 1000 through 160,000,000,000 bits per second. The delay-buffer rate controls latency. For more information, see "Oversubscribing Interface Bandwidth" on page 153 and "Providing a Guaranteed Minimum Rate" on page 160.

4. Include the guaranteed-rate statement at the [edit class-of-service traffic-control-profiles *profile-name*] hierarchy level:

   ```
   [edit class-of-service traffic-control-profiles profile-name]
   guaranteed-rate (percent percentage | rate);
   ```

You can configure the guaranteed rate as a percentage from 1 through 100 or as an absolute rate from 1000 through 160,000,000,000 bits per second. The guaranteed rate corresponds to a committed information rate (CIR). For more information, see "Providing a Guaranteed Minimum Rate" on page 160.

To share an instance of the traffic-control profile, perform the following steps:

5. Include the **shared-scheduler** statement at the [**edit interfaces** *interface-name*] hierarchy level:

   ```
   [edit interfaces interface-name]
   shared-scheduler;
   ```

   This statement enables logical interfaces belonging to the same physical port to share one set of shaping and scheduling resources.

NOTE: On each physical interface, the **shared-scheduler** and **per-unit-scheduler** statements are mutually exclusive. Even so, you can configure one logical interface for each shared instance. This effectively provides the functionality of per-unit scheduling.

6. To apply the traffic-control profile to an input interface, include the **input-traffic-control-profile** and **shared-instance** statements at the [**edit class-of-service interfaces** *interface-name* **unit** *logical-unit-number*] hierarchy level:

   ```
   [edit class-of-service interfaces interface-name unit logical-unit-number]
   input-traffic-control-profile profile-name shared-instance instance-name;
   ```

   These statements are explained in Step 7.

7. To apply the traffic-control profile to an output interface, include the **output-traffic-control-profile** and **shared-instance** statements at the [**edit class-of-service interfaces** *interface-name* **unit** *logical-unit-number*] hierarchy level:

   ```
   [edit class-of-service interfaces interface-name unit logical-unit-number]
   output-traffic-control-profile profile-name shared-instance instance-name;
   ```

   The profile name references the traffic-control profile you configure in steps 1 through 4. The shared-instance name does not reference a configuration. It can be any text string you wish to apply to multiple logical interfaces that you want to share the set of resources configured in the traffic-control profile. Each logical interface shares a set of scheduling and shaping resources with other logical interfaces that are on the same physical port and that have the same shared-instance name applied.

   This concept is demonstrated in "Examples: Shaping Input and Output Traffic on Ethernet IQ2 Interfaces" on page 207.

NOTE: You cannot include the **output-traffic-control-profile** statement in the configuration if any of the following statements are included in the logical interface configuration: **scheduler-map**, **shaping-rate**, **adaptive-shaper**, **virtual-channel-group**.

Differences Between Per-Unit Scheduling and Shared Scheduling

Shared scheduling allows you to allocate separate pools of shared resources to subsets of logical interfaces belonging to the same physical port.

Per-unit scheduling enables one set of output queues for each logical interface configured under the physical interface.

An *unconfigured logical interface* (in the context of CoS) is a logical interface that you configure at the [edit interfaces *interface-name* unit *logical-unit-number*] hierarchy level, but do not configure at the [edit class-of-service interfaces *interface-name* unit *logical-unit-number*] hierarchy level.

The differences between per-unit scheduling and shared-scheduling are as follows:

- With per-unit scheduling, an unconfigured logical interface receives its own set of output queues.
- With shared scheduling, an unconfigured logical interface receives its own set of output queues only if there is some configuration for that logical interface at the [edit class-of-service interfaces *interface-name* unit *logical-unit-number*] hierarchy level.
- When you configure shared scheduling, you can include the shared-instance statement with the traffic-control profile. The shared-instance statement is not supported with per-unit scheduling.
- When you configure shared scheduling, a dedicated scheduler is assigned to a logical interface on the output direction only, if you configure one or more of the following: a scheduler map, a shaping-rate, a guaranteed rate, or a traffic-control profile. All the other logical interfaces use the same set of queues in the output direction. Similarly, a dedicated scheduler is assigned to a logical interface on the input direction only, if you configure one or more of the following: an input scheduler map, an input shaping rate, or an input traffic-control profile. All other logical interfaces use the same set of queues in the input direction.

Configuring Separate Input Schedulers

As an alternative to shared input traffic-control profiles, you can configure each interface to use its own input scheduler. For each physical interface, you can apply an input scheduler map to the physical interface or its logical interfaces, but not both.

For information about configuring schedulers and scheduler maps, see “Configuring a Scheduler” on page 119 and “Configuring the Scheduler Map” on page 139. Gigabit Ethernet IQ2 interfaces support up to eight forwarding classes and queues. For more information, see “Configuring Up to Eight Forwarding Classes” on page 98.

To configure a separate input scheduler on the physical interface, include the input-scheduler-map statement at the [edit class-of-service interfaces *interface-name*] hierarchy level:

```
[edit class-of-service interfaces interface-name]
input-scheduler-map map-name;
```

To configure a separate input scheduler on a logical interface, perform the following steps:

1. Include the input-scheduler-map statement at the [edit class-of-service interfaces *interface-name* unit *logical-unit-number*] hierarchy level:

   ```
   [edit class-of-service interfaces interface-name unit logical-unit-number]
   input-scheduler-map map-name;
   ```

2. For the corresponding physical interface, you must also include the per-unit-scheduler statement at the [edit interfaces *interface-name*] hierarchy level:

   ```
   [edit interfaces interface-name]
   per-unit-scheduler;
   ```

   The per-unit-scheduler statement enables one set of output queues for each logical interface configured under the physical interface.

   On Gigabit Ethernet IQ2 PIC interfaces, configuration of the per-unit-scheduler statement requires that you configure VLAN tagging also. When you include the per-unit-scheduler statement, the maximum number of VLANs supported is 767 on a single-port Gigabit Ethernet IQ PIC. On a dual-port Gigabit Ethernet IQ PIC, the maximum number is 383.

Configuring Hierarchical Input Shapers

You can apply input shaping rates to both the physical interface and its logical interfaces. The rate specified at the physical level is distributed among the logical interfaces based on their input shaping-rate ratio.

To configure an input shaper on the physical interface, include the input-shaping-rate statement at the [edit class-of-service interfaces *interface-name*] hierarchy level:

```
[edit class-of-service interfaces interface-name]
input-shaping-rate rate;
```

To configure an input shaper on the logical interface, include the input-shaping-rate statement at the [edit class-of-service interfaces *interface-name* unit *logical-unit-number*] hierarchy level:

```
[edit class-of-service interfaces interface-name unit logical-unit-number]
input-shaping-rate (percent percentage | rate);
```

For each logical interface, you can specify a percentage of the physical rate or an actual rate. The software converts actual rates into percentages of the physical rate.

Examples: Shaping Input and Output Traffic on Ethernet IQ2 Interfaces

This section includes the following examples:

Configuring Both a CIR and a PIR

On Gigabit Ethernet IQ2 interfaces, you can configure a CIR (guaranteed rate) and a PIR (shaping rate) on a single logical interface. In the following example, logical unit 0 has a CIR equal to 30 Mbps and a PIR equal to 200 Mbps. Logical unit 1 has a PIR equal to 300 Mbps. Logical unit 2 has a CIR equal to 100 Mbps and a PIR that is unspecified. For logical unit 2, the software causes the PIR to be 100 Mbps (equal to the CIR) because the PIR must be equal to or greater than the CIR.

Excess bandwidth is the leftover bandwidth on the port after meeting all the guaranteed rate requirements of the logical interfaces. For each port, excess bandwidth is shared as follows:

- Proportional to the guaranteed rate—This method is used if you configure one or more logical interfaces on a port to have a guaranteed rate.
- Proportional to the shaping rate—This method is used if you configure none of the logical interfaces on a port to have a guaranteed rate.

In this example, bandwidth is shared proportionally to the guaranteed rate because at least one logical interface has a guaranteed rate.

```
class-of-service {
    traffic-control-profiles {
        profile1 {
            shaping-rate 200m;
            guaranteed-rate 30m;
            delay-buffer-rate 150m;
            scheduler-map sched-map;
        }
        profile2 {
            shaping-rate 300m;
            delay-buffer-rate 500k;
            scheduler-map sched-map;
        }
        profile3 {
            guaranteed-rate 100m;
            scheduler-map sched-map;
        }
    }
    interface ge-3/0/0 {
        unit 0 {
            output-traffic-control-profile profile1;
        }
        unit 1 {
            output-traffic-control-profile profile2;
        }
```

```
        unit 2 {
            output-traffic-control-profile profile3;
        }
    }
}
```

Configuring Shared Resources

For input traffic on physical interface ge-1/2/3, logical interface units 1, 2, and 3 are sharing one set of scheduler-shaper resources, defined by traffic-control profile s1. Logical interface units 4, 5, and 6 are sharing another set of scheduler-shaper resources, defined by traffic-control profile s1.

For output traffic on physical interface ge-1/2/3, logical interface units 1, 2, and 3 are sharing one set of scheduler-shaper resources, defined by traffic-control profile s2. Logical interface units 4, 5, and 6 are sharing another set scheduler-shaper resources, defined by traffic-control profile s2.

For each physical interface, the shared-instance statement creates one set of resources to be shared among units 1, 2, and 3 and another set of resources to be shared among units 4, 5, and 6. An input and output shaping rate is configured at the physical interface level, which demonstrates the hierarchical shaping capability of the Gigabit Ethernet IQ2 PIC.

```
class-of-service {
    traffic-control-profiles {
        s1 {
            scheduler-map map1;
            shaping-rate 100k;
        }
        s2 {
            scheduler-map map1;
            shaping-rate 200k;
        }
    }
    interfaces {
        ge-1/2/3 {
            input-shaping-rate 500m;
            shaping-rate 500m; # Output shaping rate
            unit 1 {
                input-traffic-control-profile s1 shared-instance 1;
                output-traffic-control-profile s2 shared-instance 1;
            }
            unit 2 {
                input-traffic-control-profile s1 shared-instance 1;
                output-traffic-control-profile s2 shared-instance 1;
            }
            unit 3 {
                input-traffic-control-profile s1 shared-instance 1;
                output-traffic-control-profile s2 shared-instance 1;
            }
            unit 4 {
                input-traffic-control-profile s1 shared-instance 2;
                output-traffic-control-profile s2 shared-instance 2;
            }
```

```
            unit 5 {
              input-traffic-control-profile s1 shared-instance 2;
              output-traffic-control-profile s2 shared-instance 2;
            }
            unit 6 {
              input-traffic-control-profile s1 shared-instance 2;
              output-traffic-control-profile s2 shared-instance 2;
            }
          }
        }
        forwarding-classes { # Map one forwarding class to one queue.
          queue 0 fc-be;
          queue 1 fc-be1;
          queue 2 fc-ef;
          queue 3 fc-ef1;
          queue 4 fc-af11;
          queue 5 fc-af12;
          queue 6 fc-nc1;
          queue 7 fc-nc2;
        }
        classifiers { # Map 802.1p bits to forwarding-class and loss-priority.
          ieee-802.1 ieee-8021p-table {
            forwarding-class fc-nc2 {
              loss-priority low code-points [111];
            }
            forwarding-class fc-nc1 {
              loss-priority low code-points [110];
            }
            forwarding-class fc-af12 {
              loss-priority low code-points  [101];
            }
            forwarding-class fc-af11 {
              loss-priority low code-points  [100];
            }
            forwarding-class fc-ef1 {
              loss-priority low code-points [011];
            }
            forwarding-class fc-ef {
              loss-priority low code-points [010];
            }
            forwarding-class fc-be1 {
              loss-priority low code-points [001];
            }
            forwarding-class fc-be {
              loss-priority low code-points [000];
            }
          }
        }
}
interfaces { # Apply behavior aggregate classifier to an interface.
        ge-1/2/3 {
          unit 0 {
          classifier {
            ieee-802.1 ieee-8021p-table ;
          }
        }
}
```

Simple Filter Configure a simple filter that overrides the classification derived from the lookup of the Layer 2 fields.

```
firewall {
    family inet {
        simple-filter sf-1 {
            term 1 {
                source-address 172.16.0.0/16;
                destination-address 20.16.0.0/16;
                source-port 1024-9071;
            }
            then { # Action with term-1
                forwarding-class fc-be1;
                loss-priority high;
            }
            term 2 {
                source-address 173.16.0.0/16;
                destination-address 21.16.0.0/16;
            }
            then { # Action with term-2
                forwarding-class fc-ef1;
                loss-priority low;
            }
        }
    }
}
interfaces { # Apply the simple filter.
    ge-1/2/3 {
        unit 0 {
            family inet {
                simple-filter {
                    input sf-1 ;
                }
            }
        }
    }
}

class-of-service {
    scheduler-maps { # Configure a custom scheduler map.
        map1 {
            forwarding-class fc-be scheduler sch-Q0;
            forwarding-class fc-be1 scheduler sch-Q1;
            forwarding-class fc-ef scheduler sch-Q2;
            forwarding-class fc-ef1 scheduler sch-Q3;
            forwarding-class fc-af11 scheduler sch-Q4;
            forwarding-class fc-af12 scheduler sch-Q5;
            forwarding-class fc-nc1 scheduler sch-Q6;
            forwarding-class fc-nc2 scheduler sch-Q7;
        }
    }
```

```
    schedulers { # Define schedulers.
      sch-Q0 {
        transmit-rate percent 25;
        buffer-size percent 25;
        priority low;
        drop-profile-map loss-priority any protocol both drop-default;
      }
      sch-Q1 {
        transmit-rate percent 5;
        buffer-size temporal 2000;
        priority high;
        drop-profile-map loss-priority any protocol both drop-ef;
      }
      sch-Q2 {
        transmit-rate percent 35;
        buffer-size percent 35;
        priority low;
        drop-profile-map loss-priority any protocol both drop-default;
      }
      sch-Q3 {
        transmit-rate percent 5;
        buffer-size percent 5;
        drop-profile-map loss-priority any protocol both drop-default;
      }
      sch-Q4 {
        transmit-rate percent 5;
        priority high;
        drop-profile-map loss-priority any protocol both drop-ef;
      }
      sch-Q5 {
        transmit-rate percent 10;
        priority high;
        drop-profile-map loss-priority any protocol both drop-ef;
      }
      sch-Q6 {
        transmit-rate remainder;
        priority low;
        drop-profile-map loss-priority any protocol both drop-default;
      }
      sch-Q7 {
        transmit-rate percent 5;
        priority high;
        drop-profile-map loss-priority any protocol both drop-default;
      }
    }
}
```