PfR Internet Presence - Outbound Load Balancing

Navigation

- Go to PfR home page
- Go to PfR Solution Guides home page

Contents

Internet Presence and Load Balancing

This use case is the most common deployment scenario as this is the primary customer use case the PfR technology was developed to address; optimization of large numbers of client devices sourced from several ISP connections. In terms of megabits per second, the bulk of the user traffic is from server to client. PfR, therefore, is configured and addresses the path selection from server to client over two or more links to typically multiple ISPs. It's also possible to load-balance on ingress using BGP.

Applications continue to rely increasingly on distributed sources of data and information and they consume more and more bandwidth. Performance Routing provides bi-directional, traffic-class performance and link utilization based load distribution to enhance network and application performance.

Key Advantages of using PfR for Load balancing:

- Utilization based load-balancing: PfR takes real-time link utilization into account when load balancing the links. This will ensure that a link will not go beyond a certain percentage of its maximum capacity (75% by default).
- Application Performance based Load Balancing: PfR does not randomly forward traffic through one link or another. It takes application performance requirements into consideration and then forwards the traffic through a link which meets the performance policy requirements. PfR also load balances the link at the same time.
- Bi-directional Solution: PfR is a bi-directional load balancing solution which influences outbound as well as in-bound traffic.
- Consolidated Centralized View: PfR offers consolidated and centralized view of the state of all external links in the network. At any given time, the network administrator can see the current link utilization (in kbps and percentage of its capacity), maximum link threshold, and the policies applied to the links in the network.

PfR Solution Used

The PfR configuration deployed is simple passive monitoring of traffic and dedicated chassis for the control function. The load-balancing used here is taking place between external interfaces and will be based on the top talker prefixes.

This deployment does not use active probes. For PfR to verify reachability for a destination network prefix, TCP traffic must be observed on more than one exit interface so PfR has more than one exit with validated

reachability to the target network prefix.

BGP is used between the site and all upstream ISPs, using:

- Default routing or
- Partial routes or
- Full routes

In all the above 3 cases, the PfR configuration will remain the same.

PfR Solution Highlights:

- Learning phase: automatic. Top talkers based on Netflow reports from the Border Routers
- Measuring phase: mode monitor passive. which means PfR only supports passive data collection based on Netflow monitoring to optimize outbound traffic.
- Policies:

  utilization: Link utilization for each exit interface on the Border Routers should not exceed 90% of the defined bandwidth
  range: traffic should be load-balanced so that the average bandwidth range between external interfaces should be maintain within 10%.

- Enforcement: load-balancing based on prefixes and BGP used on the uplinks toward the Service Providers. PfR will enforce the path my modifying the BGP local-pref attribute for controlled prefixes.

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.

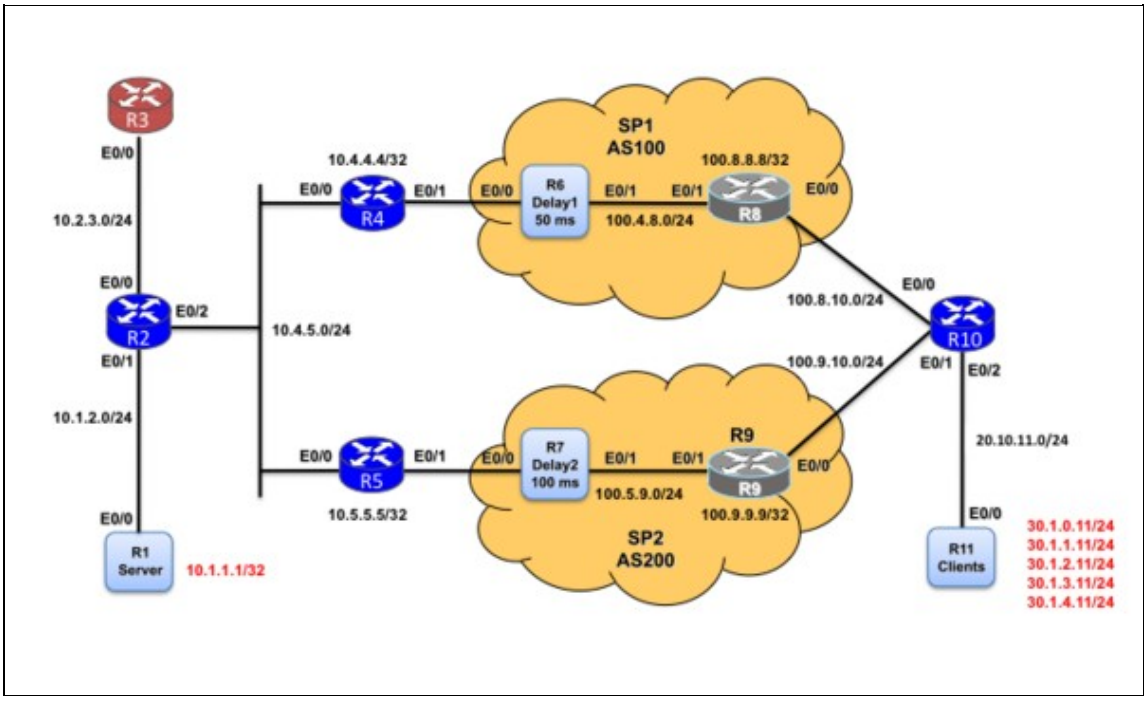
Range

Usage of this policy keeps some or 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 with in 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.

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. For an Internet Presence solution, it may be recommended to have a dedicated Master Controller given the possible high number of prefixes that have to be optimized and managed.

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



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. The following configuration is just an example of a flow monitor definition.

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:                              34
  High Watermark:                              433

  Flows added:                              3709977
  Flows aged:                               3709943
    - Active timeout      (  1800 secs)       9092
    - Inactive timeout    (    15 secs)    3700851
    - 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
===============  ===============  =============  =============  ====================  =======  ===
10.2.3.254       10.4.5.4                 3949          31294  Et0/0                       6  Et0
10.2.3.254       10.4.5.5                 3949          48078  Et0/0                       6  Et0
10.2.3.254       10.4.5.4                 3949          12936  Et0/0                       6  Et0
10.1.1.1         30.1.0.11                7000           7008  Et0/1                       6  Et0
10.1.1.2         30.1.1.11                  25           1032  Et0/1                       6  Et0
10.1.1.3         30.1.2.11                  80           2002  Et0/1                       6  Et0
10.2.3.254       10.4.5.5                 3949          58627  Et0/0                       6  Et0
10.1.1.4         30.1.3.11                3007           3012  Et0/1                       6  Et0
10.1.1.4         40.1.0.11                4004           4000  Et0/1                       6  Et0
10.1.1.3         30.1.4.11                  80           2003  Et0/1                       6  Et0
10.1.1.1         30.1.0.11                7000           7009  Et0/1                       6  Et0
```

```
10.1.1.2        30.1.2.11                 25          1033  Et0/1                     6  Et0
10.1.1.3        30.1.3.11                 80          2004  Et0/1                     6  Et0
10.1.1.4        40.1.0.11               4005          4000  Et0/1                     6  Et0
10.1.1.3        30.1.4.11                 80          2005  Et0/1                     6  Et0
10.1.1.1        30.1.0.11               7000          7010  Et0/1                     6  Et0

[snip]
```

Display Routing Table

Let's have a look at the routing table before applying a PfR configuration. Remote sites have prefixes 30.1.0.0/16 and 40.1.0.0/16. For clarity, only interesting part matching the destination prefixes of the routing tables are displayed.

On R2:

```
R2#sh ip route bgp
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      30.0.0.0/24 is subnetted, 11 subnets
B        30.1.0.0 [200/0] via 100.5.9.9, 00:06:33
B        30.1.1.0 [200/0] via 100.5.9.9, 00:06:33
B        30.1.2.0 [200/0] via 100.5.9.9, 00:06:33
B        30.1.3.0 [200/0] via 100.5.9.9, 00:06:33
B        30.1.4.0 [200/0] via 100.5.9.9, 00:06:33
B        30.1.5.0 [200/0] via 100.5.9.9, 00:06:33
B        30.1.6.0 [200/0] via 100.5.9.9, 00:06:33
B        30.1.7.0 [200/0] via 100.5.9.9, 00:06:33
B        30.1.8.0 [200/0] via 100.5.9.9, 00:06:33
B        30.1.9.0 [200/0] via 100.5.9.9, 00:06:33
B        30.1.10.0 [200/0] via 100.5.9.9, 00:06:33
      40.0.0.0/24 is subnetted, 1 subnets
B        40.1.0.0 [200/0] via 100.5.9.9, 00:06:33
R2#
```

This table clearly shows that only one exit point is used. BGP chooses one best path for the remote prefixes.

On the Border Router R4:

```
R4#sh ip route bgp
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
```

```
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      30.0.0.0/24 is subnetted, 11 subnets
B        30.1.0.0 [200/0] via 100.5.9.9, 00:10:21
B        30.1.1.0 [200/0] via 100.5.9.9, 00:10:21
B        30.1.2.0 [200/0] via 100.5.9.9, 00:10:21
B        30.1.3.0 [200/0] via 100.5.9.9, 00:10:21
B        30.1.4.0 [200/0] via 100.5.9.9, 00:10:21
B        30.1.5.0 [200/0] via 100.5.9.9, 00:10:21
B        30.1.6.0 [200/0] via 100.5.9.9, 00:10:21
B        30.1.7.0 [200/0] via 100.5.9.9, 00:10:21
B        30.1.8.0 [200/0] via 100.5.9.9, 00:10:21
B        30.1.9.0 [200/0] via 100.5.9.9, 00:10:21
B        30.1.10.0 [200/0] via 100.5.9.9, 00:10:21
      40.0.0.0/24 is subnetted, 1 subnets
B        40.1.0.0 [200/0] via 100.5.9.9, 00:10:21
R4#
```

On the Border Router R5:

```
R5#sh ip route bgp
Codes: L - local, C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route, H - NHRP, l - LISP
       + - replicated route, % - next hop override

Gateway of last resort is not set

      30.0.0.0/24 is subnetted, 11 subnets
B        30.1.0.0 [20/0] via 100.5.9.9, 00:11:20
B        30.1.1.0 [20/0] via 100.5.9.9, 00:11:20
B        30.1.2.0 [20/0] via 100.5.9.9, 00:11:20
B        30.1.3.0 [20/0] via 100.5.9.9, 00:11:20
B        30.1.4.0 [20/0] via 100.5.9.9, 00:11:20
B        30.1.5.0 [20/0] via 100.5.9.9, 00:11:20
B        30.1.6.0 [20/0] via 100.5.9.9, 00:11:20
B        30.1.7.0 [20/0] via 100.5.9.9, 00:11:20
B        30.1.8.0 [20/0] via 100.5.9.9, 00:11:20
B        30.1.9.0 [20/0] via 100.5.9.9, 00:11:20
B        30.1.10.0 [20/0] via 100.5.9.9, 00:11:20
      40.0.0.0/24 is subnetted, 1 subnets
B        40.1.0.0 [20/0] via 100.5.9.9, 00:11:20
R5#
```

In order to achieve load-balancing, one can look at the prefix utilization and manually apply route-map to try to load-balance the traffic across all exits. While this is possible, PfR brings a lot of advanced features while keeping the configuration very simple. Load-balancing is achieved by a few lines of configuration and PfR takes care of the dynamic bandwidth utilization.

PfR Components Configuration

Here is the Master Controller configuration:

```
!
key chain pfr
 key 0
  key-string cisco
!
pfr master
 max-range-utilization percent 10
 logging
 !
 ! ========================
 ! Border Routers Definition
 ! ========================
 !
 border 10.4.5.5 key-chain pfr
  interface Ethernet0/1 external
   max-xmit-utilization percentage 90
  interface Ethernet0/0 internal
 !
 border 10.4.5.4 key-chain pfr
  interface Ethernet0/1 external
   max-xmit-utilization percentage 90
  interface Ethernet0/0 internal
 !
 !
 ! ===================================
 ! Learning based on highest throughput
 ! ===================================
 !
 learn
  throughput
  periodic-interval 0
  monitor-period 1
  prefixes 1000 applications 0
  expire after time 60
 !
 ! ========================
 ! Policies
 ! ========================
 !
 max prefix total 20000 learn 20000
 mode route control
 periodic 600
 mode route control
 mode monitor passive
 periodic 600
 resolve utilization priority 1 variance 5
 resolve range priority 2
 no resolve delay
 !
```

GLOBAL

- logging : Enable PfR Syslogs (can be checked using show logging).

LEARN

- Learn mode is configured on the master controller. The border routers learns network prefixes from the NetFlow cache and break this learning step into one minute intervals specified by the monitor-period keyword.
- The periodic-interval value of 0 indicates the border routers immediately begin relearning prefixes after the monitor-period has expired and the collection of prefixes have been reported to the master controller.
- The border routers summarize or aggregate flows into memory based on the value specified by the 'aggregation-type' keyword. By default prefixes are summarized on a length of /24.

- throughput: learn top prefixes based on throughput
- monitor-period 1: BRs exports statistics every 1 minute
- periodic-interval 0: infinite
- prefixes 1000: limit the number of prefixes to learn per period to 1000
- expire after time 60: delete prefix if not relearned in 60 Minutes

OPTIMIZATION

- As the master controller has received the observed flows from all the border routers, there are several other keywords that govern how these prefixes are managed.
- The master controller must sort the collection of learned prefixes from the most current period from all border routers along with prefixes learned previously. The max prefix keyword determines the total number of prefixes stored by the master controller, as well as the maximum number of learned prefixes.
- The other method a prefix may be managed is by reference in a pfr-map through a prefix-list referenced from a match traffic-class statement. This is the static configuration method. This would account for the difference between the total value and the learned value. The expire after time keyword is a means of removing prefixes from the collection if no new traffic is observed in the number of minutes defined on the keyword. In other words, it is a means of aging and removing older entries which are no longer active.

- mode route control: PfR control the routes
- mode monitor passive: using active probes over the Internet is not a recommended solution as may be dropped along the path
- max-range-utilization percent 10 : Start Load balancing when exits links utilization differ more than 10%.
- max-xmit-utilization percentage 90 : upper threshold on the amount of traffic a specific link can carry.
- max prefix total 20000: the maximum number of prefixes to manage in the PfR database

- resolve: Check utilization then range.

  if utilization on a given external link is more than 90%, then the link is Out Of Policy
  if utilization on an external link is more than 10% greater than the range across all of the links, PfR will detect a Range OOP condition.

- periodic 600: policies are reevaluated every 600 sec

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.

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).

You may have to wait a few cycles for the MC to be able to learn a few prefixes and then issue the command:

```
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: 7 (max 20000)
  Max prefixes: total 20000 learn 20000
  Prefix count: total 7, learn 7, cfg 0
  PBR Requirements met
  Nbar Status: Inactive

Border           Status   UP/DOWN             AuthFail  Version
10.4.5.5         ACTIVE   UP       01:20:57          0  3.0
10.4.5.4         ACTIVE   UP       01:22:03          0  3.0

Global Settings:
  max-range-utilization percent 10 recv 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 300
  periodic 600
  probe frequency 56
  number of jitter probe packets 100
  mode route control
  mode monitor passive
  mode select-exit good
  loss relative 10
  jitter threshold 20
  mos threshold 3.60 percent 30
  unreachable relative 50
  resolve utilization priority 1 variance 5
  resolve range priority 2 variance 0

Learn Settings:
  current state : STARTED
  time remaining in current state : 90 seconds
  throughput
  no delay
  no inside bgp
  monitor-period 1
  periodic-interval 0
  aggregation-type prefix-length 24
  prefixes 1000 appls 0
  expire after time 60
MC#
```

What to check:

- Both Border Routers are up and running
- Number of Monitored Prefixes: 7 (max 20000) - 7 prefixes monitored for a maximum of 20000
- Max prefixes: total 20000 learn 20000 - The max number of monitored prefixes, the max number of learnt prefixes.
- Prefix count: total 7, learn 7, cfg 0 - 7 prefixes automatically learnt using NetFlow, 0 statically configured
- Global Settings: max-range-utilization is 10 (percentage of bandwidth difference between all exit interfaces)
- Learn Settings: prefixes 1000 appls 0 - 1000 prefixes max automatically learnt per cycle
- All default policy settings are displayed
- Learn is started (current state : STARTED)

On the Master Controller, check the details for each Border Routers:

```
MC#sh pfr master border detail
Border           Status   UP/DOWN             AuthFail  Version
10.4.5.4         ACTIVE   UP       00:15:44          0  3.0
 Et0/0           INTERNAL UP
 Et0/1           EXTERNAL UP

 External            Capacity      Max BW   BW Used    Load Status          Exit Id
 Interface            (kbps)       (kbps)    (kbps)    (%)
 ---------           --------      ------   ------- ------- ------          ------
```

```
 Et0/1           Tx       300         270       199      66 UP                    4
                 Rx                   300        55      18
------------------------------------------------------------------------------------
Border           Status   UP/DOWN             AuthFail  Version
10.4.5.5         ACTIVE   UP       00:15:47          0  3.0
 Et0/0           INTERNAL UP
 Et0/1           EXTERNAL UP

 External            Capacity      Max BW   BW Used    Load Status          Exit Id
 Interface            (kbps)       (kbps)    (kbps)    (%)
 ---------           --------      ------   ------- ------- ------           ------
 Et0/1           Tx       300         270       225      75 UP                    3
                 Rx                   300         0       0
MC#
```

Border Routers

On the Border Routers, you can also check the connectivity with the Master Controller:

```
R4#sh pfr border
OER BR 10.4.5.4 ACTIVE, MC 10.2.3.254 UP/DOWN: UP 00:14:50,
  Auth Failures: 0
  Conn Status: SUCCESS
  OER Netflow Status: ENABLED, PORT: 3949
  Version: 3.0  MC Version: 3.0
  Exits
  Et0/0           INTERNAL
  Et0/1           EXTERNAL
R4#
```

```
R5#sh pfr border
OER BR 10.4.5.5 ACTIVE, MC 10.2.3.254 UP/DOWN: UP 00:15:31,
  Auth Failures: 0
  Conn Status: SUCCESS
  OER Netflow Status: ENABLED, PORT: 3949
  Version: 3.0  MC Version: 3.0
  Exits
  Et0/0           INTERNAL
  Et0/1           EXTERNAL
R5#
```

Policy Configuration

After the Master Controller and Traffic Classes verification, the third step is to check the policies associated with the Traffic Classes. In this solution guide, we have defined global policies that apply to all Traffic

Classes:

```
MC#sh pfr master policy
Default Policy Settings:
  backoff 300 3000 300
  delay relative 50
  holddown 300
  periodic 600
  probe frequency 56
  number of jitter probe packets 100
  mode route control
  mode monitor passive
  mode select-exit good
  loss relative 10
  jitter threshold 20
  mos threshold 3.60 percent 30
  unreachable relative 50
  resolve utilization priority 1 variance 5
  resolve range priority 2 variance 0
MC#

What to check:
* Check that mode route control is on. ie PfR is controlling the routes
* Check the policies (utilization then range).
```

Verify Load balancing

As soon as you see:

```
MC#
*Mar 25 15:14:33.544: %OER_MC-5-NOTICE: Prefix Learning WRITING DATA
*Mar 25 15:14:33.618: %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 R5 (10.4.5.5) is currently the only exit point while R4 (10.4.5.4) has no traffic:

```
MC#
*Mar 25 15:14:44.360: %OER_MC-5-NOTICE: Range OOP BR 10.4.5.5, i/f Et0/1, percent 93. Other BR 10.
*Mar 25 15:14:44.360: %OER_MC-5-NOTICE: Load OOP BR 10.4.5.5, i/f Et0/1,  load 279 policy 270
*Mar 25 15:14:44.360: %OER_MC-5-NOTICE: Exit 10.4.5.5 intf Et0/1 OOP, Tx BW 279, Rx BW 0, Tx Load
MC#
*Mar 25 15:14:58.389: %OER_MC-5-NOTICE: Route changed Prefix 30.1.3.0/24, BR 10.4.5.4, i/f Et0/1,
MC#
```

Bandwidth used on exit links

The first step is probably to verify the accuracy of the load-balancing scheme on both Border Routers R4 and R5.

```
MC#sh pfr master border detail
Border           Status   UP/DOWN             AuthFail  Version
10.4.5.4         ACTIVE   UP       00:37:31          0  3.0
 Et0/0           INTERNAL UP
 Et0/1           EXTERNAL UP

 External            Capacity      Max BW   BW Used    Load Status          Exit Id
 Interface            (kbps)       (kbps)    (kbps)    (%)
 ---------           --------      ------   ------- ------- ------           ------
 Et0/1           Tx       300         270       199      66 UP                    4
                 Rx                   300        58      19
-------------------------------------------------------------------------------------
Border           Status   UP/DOWN             AuthFail  Version
10.4.5.5         ACTIVE   UP       00:37:33          0  3.0
 Et0/0           INTERNAL UP
 Et0/1           EXTERNAL UP

 External            Capacity      Max BW   BW Used    Load Status          Exit Id
 Interface            (kbps)       (kbps)    (kbps)    (%)
 ---------           --------      ------   ------- ------- ------           ------
 Et0/1           Tx       300         270       208      69 UP                    3
                 Rx                   300         0       0
MC#
```

Traffic Classes

On the Master Controller, you have all the Traffic Classes (in this case prefixes) learnt as well as statistics. 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
-------------------------------------------------------------------------------
30.1.0.0/24              N    N    N           N           N N
                       INPOLICY*       77          10.4.5.5 Et0/1           U
             104      104        0        0        0        0       55        6
               N        N        N        N        N        N        N        N

40.1.0.0/24              N    N    N           N           N N
                       INPOLICY      254          10.4.5.4 Et0/1           BGP
              52       53        0        0        0        0       65        7
               N        N        N        N        N        N        N        N

30.1.1.0/24              N    N    N           N           N N
                       INPOLICY*       68          10.4.5.5 Et0/1           U
```

```
             104      104        0        0        0        0       55        6
               N        N        N        N        N        N        N        N

30.1.2.0/24              N    N    N          N           N N
                       INPOLICY*       76          10.4.5.5 Et0/1           U
             104      104        0        0        0        0       53        6
               N        N        N        N        N        N        N        N

30.1.3.0/24              N    N    N          N           N N
                       INPOLICY      342          10.4.5.4 Et0/1           BGP
              52       54        0        0        0        0       66        7
               N        N        N        N        N        N        N        N

30.1.4.0/24              N    N    N          N           N N
                       INPOLICY*       62          10.4.5.5 Et0/1           U
             104      104        0        0        0        0       55        6
               N        N        N        N        N        N        N        N

30.1.5.0/24              N    N    N          N           N N
                       INPOLICY      314          10.4.5.4 Et0/1           BGP
              52       53        0        0        0        0       67        7
               N        N        N        N        N        N        N        N

MC#
```

A closer look at the results

If we look at 2 specific prefixes 30.1.0.0/24 and 40.1.0.0/24, we can see that PfR controls one, while the other is under the parent route control.

From the command ?show oer master traffic-class", let?s focus on the prefix 30.1.0.0/24 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),
 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
-------------------------------------------------------------------------------
30.1.0.0/24              N    N    N          N           N N
                       INPOLICY*       77          10.4.5.5 Et0/1           U
             104      104        0        0        0        0       55        6
               N        N        N        N        N        N        N        N

[SNIP]

MC#
```

In looking at the detail display of the prefix, several items bear notice:

- Line1: prefix
- Line2: State of INPOLICY*

  The asteric (*) indicates this prefix is uncontrolled by PfR (the parent route controls routing), but is currently inpolicy.

- **Line3: Passive results**
- Line3 (PasSDly, PasLDly): short-term and long-term passive delay, measured from the TCP Syn/Ack. TCP Syn/Ack messages 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 ISP2 (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** - Not Applicable here

Now, let?s focus on the prefix 40.1.0.0/24:

```
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
------------------------------------------------------------------------------

[snip]

40.1.0.0/24               N    N    N           N           N N
                          INPOLICY      254          10.4.5.4 Et0/1           BGP
              52       53        0        0        0        0       65        7
              N        N         N        N        N        N        N        N


[snip]

MC#
```

In looking at the detail display of the prefix, several items bear notice:

- Line1: prefix
- Line2: State of INPOLICY?this prefix is controlled by PfR and is currently inpolicy. BGP is used to enforce the path.

- **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 50 ms, which is the delay through ISP1 (BR R4).
- 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** - Not applicable here

Verify Enforcement

BGP Route Table on R2

R2 is an iBGP peer for both border routers and as such as the BGP route table. After a few cycles, PfR controls the prefixes and modifies BGP local-pref to enforce the path to the appropriate Border Routers.

```
R2#sh ip bgp
BGP table version is 83, 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, m multipath, b backup-path, x best-external
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path

[snip]

*>i30.1.0.0/24      100.5.9.9                0    500      0 200 20 i
*>i30.1.1.0/24      100.5.9.9                0    500      0 200 20 i
*>i30.1.2.0/24      100.5.9.9                0    500      0 200 20 i
*>i30.1.3.0/24      100.4.8.8                0   5000      0 100 20 i
*>i30.1.4.0/24      100.5.9.9                0    500      0 200 20 i
*>i30.1.5.0/24      100.4.8.8                0   5000      0 100 20 i
*>i30.1.6.0/24      100.5.9.9                0    500      0 200 20 i
*>i30.1.7.0/24      100.5.9.9                0    500      0 200 20 i
*>i30.1.8.0/24      100.5.9.9                0    500      0 200 20 i
*>i30.1.9.0/24      100.5.9.9                0    500      0 200 20 i
*>i30.1.10.0/24     100.5.9.9                0    500      0 200 20 i
*>i40.1.0.0/24      100.4.8.8                0   5000      0 100 20 i
```

```
[snip]

R2#
```

- The prefix 30.1.0.0/24 uncontrolled by PfR has a local-preference of 500 (which is the one defined on the R5 bgp peer) toward exit BR R5. Thus value is unchanged by PfR. Depending of the IOS release used, PfR will sometimes always control the prefixes even if the exit BR is the default one calculated by BGP. If this is the case, then a local-pref of 5000 would be assigned toward BR R5.
- The prefix 40.1.0.0/24 controlled by PfR has a local-preference of 5000 (this is the default value assigned by PfR and can be changed in the PfR global configuration) toward exit BR R4.

Border Routers

Let?s have a look at the border routers.

Let?s begin with R5:

```
R5#sh pfr border routes bgp
BGP table version is 92, local router ID is 10.5.5.5
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale, m multipath, b backup-path, x best-external
Origin codes: i - IGP, e - EGP, ? - incomplete
OER Flags: C - Controlled, X - Excluded, E - Exact, N - Non-exact, I - Injected

   Network          Next Hop        OER    LocPrf Weight Path
*>i30.1.3.0/24      100.4.8.8       XN      5000      0 100 20 i
*>i30.1.5.0/24      100.4.8.8       XN      5000      0 100 20 i
*>i40.1.0.0/24      100.4.8.8       XN      5000      0 100 20 i
R5#
```

The ?X? under the OER column for the 40.1.0.0/24 route on R5 means that the route is not locally controlled. Meaning that the local preference 5000 is being injected from another router. When the ?X? attribute is set, the exact vs. non-exact is meaningless.

If we look at this route on R4, we will see that it is locally controlled, and the exact route is controlled. The ?exact? means that the 40.1.0.0/24 route was found in the BGP table and there are no more specific subnets underneath:

```
R4#sh pfr border route bgp
BGP table version is 89, local router ID is 10.4.4.4
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              r RIB-failure, S Stale, m multipath, b backup-path, x best-external
Origin codes: i - IGP, e - EGP, ? - incomplete
OER Flags: C - Controlled, X - Excluded, E - Exact, N - Non-exact, I - Injected

   Network          Next Hop        OER    LocPrf Weight Path
*> 30.1.3.0/24      100.4.8.8       CE              0 100 20 i
*> 30.1.5.0/24      100.4.8.8       CE              0 100 20 i
*> 40.1.0.0/24      100.4.8.8       CE              0 100 20 i
R4#
```

Conclusion

One of the basic solution provided by PfR is to be able to load-balance traffic among multiple exit interface. The configuration is straightforward and very efficient. Based on this simple load-balancing configuration, it's possible to add more complex policies that take into account the Traffic Class performance requirements. Delay or loss are two key policies that may added to the existing utilization and range policies.