

Cisco Performance Routing (PfR) Solution Guides

PfR Advanced Load Balancing using BGP Policies based on range, link utilization and delay Path Enforcement using PBR

Navigation

- [Go to PfR home page](#)
- [Go to PfR Solution Guides home page](#)

Contents

- [1 PfR Features that Enable Load Balancing](#)
 - ◆ [1.1 Link Utilization](#)
 - ◆ [1.2 Range](#)
 - ◆ [1.3 Traffic Class Performance](#)
- [2 Enterprise Needs](#)
- [3 PfR Solution Used](#)
- [4 PfR Network Topology Used](#)
- [5 PfR Components Configuration](#)
- [6 Flexible Netflow](#)
- [7 Checking Statistics and Flows](#)
- [8 Master Controller Verification](#)
 - ◆ [8.1 Master Controller and Traffic Classes](#)
 - ◆ [8.2 Display Learn-list](#)
 - ◆ [8.3 Policy Configuration](#)
- [9 Verify Traffic Classes Statistics](#)
 - ◆ [9.1 Traffic Classes](#)
 - ◆ [9.2 A closer look at the results](#)
 - ◆ [9.3 Active Probes](#)
- [10 Verify Enforcement](#)
 - ◆ [10.1 Policy-Based Routing used](#)

- ◆ 10.2 BGP Routes by default
- ◆ 10.3 Dynamic Route Maps generated
- 11 Add Delay on SP1
 - ◆ 11.1 Normal delay across SP1
 - ◆ 11.2 Adding a delay of 500ms on SP1
 - ◆ 11.3 Move TC to new exit ?
HOLDDOWN State
- 12 Conclusion

PfR Features that Enable Load Balancing

Link Utilization

Usage of this policy sets an upper threshold on the amount of traffic a specific link can carry. For example, if the upper threshold for a link is 90 % of total bandwidth, and it is currently running at 95 % of bandwidth, the link is Out-of-Policy (OOP). Cisco PfR will attempt to bring the link back into policy by repeatedly moving prefixes from the over-used link onto other exit links.

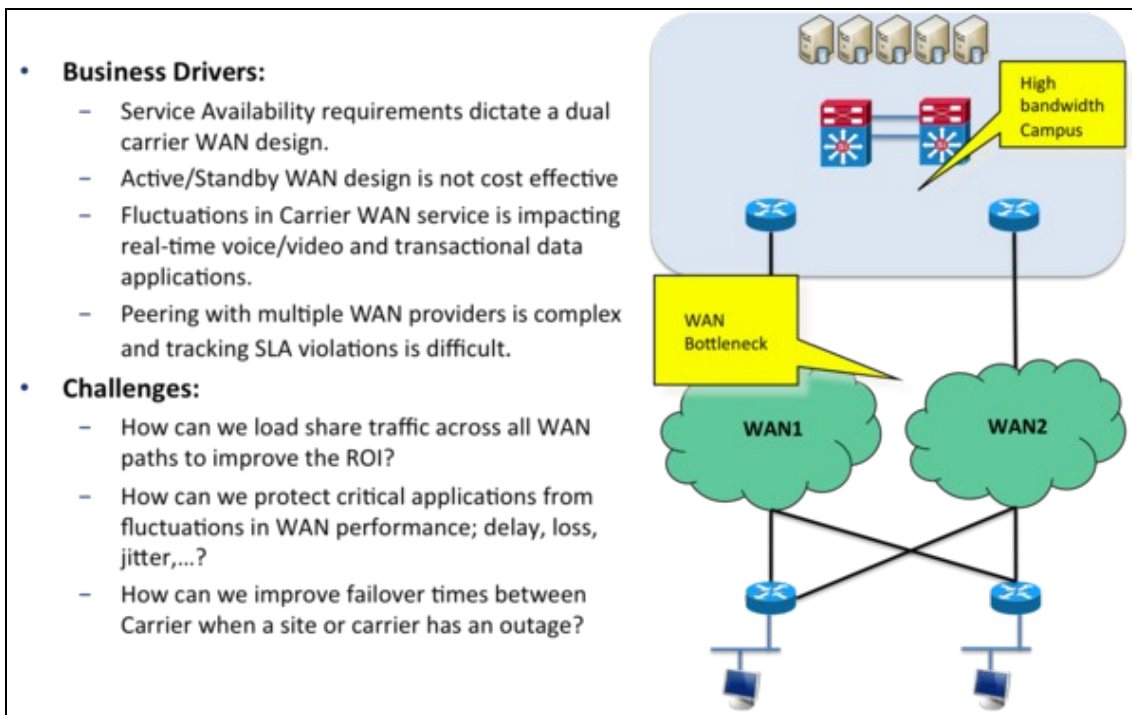
Range

Usage of this policy keeps all WAN links within a certain utilization range, relative to each other in order to ensure fair load-sharing across all concerned links. The range functionality allows the network administrator to instruct Cisco PfR to keep the usage on a set of exit links within a certain percentage range of each other. If the difference between the links becomes too great, Cisco PfR will attempt to bring the link back in to policy by distributing data traffic among the available exit links.

Traffic Class Performance

Usage of this policy enables the customer to define multiple paths that a set of traffic (ie voice traffic) could use as long as all the paths maintain the performance SLA ?s that are needed for that set of traffic. Hence, a policy that determines voice traffic to have a delay threshold of less than 250 msec can utilize multiple paths in the network if available, as long as all the paths deliver the traffic within its performance bounds.

Enterprise Needs



This solution is for an Enterprise which is dual-attached to two IP-VPN Service Providers. The load-balancing used here is taking place between external interfaces and specific policies are to be applied depending on the traffic type. In most (all) cases, Qos is already in place and classification/marketing procedures were studied a while back. Therefore packets entering on the Border Router are already classified and marked directly on the access switch connecting the station, IP Phone or multimedia terminal. That means PfR can probably use the dscp field as the most efficient way for the traffic profiling phase.

- **BUSINESS:** this group encompass all the critical applications. In most cases, these are transactional applications and by nature are delay intolerant. One of the key goal is to protect these applications and to be able to track variations in SLA. Traffic in this group is marked with DSCP=AF31.
- **BEST-EFFORT (BE):** traffic that has a low priority but could have a high bandwidth. Traffic in this group is marked with DSCP=0.
- **VOICE:** voice traffic that is delay intolerant and UDP based, which means that passive monitoring cannot be used here. Traffic in this group is marked with DSCP=EF.

PfR Solution Used

This solution describes a more advanced load sharing and will be based on automatic application classification based on DSCP, learning-list and specific policies per group of traffic.

Traffic is divided in 3 majors groups which will have specific policies applied. The main PfR features used in this configuration are the following (there are others but these ones are the most important):

Traffic Profiling:

PfR:Solutions:AdvancedLoadBalancing

- Learn-list: a flexible way to define the Traffic Classes, in this case BUSINESS, BE and VOICE
- Learning phase: automatic. Top talkers based on Netflow reports from the Border Routers
- BUSINESS group: match the critical traffic. PfR policies are based on utilization and delay. That's why we use resolve delay, resolve utilization and range. Also define the delay and utilization thresholds. The values defined here are just for lab purpose and are not firm recommendations.
- BE Group: match the low priority traffic. PfR policies are based on range and link utilization and disable delay.
- VOICE Group: match the voice traffic. PfR policies are based on delay only and disable the others.

Note: If the marking is done on the Border Router and it is also used as the classifier in PfR then it will not work. Because Netflow will see the packet before the marking and hence would not account for the BW and performance under correct TC. So marking should be done before it reaches to BR.

Measurement and optimization:

- BUSINESS

Monitoring mode is set to both which is the default. Passive monitoring based on Netflow and active probes when choosing new exits.

Policies based on delay, then range and utilization

- BE

Monitoring mode is set to both which is the default. Passive monitoring based on Netflow and active probes when choosing new exits.

Policies based on range then utilization

- VOICE

Monitoring mode is set to active throughput. Active probes are generated to check the path and netflow is used to gather the bandwidth usage per Traffic Class

Policies based on delay

Enforcement:

Compared to the Basic Load Balancing Solution (see [PfR:Solutions:BasicLoadBalancing](#)) where classification is based on prefixes, the classification described here is based on DSCP, therefore the only enforcement that can be used is Policy-based Routing.

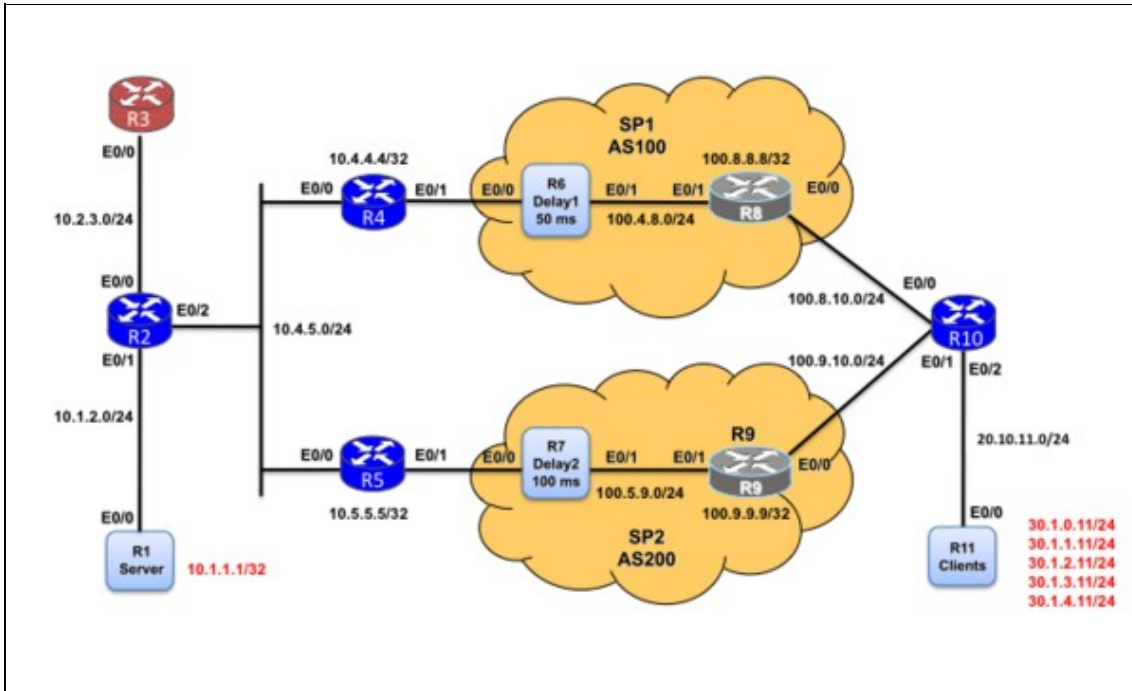
Performance Routing is Application Based

PfR will then dynamically generate dynamic route-map on the Border Routers.

PfR Network Topology Used

The central site has two Border Routers, connected to two separate Service Providers using eBGP. R2, R4 and R5 are iBGP peers.

- R3 is the Master Controller
- R4 and R5 the Border Routers
- R1 and R11 are packet generators that send/receive traffic (http, ssh, etc).
- Traffic Simulator tool is used between R1 and R11 to emulates traffic, both TCP and UDP.
- R6 and R7 are delay generators that add delay/loss to the path through SP1 and SP2. By default, 100 ms through SP1 and 50 ms through SP2.



PfR Components Configuration

```
pfr master
! =====
! For Each Traffic Class, Associate a Policy
! =====
!
policy-rules MYMAP
!
max-range-utilization percent 10
logging
!
! =====
! Border Routers Definition
! =====
!
border 10.4.5.4 key-chain pfr
interface Ethernet0/0 internal
interface Ethernet0/1 external
max-xmit-utilization percentage 90
!
border 10.4.5.5 key-chain pfr
interface Ethernet0/0 internal
interface Ethernet0/1 external
max-xmit-utilization percentage 90
```

PfR:Solutions:AdvancedLoadBalancing

```
!  
! =====  
! Learning based on learn-list  
! =====  
!  
learn  
  throughput  
  delay  
  periodic-interval 0  
  monitor-period 1  
  list seq 10 refname BUSINESS  
    traffic-class access-list BUSINESS  
    throughput  
  list seq 20 refname BE  
    traffic-class access-list BE  
    throughput  
  list seq 30 refname VOICE  
    traffic-class access-list VOICE  
    throughput  
!  
! =====  
! For test purpose, decrease the hold timer  
! =====  
!  
holddown 90  
mode route control  
periodic 180  
!  
!  
! =====  
! ACL used by learn-list  
! =====  
!  
ip access-list extended BE  
  permit ip any any dscp default  
ip access-list extended BUSINESS  
  permit ip any any dscp af31  
ip access-list extended VOICE  
  permit ip any any dscp ef  
!  
!  
! =====  
! Policies applied to Business TCs  
! resolve based on delay, range, then utilization  
! monitoring mode is both  
! =====  
!  
pfr-map MYMAP 10  
  match pfr learn list BUSINESS  
  set mode select-exit good  
  set delay threshold 200  
  set mode route control  
  set mode monitor both  
  set resolve delay priority 1 variance 20  
  set resolve range priority 5  
  set resolve utilization priority 10 variance 20  
  set probe frequency 30  
!  
!  
! =====  
! Policies applied to BE TCs  
! resolve based on range, then utilization  
! monitoring mode is both  
! =====
```

```
!  
pfr-map MYMAP 20  
  match pfr learn list BE  
  set mode select-exit good  
  set mode route control  
  set mode monitor both  
  set resolve range priority 5  
  set resolve utilization priority 10 variance 20  
  no set resolve delay  
  set probe frequency 30  
!  
!  
! =====  
! Policies applied to VOICE TCs  
! resolve based on delay only  
! monitoring mode is active throughput  
! =====  
!  
pfr-map MYMAP 30  
  match pfr learn list VOICE  
  set mode select-exit good  
  set delay threshold 200  
  set mode route control  
  set mode monitor active throughput  
  set resolve delay priority 1 variance 20  
  no set resolve range  
  no set resolve utilization  
  set probe frequency 30  
!
```

Flexible Netflow

While configuring Netflow is not a mandatory task for PfR to work, it allows to have a good understanding of the traffic flows across the border routers.

Flow Record Definition

```
!  
flow record MYRECORD  
  match ipv4 protocol  
  match ipv4 source address  
  match ipv4 destination address  
  match transport source-port  
  match transport destination-port  
  match interface input  
  collect ipv4 dscp  
  collect interface output  
  collect counter bytes  
  collect counter packets  
!
```

Flow Monitor Definition

```
flow monitor MYMONITOR  
  record MYRECORD  
!
```

And then apply the FNF Monitor on the interface

```
interface Ethernet0/0
 ip flow monitor MYMONITOR input
!
```

Checking Statistics and Flows

As explained before, explicitly enabling Netflow is not required for PfR to run but is a good practice to check active flows crossing the Border Routers, verify the ingress/egress interfaces used (must be internal to external or vice-versa).

Here is the output on R2 which sees all flows:

```
R2# sh flow monitor MYMONITOR cache format table
Cache type:                Normal
Cache size:                 4096
Current entries:           148
High Watermark:           169

Flows added:                38770
Flows aged:                38622
- Active timeout          ( 1800 secs)    14
- Inactive timeout        (   15 secs)   38608
- Event aged              0
- Watermark aged         0
- Emergency aged         0
```

| IPV4 SRC ADDR | IPV4 DST ADDR | TRNS SRC PORT | TRNS DST PORT | INTF INPUT | IP PROT | int |
|---------------|---------------|---------------|---------------|------------|---------|-------|
| ===== | ===== | ===== | ===== | ===== | ===== | ===== |
| 10.2.3.3 | 10.4.5.4 | 3949 | 31739 | Et0/0 | 6 | EtC |
| 10.2.3.3 | 10.4.5.5 | 3949 | 48115 | Et0/0 | 6 | EtC |
| 10.1.1.1 | 30.1.1.11 | 6001 | 51642 | Et0/1 | 17 | EtC |
| 10.1.1.2 | 30.1.2.11 | 6002 | 61192 | Et0/1 | 17 | EtC |
| 10.1.1.3 | 30.1.3.11 | 6003 | 50802 | Et0/1 | 17 | EtC |
| 10.1.1.4 | 30.1.4.11 | 6004 | 61090 | Et0/1 | 17 | EtC |
| 10.1.1.1 | 30.1.2.11 | 80 | 2012 | Et0/1 | 6 | EtC |
| 10.1.1.1 | 30.1.0.11 | 7000 | 7063 | Et0/1 | 6 | EtC |
| 10.1.1.1 | 30.1.2.11 | 7000 | 7064 | Et0/1 | 6 | EtC |

[snip]

Master Controller Verification

Before starting to look at the results, a few checks are needed to verify that the configuration is correct, that the learning is correctly defined and that policies are well defined and associated with learn-list.

Goal:

- Verify that the MC is active
- Verify that the learning process is enabled on the Master Controller
- Display the policy settings as well as the learning parameters and global timers.

Master Controller and Traffic Classes

The first step is to check the master controller configuration, verify the border routers, verify the parameters used (default and configured) and check the learn-list.

```
MC#sh pfr master
OER state: ENABLED and ACTIVE
  Conn Status: SUCCESS, PORT: 3949
  Version: 3.0
  Number of Border routers: 2
  Number of Exits: 2
  Number of monitored prefixes: 22 (max 5000)
  Max prefixes: total 5000 learn 2500
  Prefix count: total 22, learn 15, cfg 0
  PBR Requirements met
  Nbar Status: Inactive

Border          Status  UP/DOWN      AuthFail  Version
10.4.5.4        ACTIVE  UP           01:45:35  0        3.0
10.4.5.5        ACTIVE  UP           01:45:35  0        3.0

Global Settings:
max-range-utilization percent 10 rcv 0
mode route metric bgp local-pref 5000
mode route metric static tag 5000
trace probe delay 1000
logging
exit holddown time 60 secs, time remaining 0

Default Policy Settings:
backoff 300 3000 300
delay relative 50
holddown 90
periodic 180
probe frequency 56
number of jitter probe packets 100
mode route control
mode monitor both
mode select-exit good
loss relative 10
jitter threshold 20
mos threshold 3.60 percent 30
unreachable relative 50
resolve delay priority 11 variance 20
resolve range priority 12 variance 0
resolve utilization priority 13 variance 20
```

Learn Settings:

PfR:Solutions:AdvancedLoadBalancing

```
current state : STARTED
time remaining in current state : 67 seconds
throughput
delay
no inside bgp
monitor-period 1
periodic-interval 0
aggregation-type prefix-length 24
prefixes 100 appls 100
expire after time 720
```

```
Learn-List seq 10 refname BUSINESS
Configuration:
  Traffic-Class Access-list: BUSINESS
  Aggregation-type: prefix-length 24
  Learn type: throughput
  Session count: 50 Max count: 100
  Policies assigned: 10
  Status: ACTIVE
```

Stats:

```
Traffic-Class Count: 4
```

```
Learn-List seq 20 refname BE
```

```
Configuration:
  Traffic-Class Access-list: BE
  Aggregation-type: prefix-length 24
  Learn type: throughput
  Session count: 50 Max count: 100
  Policies assigned: 20
  Status: ACTIVE
```

Stats:

```
Traffic-Class Count: 7
```

```
Learn-List seq 30 refname VOICE
```

```
Configuration:
  Traffic-Class Access-list: VOICE
  Aggregation-type: prefix-length 24
  Learn type: throughput
  Session count: 50 Max count: 100
  Policies assigned: 30
  Status: ACTIVE
```

Stats:

```
Traffic-Class Count: 4
```

MC#

What to check:

- Both Border Routers are up and running
- Number of Monitored Prefixes: 22 and 15 are learned
- Global Settings: max-range-utilization is 10 (percentage of bandwidth difference between all exit interfaces)
- All default policy settings are displayed
- Learn is started (current state : STARTED)
- Then all learn-list are displayed
- For each learn-list, check the Traffic Class access-list, the policy number associated (which is the pfr-map it refers to).
- The Traffic Class count is also displayed which allows to check whether the learning process works well.

Display Learn-list

Because we use learn-list, there is a specific command to have a more detailed view:

```
MC#sh pfr master learn list

Learn-List seq 10 refname BUSINESS
Configuration:
  Traffic-Class Access-list: BUSINESS
  Aggregation-type: prefix-length 24
  Learn type: throughput
  Session count: 50 Max count: 100
  Policies assigned: 10
  Status: ACTIVE
Stats:
  Traffic-Class Count: 4
  Traffic-Class Learned:
    Appl Prefix 30.1.3.0/24 af31 256
    Appl Prefix 30.1.1.0/24 af31 256
    Appl Prefix 30.1.2.0/24 af31 256
    Appl Prefix 30.1.0.0/24 af31 256
Learn-List seq 20 refname BE
Configuration:
  Traffic-Class Access-list: BE
  Aggregation-type: prefix-length 24
  Learn type: throughput
  Session count: 50 Max count: 100
  Policies assigned: 20
  Status: ACTIVE
Stats:
  Traffic-Class Count: 7
  Traffic-Class Learned:
    Appl Prefix 20.10.11.0/24 defa 256
    Appl Prefix 30.1.0.0/24 defa 256
    Appl Prefix 30.1.2.0/24 defa 256
    Appl Prefix 30.1.3.0/24 defa 256
    Appl Prefix 30.1.1.0/24 defa 256
    Appl Prefix 30.1.5.0/24 defa 256
    Appl Prefix 30.1.4.0/24 defa 256
Learn-List seq 30 refname VOICE
Configuration:
  Traffic-Class Access-list: VOICE
  Aggregation-type: prefix-length 24
  Learn type: throughput
  Session count: 50 Max count: 100
  Policies assigned: 30
  Status: ACTIVE
Stats:
  Traffic-Class Count: 4
  Traffic-Class Learned:
    Appl Prefix 30.1.3.0/24 ef 256
    Appl Prefix 30.1.4.0/24 ef 256
    Appl Prefix 30.1.2.0/24 ef 256
    Appl Prefix 30.1.1.0/24 ef 256
MC#
```

For each group, this command displays the prefixes that are dynamically learned.

Policy Configuration

After the Master Controller and Traffic Classes verification, the third step is to check the policies associated with the Traffic Classes.

```
MC#sh pfr master policy
Default Policy Settings:
  backoff 300 3000 300
  delay relative 50
  holddown 90
  periodic 180
  probe frequency 56
  number of jitter probe packets 100
  mode route control
  mode monitor both
  mode select-exit good
  loss relative 10
  jitter threshold 20
  mos threshold 3.60 percent 30
  unreachable relative 50
  resolve delay priority 11 variance 20
  resolve range priority 12 variance 0
  resolve utilization priority 13 variance 20
oer-map MYMAP 10
  sequence no. 8444249301975040, provider id 1, provider priority 30
  host priority 0, policy priority 10, Session id 0
  match oer learn list BUSINESS
  backoff 300 3000 300
  *delay threshold 200
  holddown 90
  periodic 180
  *probe frequency 30
  number of jitter probe packets 100
  *mode route control
  *mode monitor both
  *mode select-exit good
  loss relative 10
  jitter threshold 20
  mos threshold 3.60 percent 30
  unreachable relative 50
  next-hop not set
  forwarding interface not set
  *resolve delay priority 1 variance 20
  *resolve range priority 5 variance 0
  *resolve utilization priority 10 variance 20
oer-map MYMAP 20
  sequence no. 8444249302630400, provider id 1, provider priority 30
  host priority 0, policy priority 20, Session id 0
  match oer learn list BE
  backoff 300 3000 300
  delay relative 50
  holddown 90
  periodic 180
  *probe frequency 30
  number of jitter probe packets 100
  *mode route control
  *mode monitor both
  *mode select-exit good
  loss relative 10
  jitter threshold 20
  mos threshold 3.60 percent 30
```

PfR:Solutions:AdvancedLoadBalancing

```
unreachable relative 50
next-hop not set
forwarding interface not set
*resolve range priority 5 variance 0
*resolve utilization priority 10 variance 20
oer-map MYMAP 30
  sequence no. 8444249303285760, provider id 1, provider priority 30
  host priority 0, policy priority 30, Session id 0
  match oer learn list VOICE
  backoff 300 3000 300
*delay threshold 200
  holddown 90
  periodic 180
*probe frequency 30
  number of jitter probe packets 100
*mode route control
*mode monitor active throughput
*mode select-exit good
  loss relative 10
  jitter threshold 20
  mos threshold 3.60 percent 30
  unreachable relative 50
  next-hop not set
  forwarding interface not set
*resolve delay priority 1 variance 20

* Overrides Default Policy Setting
MC#
```

Verify Traffic Classes Statistics

As soon as you see:

```
MC#
*Sep  3 16:01:22.924: %OER_MC-5-NOTICE: Prefix Learning WRITING DATA
*Sep  3 16:01:22.966: %OER_MC-5-NOTICE: Prefix Learning STARTED
MC#
```

You should be able to see the traffic classes on the Master Controller. You will need a few cycles before having all prefixes in INPOLICY state. You will start getting OOP (out of policy messages) from PfR regarding range, because R4 (10.4.5.4) is currently the only exit point while R5 (10.4.5.5) has no traffic:

```
MC#
*Sep  3 16:01:39.232: %OER_MC-5-NOTICE: Range OOP BR 10.4.5.4, i/f Et0/1, percent 76. Other BR 10.
*Sep  3 16:01:39.232: %OER_MC-5-NOTICE: Exit 10.4.5.4 intf Et0/1 OOP, Tx BW 383, Rx BW 54, Tx Load
MC#
```

Traffic Classes

On the Master Controller, you have all the Traffic Classes (in this case based on DSCP values) learnt as well as statistics:

PfR:Solutions:AdvancedLoadBalancing

Using show pfr master traffic-class:

```
MC#sh pfr master traffic-class
OER Prefix Statistics:
Pas - Passive, Act - Active, S - Short term, L - Long term, Dly - Delay (ms),
P - Percentage below threshold, Jit - Jitter (ms),
MOS - Mean Opinion Score
Los - Packet Loss (packets-per-million), Un - Unreachable (flows-per-million),
E - Egress, I - Ingress, Bw - Bandwidth (kbps), N - Not applicable
U - unknown, * - uncontrolled, + - control more specific, @ - active probe all
# - Prefix monitor mode is Special, & - Blackholed Prefix
% - Force Next-Hop, ^ - Prefix is denied

DstPrefix      Appl_ID Dscp Prot      SrcPort      DstPort SrcPrefix
      Flags          State      Time          CurrBR  CurrI/F Protocol
      PasSDly PasLDly  PasSUn  PasLUn  PasSLos  PasLLos  EBw  IBw
      ActSDly ActLDly  ActSUn  ActLUn  ActSJit  ActPMOS  ActSLos ActLLos
-----
330.1.0.0/24          N af31  256          N          N 0.0.0.0/0
      INPOLICY          76          10.4.5.5 Et0/1          PBR
      52          52          0          0          0          0          65          7
      52          51          0          0          N          N          N          N

30.1.1.0/24          N ef  256          N          N 0.0.0.0/0
      # INPOLICY*          @39          10.4.5.4 Et0/1          U
      U          U          0          0          0          0          9          9
      56          52          0          0          N          N          N          N

30.1.1.0/24          N defa  256          N          N 0.0.0.0/0
      INPOLICY          @33          10.4.5.4 Et0/1          PBR
      53          53          0          0          0          0          31          3
      U          54          0          0          N          N          N          N

30.1.1.0/24          N af31  256          N          N 0.0.0.0/0
      INPOLICY          @47          10.4.5.4 Et0/1          PBR
      53          53          0          0          0          0          35          4
      U          53          0          0          N          N          N          N

[snip]
MC#
```

A closer look at the results

Let's have a look at a few entries.

From the command ?show oer master traffic-class. Let's focus on a specific TC which belongs to BUSINESS group to understand the interesting values reported here:

```
MC#sh pfr master traffic-class
OER Prefix Statistics:
Pas - Passive, Act - Active, S - Short term, L - Long term, Dly - Delay (ms),
P - Percentage below threshold, Jit - Jitter (ms),
```

PfR:Solutions:AdvancedLoadBalancing

MOS - Mean Opinion Score

Los - Packet Loss (packets-per-million), Un - Unreachable (flows-per-million),

E - Egress, I - Ingress, Bw - Bandwidth (kbps), N - Not applicable

U - unknown, * - uncontrolled, + - control more specific, @ - active probe all

- Prefix monitor mode is Special, & - Blackholed Prefix

% - Force Next-Hop, ^ - Prefix is denied

| DstPrefix | Flags | Appl_ID | Dscp | Prot | SrcPort | DstPort | SrcPrefix | Protocol |
|-------------|---------|---------|----------|--------|---------|---------|----------------|----------|
| | PasSDly | PasLDly | PasSUn | PasLUn | PasSLos | PasLLos | EBw | IBw |
| | ActSDly | ActLDly | ActSUn | ActLUn | ActSJit | ActPMOS | ActSLos | ActLLos |
| 30.1.0.0/24 | | | N af31 | 256 | | N | N 0.0.0.0/0 | |
| | | | INPOLICY | | 76 | | 10.4.5.5 Et0/1 | PBR |
| | 52 | 52 | 0 | 0 | 0 | 0 | 65 | 7 |
| | 52 | 51 | 0 | 0 | N | N | N | N |

- Line1: prefix
- Line2: State of INPOLICY?this prefix is controlled by PfR and is currently inpolicy. PBR is used to enforce the path. Forwarding decisions based on other packet fields (such as TCP port numbers, DSCP field) cannot be done via the traditional routing table. For this reason, PfR will create a dynamic Policy Based Routing (PBR) policy and apply it to the PfR internal interfaces of the border routers.
- Line2: R5 is the exit point (SP2)

• Line3: Passive results

- Line3 (PasSDly, PasLDly): short-term and long-term passive delay, measured from the TCP Syn/Ack. As explained previously, TCP Syn/Ack are used to check the reachability of the prefix and to collect the delay and loss information. The delay is around 100 ms, which is the delay through SP2 (BR R5).
- Line3 (PasSUn, PasLUn): short-term and long-term statistics for Unreachable.
- Line3 (PasSLos, PasLLos): short-term and long-term statistics for Loss.
- Line3 (EBw/IBw): the bandwidth for this prefix in egress and ingress direction.

• Line4: Active results

- Line4 (ActSDly, ActLDly): short-term and long-term active delay, measured by active probes that are automatically learned.
- Line4 (ActSUn, ActLUn): short-term and long-term Unreachable results.
- Line4 (ActSLos, ActLLoss): short-term and long-term Loss results.

From the command ?show oer master traffic-class. Let?s focus on a specific TC which belongs to VOICE group to understand the interesting values reported here. Remember that the monitor mode used for this group is "active throughput".

PfR:Solutions:AdvancedLoadBalancing

```
MC#sh pfr master traffic-class
OER Prefix Statistics:
Pas - Passive, Act - Active, S - Short term, L - Long term, Dly - Delay (ms),
P - Percentage below threshold, Jit - Jitter (ms),
MOS - Mean Opinion Score
Los - Packet Loss (packets-per-million), Un - Unreachable (flows-per-million),
E - Egress, I - Ingress, Bw - Bandwidth (kbps), N - Not applicable
U - unknown, * - uncontrolled, + - control more specific, @ - active probe all
# - Prefix monitor mode is Special, & - Blackholed Prefix
% - Force Next-Hop, ^ - Prefix is denied
```

| DstPrefix | Flags | Appl_ID | Dscp | Prot | SrcPort | DstPort | SrcPrefix | Protocol |
|-------------|---------|---------|-----------|--------|---------|----------|-----------|----------|
| | PasSDly | PasLDly | PasSUn | PasLUn | PasSJos | PasLJos | EBw | IBw |
| | ActSDly | ActLDly | ActSUn | ActLUn | ActSJit | ActPMOS | ActSJos | ActLJos |
| 30.1.1.0/24 | | | N ef 256 | | N | N | 0.0.0.0/0 | |
| | # | | INPOLICY* | @39 | | 10.4.5.4 | Et0/1 | U |
| | U | U | 0 | 0 | 0 | 0 | 9 | 9 |
| | 56 | 52 | 0 | 0 | N | N | N | N |

- Line1: prefix
- Line2: State of INPOLICY*?The asteric (*) indicates this prefix is uncontrolled by PfR (the parent route controls routing) , but is currently inpolicy. The '#' indicates that mode active. The ?at sign? (@) on the Time Remaining value means the prefix is being actively probed. The numerical value is a countdown timer indicating when this state will expire. throughput is used here.
- Line2: R4 is the exit point (SP1)
- Line2: Not Applicable, this TC is uncontrolled and therefore follows the routing table for this prefix.

• Line3: Passive results

- Line3 (PasSDly, PasLDly): active mode used, so Not Applicable here.
- Line3 (PasSUn, PasLUn): active mode used, so Not Applicable here..
- Line3 (PasSJos, PasLJos): active mode used, so Not Applicable here.
- Line3 (EBw/IBw): mode active throughput is used, therefore the bandwidth for this prefix is reported in egress and ingress direction.

• Line4: Active results

- Line4 (ActSDly, ActLDly): short-term and long-term active delay, measured by active probes that are automatically learned here.
- Line4 (ActSUn, ActLUn): short-term and long-term Unreachable results.
- Line4 (ActSJos, ActLJos): short-term and long-term Loss results.

Active Probes

In this configuration active probes are automatically calculated and generated, but active probes can be manually defined per TC in the pfr-map section.

PfR:Solutions:AdvancedLoadBalancing

```
MC#sh pfr master active-probes
      OER Master Controller active-probes
Border   = Border Router running this Probe
State    = Un/Assigned to a Prefix
Prefix   = Probe is assigned to this Prefix
Type     = Probe Type
Target   = Target Address
TPort    = Target Port
How      = Was the probe Learned or Configured
N - Not applicable
```

The following Probes exist:

| State | Prefix | Type | Target | TPort | How | Codec |
|----------|---------------|------|-------------|-------|------|-------|
| Assigned | 20.10.11.0/24 | echo | 20.10.11.11 | N | Lrnd | N |
| Assigned | 30.1.2.0/24 | echo | 30.1.2.11 | N | Lrnd | N |
| Assigned | 30.1.3.0/24 | echo | 30.1.3.11 | N | Lrnd | N |
| Assigned | 30.1.1.0/24 | echo | 30.1.1.11 | N | Lrnd | N |
| Assigned | 30.1.5.0/24 | echo | 30.1.5.11 | N | Lrnd | N |
| Assigned | 30.1.4.0/24 | echo | 30.1.4.11 | N | Lrnd | N |
| Assigned | 30.1.0.0/24 | echo | 30.1.0.11 | N | Lrnd | N |

The following Probes are running:

| Border | State | Prefix | Type | Target | TPort |
|--------|-------|--------|------|--------|-------|
|--------|-------|--------|------|--------|-------|

MC#

Verify Enforcement

Policy-Based Routing used

Forwarding decisions based on other packet fields (in this case DSCP field) cannot be done via the traditional routing table. For this reason, PfR will create a dynamic Policy Based Routing (PBR) policy and apply it to the PfR internal interfaces of the border routers. But for PBR to be successful, BRs have to be adjacent either through a direct connection or via a GRE tunnel.

This requirement can be verified with the "show pfr master" command and looking for: PBR Requirements met

```
MC#sh pfr master
OER state: ENABLED and ACTIVE
Conn Status: SUCCESS, PORT: 3949
Version: 3.0
Number of Border routers: 2
Number of Exits: 2
Number of monitored prefixes: 22 (max 5000)
Max prefixes: total 5000 learn 2500
Prefix count: total 22, learn 15, cfg 0
PBR Requirements met          <---- Check this
Nbar Status: Inactive
```

PfR:Solutions:AdvancedLoadBalancing

| Border | Status | UP/DOWN | | AuthFail | Version |
|----------|--------|---------|----------|----------|---------|
| 10.4.5.4 | ACTIVE | UP | 03:09:01 | 0 | 3.0 |
| 10.4.5.5 | ACTIVE | UP | 03:09:02 | 0 | 3.0 |

BGP Routes by default

PfR being not active, the parent routes are BGP based on R2, R4 and R5. R4 is the preferred exit point as seen below:

```
R2#sh ip bgp
BGP table version is 27, local router ID is 10.2.2.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete
```

```
   Network          Next Hop          Metric LocPrf Weight Path
r i10.4.5.0/24      10.5.5.5          0      100    0 i
r>i                 10.4.4.4          0      100    0 i
*>i20.10.11.0/24    100.4.8.8         0      100    0 100 20 i
* i                 100.5.9.9         0      100    0 200 20 i
*>i20.11.11.11/32   100.4.8.8         0      100    0 100 20 i
* i                 100.5.9.9         0      100    0 200 20 i
*>i30.1.0.0/24      100.4.8.8         0      100    0 100 20 i
* i                 100.5.9.9         0      100    0 200 20 i
*>i30.1.1.0/24      100.4.8.8         0      100    0 100 20 i
* i                 100.5.9.9         0      100    0 200 20 i
*>i30.1.2.0/24      100.4.8.8         0      100    0 100 20 i
* i                 100.5.9.9         0      100    0 200 20 i
*>i30.1.3.0/24      100.4.8.8         0      100    0 100 20 i
* i                 100.5.9.9         0      100    0 200 20 i
*>i30.1.4.0/24      100.4.8.8         0      100    0 100 20 i
* i                 100.5.9.9         0      100    0 200 20 i
*>i30.1.5.0/24      100.4.8.8         0      100    0 100 20 i
   Network          Next Hop          Metric LocPrf Weight Path
* i                 100.5.9.9         0      100    0 200 20 i
r>i100.4.8.0/24     10.4.4.4          0      100    0 i
r>i100.5.9.0/24     10.5.5.5          0      100    0 i
*>i100.8.10.0/24    100.4.8.8         0      100    0 100 20 i
* i                 100.5.9.9         0      100    0 200 20 i
*>i100.9.10.0/24    100.4.8.8         0      100    0 100 20 i
* i                 100.5.9.9         0      100    0 200 20 i
R2#
```

Dynamic Route Maps generated

PfR creates several dynamic route-maps to enforce the path for BUSINESS, BE and VOICE application TCs. Depending on two dynamic ACLs, PBR enforce a next-hop to SP1 or SP2.

Let's check on the first border router R4:

Policy-Based Routing used

PfR:Solutions:AdvancedLoadBalancing

```
R4#sh route-map dynamic
route-map OER_INTERNAL_RMAP, permit, sequence 0, identifier 3640655874
  Match clauses:
    ip address (access-lists): oer#2
  Set clauses:
    ip next-hop 100.4.8.8
    interface Ethernet0/1
  Policy routing matches: 398009 packets, 232904449 bytes
route-map OER_INTERNAL_RMAP, permit, sequence 1, identifier 4060086275
  Match clauses:
    ip address (access-lists): oer#3
  Set clauses:
    ip next-hop 10.4.5.5
    interface Ethernet0/0
  Policy routing matches: 421999 packets, 250233694 bytes
Current active dynamic routemaps = 1
R4#
```

This route-map is applied on the ingress interface of the Border Router.

```
R4#sh ip policy
Interface      Route map
Ethernet0/0    OER_INTERNAL_RMAP (Dynamic)
R4#
```

The first route-map sets the next-hop to R8, which means traffic matching the ACL will be sent to SP1. The second route-map sets the next-hop to R5, which means that a packet received on R4 will be forwarded to the other BR for traffic that match the corresponding ACL.

The corresponding ACL are the following:

```
R4#sh ip access-lists dynamic
Extended IP access list oer#2
  2047 permit ip any 30.1.1.0 0.0.0.255 dscp default (14190 matches)
  9215 permit ip any 30.1.3.0 0.0.0.255 dscp default (6803 matches)
  32767 permit ip any 30.1.1.0 0.0.0.255 dscp af31 (36545 matches)
  65535 permit ip any 30.1.5.0 0.0.0.255 dscp default (74573 matches)
Extended IP access list oer#3
  2047 permit ip any 30.1.3.0 0.0.0.255 dscp af31 (2322 matches)
  4095 permit ip any 30.1.2.0 0.0.0.255 dscp af31 (6666 matches)
  8191 permit ip any 30.1.2.0 0.0.0.255 dscp default (8352 matches)
  262143 permit ip any 30.1.4.0 0.0.0.255 dscp default (73840 matches)
  1048575 permit ip any 30.1.0.0 0.0.0.255 dscp af31 (103854 matches)
R4#
```

And it's easy to understand and troubleshoot if needed.

A similar behavior can be observed on the second Border Router R5.

Add Delay on SP1

Normal delay across SP1

BUSINESS and VOICE groups have a delay resolver configured with an absolute threshold of 150 ms while BE group has only range and utilization. Increasing the delay on SP1 would then only impact business and voice traffic.

In this section, we'll take the SP1 path and add delay to it and observe what PfR does.

- For TCP sessions, PfR is able to detect the delay along the path using passive monitoring only. Because the monitoring mode is both, only the passive delay can trigger an OOP event.
- For UDP session, active mode is used and PfR generates probes. Active delay will trigger an OOP message.

PfR maintain delay statistics for the short-term delay (5 min) and the long-term delay (60 min).

Remember that there are two ways to specify what is flagged as an OOP condition in PfR:

- Relative: Relative implies the use of short/long term statistics. You specify a percentage and that is used to gauge if a TC is OOP.
- Absolute: You now have a threshold that is compared only to the short-term delay.

Before adding delay, let's verify the RTT from R4 across the SP1 link:

```
R4#ping 100.4.8.8
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 100.4.8.8, timeout is 2 seconds:
!!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 51/53/54 ms
R4#
```

Adding a delay of 500ms on SP1

Now, we are adding 500 ms delay on SP1. Again verify on R4:

```
R4#ping 100.4.8.8
Type escape sequence to abort.
Sending 5, 100-byte ICMP Echos to 100.4.8.8, timeout is 2 seconds:
!!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 502/504/508 ms
R4#
```

PfR:Solutions:AdvancedLoadBalancing

The delay threshold being defined in an absolute mode, is therefore compared to the short-term delay. By issuing "sh pfr master traffic-class", you can notice that the short-term delay (PasSDly, ActSDly) is increasing for traffic going over the SP1 path.

```
MC#sh pfr master traffic-class
OER Prefix Statistics:
Pas - Passive, Act - Active, S - Short term, L - Long term, Dly - Delay (ms),
P - Percentage below threshold, Jit - Jitter (ms),
MOS - Mean Opinion Score
Los - Packet Loss (packets-per-million), Un - Unreachable (flows-per-million),
E - Egress, I - Ingress, Bw - Bandwidth (kbps), N - Not applicable
U - unknown, * - uncontrolled, + - control more specific, @ - active probe all
# - Prefix monitor mode is Special, & - Blackholed Prefix
% - Force Next-Hop, ^ - Prefix is denied
```

| DstPrefix | Flags | Appl_ID | Dscp | Prot | SrcPort | DstPort | SrcPrefix | Protocol |
|-------------|---------|---------|-----------|--------|---------|----------|-------------|----------|
| | PasSDly | PasLDly | PasSUn | PasLUn | PasSLos | PasLLos | EBw | IBw |
| | ActSDly | ActLDly | ActSUn | ActLUn | ActSJit | ActPMOS | ActSLos | ActLLos |
| 30.1.0.0/24 | | | N af31 | 256 | | N | N 0.0.0.0/0 | |
| | | | INPOLICY | 107 | | 10.4.5.5 | Et0/1 | PBR |
| | 52 | 52 | | 0 | 0 | 0 | 66 | 7 |
| | 48 | 51 | | 0 | 0 | N | N | N |
| 30.1.1.0/24 | | | N defa | 256 | | N | N 0.0.0.0/0 | |
| | | | INPOLICY | 53 | | 10.4.5.4 | Et0/1 | PBR |
| | 169 | 70 | | 0 | 0 | 0 | 18 | 1 |
| | 507 | 143 | | 0 | 0 | N | N | N |
| 30.1.1.0/24 | | | N af31 | 256 | | N | N 0.0.0.0/0 | |
| | | | INPOLICY | @24 | | 10.4.5.4 | Et0/1 | PBR |
| | 79 | 55 | | 0 | 0 | 0 | 30 | 3 |
| | 503 | 74 | | 0 | 0 | N | N | N |
| 30.1.3.0/24 | | | N ef | 256 | | N | N 0.0.0.0/0 | |
| | # | | INPOLICY* | @43 | | 10.4.5.4 | Et0/1 | U |
| | U | U | | 0 | 0 | 0 | 9 | 9 |
| | 233 | 66 | | 0 | 0 | N | N | N |

Move TC to new exit ? HOLDDOWN State

After a few cycles we can notice that the Short Delay is now above 200 ms. The absolute delay threshold is compared against the short-term delay value and therefore trigger an OOP message for traffic classes belonging BUSINESS and VOICE groups going over the SP1 link:

```
*Oct 22 14:40:34.907: %OER_MC-5-NOTICE: Active ABS Delay OOP Appl Prefix 30.1.4.0/24 ef 256, del
*Oct 22 14:40:35.865: %OER_MC-5-NOTICE: Active ABS Delay OOP Appl Prefix 30.1.3.0/24 ef 256, del
*Oct 22 14:42:47.739: %OER_MC-5-NOTICE: Passive ABS Delay OOP Appl Prefix 30.1.2.0/24 af31 256, de
```

PfR:Solutions:AdvancedLoadBalancing

When OOP condition is detected, PfR tries to find an alternate path. When PfR makes a route change to a specific TC, this TC will move to the HOLDDOWN state to avoid erratic behavior. The timer associated is the holddown timer defined in the pfr master configuration. If the current path is the best path but it is OOPOLICY then it will go to OOPOLICY state. A syslog message indicates that a new exit is found. The primary path being OOP, PfR moves BUSINESS and VOICE Traffic Classes to the SP2 path.

```
*Oct 22 14:41:07.919: %OER_MC-5-NOTICE: Route changed Appl Prefix 30.1.2.0/24 ef 256, BR 10.4.5.
*Oct 22 14:41:32.356: %OER_MC-5-NOTICE: Route changed Appl Prefix 30.1.4.0/24 ef 256, BR 10.4.5.
*Oct 22 14:41:33.305: %OER_MC-5-NOTICE: Route changed Appl Prefix 30.1.3.0/24 ef 256, BR 10.4.5.

*Oct 22 14:43:45.040: %OER_MC-5-NOTICE: Route changed Appl Prefix 30.1.2.0/24 af31 256, BR 10.4.5.
```

As seen below, BUSINESS and VOICE traffic classes are now all on SP2 path while BE Traffic Classes stay INPOLICY and therefore stay on the SP1 path.

```
MC#sh pfr master traffic-class
```

```
OER Prefix Statistics:
```

```
Pas - Passive, Act - Active, S - Short term, L - Long term, Dly - Delay (ms),
P - Percentage below threshold, Jit - Jitter (ms),
MOS - Mean Opinion Score
Los - Packet Loss (packets-per-million), Un - Unreachable (flows-per-million),
E - Egress, I - Ingress, Bw - Bandwidth (kbps), N - Not applicable
U - unknown, * - uncontrolled, + - control more specific, @ - active probe all
# - Prefix monitor mode is Special, & - Blackholed Prefix
% - Force Next-Hop, ^ - Prefix is denied
```

| DstPrefix | Flags | Appl_ID | Dscp | Prot | SrcPort | DstPort | SrcPrefix | Protocol |
|-------------|---------|---------|----------|--------|---------|----------|-----------|----------|
| | PasSDly | PasLDly | PasSUn | PasLUn | PasSJos | PasLJos | EBw | IBw |
| | ActSDly | ActLDly | ActSUn | ActLUn | ActSJit | ActPMOS | ActSJos | ActLJos |
| 30.1.0.0/24 | | | N af31 | 256 | N | N | 0.0.0.0/0 | |
| | | | INPOLICY | 18 | | 10.4.5.5 | Et0/1 | PBR |
| | 52 | 52 | 0 | 0 | 0 | 0 | 66 | 7 |
| | 52 | 51 | 0 | 0 | N | N | N | N |
| 30.1.1.0/24 | | | N ef | 256 | N | N | 0.0.0.0/0 | |
| | # | | HOLDDOWN | 82 | | 10.4.5.5 | Et0/1 | PBR |
| | U | U | 0 | 0 | 0 | 0 | 9 | 9 |
| | U | U | 0 | 0 | N | N | N | N |
| 30.1.1.0/24 | | | N defa | 256 | N | N | 0.0.0.0/0 | |
| | | | INPOLICY | @14 | | 10.4.5.4 | Et0/1 | PBR |
| | 61 | 54 | 0 | 0 | 0 | 0 | 24 | 3 |
| | 510 | 102 | 0 | 0 | N | N | N | N |
| 30.1.2.0/24 | | | N ef | 256 | N | N | 0.0.0.0/0 | |
| | # | | HOLDDOWN | 75 | | 10.4.5.5 | Et0/1 | PBR |
| | U | U | 0 | 0 | 0 | 0 | 9 | 9 |
| | U | U | 0 | 0 | N | N | N | N |

Conclusion

Performance Routing is an advanced technology that allows multiple deployments, one of them being the one described in this document. The configuration is straightforward and very efficient. Based on that, a customer can evolve to a more complex Performance Routing solution or can fine-tune the Traffic Class definition.