BGP-THE HIGH WAY OF INTERNET

THE RECENT ADVANCEMENTS

ROHIT BOTHRA

BROCADE COMMUNICATIONS

1

AGENDA

- Introduction
- BGP infra : Advances
- BGP HA at a glance
- BGP Multipath Signaling
- BGP path diversity: Solutions
- VPN advancements
- BGP for SDN
- Summary

INTRODUCTION

- Service Providers are in constant pressure to mitigate the ever new service requirements.
- Multimedia, Mobile Internet and Cloud Services will generate massive bandwidth explosion
- Prefix growth is almost a linear curve
- Evolution of offered BGP services go from basic technologies to very advanced infrastructures.

WHY BGP ?

- **Investment protection:**
  - Well known protocol
  - Rich set of tools
- **Robustness:** Run over TCP, Years of improvements
- **Low Overhead:** Sends an update once and then remains silent
- **Scalability:** Path Vector Protocol, Route- Reflector, Controller
- **High Availability:** NSR, PIC, GSHUT
- **Simplicity:** BGP is simple (even if knobs make BGP BIG and sometimes less trivial to read)
- **Multi- protocol:** IPv4, IPv6, L2VPN, L3VPN, Multicast, SDN
- **Incremental:** Easy to extend: NLRI,Path Attribute, Community
- **Flexible:** Policy

FUTURE SERVICES: BGP ALL THE WAY

ALMOST ALL SERVICES ARE MOVING TOWARDS BGP

Solutions	Before	Present & Future
Internet (Peering)	BGP IPv4	BGP IPv4/v6
Private IP (L3VPN)	BGP IPv4	BGP IPv4/v6 + HA + Scalability
Private Multicast (Mc VPN)	PIM	BGP Multicast VPN
L2 Services (L2VPN)	LDP VPWS/VPLS	BGP VPLS/VPWS, eVPN
DDOS mitigation	CLI, ACL, PBR	BGP flowspec
Network Monitoring	SNMP	BGP monitoring protocol, BGP
Security	Filters	BGP Sec (RPKI), BGP flowspec
Proximity		BGP Link State
Data Center Scale	IGP (ISIS, OSPF) or L2 (Trill, FP, Vlan)	BGP, BGP SR
MPLS transport	LDP	BGP + Label Unicast (Unified MPLS)
SDN	PBR, OpenFlow (2013), Yang (future)	BGP flowspec, BGP Link State, BMP, BGP route controller, BGP Label Unicast, BGP Segment Routing
Overlay Transport		VxLAN BGP sign, Softwire

BGP INFRA: ADVANCEMENTS

BGP INFRA ADVANCEMENTS

- Entirely depends on BGP stack implementation on a particular product segment.
- Keepalive Enhancements
  - Loosing or delayed keep-alive message result in session flaps
  - Keepalive processing can be done with priority queuing.
- Update Generation Enhancements
  - Update generation is the most important, time-critical task
  - Can be implemented as a separate process to provide more CPU cycle
- Parallel Route Refresh
  - Significant delay in advertising incremental updates while RR is servicing route refresh requests or converging newly established peers
  - Parallelize refresh and incremental updates
- Adaptive Update Cache Size
  - Instead of using a fixed cache size, can be dynamically adapted to the address family used.
  - Update messages can also be generated and kept in circular buffers for transmission.

BGP High Availability

BGP GRACEFULL SHUTDOWN

BGP Gracefull Shutdown allows to do maintenance on router without service disruption.

This new knob allows a router to notify neighbor to redirect traffic to other paths and after some time will drop BGP sessions.

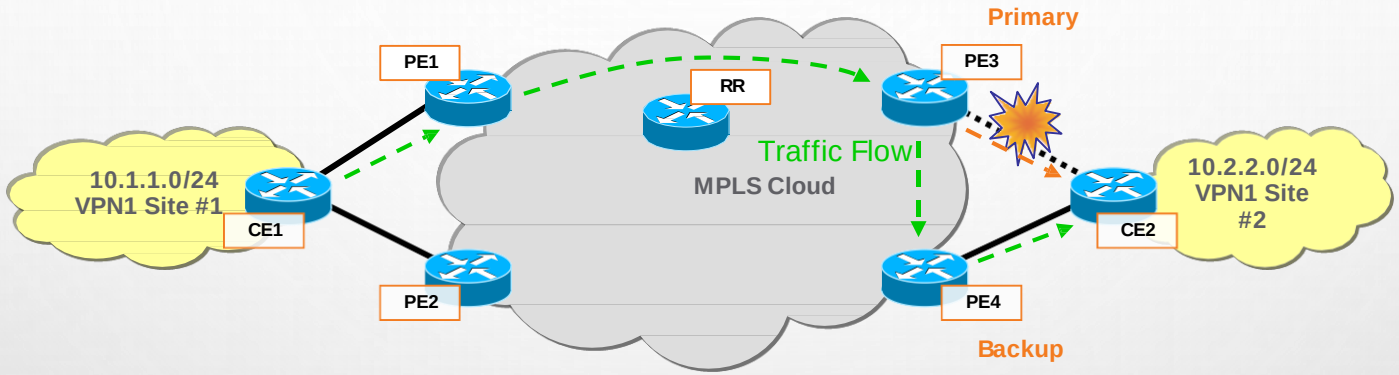
The notification could be done using Local Preference attribute or user community attribute

PIC OVERVIEW

- Internet Service Providers provide a strict SLAs to their Financial and Business VPN customers where they need to offer a sub-second convergence in the case of Core/Edge Link or node failures in their network
- Prefix Independent Convergence (PIC) is supported for CORE link failures as well as edge node failures
- BGP Best- External project provides support for advertisement of Best- External path to the iBGP/RR peers when a locally selected bestpath is from an internal peer
- BGP PIC Unipath provides a capability to install a backup path into the forwarding table to provide prefix independent convergence in case of the PE- CE link failure

PIC EDGE: LINK PROTECTION

BGP RESILIENCY/HA ENHANCEMENT



- Dataplane (via BFD or link layer mechanism) detects PE3- CE2 link failure
  - immediately swaps to repair path label Traffic shunted to PE4 and across PE4- CE2 link

BGP MULTIPATH SIGNALING

Benefits of Multi-Path

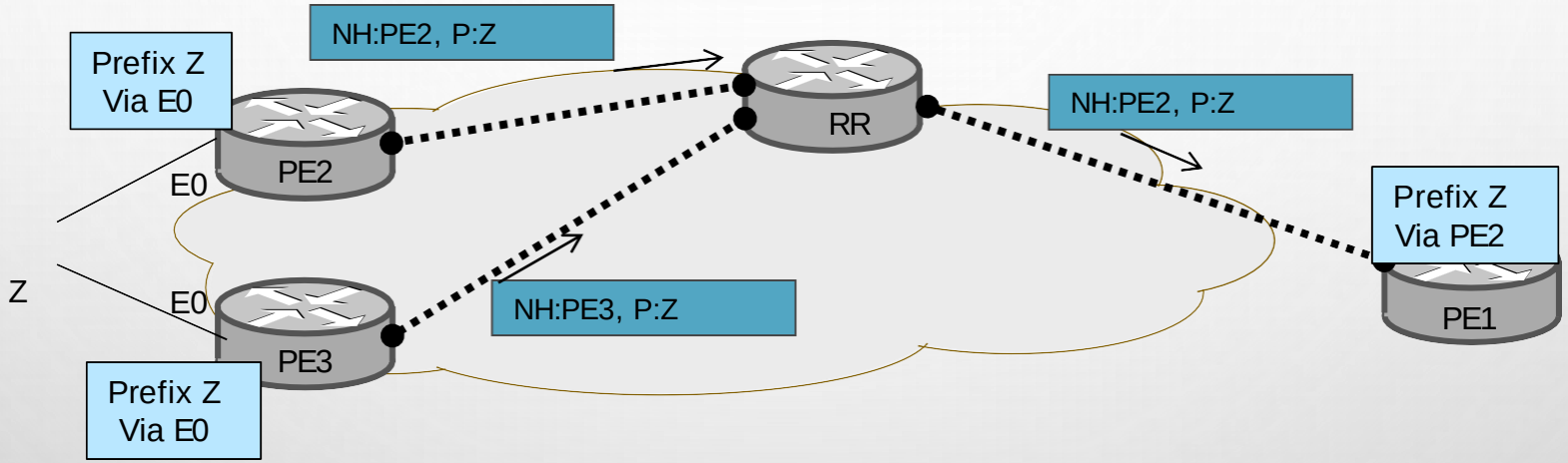
- Convergence
  - BGP fast convergence (2+ paths in local BGP DB)
  - BGP PIC edge (2+ paths ready in forwarding plane)
- Multipath load balancing
  - ECMP LB (eg in data center)
- Prevents Oscillation
- Allows Hot Potato Routing

BGP BEST PATH SELECTION

For Your Reference

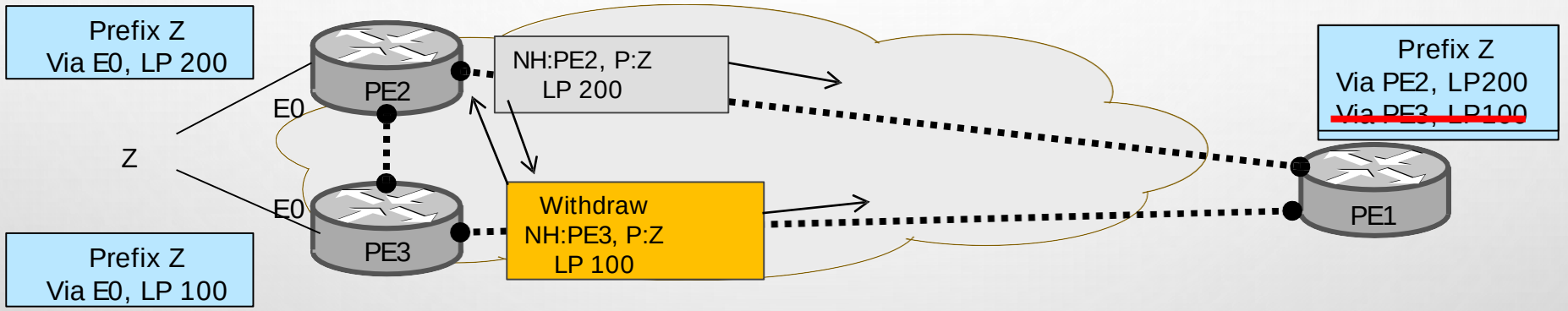
Path selection mechanism	Details
Weight	This is a Cisco-defined attribute that is assigned locally to your router and does not get carried through to the router updates. If there are multiple paths to a particular IP address (which is very common), then BGP looks for the path with the highest weight. There are several ways to set the weight parameter, such as the neighbor command, the as-path access list, or route maps.
Local Preference	This is an indicator to the AS as to which path has local preference, with the highest preference being preferred. The default is 100.
Network or Aggregate	This criterion prefers the path that was locally originated via a network or aggregate. The aggregation of specific routes into one route is very efficient and saves space on your network.
Shortest AS_PATH	BGP uses this one only when there is a “tie” comparing weight, local preference, and locally originated vs. aggregate addresses.
Lowest origin type	This deals with protocols such as Interior Gateway Protocol (IGP) being a lower preference than Exterior Gateway Protocol (EGP).
Lowest multi-exit discriminator (MED)	This is also known as the external metric of a route. A lower MED value is preferred over a higher value
eBGP over iBGP	Similar to “lowest origin type”, BGP AS Path prefers eBGP over iBGP
ieBGP Multiple paths	BGP path selection stop here for ieBGP multipath.
Lowest IGP metric	This criterion prefers the path with the lowest IGP metric to the BGP next hop.
eBGP Multiple paths	BGP path selection stop here for eBGP multipath.
External paths	When both paths are external, it prefers the path that was received first (the oldest one).
Lowest router ID	This prefers the route that comes from the BGP router with the lowest router ID.
Minimum cluster list	If the originator or router ID is the same for multiple paths, it prefers the path with the minimum cluster list length.
Lowest neighbor address	This prefers the path that comes from the lowest neighbor address

BEST PATH SELECTION IMPACT: BGP ROUTE- REFLECTOR



RR does best path selection, result..only one NLRI is annonce to RR client.

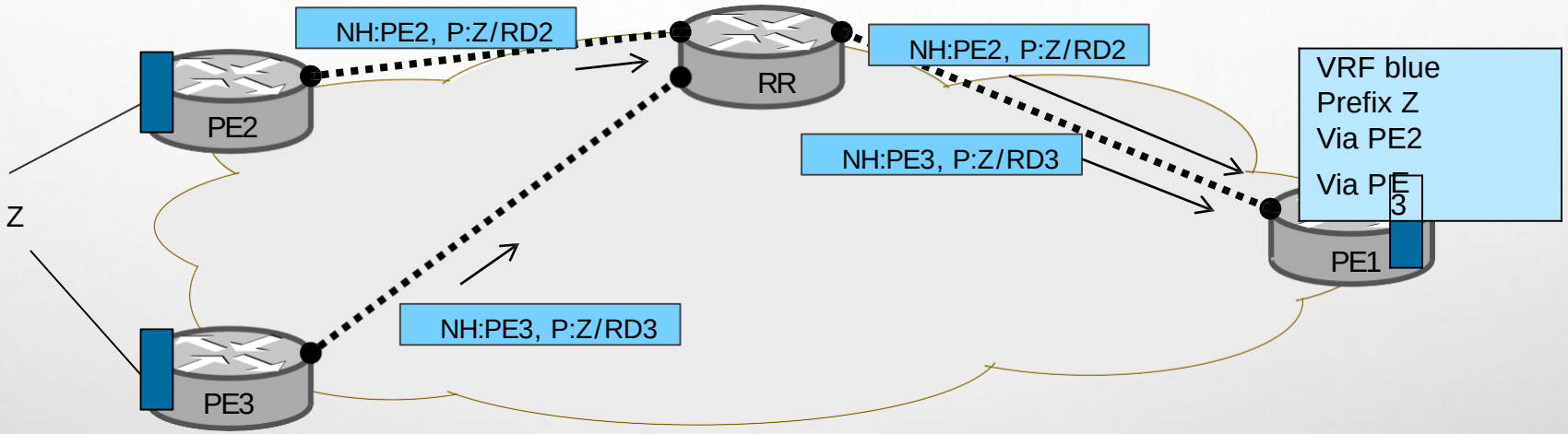
BGP POLICY IMPACT



BGP does withdraw path with lower policies (MED, Local Preference, Weight,..),… result only one NLRI is announce to BGP peer's.

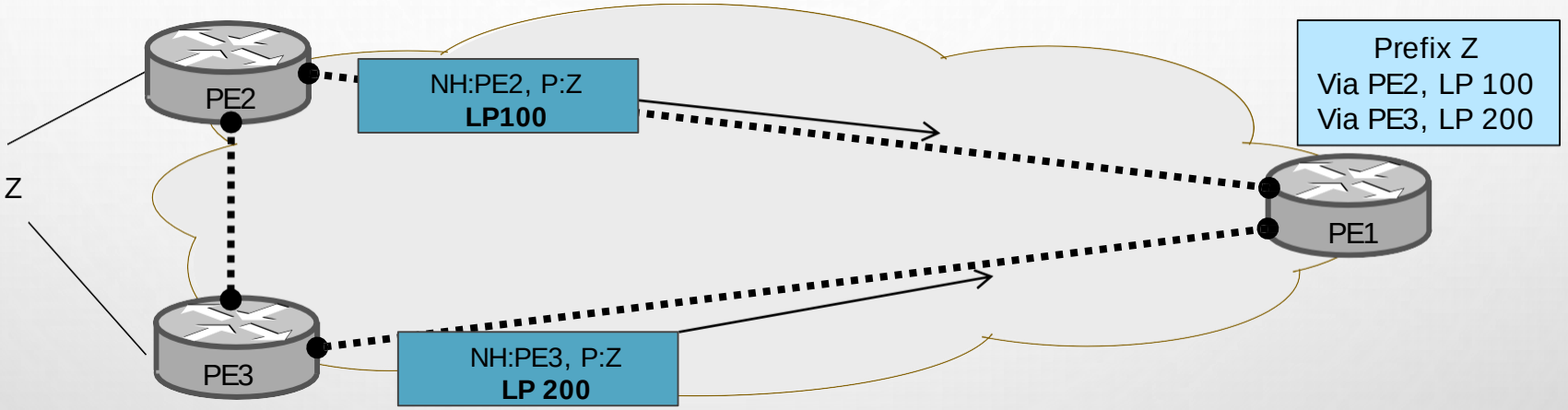
BGP Path Diversity: Solutions

UNIQUE RD FOR MPLS- VPN



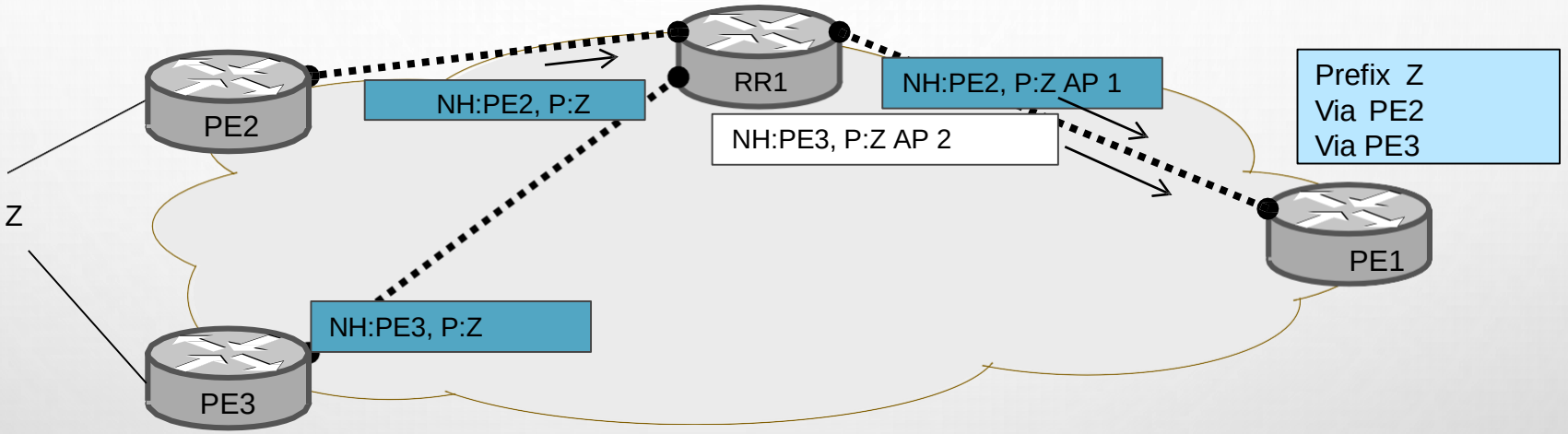
- Unique RD per VRF → Unique VPNv4/v6 NLRI
- RR does best path on two different VPNv4/v6 NLRI, both forwarded
- Recommended method for MPLS- VPN

BGP BEST EXTERNAL



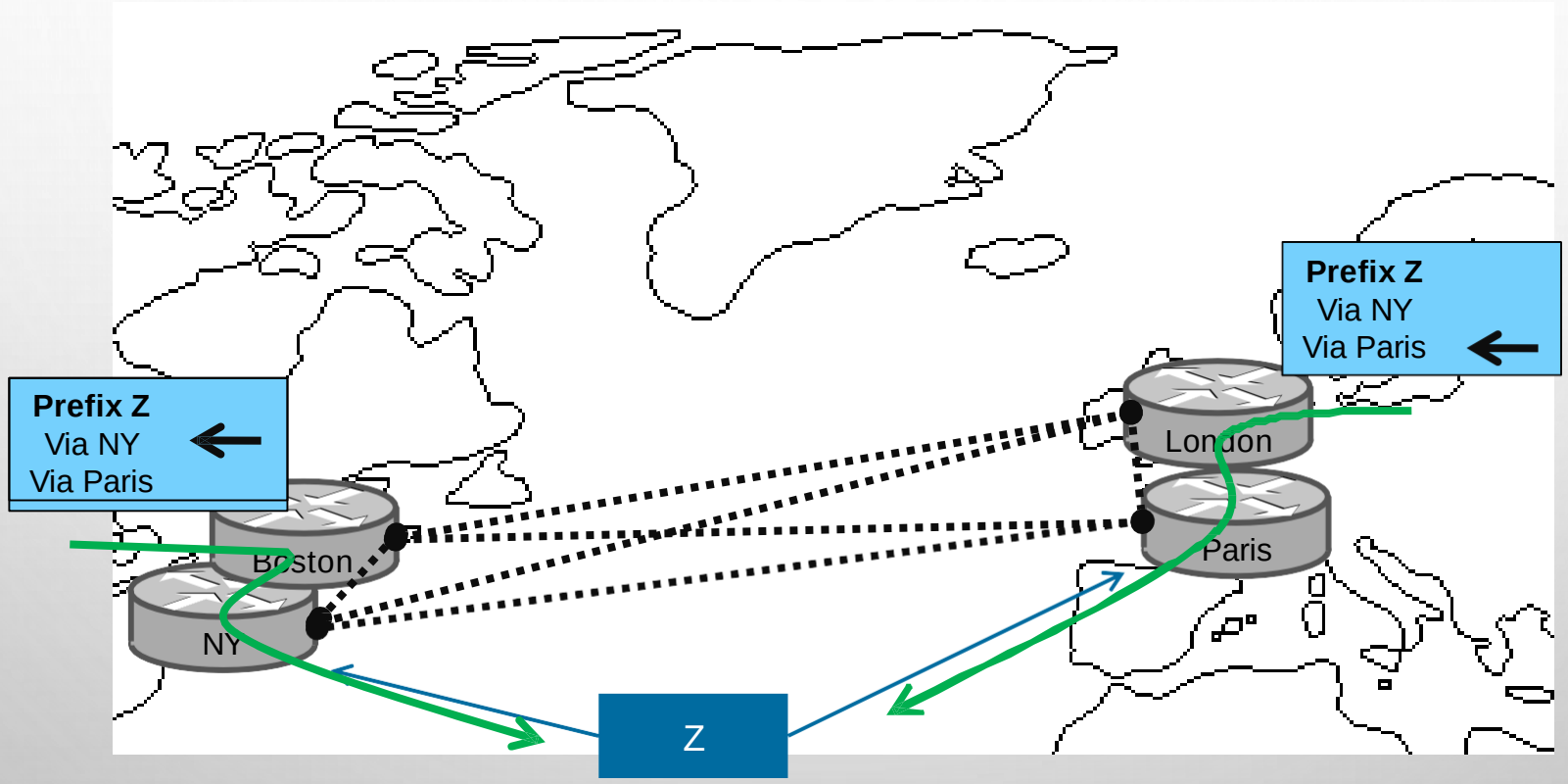
- With Best External, The backup PE (PE2 here) still propagate to the RRs or Peers its own best external path.
- PE1 and PE3 have 2 path

BGP ADD-PATH

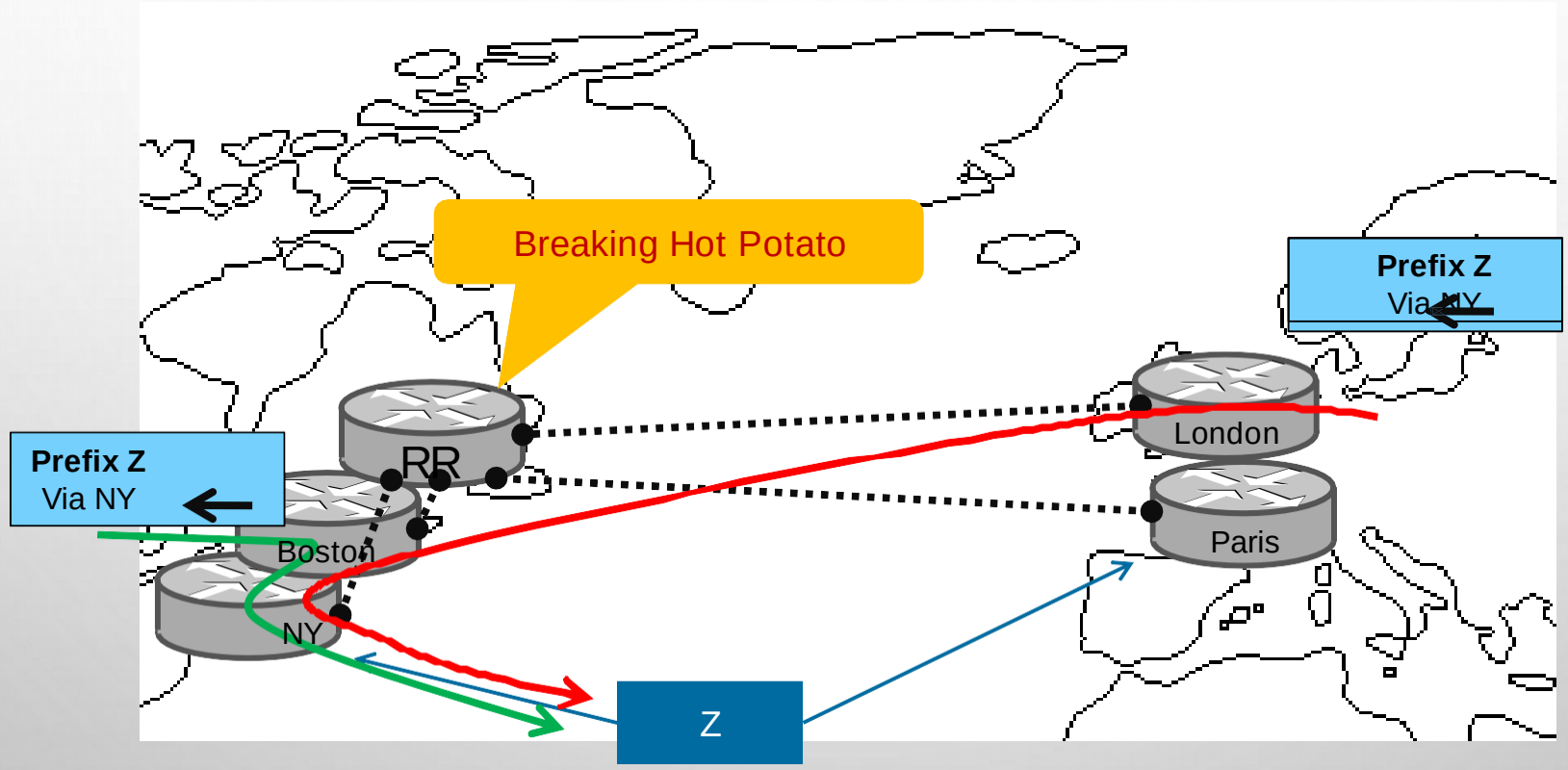


- Add-Path will signal diverse paths from 2 to X paths
- Required all Add-Path receiver BGP router to support Add-Path capability.

Hot Potato with Full Mesh



Hot Potato with RR



Optimal Route Reflection

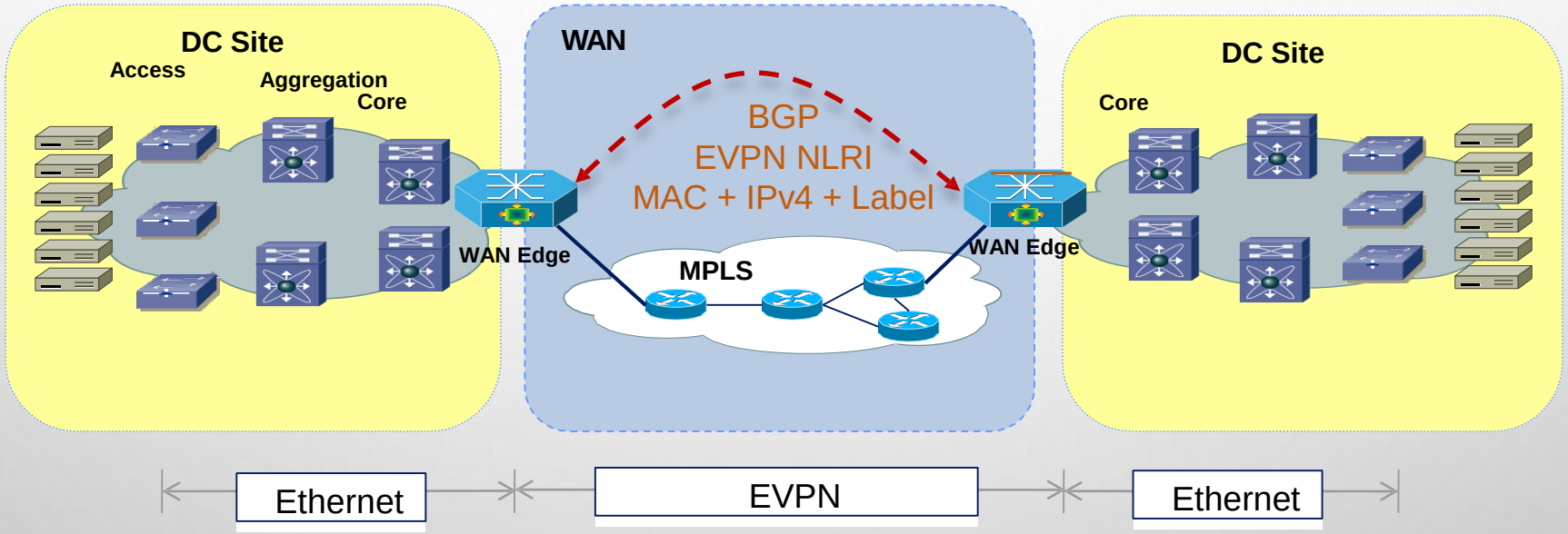
- Described in draft-ietf-idr-optimal-route-reflection-06
- Three ways of doing ORR
  - Addpath (Option 1)
  - RR based ORR (Option 2)
  - RR Client BGP assisted ORR (Option 3)

BGP VPN Enhancements

L3VPN IBGP PE-CE (RFC6368)

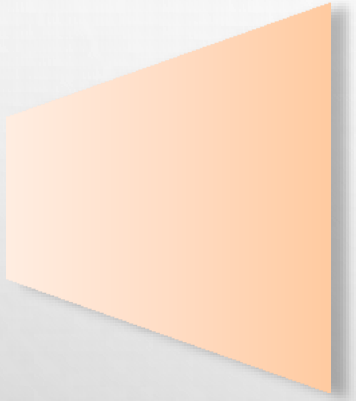
iBGP between PE and CE is supported. The PE will place the received iBGP attributes in a new attribute ATTR_SET and transport them over the Service Provider backbone. This way the Customer BGP attributes (i.e. local pref) are retained

EVPN (RFC-7432)



WHY EVPN ?

Data Center Interconnect requirements were not fully addressed by current L2VPN technologies



- Per-Flow Redundancy and Load Balancing
- Simplified Provisioning and Operation
- Optimal Forwarding
- Fast Convergence
- MAC Address Scalability

Ethernet Virtual Private Network (EVPN) and Provider Backbone Bridging EVPN (PBB-EVPN) designed to address these requirements

BGP FOR SDN

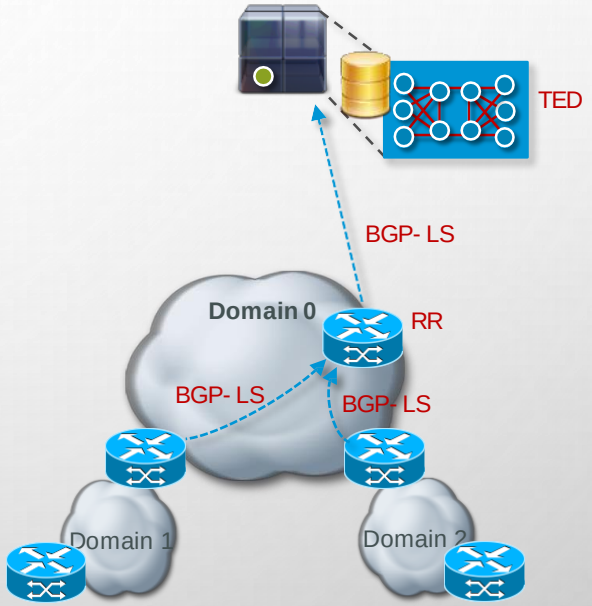
What is SDN?

(per Wikipedia definition)

Software defined networking (SDN) is an approach to building computer networks that separates and abstracts elements of these systems

High Level Perspective of BGP-Linkstate (BGP- LS)

- BGP may be used to advertise link state and link state TE database of a network (BGP- LS)
- Provides a familiar operational model to easily aggregate topology information across domains
- New link- state address family
- Support for distribution of OSPF and IS- IS link state databases
- Topology information distributed from IGP into BGP (only if changed)

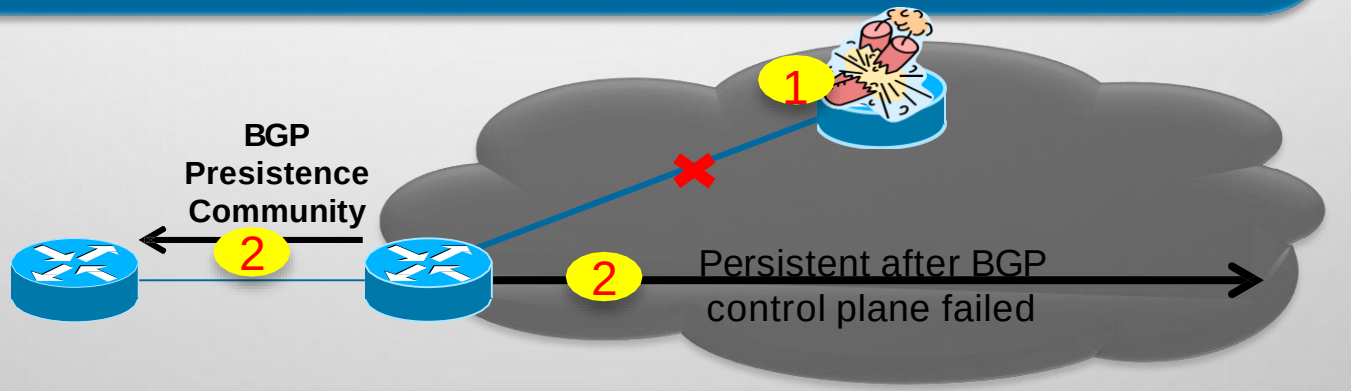


BGP PERSISTENCE (DRAFT-UTTARO-IDR-BGP-PERSISTENCE-003)

BGP persistence:
for certain AFI/SAFI combinations it is desirable that a BGP speaker be able to retain routing state learned over a session that has terminated.

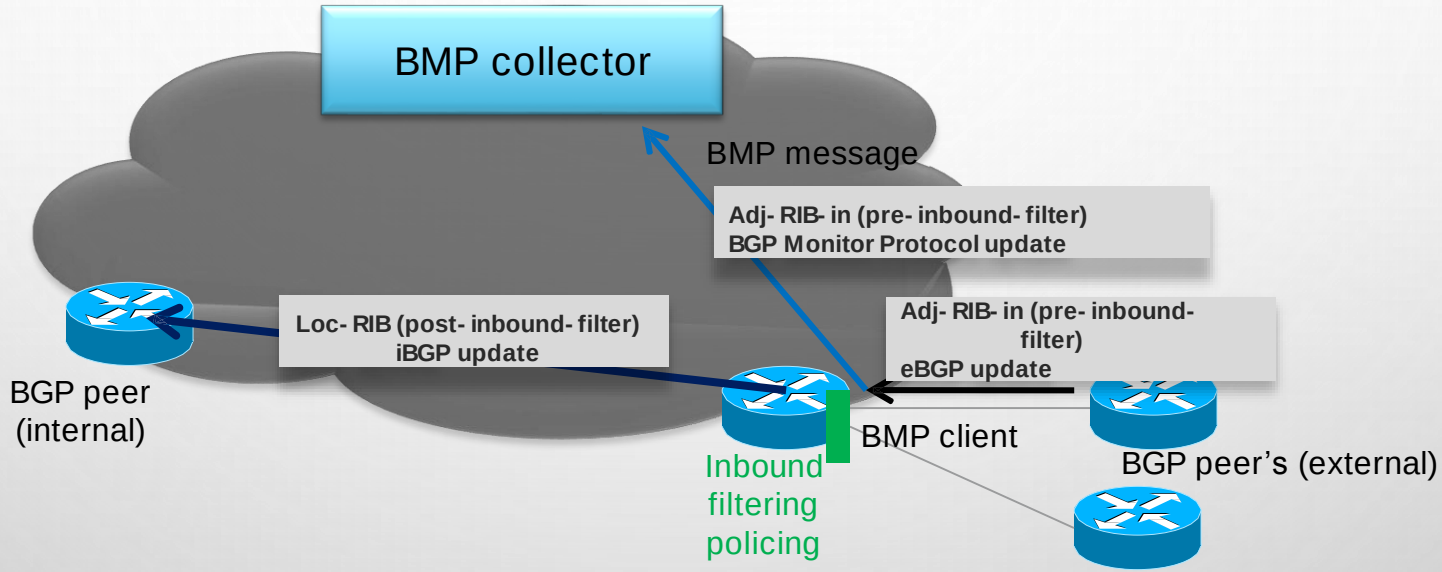
By maintaining routing state forwarding may be preserved.

draft- uttaro- idr- bgp- persistence



BMP Overview (RFC Draft)

BMP Client/Collector



BMP OVERVIEW

WHY BMP ?

There are many reasons to use BMP, but to highlight a few common ones:

1. **Looking Glasses** - IPv4, IPv6, and VPN4
2. **Route Analytics** - Track convergence times, history of prefixes as they change over time, monitor and track BGP policy changes, etc...
3. **Traffic Engineering Analytics** - Adapt dynamically to change and know what is the best shift
4. **BGP pre- policy What- Ifs** - Pre- policy routing information provides insight into all path attributes from various points in the network allowing nonintrusive what- if topology views for new policy validations

many more….

Other BGP cool features (that may interest you ☺)

- Segment Routing – BGP EPE
- Slow peer management (pure implementation)
- AIGP
- BGP flowspec
- Multi-AS BGP
- RTC & Automated RT Filtering
- Accept Own
- Attribute Filtering
- BGP origin Validation
- eiBGP multipath for non-VRF interfaces
- Per CE label

SUMMARY

- BGP is the glue which ties together the AS based internet infra.
- BGP is essential for all most all future services.
- The role of BGP has changed from the ‘protocol of the internet’ to the preferred signaling (control plane) protocol for any kind of services.
- BGP has gone through lot of new changes (adv features) to mitigate the new service requirements.
- For the future SDN or NFV based services, BGP is going to play a major role
- Service Providers need to carefully evaluate each of the features discussed before deploying the actual services on top.

REFERENCES

- BGP GR shut(IETF draft) -- https://tools.Ietf.Org/html/draft-ietf-grow-bgp-gshut-03
- BGP PIC (IETF draft) -- https://tools.Ietf.Org/html/draft-rtgwg-bgp-pic-00
- BGP multipath (RFC 6674)
  - BGP best external
  - Add-path
  - Optimal route reflection
- Segment routing with BGP EPE (ietf-draft) - https://tools.Ietf.Org/html/draft-ietf-idr-bgpls-segment-routing-epe-05
- iBGP as PE-CE protocol (RFC 6368)
- Evpn/Evpn-pbb (rfc-7432)
- BGP-LS (rfc-7752)
- BGP flowspec (RFC-5575)
- BGP persistence (RFC-5575)
- BGP MP (IETF draft) - https://tools.Ietf.Org/html/draft-ietf-grow-bmp-17

Thank You!!!