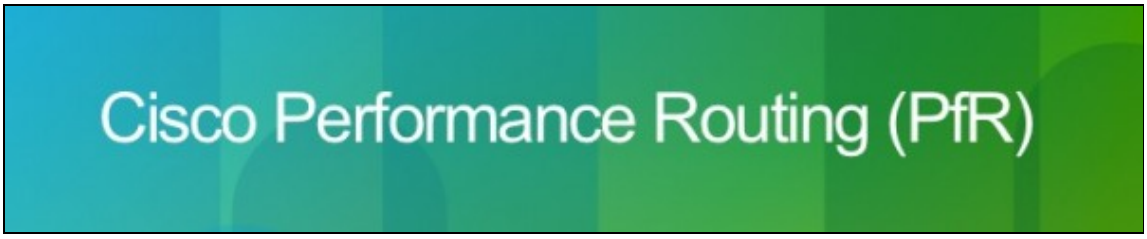


PfR Technology Overview

Navigation

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

Contents

- 1 Moving to Performance Routing
  - 1.1 Better Use of WAN Links
  - 1.2 Better Application performance and availability
  - 1.3 Application Routing
  - 1.4 Performance Routing
- 2 PfR Technical Overview
  - 2.1 General
  - 2.2 Performance Routing Policy Engine
  - 2.3 Reachability must be verified
- 3 Learning Phase
  - 3.1 Definition
  - 3.2 Manual Configuration
  - 3.3 Automatic Configuration
  - 3.4 How to use the learn-list traffic-class criteria
    - 3.4.1 Example1 - if you want to learn only 10.1.0.0/16 prefixes
    - 3.4.2 Example2 - Correct use of ACL with learn list
- 4 Measuring Phase
  - 4.1 Mode monitor passive
  - 4.2 Mode monitor active
  - 4.3 Hybrid Modes
    - 4.3.1 Mode monitor both
    - 4.3.2 Mode monitor Fast
- 5 Apply Policy Phase
  - 5.1 PfR Features that Enable Load Balancing
    - 5.1.1 Link Utilization

◊ 5.1.2 Range
◊ 5.1.3 Traffic Class Performance
♦ 5.2 PfR Policy Thresholds
♦ 5.3 Resolvers
♦ 5.4 PfR Traffic Class States
♦ 5.5 Configuration Examples
- 6 Enforce Phase
- 7 Verify Phase
- 8 Best Practices
♦ 8.1 Load Interval and Bandwidth
♦ 8.2 Passive mode

Moving to Performance Routing

Better Use of WAN Links

Many customers have multiple links to the WAN, being privately held or managed by a Service Provider. In many cases, one link is considered as primary while the other one is considered as a backup which leads to a waste of resources.

Today's challenges in the WAN:

- Traffic increase and badly load-balanced across central site uplinks
- Expensive backup links under utilized
- Load-sharing through complex BGP tricks or PBR
- Manual check of Netflow stats, IP SLA probes and Prefix utilization

What would be the solution:

- Full utilization of expensive network resources
- Efficient distribution of traffic based upon load
- Traffic optimized based upon circuit cost profiles
- Minimization of underutilized expensive WAN paths

Better Application performance and availability

Many customers would like to be aware of fluctuating service levels, Soft Errors or Brownouts. Routing protocols are not aware of such events. This could lead to a bad user experience, to application performance degradation. But still, the routing protocol is up.

- blackout: it is 100% loss of data (control plane is still UP).
- brownout: it is subjective - It is a high loss and very high delay condition.

Challenges:

- Application performance degraded

  Bad user experience
  But no routing problems reported

- Unaware of Soft Errors

  Intermittent reachability (SP flapping)
  Black holes
  New application requirements (low latency)
  But again ? no routing problems reported

- Where is the problem?

  WAN Congestion Problem?
  ISP?

Solution:

- Avoidance of network brownouts and soft errors
- Intelligent Routing based on performance
- Moving to adaptative routing based upon application delay, loss, jitter, bandwidth

Application Routing

One of the key requirement is to be able to route not only based on the prefix but also based on the type of application. Routing should take the best path for an application rather than a prefix.

Challenges:

- Critical application routed on primary link for best SLA
- Use backup links for best effort applications
- But still .. Use backup in case of performance degradation

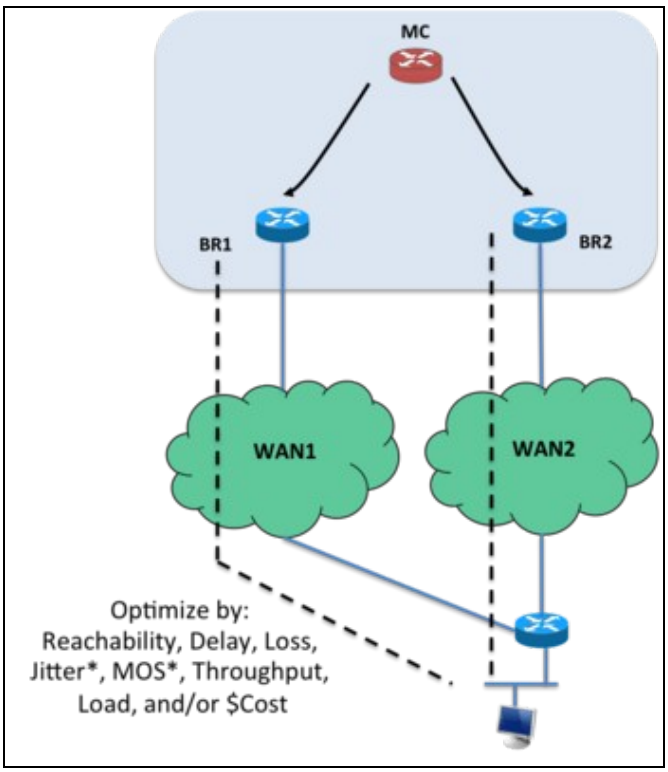
Current Deployment:

- Complex configuration based on EEM scripts and Policy Based Routing (PBR).
- Static routing

Solution:

- Responsiveness to critical application performance requirements
- Time/delay sensitive: voice, vertical apps (trading floor)
- Loss sensitive: video, circuit emulation
- Data center traffic: SAN extension, Internet ISP load balancing
- Transactional traffic: e-commerce transactions, automated B2B, ERP

Performance Routing



Traditional routing uses static metrics to provide ?reachability? information, but has no link utilization information, no data plane knowledge and is unaware of soft errors.

Cisco Performance Routing (PfR) enhances routing in order to select the best path based on user defined policy. The PfR policy can: minimize cost efficiently distribute traffic load and/or select the optimum performing path for applications

Cisco PfR enables intelligent traffic management that can dynamically route around soft errors in the Enterprise WAN or Internet and makes adaptive routing adjustments based on advanced criteria such as response time, packet loss, jitter, mean opinion score (MOS), availability, traffic load, and cost policies.

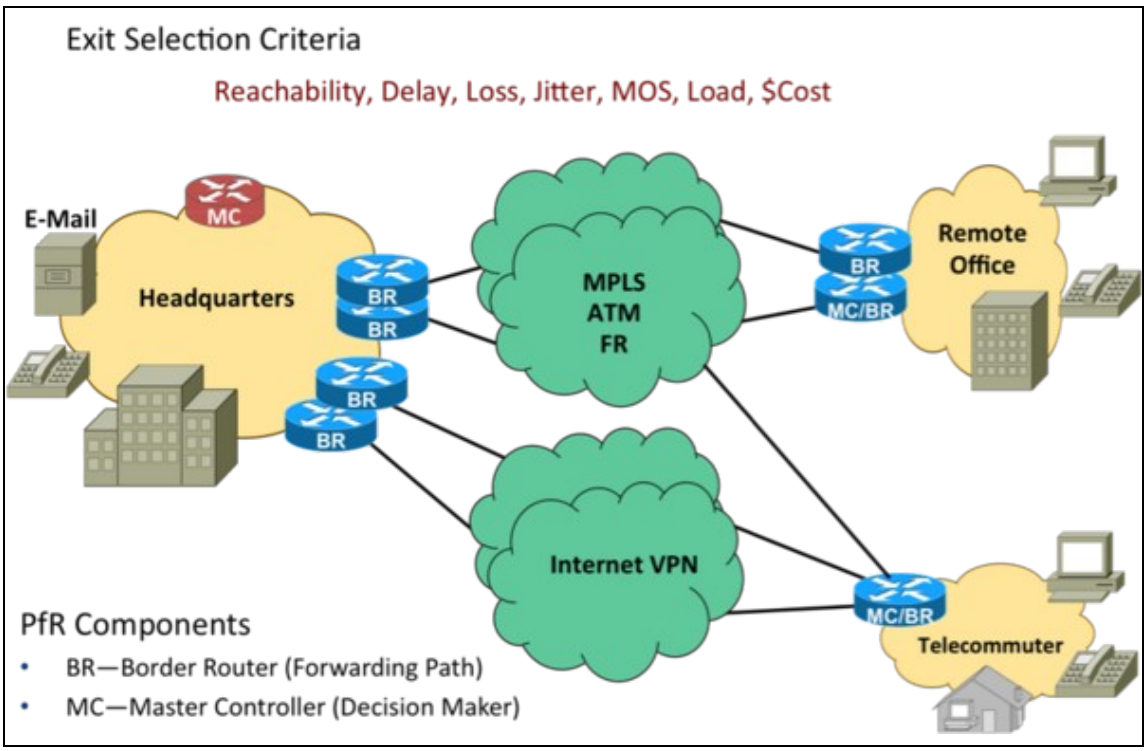
Cisco PfR augments traditional routing (BGP, EIGRP, OSPF, PBR) with realtime performance metrics in IOS instrumentation (Interfaces, Netflow, IP SLA, ?).

PfR Technical Overview

General

In any PfR implementation, a master controller (MC) and at least one border router (BR) must be configured. The MC commands and controls the BRs and maintains a central repository for the data collected by the BRs.

The MC function can be collocated (configured) on the same router as the BR, or it can be a dedicated, standalone IOS platform..



BRs are in the user traffic switching path. BRs collect data from their Netflow cache and from the IP SLA probes results, provide a degree of aggregation of this information, and influence the packet switching path to manage user traffic.

The MC is the policy decision maker. Typically, at a large site (data center or campus), the MC is a standalone chassis while at smaller branch locations the MC is typically collocated (configured) on the same platform as the BR. As a general rule, the large locations manage more network prefixes and/or applications than a branch deployment and thus consumes more CPU and memory resources for the MC function. Therefore, it makes a good design practice to dedicate a chassis for the MC at the large sites. The branch typically manages fewer active network prefixes and/or applications and due to the costs associated with dedicating a chassis at each branch, the network manager can collocate the MC and BR on the same platform. CPU and memory utilization should be monitored on platforms operating as MCs and if utilization is high, the network manager should consider an MC platform with a higher capacity CPU and memory.

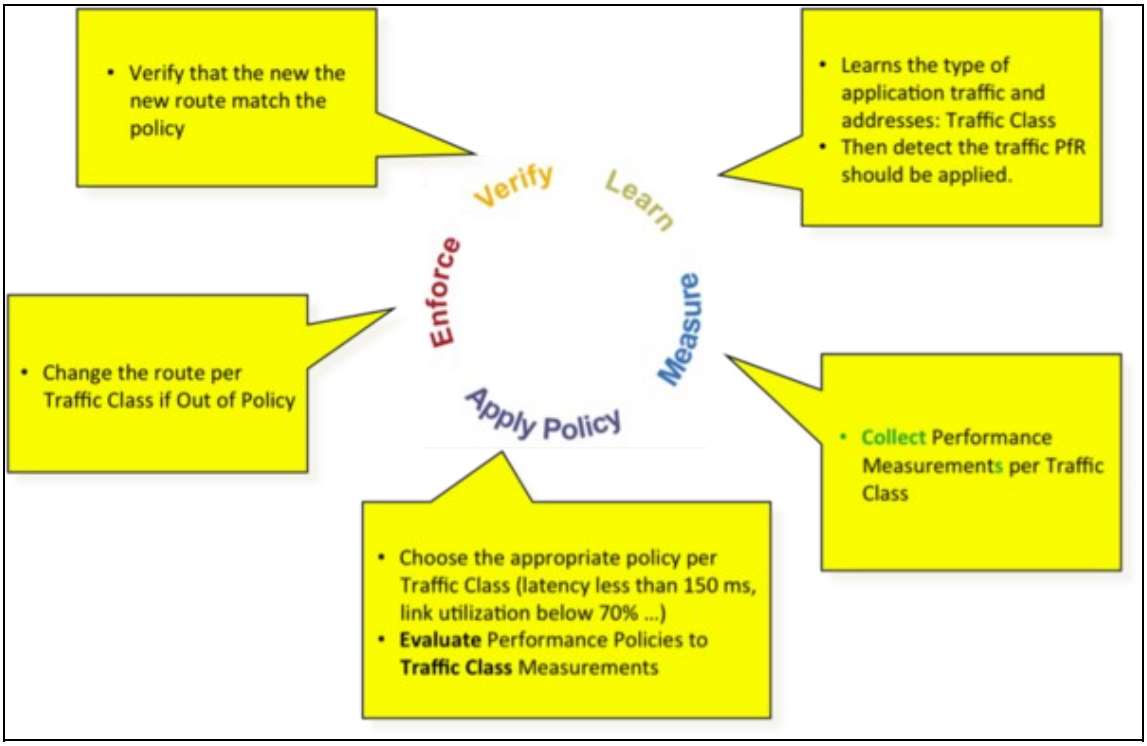
The MC communicates with the BRs over an authenticated TCP socket, but has no requirement for populating its own IP routing table with anything more than a route to reach the BRs.

Because PfR is a path selection technology, there must be at least two external interfaces under the control of PfR and at least one internal interface. There must be at least one BR configured. If there is only one BR configured, then both external interfaces are attached to the single BR. If more than one BR is configured, then the two or more external interfaces are configured across these BRs. External links, or exit points, are therefore owned by the BR; they may be logical (tunnel interfaces) or physical links (Serial, Ethernet, ...).

For PfR to be able to manage prefixes or application, traffic should cross the BRs from internal interfaces to external interfaces and vice-versa.

Performance Routing Policy Engine

Performance based routing happens in five stages:



- Learn Applications: MC tells BR to learn ?interesting? applications, called Traffic Classes.

  This could be destination prefix with or without port, dscp, source prefix or even application using NBAR.
  This profiling process can be entirely automatic based on the top talkers (using Netflow) or configured manually.

- Measure Application performance (Collects traffic class statistics for learned applications):

  Monitor Modes: Passive, Active, Both, Fast, Special (Cat6K)
  Netflow for UDP (bandwidth) and TCP flows (availability, delay, bandwidth, loss)
  IP SLA for TCP and UDP flows (Availability, delay, loss, jitter, MOS).

- Apply Policy

  Use measured application data to determine whether managed traffic-class is out of policy (OOP) and if an alternate path can meet the policy requirements.

- Enforce (re-route traffic)

  Prefix Control: Inject BGP or Static routes
  Application Control: Dynamic Route-map/PBR for traffic classes defined by ACLs, NBAR, unsupported routing protocols (OSPF, ISIS) or, BRs running a mix of routing protocols.

- Verify that the new route match the policy

Reachability must be verified

For PfR to consider an exit interface as a candidate for traffic, reachability to the target network prefix must be verified. When PfR is configured as passive mode (mode monitor passive, Netflow based only), TCP flows must be present across the exit interface to learn the validity of reachability across the exit interface.

Note that a parent route needs to be present to direct traffic for a target network out the external interfaces, in order to allow the Netflow subsystem to identify the validity of reachability through the TCP flows. Given this, if there is no TCP traffic out an exit interface, no passive measurements are available to Netflow/PfR. Or, if there are long-lived TCP flows, flows lasting longer than the PfR monitor period, no TCP SYNc and TCP SYNc/ACK are seen during the monitor period. So in this case, traffic may be active, but because the TCP SYNc and TCP SYNc/ACK is not seen during the monitor period, no delay and reachability can be deduced from this long persistent flow.

For PfR to function optimally in passive monitor mode, more TCP flows equate to more data points for the master controller to analyze and manage. As the number of TCP flows increase, the database becomes more granular, meaning more delay, loss, and reachability information is available for a given network prefix.

Learning Phase

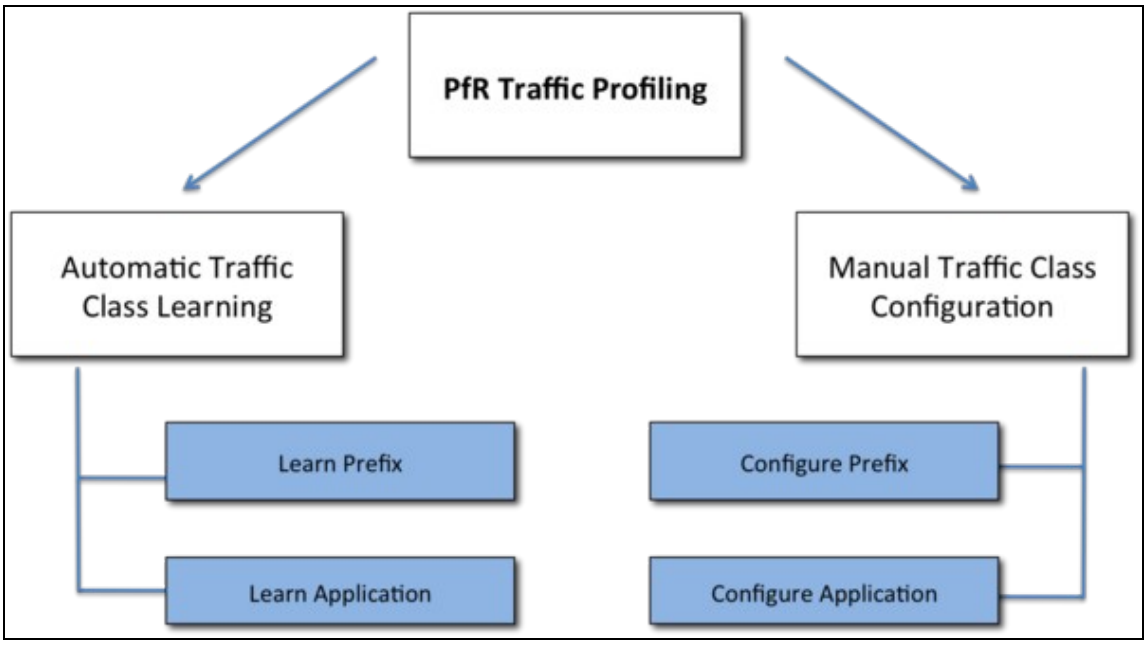
Definition

Also called Profiling phase. Before optimizing traffic, PfR has to determine the traffic classes from the traffic flowing through the border routers. To optimize traffic routing, subsets of the total traffic must be identified, and these traffic subsets are named traffic classes. The list of traffic classes entries is named a Monitored Traffic Class list.

Traffic Classes can be based on layer3 information such as prefixes, or layer4 such as port number, on DSCP values or application by using Network Based Application Recognition (NBAR).

Type		Example
Destination Prefix (Mandatory)		10.0.0.0/8 20.1.1.0/24
Application (Optional)	ACL	10.1.1.0/24 dscp ef 10.1.1.0/24 dst-port 50
	Well-Known	10.1.1.0/24 telnet 20.1.0.0/16 ssh
	NBAR	10.1.1.0/24 nbar RTP 20.1.1.0/24 nbar citrix 12.4(20)T

The entries in the Monitored Traffic Class list can be profiled either by automatically learning the traffic flowing through the device (using Flexible Netflow) or by manually configuring the traffic classes. Learned and configured traffic classes can both exist in the MTC list at the same time. The PfR profile phase includes both the learn mechanism and the configure mechanism.



Manual Configuration

In this mode, the prefixes or applications are directly and manually entered into the PfR database. There is no learning needed. As soon as you add the prefix-list, the prefixes are entered into the PfR database.

The following example shows the ability to define specific prefixes:

```
pfr master
 policy-rules MYMAP
 logging
 border 10.4.5.4 key-chain key1
  interface Ethernet0/0 internal
  interface Ethernet0/1 external
 !
 border 10.4.5.5 key-chain key1
  interface Ethernet0/0 internal
  interface Ethernet0/1 external
!
ip prefix-list BRANCH1 seq 5 permit 30.1.0.0/16
!
pfr-map MYMAP 10
 match ip address prefix-list BRANCH1
!
```

The following example shows the ability to define specific applications:

```
pfr master
 policy-rules MYMAP
 logging
 border 10.4.5.4 key-chain key1
  interface Ethernet0/0 internal
  interface Ethernet0/1 external
 !
 border 10.4.5.5 key-chain key1
  interface Ethernet0/0 internal
  interface Ethernet0/1 external
!
ip prefix-list FILTER_BRANCH1 seq 10 permit 10.1.1.0/24
ip prefix-list FILTER_BRANCH1 seq 20 permit 10.1.2.0/24
!
ip prefix-list FILTER_BRANCH2 seq 10 permit 10.1.3.0/24
ip prefix-list FILTER_BRANCH2 seq 20 permit 10.1.4.0/24
!
! Define FTP application
ip access-list extended MY_APP 10
  permit tcp any any eq 21
!
pfr-map MYMAP 10
  match traffic-class application telnet prefix-list FILTER_BRANCH1
  set mode select-exit good
  set delay threshold 2000
  set mode route control
  set mode monitor both
  no set resolve delay
  set active-probe echo 10.1.1.10
!
pfr-map MYMAP 20
  match traffic-class application http prefix-list FILTER_BRANCH2
  set mode select-exit good
  set delay threshold 2000
  set mode route control
  set mode monitor both
  no set resolve delay
!
pfr-map MYMAP 30
   traffic-class access-list MY_APP filter FILTER_BRANCH2
  set mode select-exit good
  set delay threshold 2000
  set mode route control
  set mode monitor both
  no set resolve delay
  set active-probe echo 10.1.3.10
!
```

This mode is used when you know the prefixes or applications to optimize and want to explicitly list them.

Automatic Configuration

In this mode, Performance Routing (PfR) has to determine the traffic classes from the traffic flowing through the border routers. This is the ?learn? block of the PfR configuration.

A very basic automatic learning configuration would be the following, where PfR automatically tracks the top talkers based on netflow information received from the BRs:

```
key chain key1
 key 1
 key-string cisco
!
oer master
 logging
 border 10.4.5.4 key-chain key1
  interface Ethernet0/0 internal
  interface Ethernet0/1 external
 !
 border 10.4.5.5 key-chain key1
  interface Ethernet0/0 internal
  interface Ethernet0/1 external
 !
 learn
  throughput
  monitor 1
  periodic 0
!
```

learn-list

PfR supports a learn list configuration mode to simplify the learning of traffic classes (TC) and to provide greater flexibility. In each learn list, different criteria including prefixes, application definitions, filters, and aggregation parameters for learning traffic classes can be configured. A traffic class is automatically learned by PfR based on each learn list criteria, and each learn list is configured with a sequence number. The sequence number determines the order in which learn list criteria are applied. Learn lists allow different PfR policies to be applied to each TC.

```
ip prefix-list BRANCH1 seq 5 permit 30.1.0.0/16
!
oer master
 max-range-utilization percent 10
 policy-rules MYMAP
 logging
 learn
  throughput
  delay
  periodic-interval 0
  monitor-period 1
  list seq 10 refname BRANCH_BUSINESS
   traffic-class application ssh filter BRANCH1
   throughput
  list seq 20 refname BRANCH_BE
   traffic-class prefix-list BRANCH1
   throughput
 holddown 180
 mode select-exit best
 periodic 180
!
```

How to use the learn-list traffic-class criteria

Access-list under learn-list are meant for defining application signatures and not for filtering.

- if you want to learn a specific subset of prefixes, then you can *only* use prefix-list.

- if you want to learn a specific subset of applications, then you will use access-list with an optional filter list for prefix

Example1 - if you want to learn only 10.1.0.0/16 prefixes

Correct

```
!
ip prefix-list FILTER seq 10 permit 10.1.0.0/16
!
pfr master
 learn
  list seq 10 refname TEST
   traffic-class prefix-list FILTER
!
```

Incorrect

```
!
ip access-list extended FILTER_ACL 10
 permit ip any 10.1.0.0 0.0.255.255
!
pfr master
 learn
  list seq 10 refname TEST
   traffic-class access-list FILTER
!
```

Example2 - Correct use of ACL with learn list

access-list are used to define applications.

```
! Configure App signature only once.
! FTP application
ip access-list extended MY_APP 10
  permit tcp any any eq 21
!
ip prefix-list FILTER_BRANCH1 seq 10 permit 10.1.1.0/24
ip prefix-list FILTER_BRANCH1 seq 20 permit 10.1.2.0/24
!
ip prefix-list FILTER_BRANCH2 seq 10 permit 10.1.3.0/24
ip prefix-list FILTER_BRANCH2 seq 20 permit 10.1.4.0/24
!
!
pfr master
 learn
  list seq 10 refname BRANCH1
```

```
   traffic-class access-list MY_APP filter FILTER_BRANCH1
  list seq 10 refname BRANCH2
   traffic-class access-list MY_APP filter FILTER_BRANCH2
!
```

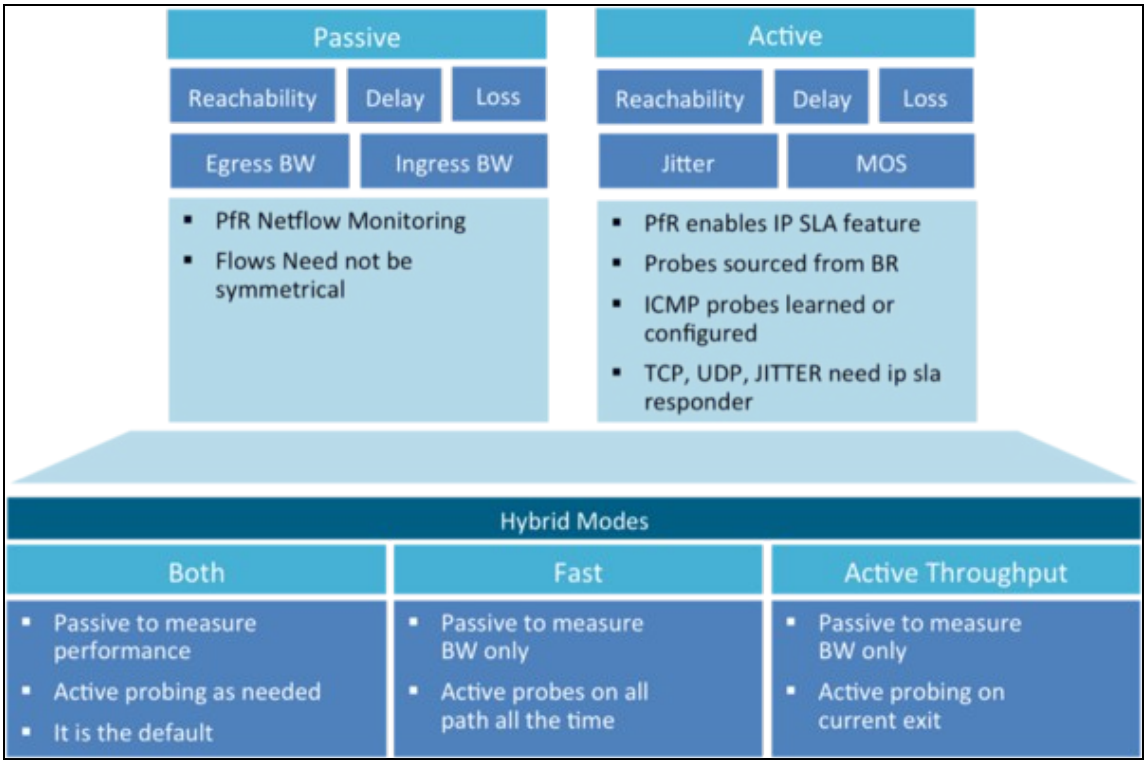
Measuring Phase

The PfR measure phase is the second step in the PfR performance engine. The Traffic Class list is now full of traffic class entries and PfR must measure the performance metrics of these traffic class entries.

Measuring is defined in PfR as the act of measurement performed periodically over a set interval of time where the measurements are compared against a user defined threshold.

PfR measures the performance of traffic classes using active and passive monitoring techniques and it also measures, by default, the utilization of links.

The border routers collect passive monitoring and active monitoring statistics and then transmit this information to the master controller. The PfR measure phase is complete when each traffic class entry in the MTC list has associated performance metric measurements.

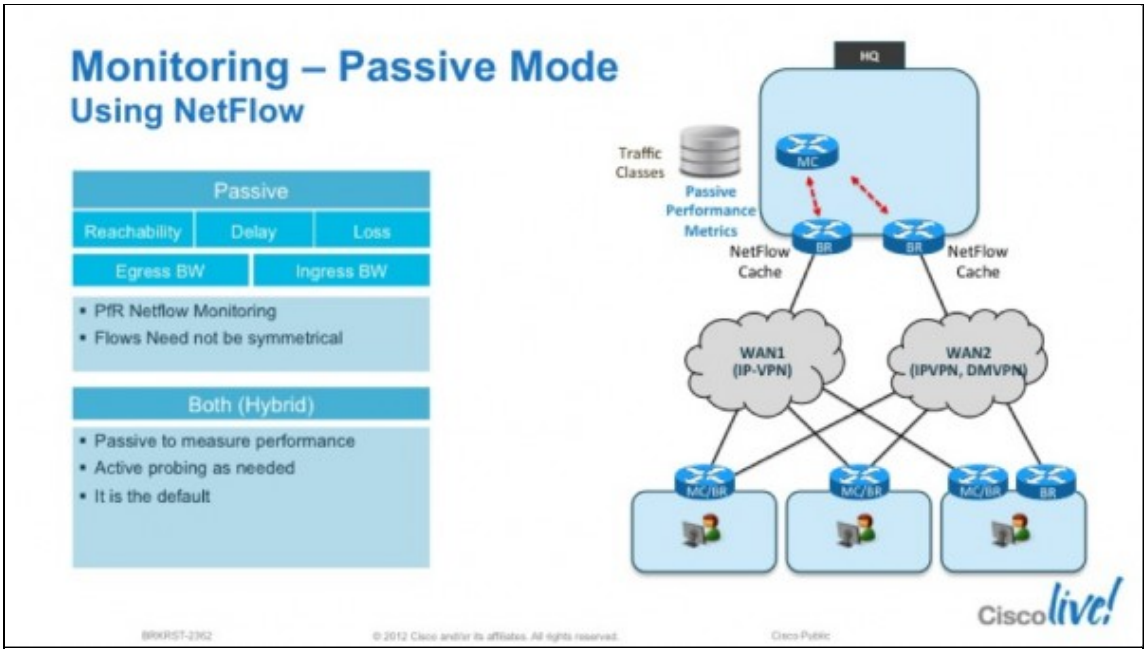


- ?TCP, UDP, Jitter need IP SLA responder targets defined?
- Mode monitor Both: It is the default (recommended)
- Mode monitor Fast: recommended for fast convergence

Mode monitor passive

Passive monitoring is the act of PfR gathering information on user packets assembled into flows by Netflow. Passive monitoring is typically only recommended in Internet edge deployments because active probing is ineffective because of security policies that block probing. PfR, when enabled, automatically enables Netflow on the managed interfaces on the Border Routers. By aggregating this information on the Border Routers and periodically reporting the collected data to the Master Controller, the network prefixes and applications in use can automatically be learned.

The border routers report traffic flows identified by Netflow to the master controller. The average delay of the flows, packet loss, and reachability along with the outbound throughput in terms of bits per second is determined for the destination IP prefixes observed in the Netflow data.



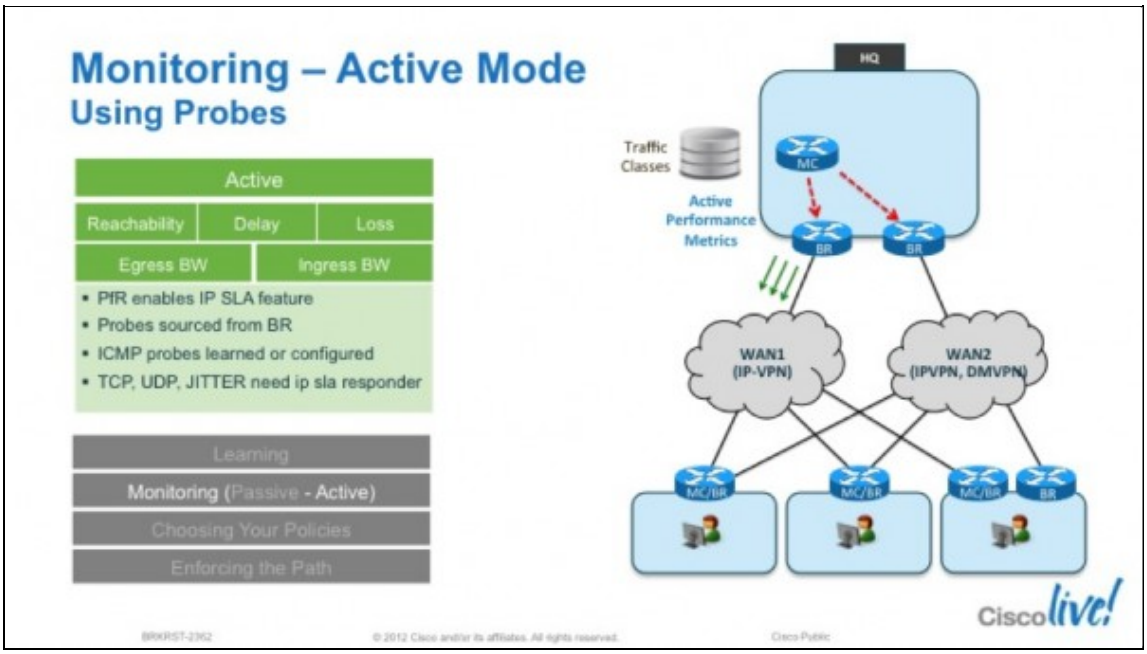
Measurements of the TCP traffic flows are characterized by:

- Delay?Time between TCP SYN and TCP SYN/ACK in a TCP three-way handshake.
- Loss?TCP sequence numbers are tracked, loss can estimated when lower sequence numbers than the highest sequence number observed are seen.
- Reachability?Repeated TCP SYNs without an accompanying TCP SYN/ACK identify reachability failures.
- Throughput?Throughput is calculated from NetFlow and measured in bits per second (bps). Measurements of non-TCP traffic flows is characterized by throughput only.

Mode monitor active

Active monitoring is the act of generating Cisco IOS IP Service Level Agreements (SLAs) probes to generate test traffic for the purpose of obtaining information regarding the characteristics of the WAN links. PfR can

either implicitly generates active probes when passive monitoring has identified destination hosts, or the network manager can explicitly configured probes in the PfR configuration.



In this mode, IP SLAs probes are generated by the border routers through the current exits and transmitted at the configured probe frequency value. Probes are only generated through alternate paths (exits) in the event the current path is out-of-policy.

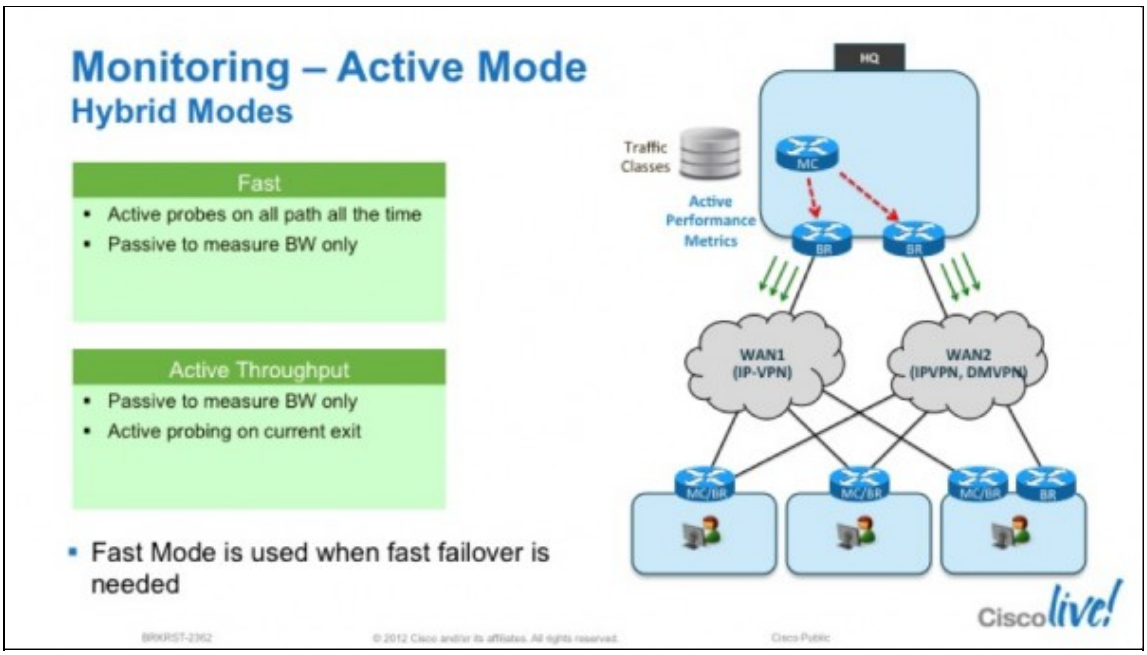
The ICMP ECHO requests that are generated by default constitute additional background traffic on the network. When used on the Internet, activating probing may not be desirable in that ICMP packets may be blocked or administratively prohibited and may be considered a threatening or abusive posture to the target hosts.

By default, an active probe is of the type of ICMP echo. If VoIP is to be characterized, the network manager may choose to explicitly configure an active probe. Following is an example from an *pfr-map* using a traffic-class that matches VoIP streams with probes sent every 2 seconds.

```
set active-probe jitter 30.1.0.11 target-port 33033 codec g729a
set probe frequency 2
```

It is possible, and practical, to use active monitoring for specific traffic-class while using a global configuration option defaulting to passive monitoring of all other traffic through the TCP flows.

Hybrid Modes



Mode monitor both

Mode monitor both is the default value and combines the capabilities of passive and active monitoring. Up to five IP addresses are actively probed for each destination prefix network learned through passive monitoring. By default, an IP SLA ICMP ECHO probe is automatically generated for the learned IP addresses.

In this mode, IP SLAs probes are generated by the border routers through the current exits and transmitted at the configured (or default) probe frequency value. Unlike mode monitor fast, which is described in a later section, active probing does not probe all exit points continuously. It probes only the current exit point provided the status is INPOLICY and probes are generated after the prefix timer value is exhausted. Probes are only generated through alternate paths (exits) in the event the current path is out-of-policy.

By monitoring both actively and passively, additional data points regarding a network prefix can be obtained through two separate and distinct tools; Netflow for passive measurements and IP SLA for the active measurements. However, the inclusion of active probing also has disadvantages. Mode monitor both is best suited for use within the private internal network of the enterprise.

Important note: With mode monitor both, passive measurements are used to trigger the Out of Policy (OOP) state but the active probing results from the exit interfaces are used with the passive throughput measurement for the traffic class, to select the exit interface.

Mode monitor Fast

This feature was introduced in Cisco IOS Release 12.4(15)T as a key component to the Fast Reroute feature.

This mode generates active probes through all exists continuously at the configured probe frequency. This differs from either active or both modes in that these modes only generate probes through alternate paths (exits) in the event the current path is out-of-policy.

With Fast Reroute, the characteristics of the alternative paths are always known, allowing immediate use as required. If unreachable is determined to be out-of-policy for the current exit, the alternate exit that is in policy is selected as the new current exit.

The unreachable threshold is calibrated in number of failed probes per million probe attempts. If the unreachable value is set to 1, a single probe fails on the current exit, an attempt is made to locate a alternate exit. However, if the alternate exits also have a single failed probe, they are not selected because they too are out-of-policy.

The Fast Reroute feature, therefore, allows rerouting actions to be taken, at an interval approaching the configured probe frequency value. Probe frequency can now be set as low as 2 seconds if fast mode is configured. This allows re-routing at slightly more than the configured probe frequency value. The Fast Reroute feature can reroute OOP traffic in as little as 3 seconds.

The obvious drawback to this feature is the potential for adding additional network traffic overhead associated with the probes themselves and additional CPU resources to the PfR border routers, the source of the active probes.

Probes are always generated unless the prefix is deleted or in the default state,.

The active probe results are used for out-of-policy and to control routing. Passive data collected is for information only, the throughput transmit and receive Kbps values (show oer mast border detail) are used for load balancing.

Apply Policy Phase

The apply policy phase compares the measured performance metrics against well-known or configured thresholds to determine if the traffic is meeting specified levels of service, or if some action is required. If the performance metric does not conform to the threshold, a decision is made by PfR to move the traffic class or exit into another state.

Traffic Class		Link	
Performance	Availability	Performance	Administrative
▪ Delay ▪ Loss ▪ Reachability ▪ MOS ▪ Jitter	▪ Sinkhole ▪ Blackhole	▪ Load balancing ▪ Max utilization	▪ Link grouping ▪ $Cost
Scope	Global or per Policy		

Policies can be applied globally, per Traffic Class or even per link.

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. Both exit link traffic and entrance link traffic load thresholds can be configured as a PfR policy. 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 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.

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 forth at set of traffic. Hence, a policy that determines voice traffic to have a delay threshold of less than 250 msecs can utilize multiple paths in the network if available, as long as all the paths deliver the traffic within its performance bounds.

PfR Policy Thresholds

PfR policies are defined by comparing measurements for each Traffic Class against default or user-defined thresholds. PfR allows to define different policies for each Traffic Class. One TC can have a policy based on delay and loss (typically for critical applications) whereas another TC would have a policy based on load-sharing across exit points.

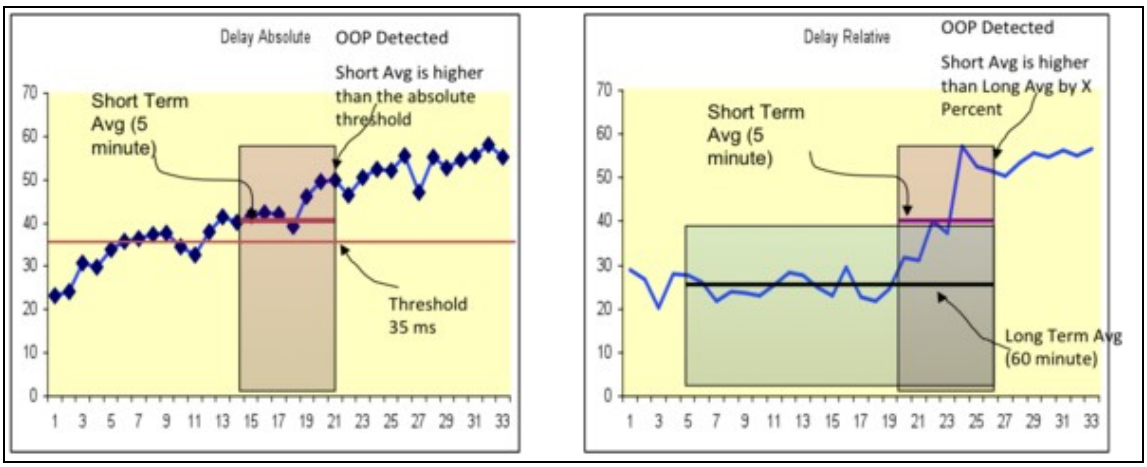
- **Passive Mode**:

  As a general rule, Cisco PfR maintains a short-term counter (statistics from the last 5 minutes) and a long-term counter (statistics from the last 60 minutes).

- **Active Mode**:

  As a general rule, Cisco PfR maintains a short-term counter (statistics from the last 5 probe results) and a long-term counter (statistics from the last 60 probes results).

Cisco PfR Thresholds can be defined in a relative or absolute mode.



- **Absolute**

  An absolute threshold is compared against the short term delay value which is 5 min for passive mode, or last 5 probes for active mode.
  Short-term counter exceeds the threshold of N msec.
  Used for delay, loss, mos, unreachable, jitter

- **Relative**

  A relative threshold is compared against the short and the long-term value
  Short-term counter exceeds the Long-term counter by X percent.
  Used for delay, loss, unreachable
  Example for delay: if the short term delay = 120ms and the long term delay = 100 ms, then the percent diff is 20. This value is compared to the threshold value configured in the oer-map (or against the default value)

The following tables gives the base units used for each parameter than can be configured in a Traffic Class policy.

	Unit	Passive	Active	
Reachability	Flows per million	100 unreachable out of 10000 TCP flows = 10,000 fpm	1 probe fail out 5 attempted = 200,000 fpm	
Delay	millisecond	Delay between TCP SYN and ACK	Round Trip Delay of Probe packet	
Loss	Packets per million	100 lost out of 10000 TCP packet = 10,000 ppm	1 loss out 100 probe packet = 10000 ppm	
Jitter	millisecond	NA	Inter-arrival packet jitter	Jitter Probe Required
MOS	Percentage of MOS below threshold	NA	3 probes out 5 have MOS below 3.85 ~ 60% MOS below 3.85	

Let's take the loss and reachability policy as an example:

- The loss policy -is expressed in Packets per million

- In passive mode: 100 lost out of 10000 TCP packets equal to 10,000 ppm
- In Active mode: 1 loss out 100 probe packets equal to 10000 ppm

- Reachability is expressed in Flows per million

- Passive mode: 100 unreachable out of 10000 TCP flows means 10,000 fpm
- Active mode: 1 probe fail out 5 attempted means 200,000 fpm

Resolvers

Cisco PfR allows multiple policies to be defined for each Traffic Class. To resolve the potential conflict of which policy to run, PfR uses a resolve function which allows to assign a priority for each policy.

By default, PfR assigns the highest priority to delay policies, followed by utilization policies, but unreachable policy always applies and has a default priority of 0 (cannot be changed).

Here is a list of the available policies:

```
MC(config-pfr-mc)#resolve ?
  cost         Specify OER cost policy resolver settings
  delay        Specify OER delay policy resolver settings
  jitter       Specify OER jitter policy resolver settings
  loss         Specify OER loss policy resolver settings
  mos          Specify OER MOS policy resolver settings
  range        Specify OER range policy resolver settings
  utilization  Specify OER utilization policy resolver settings

MC(config-pfr-mc)#
```

Each resolve is defined with a priority and inherits priorities from the global configuration mode You can change from the global inside pfr-map and Unreachable resolver has always the highest priority as seen before.

PfR Selects a Policy Conforming Exit by following each of the following steps:

- Gather traffic class measurements for all exits
- Gather link utilization for all external interfaces
- Exits with no measurements ignored
- Measurements applied using priority with variance
- Exits within variance are candidates

After All Priorities Examined:

- If a single candidate - Use single candidate
- If multiple candidates includes current exit - Choose current exit

- Else, randomly choose a candidate

An example of resolvers defined in PfR global configuration mode:

```
pfr master
 resolve delay priority 4 variance 20
 resolve loss priority 6 variance 20
 resolve util priority 8 variance 20
```

PfR Traffic Class States

The table below summarizes the possible TC states:

DEFAULT	A network prefix may be shown in the default state if it is manually configured or learned, but has not been determined to be in or out-of-policy. Prefixes may revert back to default state if, for some reason, OER can no longer control the prefix. This may happen if all the exits are out-of-policy. The default state means that the parent IP routes control the exit for this destination prefix. This would be the same behavior as if OER were not configured or shutdown.
INPOLICY	The prefix is inpolicy, which means that it meets the policy associated with this prefix or application. The prefix can be inpolicy and being controlled by OER, or inpolicy* and not controlled by OER. The presence of the asterisk (*) on the state attribute indicates the network is known to OER, but is under the control of the parent route. When no asterisk (*) is present, the prefix is being controlled by OER. The state of inpolicy is considered to be a desirable state. After a few cycles, the Traffic Classes will typically be in the INPOLICY State.
OOPOLICY	The prefix or application has been identified as failing to meet its respective policy. If traffic is identified as being out-of-policy, OER moves the traffic to an alternative exit to bring the traffic inpolicy or unmanages the traffic, allowing it to revert back to the default exits as determined by the parent routes in the IP routing table. If the traffic reverts back to the default state, OER will again cycle this traffic, like all other traffic on the network, in an attempt to optimize based on the configured or default OER policy. The state of out-of-policy is considered undesirable. The prefix or application has been identified as failing to meet its respective policy. The policy parameters (ie. Delay, Loss, etc) are defined in the oer-maps for each Traffic Class.
HOLDDOWN	The holddown state is enabled when a traffic class is initially controlled by OER. This holddown concept is applied to prevent churning or erratic behavior of OER managed routes from being injected and withdrawn from the IP routing table (and subsequently being redistributed by some IGP) or BGP tables. Once a prefix has been changed, it enters holddown for the specified (holddown) period, before it can be deemed in or out-of-policy. A network prefix can leave holddown state before the timer expires, if the current exit point experiences an unreachable out-of-policy condition. All other out-of-policy conditions are ignored during holddown state. 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.

There are two ways to specify what is flagged as an Out Of Policy (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.
Example: the short-term delay is 120ms and the long term delay is 90ms, then the percent difference is 33 which is compared to the value that you would configure.

Absolute

You now have a threshold that is compared only to the short-term delay.

Note:

The short and long term delay are defined as 5min (with fast mode it is 5 probe cycles) and 60min (fast mode definition 60 probe cycles) averages or till the last time there was an exit change.

Configuration Examples

```
!
! Policies for a Business Traffic Class
!
pfr-map MYMAP 20
 match pfr learn list BRANCH_BUSINESS
 set delay threshold 150
 set mode route control
 set mode monitor both
 set resolve delay priority 1 variance 20
 set resolve utilization priority 3 variance 20
 no set resolve range
!
!
! Policies for a Best Effort Class
!
!
pfr-map MYMAP 30
 match pfr learn list BRANCH_BE
 set mode route control
 set mode monitor both
 set resolve utilization priority 1 variance 20
 no set resolve delay
 no set resolve range
!
```

Enforce Phase

PfR, by default, operates in an observation mode for the PfR learn, measure, and apply policy phases. In observe mode, the master controller monitors traffic classes and exit links based on default and user-defined policies and then reports the status of the network including out-of-policy (OOP) events and the decisions that should be made, but does not implement any changes. The PfR enforce phase operates in control mode, not observe mode, and control mode must be explicitly configured using the **mode route control** command. In control mode, the master controller coordinates information from the borders routers in the same way as observe mode, but commands are sent back to the border routers to alter routing in the PfR managed network to implement the policy decisions.

PfR initiates route changes when one of the following occurs:

A traffic class goes OOP, an exit link goes OOP, or the periodic timer expires and the select exit mode is configured as select best mode.

During the PfR enforce phase, the master controller continues to monitor in-policy traffic classes that conform to the desired performance characteristics, to ensure that they remain in-policy. Changes are only implemented for OOP traffic classes and exits in order to bring them in-policy. To achieve the desired level of performance in your network, you must be aware of the configuration options that can affect the policy decisions made by the master controller.

Another configuration issue to consider when deploying PfR is that if aggressive delay or loss policies are defined, and the exit links are also seriously over-subscribed, it is possible that PfR will find it impossible to bring a traffic class in-policy. In this case, the master controller will either choose the link that most closely conforms to the performance policy, even though the traffic class still remains OOP, or it will remove the prefix from PfR control. PfR is designed to allow you to make the best use of available bandwidth, but it does not solve the problem of over-subscribed bandwidth.

Verify Phase

The last phase of the PfR performance loop is to verify that the actions taken during the PfR control phase control actually change the flow of traffic and that the performance of the traffic class or link does move to an in-policy state. PfR uses NetFlow to automatically verify the route control. The master controller expects a Netflow update for the traffic class from the new link interface and ignores Netflow updates from the previous path. If a Netflow update does not appear after two minutes, the master controller moves the traffic class into the default state. A traffic class is placed in the default state when it is not under PfR control.

In addition to the NetFlow verification used by PfR, there are two other methods you can use to verify that PfR has initiated changes in the network:

Syslog report-The **logging** command can be configured to notify you of all the main PfR state changes, and a syslog report can be run to confirm that PfR changes have occurred. The master controller is expecting bidirectional traffic, and a syslog report delimited for the specified prefix associated with the traffic class can confirm this.

PfR show commands-PfR show commands can be used to verify that network changes have occurred and that traffic classes are in-policy. Use the **show pfr master prefix** command to display the status of monitored prefixes. The output from this command includes information about the current exit interface, prefix delay, egress and ingress interface bandwidth, and path information sourced from a specified border router. Use the

show pfr border routes command to display information about PfR controlled routes on a border router.

Best Practices

Load Interval and Bandwidth

To provide the most granular and accurate information to the master controller, configure the load-interval on internal and external interfaces on the border routers to the minimum value of 30 seconds. Additionally, the bandwidth statement on the interface should also be appropriately configured.

Passive mode

When Cisco PfR is configured in learn mode/passive, TCP flows must be observed by the border routers to manage prefixes. This means to test PfR in a lab environment, some tool to generate actual TCP/UDP traffic and another to introduce delay, loss, etc,. is necessary to observe meaningful results.