The bridge to possible

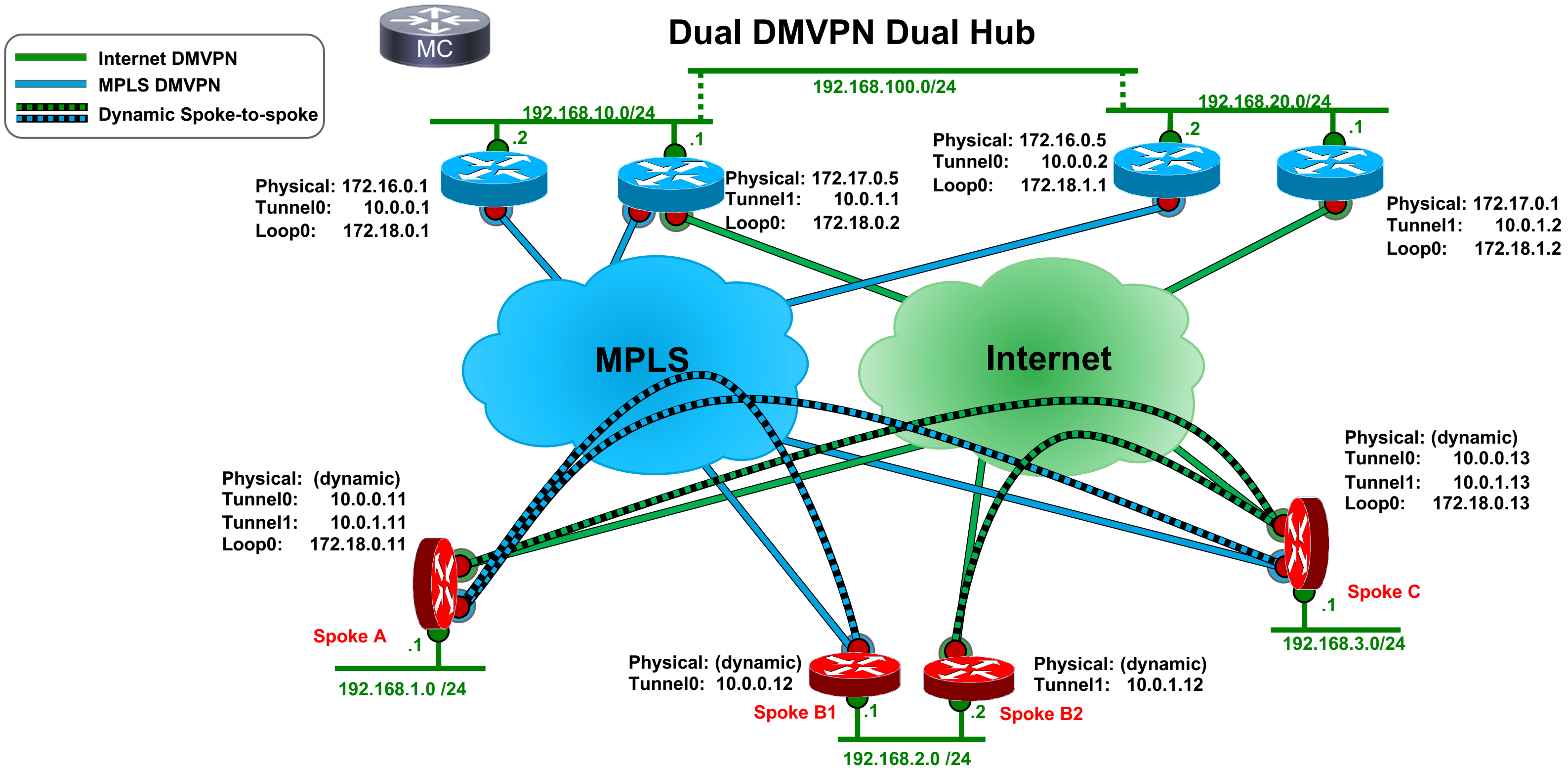
Сетевой марафон Cisco: Классика WAN

День 2. Performance Routing

Денис Коденцев
Старший Архитектор, CCIE

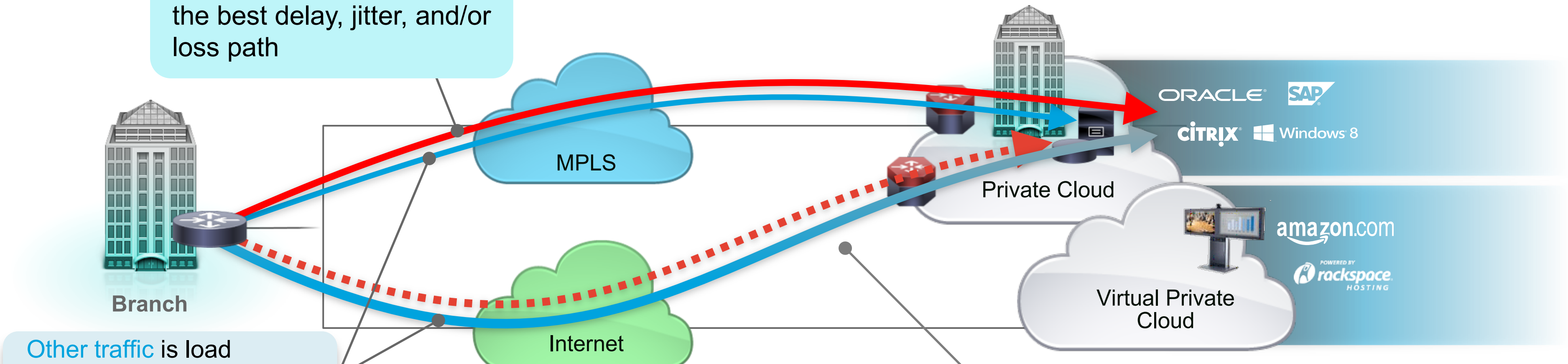
20 апреля 2021

Basic DMVPN Design for DMVPN/PfR



Intelligent Path Control

Performance Routing



Voice/Video/Critical take the best delay, jitter, and/or loss path

Other traffic is load balanced to maximize bandwidth

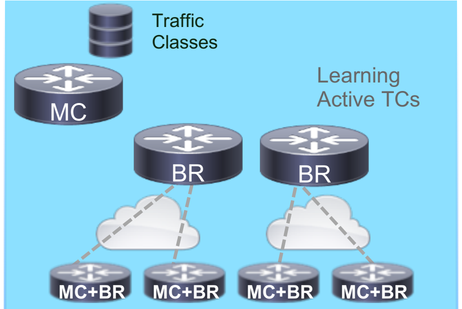
Voice/Video/Critical will be rerouted if the current path degrades below policy thresholds

- PfR monitors network performance and routes applications based on application performance policies
- PfR load balances traffic based upon link utilization levels to efficiently utilize all available WAN bandwidth

How PfR Works – Key Operations

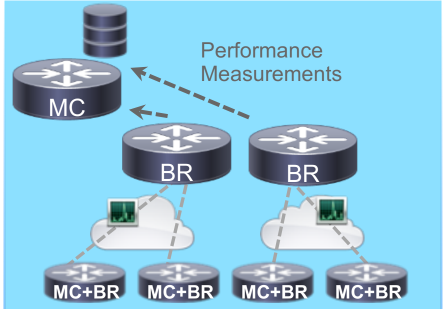
Define Your Traffic Policy

Define Traffic Classes and service level Policies based on Applications or DSCP



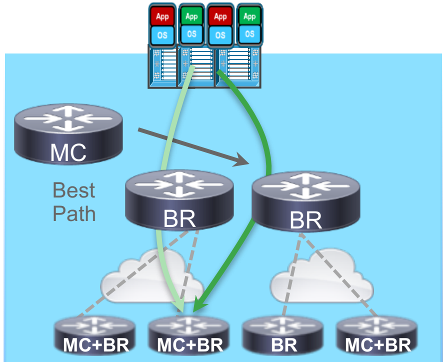
Learn the Traffic

Border Routers learn current traffic classes going to the WAN based on classifier definitions



Measurement

Measure the traffic flow and network performance and report metrics to the Master Controller

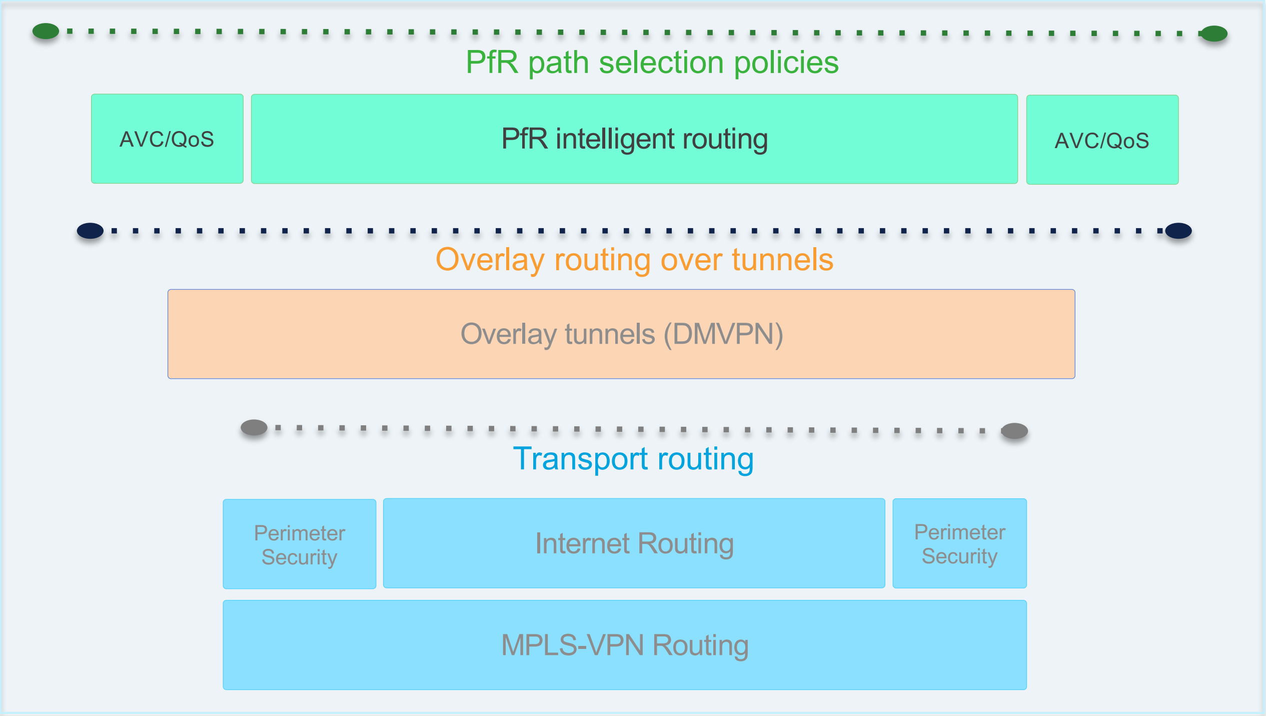


Path Enforcement

Master Controller commands path changes based on traffic class policy definitions

PfR/DMVPN Layered Solution

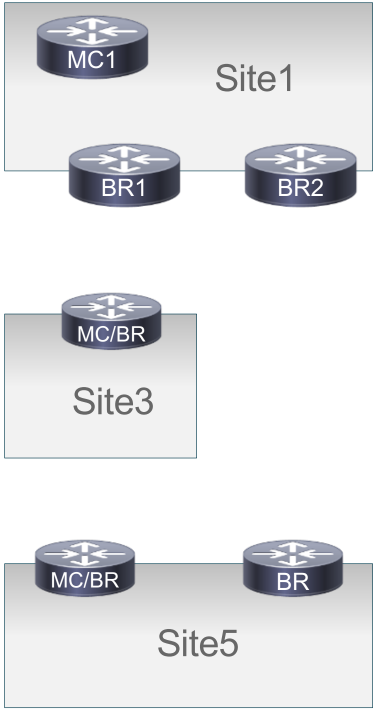
- CPE-to-CPE overlay enables separation of transport (underlay) and VPN service (overlay)
- Point to multipoint WAN connections with secure tunnel overlay architecture
- Intelligent policy routing to provide cost optimization and dynamic load balancing



How PfR works?

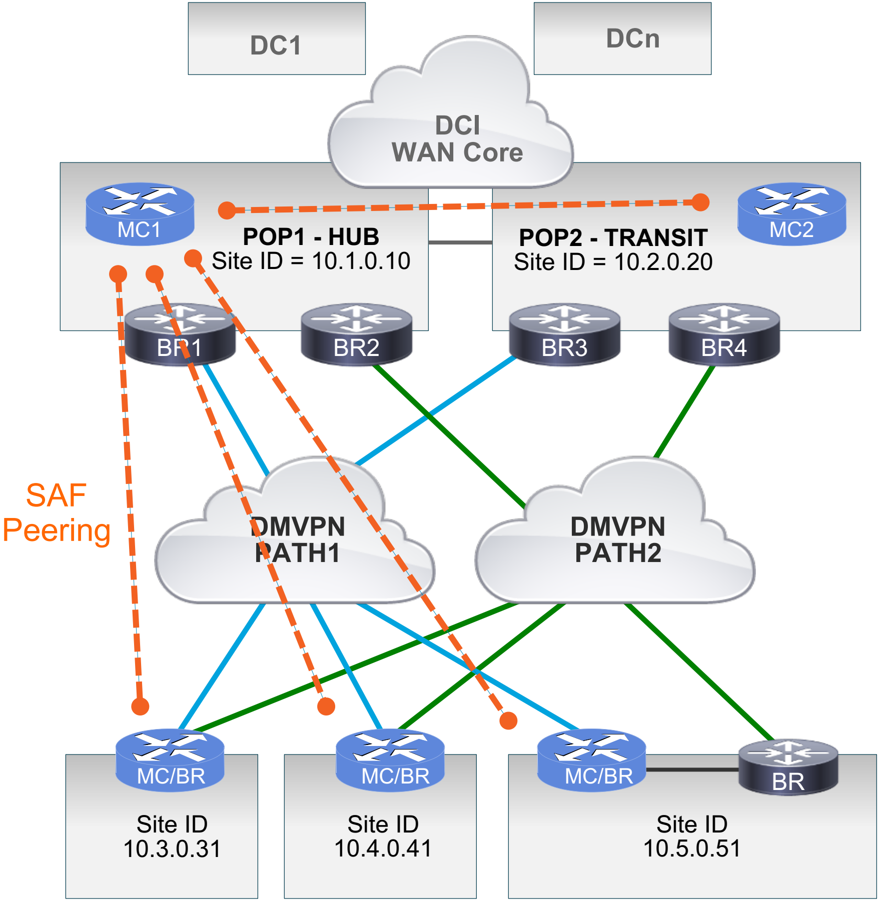
PfR Components

- The Decision Maker: Master Controller (MC)
  - Apply policy, verification, reporting
  - No packet forwarding/ inspection required
  - Standalone of combined with a BR
  - VRF Aware
  - IPv4 only (IPv6 Future)
- The Forwarding Path: Border Router (BR)
  - Gain network visibility in forwarding path (Learn, measure)
  - Enforce MC’s decision (path enforcement)
  - VRF aware
  - IPv4 only (IPv6 Future)
  - The BRs automatically build a tunnel (known as an auto-tunnel) between other BRs at a site. If the MC instructs a BR to redirect traffic to a different BR, traffic is forwarded across the auto-tunnel to reach the other BR



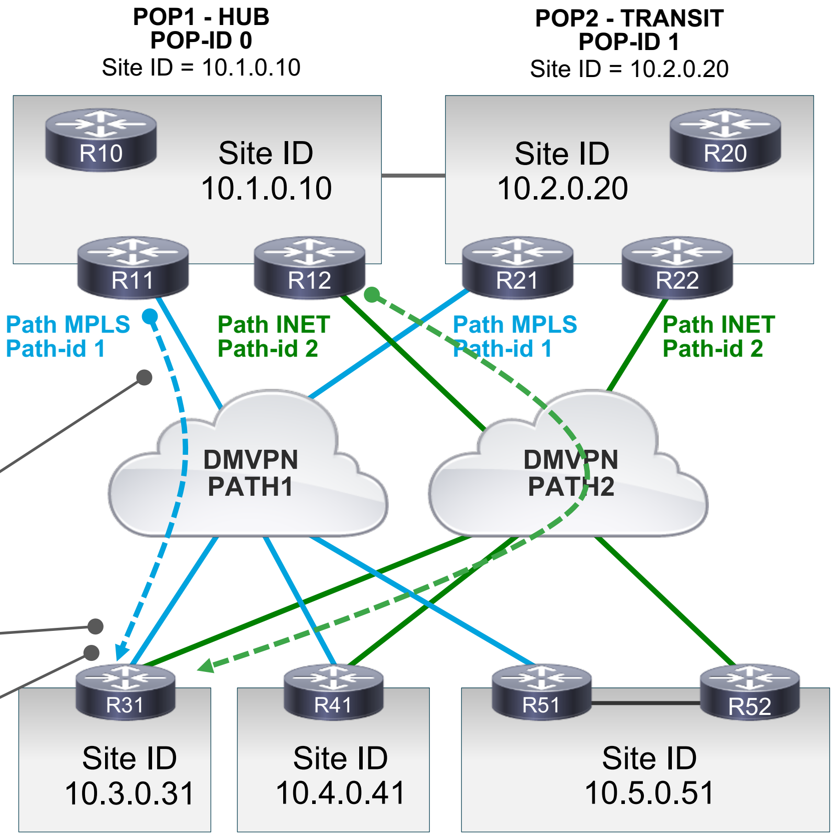
PfR Domain

- Each site runs PfR
- The local MC peers with the logical domain controller (aka Hub MC) to get its policies, and monitoring guidelines.
- Local MC gets its path control configuration and policies from the logical domain controller through the SAF Peering Service
- Peering based on Service Announcement Framework (SAF)



Path Discovery

- Assign Path Name and Path Id to every tunnel on Hub/Transit BRs
- Hub and Transit BRs send Discovery Packet with path names from to all discovered sites



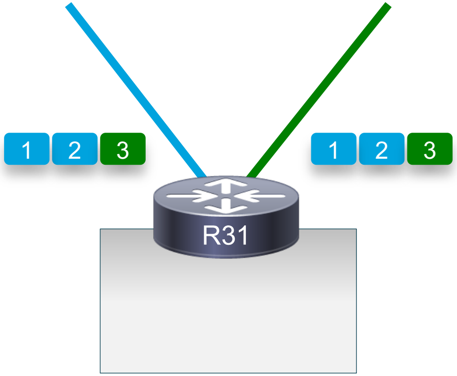
1. Discovery Probes

2. WAN Path Discovered

3. Three Performance Monitors dynamically Assigned

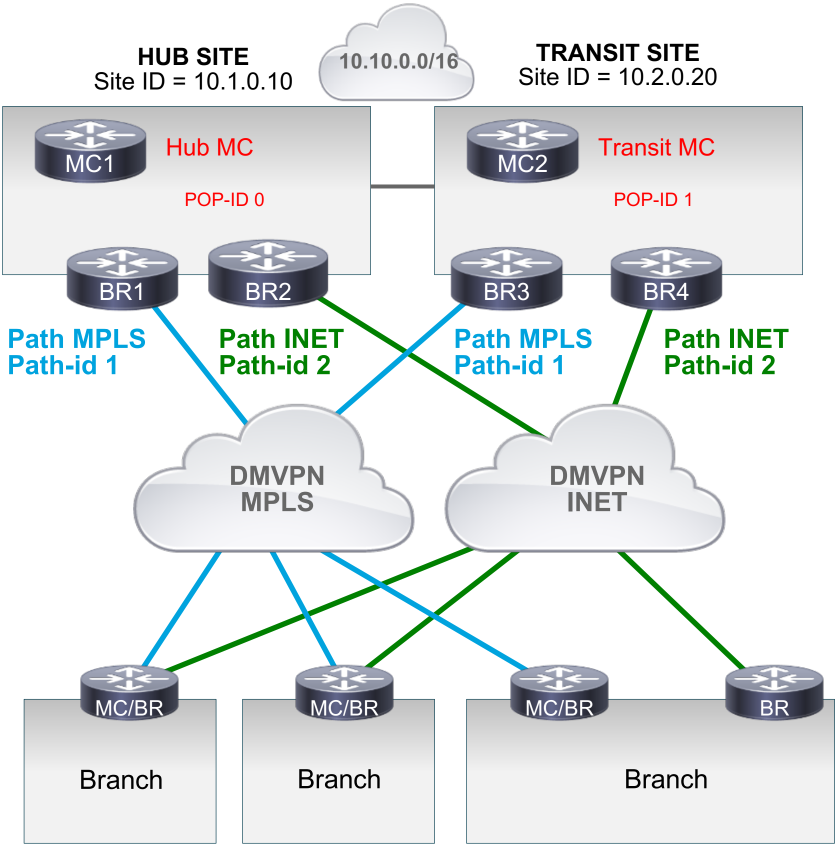
WAN Interface – Performance Monitors

- PfR automatically configures 3 Performance Monitors instances (PMI) over external interfaces
  - Monitor1 – Site Prefix Learning (egress direction)
  - Monitor2 – Aggregate Bandwidth per Traffic Class (egress direction)
  - Monitor3 – Performance measurements (ingress direction)



Site Prefix

SITE1
PfR Site-Prefix
Configured
10.1.0.0/16
10.10.0.0/16



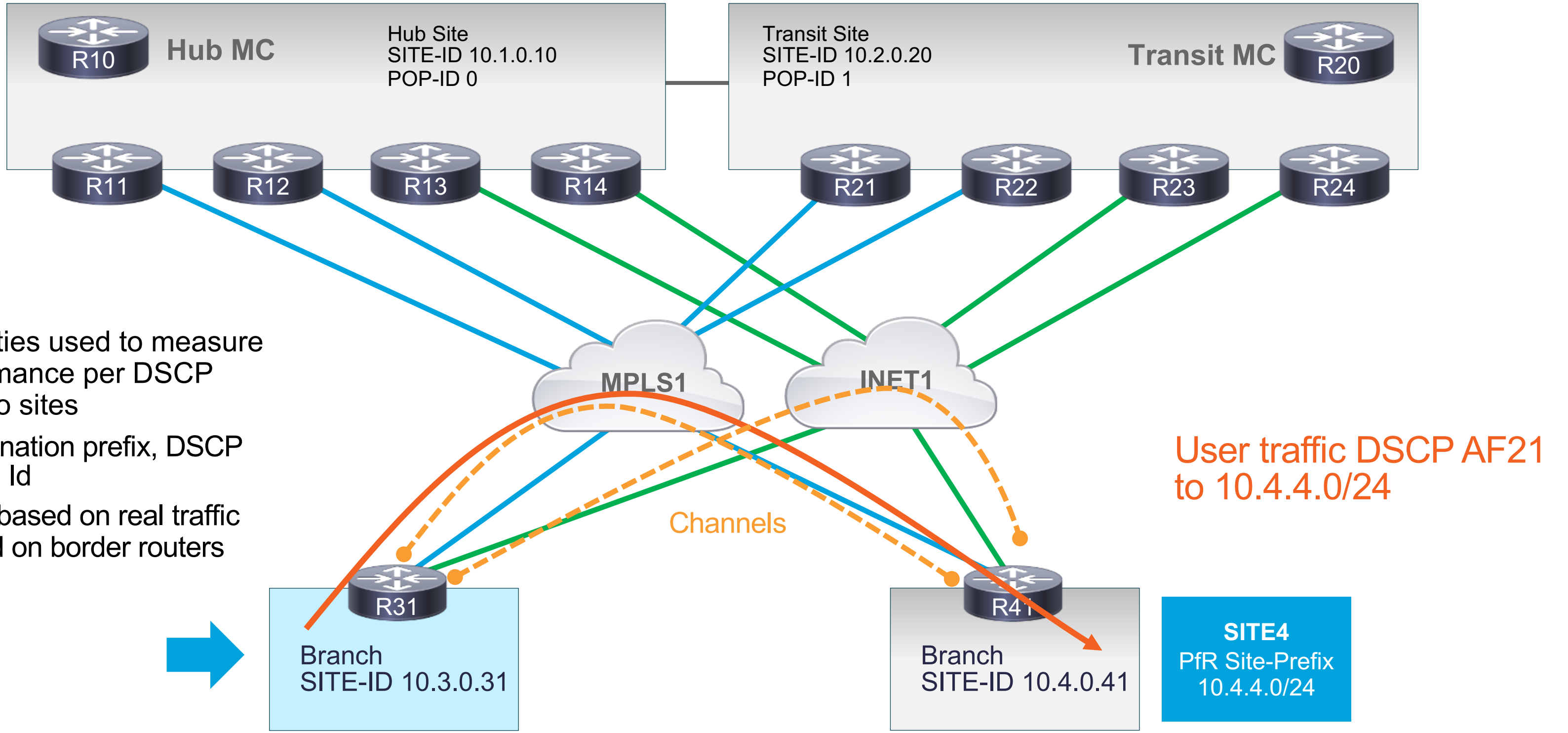
SITE2
PfR Site-Prefix
Configured
10.2.0.0/16
10.10.0.0/16

Hub/Transit sites: static definition of site prefixes is mandatory

Branch sites: static definition of site prefixes is optional but recommended

SITE3
PfR Site-Prefix
Automatic
10.3.0.0/16

PfR Channels



- Logical entities used to measure path performance per DSCP between two sites
  - Per destination prefix, DSCP and Path Id
  - Created based on real traffic observed on border routers

Performance Routing Principles

Define PfR Traffic Policies

Hub MC

Define your Traffic Policy

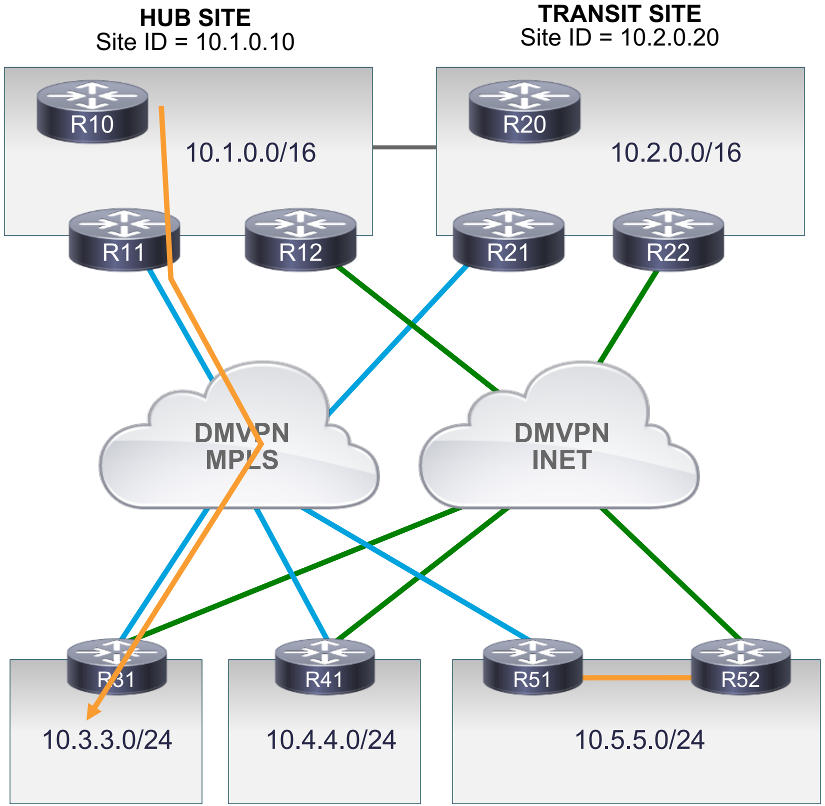
- Identify Traffic Classes based on Application or DSCP
- Performance thresholds (loss, delay and Jitter), Preferred Path
- Centralized on a Domain Controller

CLASS	MATCH	ADMIN	PERFORMANCE
Voice	DSCP Application	Preferred: MPLS Fallback: INET Next Fallback: 4G	Delay threshold Loss threshold Jitter threshold
Interactive Video	DSCP Application	Preferred: MPLS Fallback: INET	Delay threshold Loss threshold Jitter threshold
Critical Data	DSCP Application	Preferred: MPLS Fallback: INET	Delay threshold Loss threshold Jitter threshold
Best Effort	DSCP Application	-	Delay threshold Loss threshold Jitter threshold

Traffic Class – DSCP Based

DSCP Based Policies				
Prefix	**DSCP**	**AppID**	**Dest Site**	**Next-Hop**
10.3.3.0/24	EF	N/A	Site 3	?
10.3.3.0/24	AF41	N/A	Site 3	?
10.3.3.0/24	AF31	N/A	Site 3	?
10.3.3.0/24	0	N/A	Site 3	?
10.4.4.0/24	EF	N/A	Site 4	?
10.4.4.0/24	AF41	N/A	Site 4	?
10.4.4.0/24	AF31	N/A	Site 4	?
10.4.4.0/24	0	N/A	Site 4	?
10.5.5.0/24	EF	N/A	Site 5	?
10.5.5.0/24	AF41	N/A	Site 5	?
10.5.5.0/24	AF31	N/A	Site 5	?
10.5.5.0/24	0	N/A	Site 5	?

Traffic with EF, AF41, AF31 and 0



Traffic Class

- Destination Prefix
- DSCP Value
- Application (N/A when DSCP policies used)

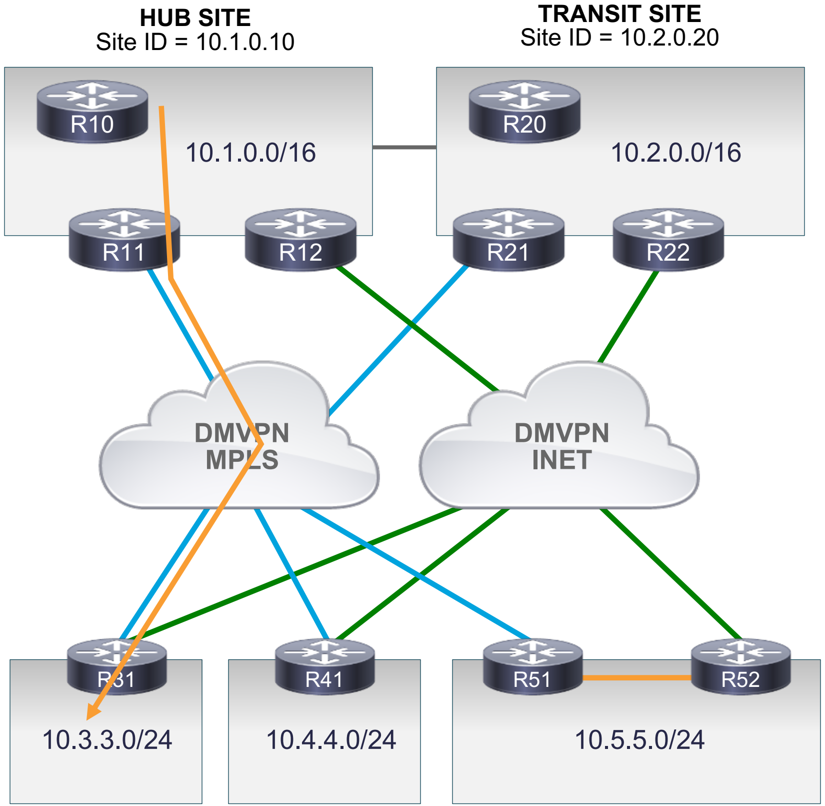
Traffic Class– Application Based

Application based Policies				
Prefix	**DSCP**	**AppID**	**Dest Site**	**Next-Hop**
10.3.3.0/24	EF	N/A	Site 3	?
10.3.3.0/24	AF41	App1	Site 3	?
10.3.3.0/24	AF41	App2	Site 3	?
10.3.3.0/24	AF41	N/A	Site 3	?
10.3.3.0/24	AF31	N/A	Site 3	?
10.3.3.0/24	0	N/A	Site 3	?
10.4.4.0/24	EF	N/A	Site 4	?
10.4.4.0/24	AF41	App1	Site 4	?
10.4.4.0/24	AF31	N/A	Site 4	?
10.4.4.0/24	0	N/A	Site 4	?
10.5.5.0/24	EF	N/A	Site 5	?
10.5.5.0/24	AF41	App2	Site 5	?
10.5.5.0/24	AF31	N/A	Site 5	?
10.5.5.0/24	0	N/A	Site 5	?

Traffic with EF, AF41, AF31 and 0
App1, App2, etc

Traffic Class

- Destination Prefix
- DSCP Value
- Application (N/A when DSCP policies used)



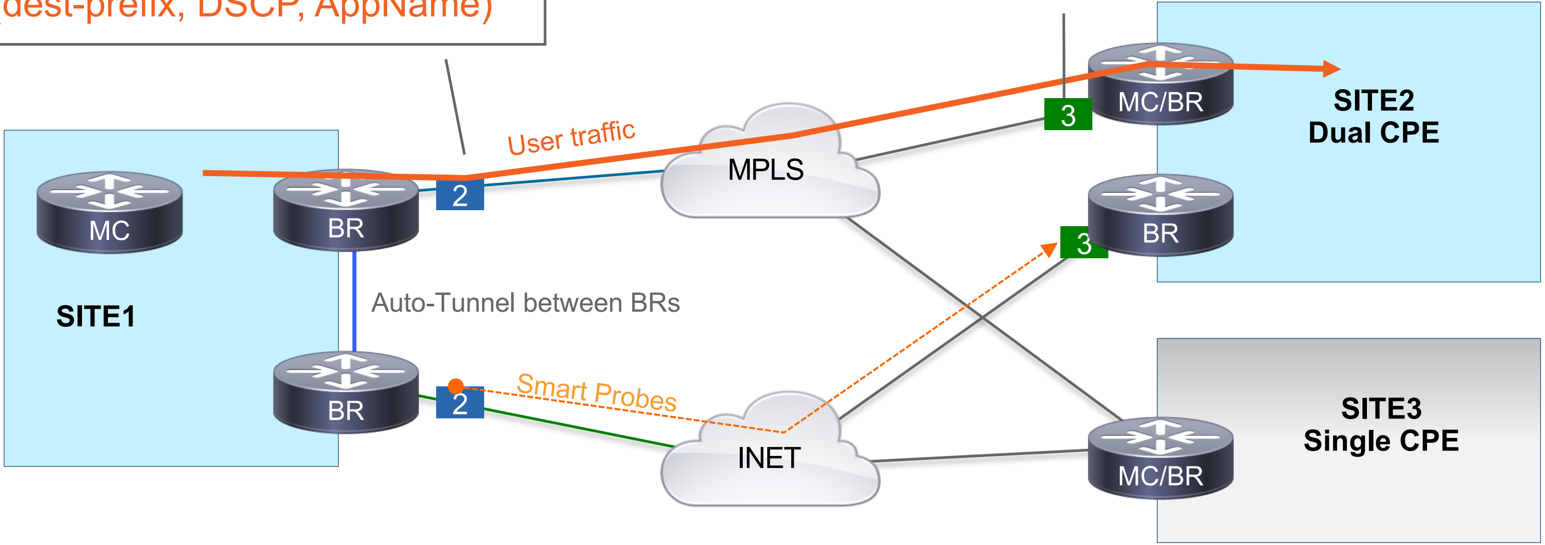
Performance Monitoring

Performance Monitor Egress

- collects Bandwidth
- Per Traffic Class (dest-prefix, DSCP, AppName)

Performance Monitor Ingress

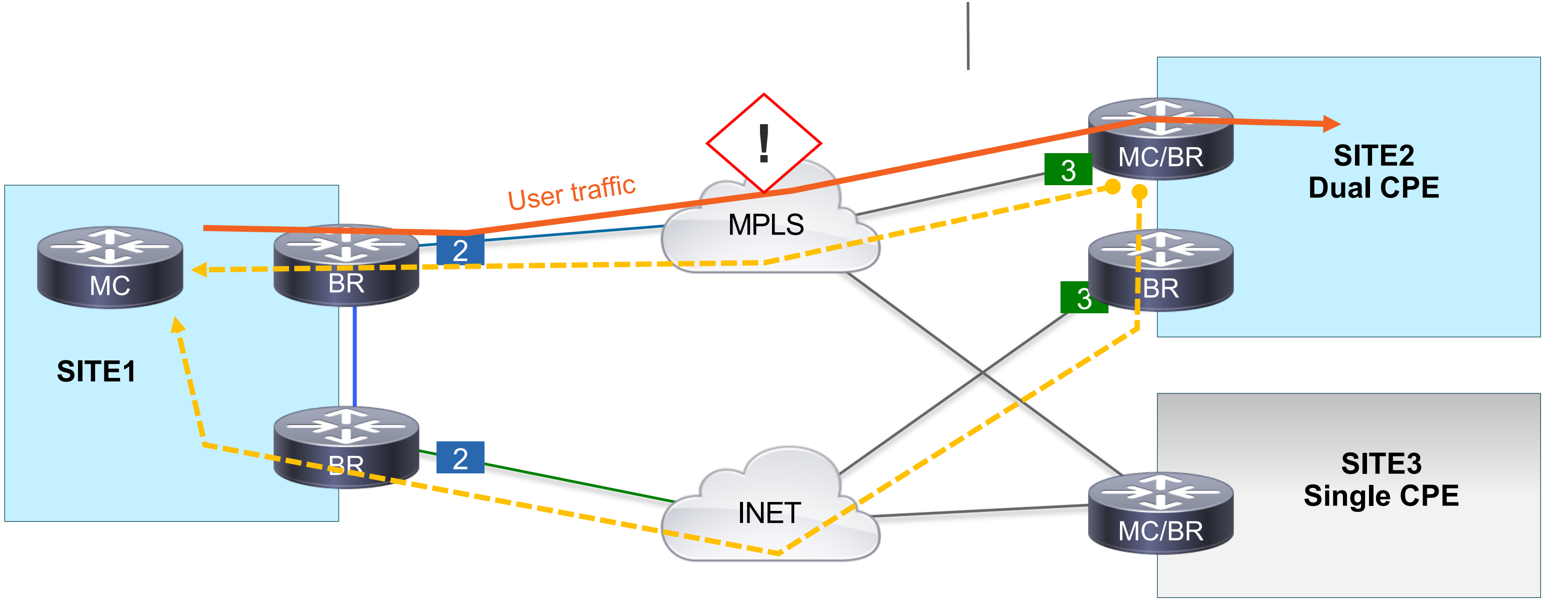
- Collect Performance Metrics
- Per Channel
  - Per DSCP
  - Per Source and Destination Site
  - Per Interface



Performance Violation

ALERT – Threshold Crossing Alert (TCA)

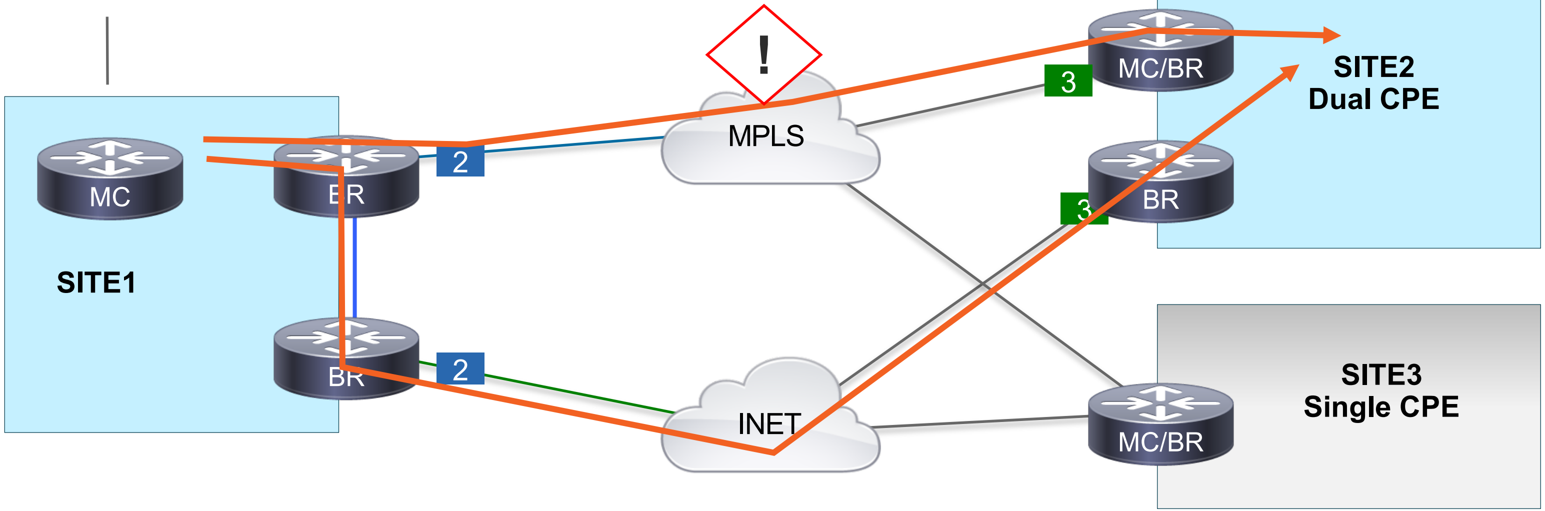
- From destination site
- Sent to source site
- Loss, delay, jitter, unreachable



Policy Decision

MC Instructs BRs

- Reroute Traffic to a Secondary Path across the auto-tunnel
- PfR Dataplane Path Control

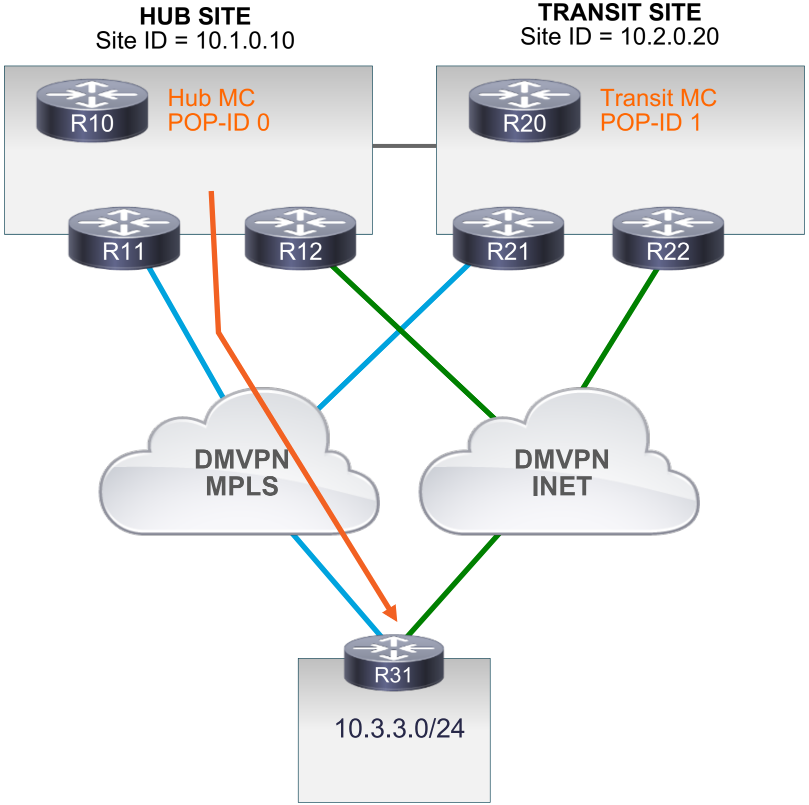


Next Hop Selection Logic

Path Selection

From POPs to Spokes

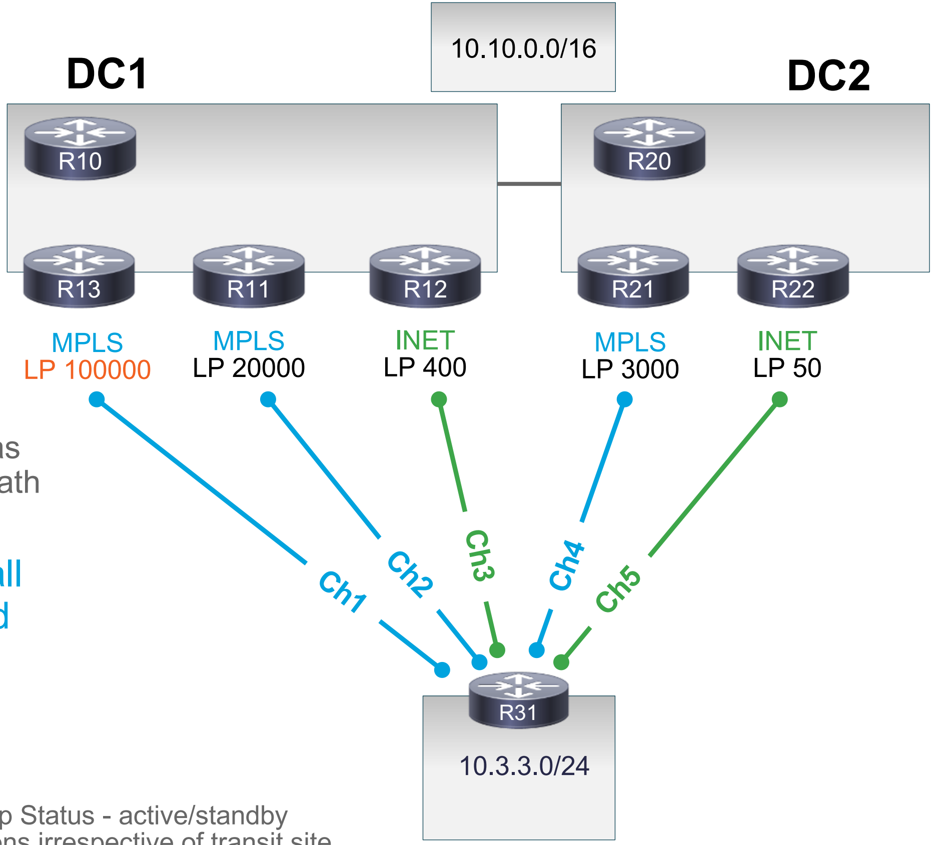
- Each POP is a unique site by itself and so it will only control traffic towards the spoke on the WAN's that belong to that POP.
- PfRv3 will NOT be redirecting traffic between POP across the DCI or WAN Core. If it is required that all the links are considered from POP to spoke, then the customer will need to use a single MC.
- Only one next hop (on branch) per DMVPN network



Path Selection

From Spokes to POPs

- The spoke considers all the paths (multiple NH's) towards the POPs
- The concept of "active" and "standby" next hops based on best metrics in routing is used to gather information about the preferred POP for a given prefix.
  - We moved away from tagging a next hop individually as active/standby and moved towards tagging a whole DC as active/standby. Path-preference is used to choose one path over other.
- If the best metric for a given prefix is on DC1 then all the next hops on that DC for all the ISPs are tagged as active (only for that prefix).
  - Best Metrics:
    - Advertised mask length
    - BGP Weight and Local Preference
    - EIGRP FD and Successor FD



Note Next Hop Status - active/standby tagging happens irrespective of transit site affinity enabled/disabled

For Your Reference

Next hop status for prefix – Details

- Active next hop: A next hop is considered active if it is located at the POP site which has the next hop with the best routing metric for a given prefix
- Standby next hop: A next hop is considered standby if it is located at the POP site which advertises a route for prefix but does not have any next hop with best metric.
- Routable* next hop: A next hop is considered routable for a given prefix if it advertises one or more routes for the prefix and it was not a candidate channel for any traffic class
- Unreachable next hop: A next hop is considered unreachable for a given prefix if it is down or does not advertise any route for the prefix
- The sorting for active/standby considers all the channels/next hops on all WAN interfaces which are “Routable”.

Note: Routable is a new status visible starting from XE 3.16.1/15.5(3)M. On the border prior to XE 3.16.1/15.5(3)M active, standby and unreachable were supported.

For Your Reference

PfRv3 and Routing Best Metrics

- A next hop in a given list is considered to have a best metric based on following metrics/criteria:
  - Advertised mask length (↑)
  - BGP: Weight(↑) , Preference length (↑)
  - EIGRP : FD (↓) , Successor FD (↓)
- Mask length takes precedence. Only if advertised mask lengths are equal, the protocol specific metrics are used.

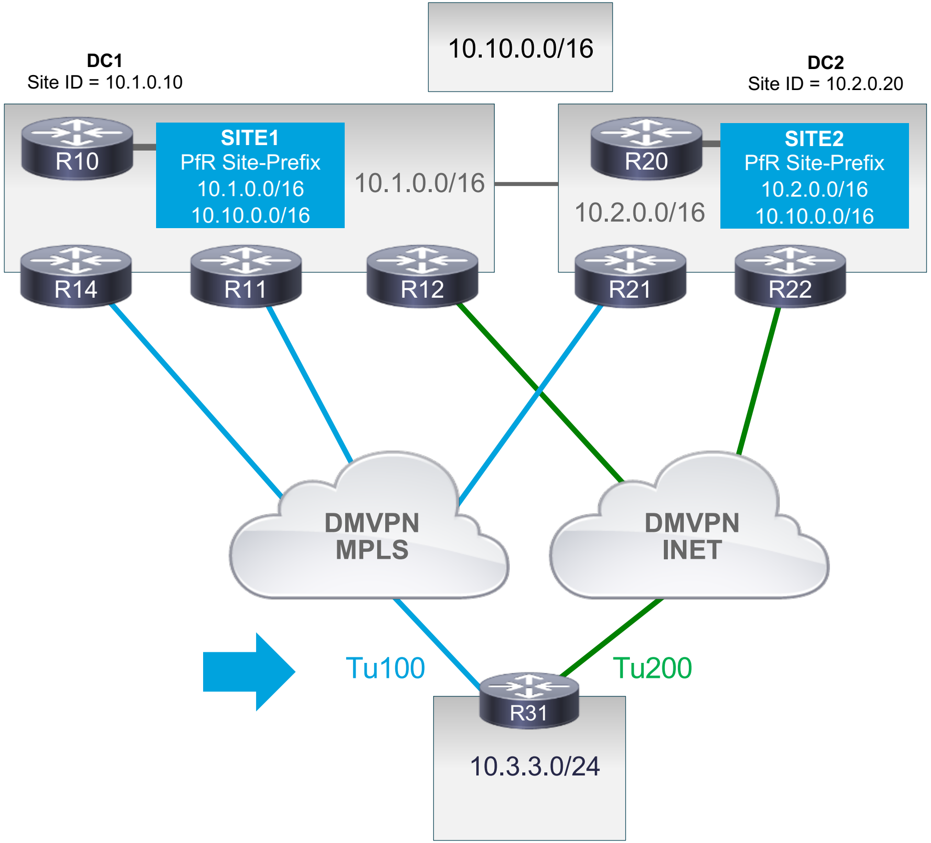
Path Preference

- With Path Preference configured, PfR will then first consider all the links belonging to the preferred path preference (i.e it will include the active and the standby links belonging to the preferred path) and will then use the fallback provider links.
- Without Path Preference configured PfR will give preference to the active channels and then the standby channels (active/standby will be per prefix) with respect to the performance and policy decisions
  - Note that the Active and Standby channels per prefix will span across the POP's.
  - Spoke will randomly (hash) choose the active channel

Load Balancing

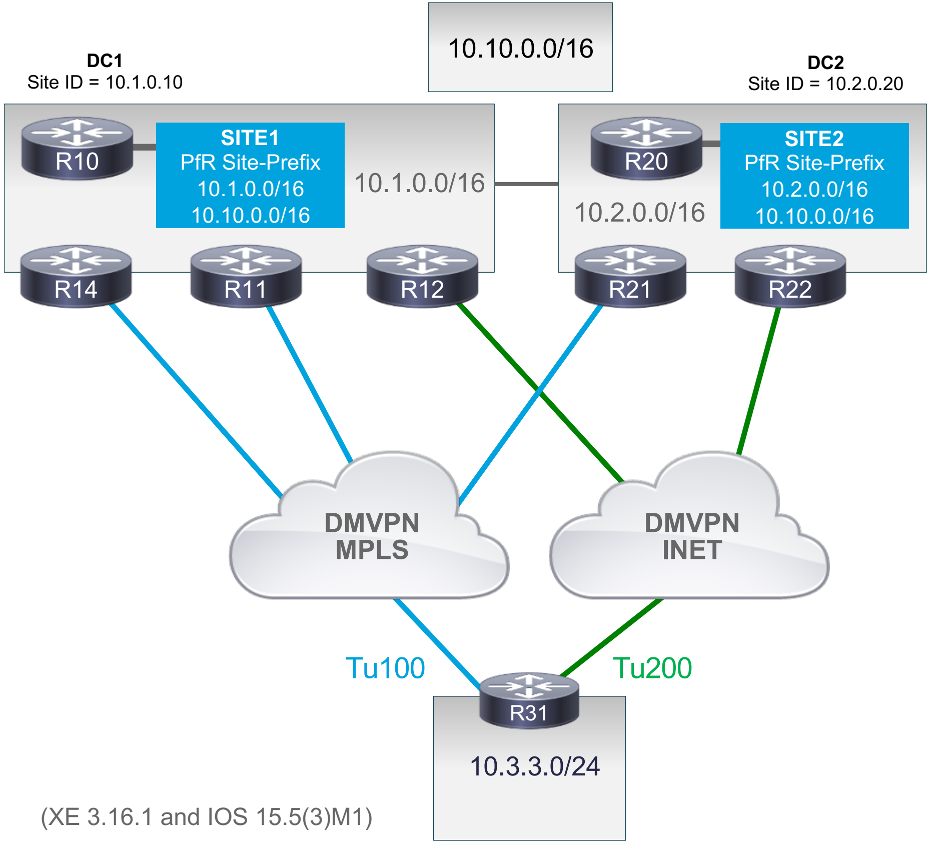
- Load balancing (LB) works on physical interface
  - Looks at the local interface bandwidth utilization and selects Path/local interface
  - Tu100 vs Tu200
- Non Performance TC
  - Load balancing at any time (not only at creation time).
  - TC will be moved to ensure bandwidth on all links is within the defined range
- Performance TC
  - Load balances only at creation time
  - TC will NOT be moved to ensure bandwidth on all links is within the defined range
  - PfR does not account for the Performance TCs getting fatter

Option to prevent placing non-performance based traffic classes on certain path



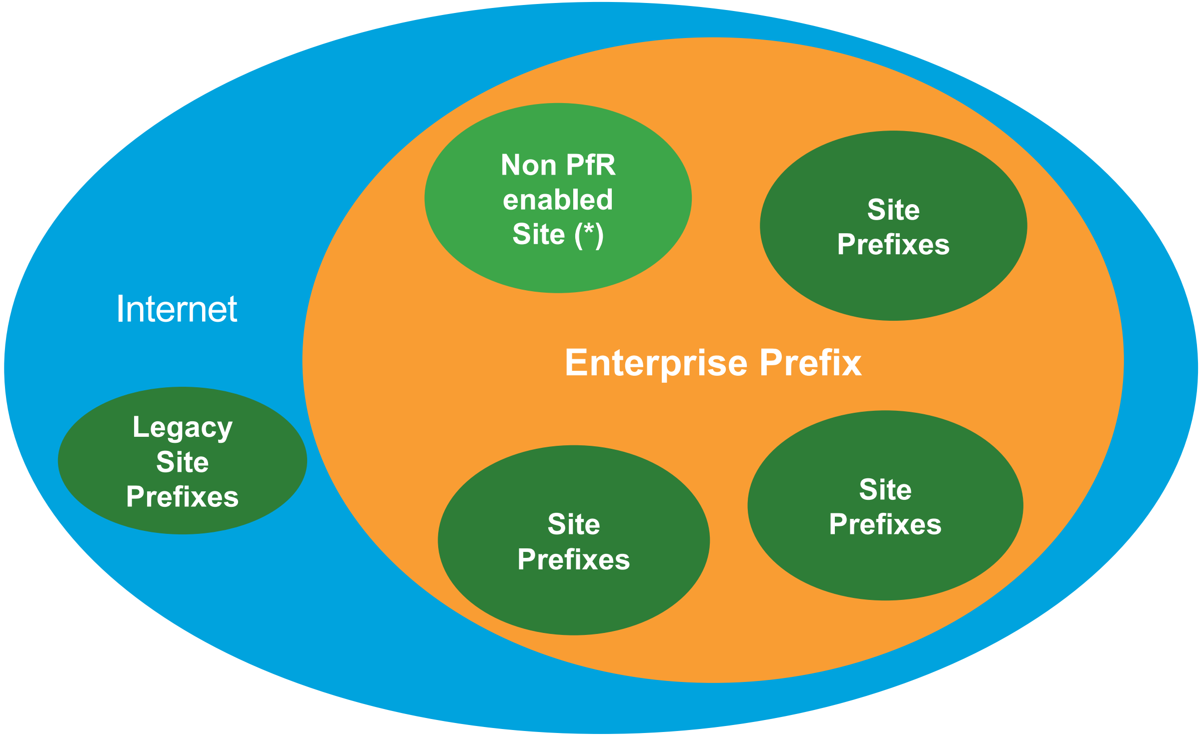
Load Sharing Next Hops

- Load Sharing (LS) works on next hops (NHs) on the same DMVPN network
- Looks at remote next hops of same Path at the hub sites: R14/R11/R21 and R12/R22
- Load-share among the equals (ie matching datacenter preference, path-preference and path)
- Statistically distribute the load among NHs on the same Path (hashing algorithm)
- Applicable only for branch-to-hub traffic



(XE 3.16.1 and IOS 15.5(3)M1)

PfR Enterprise & Site Prefix



Without enterprise-prefix: all the traffic to Non-PfR enabled will be learnt as internet traffic class and therefore subjected to load balancing.

Site prefixes for particular sites with PFRv3 enabled

Branches learn Site Prefixes Dynamically

Hubs act as transit sites –site-prefix statically defined

(*) Only routing is used between non-PfR enabled site in Enterprise Prefix

Enterprise Prefix List

- The main use of the enterprise prefix list is to determine the **enterprise boundary**.
- The enterprise prefix prefix-list defines the boundary for all the internal enterprise prefixes.
- A prefix that is not from the prefix-list is considered as internet prefix and is load balanced over the DMVPN tunnels.
- The enterprise prefix prefix-list is defined only on the Hub MC under the master controller configuration with the command **enterprise-prefix prefix-list** *prefix-list-name*.

```
pfr master
   enterprise-prefix  prefix-list ENTERPRISE_PREFIX
!
ip prefix-list ENTERPRISE_PREFIX seq 10 permit 10.0.0.0/8
```

Site Prefix List

- The site-prefix prefix-list defines static site-prefix for the local site and disables automatic site-prefix learning on the border router.
- The static-site prefix list is only required for Hub and Transit MCs.
- A site-prefix prefix-list is optional on Branch MCs.
- The site prefix is defined under the master controller configuration with the command site-prefixes prefix-list prefix-list-name

```
pfr master
   site-prefixes prefix-list SITE_PREFIX
!
ip prefix-list SITE_PREFIX seq 10 permit 10.1.0.0/16
ip prefix-list SITE_PREFIX seq 20 permit 10.2.0.0/16
!
```

For Your Reference

PfR Policies – Built-in Policy Templates

Pre-defined Template	Threshold Definition
Voice	priority 1 one-way-delay threshold 150 threshold 150 (msec) priority 2 packet-loss-rate threshold 1 (%) priority 2 byte-loss-rate threshold 1 (%) priority 3 jitter 30 (msec)
Real-time-video	priority 1 packet-loss-rate threshold 1 (%) priority 1 byte-loss-rate threshold 1 (%) priority 2 one-way-delay threshold 150 (msec) priority 3 jitter 20 (msec)
Low-latency-data	priority 1 one-way-delay threshold 100 (msec) priority 2 byte-loss-rate threshold 5 (%) priority 2 packet-loss-rate threshold 5 (%)

Pre-defined Template	Threshold Definition
Bulk-data	priority 1 one-way-delay threshold 300 (msec) priority 2 byte-loss-rate threshold 5 (%) priority 2 packet-loss-rate threshold 5 (%)
Best-effort	priority 1 one-way-delay threshold 500 (msec) priority 2 byte-loss-rate threshold 10 (%) priority 2 packet-loss-rate threshold 10 (%)
scavenger	priority 1 one-way-delay threshold 500 (msec) priority 2 byte-loss-rate threshold 50 (%) priority 2 packet-loss-rate threshold 50 (%)

Unreachable Timer

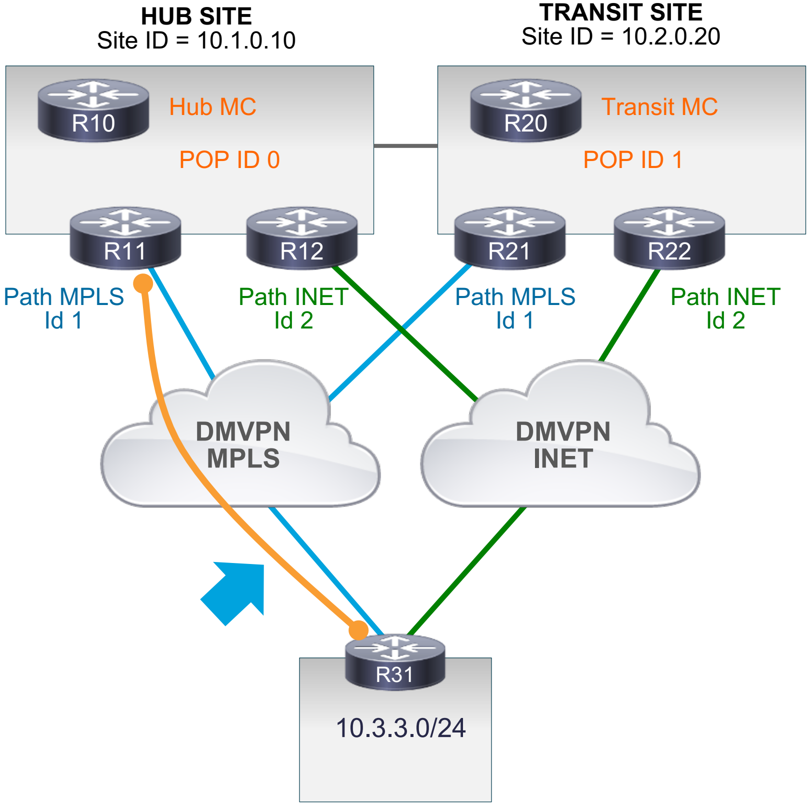
- Channel Unreachable
  - PfRv3 considers a channel reachable as long as the site receives a PACKET on that channel
  - A channel is declared unreachable in both direction if
    - There is NO traffic on the Channel, probes are the only way of detecting unreachability. So if no probe is received within 1 sec, PfR detects unreachability.
    - When there IS traffic on the channel, if PfR does not see any packet for more than a second on a channel PfR detects unreachability.

Default: 1 Sec

Recommended: 4 sec

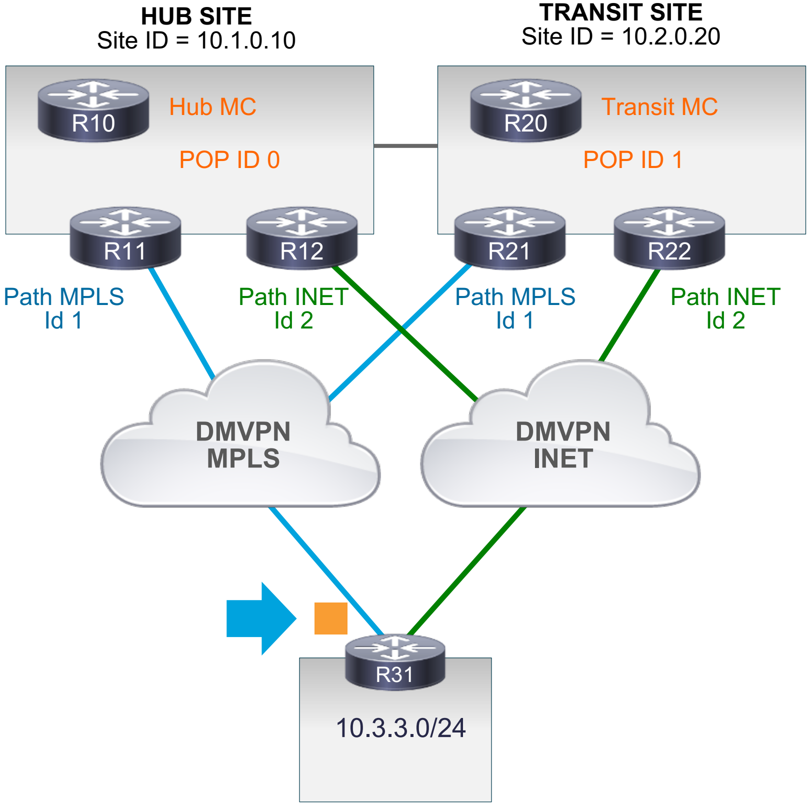
Advanced options – with 3.16 15.5(3)S / 15.5(3)M

channel-unreachable-timer 4



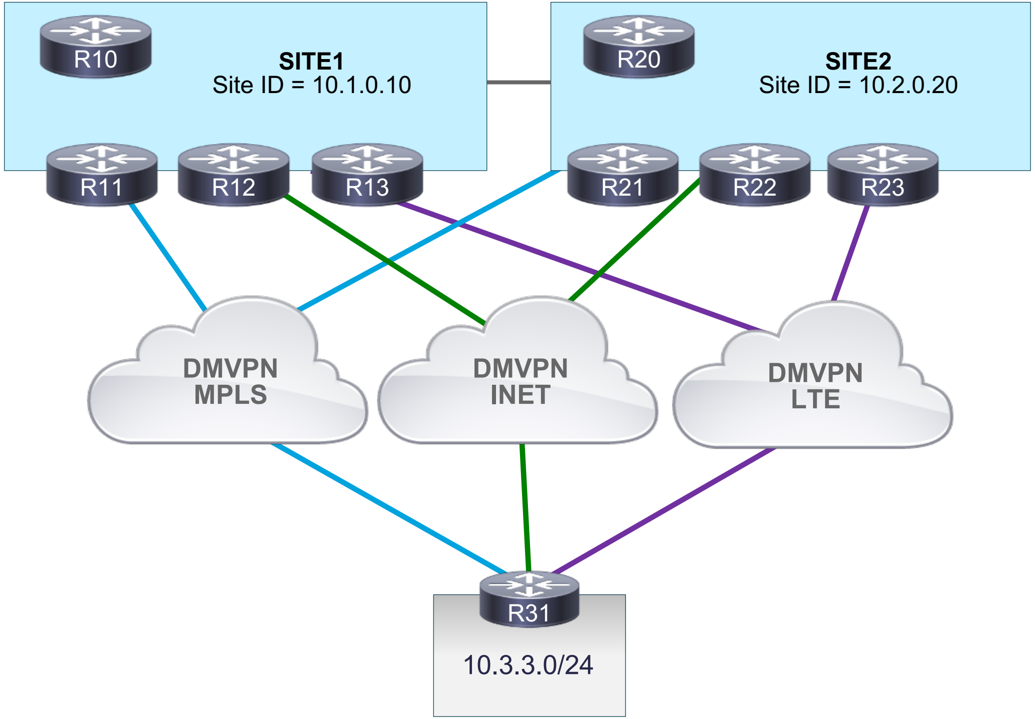
Failover Time

- Ingress Performance Violation detected
  - Delay, loss or jitter thresholds
  - Based on Monitor-interval (30 sec default)
  - Quick Monitor for fast failover



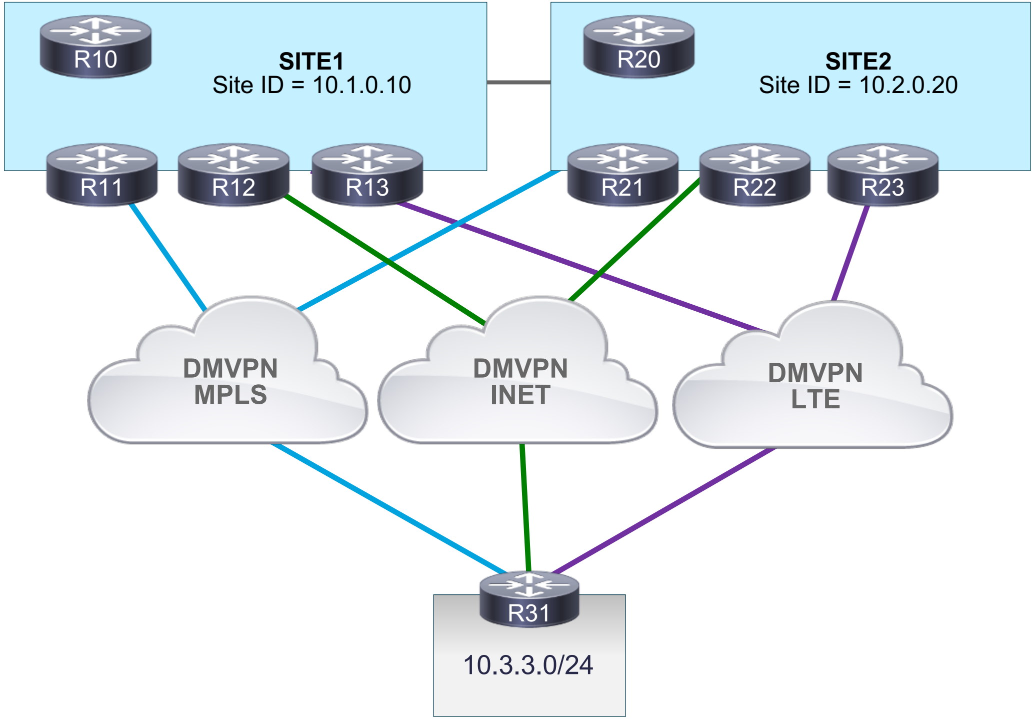
Zero SLA Support

- Zero-sla added on the WAN interface path configuration
- PfR will only probe the default channel (DSCP 0).
  - It will mute all other smart-probes besides the default channel



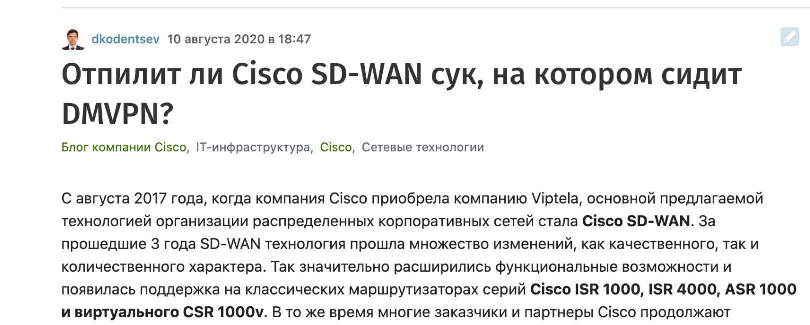
Path of Last Resort

- Path of last resort (PLR) option for metered links
- PLR Channels muted when in standby mode
- Once it is active, smart probes will only be sent on dscp 0 (zero sla) to conserve bandwidth
- Smart probe frequency will be reduced to 1 packet every 10 secs from 20 packets per secs.
- Unreachable detection will be extended to 60 secs



Вместо заключения DMVPN/PFR vs SD-WAN

https://habr.com/ru/company/cisco/blog/514616/



dkodentsev 10 августа 2020 в 18:47

Отпилит ли Cisco SD-WAN сук, на котором сидит DMVPN?

Блог компании Cisco, IT-инфраструктура, Cisco, Сетевые технологии

С августа 2017 года, когда компания Cisco приобрела компанию Viptela, основной предлагаемой технологией организации распределенных корпоративных сетей стала **Cisco SD-WAN**. За прошедшие 3 года SD-WAN технология прошла множество изменений, как качественного, так и количественного характера. Так значительно расширились функциональные возможности и появилась поддержка на классических маршрутизаторах серий **Cisco ISR 1000, ISR 4000, ASR 1000 и виртуального CSR 1000v**. В то же время многие заказчики и партнеры Cisco продолжают

Thank You