TOMORROW starts here.
Advanced in BGP

BRKRST3371

Bertrand Duvivier
bertrand@cisco.com
Agenda

• Introduction
• BGP infrastructure
• VPN enhancement
• High Availability
• Virtual Route Reflection
• Multipath Signaling
• BGP enabler for SDN
• Egress Peering Traffic Engineering
• Summary
Motivation and Development of BGP

When the Internet grew and moved to an autonomous system (AS) mesh architecture it was needed to have stable, non-chatty and low CPU consuming protocol to connect all of these AS’s together.

In June 1989, the first version of this new routing protocol was formalized, with the publishing of RFC 1105, *A Border Gateway Protocol (BGP)*.
Service Provider Routing and Services progress

- 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
## Control-Plane Evolution

Almost all services are moving towards BGP

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

- **Investment protection:**
  - Well know 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
BGP infrastructure enhancement
Infrastructure enhancements

• Keepalive Enhancements
  – Loosing or delayed keep-alive message result in session flaps
  – Hence keep-alive processing is now placed into a separate process using priority queuing mechanism

• Update Generation Enhancements
  – Update generation is the most important, time-critical task
  – Is now a separate process, to provide more CPU Quantum

• Parallel Route Refresh
  – Significant delay (up to 15–30 minutes) seen 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, the new code dynamically adapts to the address family used, the available router memory and the number of peers in an update group
Security Enhancement

BGP path attribute detection

- Malformed
  - Invalid content
  - Invalid length
- Transitive
  - Unknown
  - Unwanted

BGP path attribute error handling

- Treat-as: BGP session reset
- Treat as: BGP withdraw mechanism
- Path attribute Filtering

New features:
- IOS-XR 4.2.2
- IOS-XE 3.7
- NX-OS radar
Scalability enhancement: 64bit OS strategy

• $2^{32} = 4$ Giga bits addressable memory
  – Limit: $1.5 - 4$ BGP Millions prefixes
• $2^{64} = 16$ Eta bits addressable memory
  – « unlimited »

• IOS – XR
  – NCS6000 runs XR o Linux 64b
  – ASR9000 runs XR o QNX 32b → plan to support XR o Linux 64b in 2015
  – XRv runs XR o QNX 32b → plan to support XR o Linux 64b in 2015

• IOS – XE
  – ASR1000, CSR1000v, ISR4400 runs XE o Linux 64b
  – Others runs IOS-OS 32b
Performance: Multi-core strategy

• CPU clock: speed doesn’t increase anymore.
• CPU core: number of core per CPU are increasing. 2, 4, 8, 16, 64
• IOS XR:
  - Most of components runs in separate thread (eg IGP)
  - XR BGP run over 16+ thread’s, all synchronized
  - Optimized for multi-core CPU’s
• IOS XE:
  - IOSd runs within single thread
• NX-OS
  - BGP runs within single thread
BGP VPN enhancements
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.

Device(config)# router bgp 100
Device(config-router)# address-family ipv4 vrf blue
Device(config-router-af)# neighbor 10.0.0.1 internal-vpn-client
Multicast L3VPN Solution Space
(complete solution is now available)

- Service
  - IPv4: Native
  - IPv6: Native
  - IPv4: mVPN
  - IPv6: mVPN

- C-Multicast Signaling
  - PIM
  - PORT

- Core Tree Signaling
  - PIM (pt-mpt)
  - MLDP (pt-mpt | mpt-mpt)
  - P2MP TE (pt-mpt)

- Encapsulation /Forwarding
  - IP/GRE
  - LSM

VR 4.3.0
XE 3.8
NX-OS: radar
VPLS Label Switch Multicast

- VPLS LSM enables to multicast Unknown, Broadcast or Multicast Ethernet traffic over VPLS by an efficient packet replication method.
- BGP does signal VPLS leaves to source Mc VPLS PE.
- VPLS LSM leverages RSVP P2MP-TE LSP multicast distribution tree in the provider core for.

VPLS LSM: A Way of Efficient Replication

XR 5.1.0
XE: no plan
NX-OS: no plan
Ingress Replication uses unicast LSP in the core. Unicast LSP can be LDP, MP2P LSP, RSVP-TE, or LDP-over-RSVP-TE LSP. It supports BGP AD as well as PIM/BGP C-multicast routing. FRR can also be used in IR. Key benefit of IR is that core network does not need multicast support.
PBB-EVPN / EVPN

DC Site
Access
Aggregation
Core

WAN
BGP
EVPN NLRI
MAC + IPv4 + Label

MPLS

WAN Edge

XR 5.2.0
XE 3.15 (RR only)
NX-OS: CY2015

DC Site
Core

Ethernet

PBB-EVPN / EVPN

Ethernet
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 High Availability
## Bandwidth Impact on High Availability

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2015</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BGP BW</td>
<td>E1/T1</td>
<td>100GE</td>
<td>Moving to 100G bundle’s</td>
</tr>
<tr>
<td>Packet lost / sec</td>
<td>400</td>
<td>400.000.000</td>
<td>1B in future</td>
</tr>
<tr>
<td>BGP convergence</td>
<td>5-10 min</td>
<td>50-200 msec</td>
<td>BGP PIC (*)</td>
</tr>
<tr>
<td>OSPF/ISIS convergence</td>
<td>1 min</td>
<td>50 msec</td>
<td>LFA FRR (*)</td>
</tr>
<tr>
<td>Multi-path</td>
<td>No</td>
<td>32</td>
<td>moving to 64, 128 (*)</td>
</tr>
<tr>
<td>Box HA</td>
<td>None</td>
<td>Full HA</td>
<td>NSR( *), ISSU, BGP GSHUT</td>
</tr>
</tbody>
</table>

(*) Cisco Innovation
PIC Edge Feature 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) has been supported in IOS-XR/IOS for a while 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 projects 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.

XR 4.2.1
XE 3.8
NX-OS 6.2.8 (IP only)
PIC Edge: Link Protection
BGP Resiliency/HA Enhancement

- CEF (via BFD or link layer mechanism) detects PE3-CE2 link failure
  - CEF immediately swaps to repair path label
  Traffic shunted to PE4 and across PE4-CE2 link
• PE1 detects loss of PE3’s /32 host route in IGP
  – CEF immediately swaps forwarding destination label from PE3 to PE4 using backup path

• BGP on PE1 computes a new bestpath later, choosing PE4
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.

Traffic is redirected
BGP virtual route reflection
Virtual Route-Reflector.

- Number of control planes are growing due to integration of more BGP services, driving the need to optimize/virtualize BGP route-reflector functions.

- Without compromising:
  - Scalability (32/64b OS) and performance (Multi-core support)
  - Independence of operations: reload/update/changes VM’s or Container’s
  - Same BGP implementation and software version as deployed on the Edge (XR, XE)
  - Management (Hypervisor)

XR: XRv
XE: CSR1100v
NX-OS: radar
Evolution of IOS BGP Route-Reflector’s

From 7200 to vRR
Why having multiple 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)

• Prevent oscillation

• Allow hot potato routing
# BGP Best Path Selection

## 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 announce 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.
Solutions

- VPN unique RD
- BGP Best External
- BGP Add-Path
- BGP Optimal Route Reflection
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

IOS-XR
IOS-XE
NX-OS

VRF blue Prefix Z Via PE2 Via PE3

NH:PE2, P:Z/RD2
NH:PE2, P:Z/RD2
NH:PE3, P:Z/RD3
NH:PE3, P:Z/RD3

• 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.

* Next slides for more details

XR 4.3.1*
XE 3.10*
NX-OS 6.2.8*
### BGP Add-path flavors

IETF draft define few flavors of Add-x-Path:

- **Add-n-path**: with add-n-path the route reflector will do best path computation for all paths and send n best to BR/PE.
  
  **Usecase**: Primary + n-1 Backup scenario. (n= 3 IOS-XE, n=2 IOS XR)

- **Add-all-path**: with add-all-path, the route reflector will do the primary best path computation (only on first path) and then send all path to BR/PE. (supported on all OS: IOS-XE, IOS-XR and NX-OS)
  
  **Usecase**: hot potato routing scenario, large DC ECMP load ballancing

- **Add-all-multipath+backup**: with add-all-multipath, the route reflector will do the primary best path computation and send all equal-path (according to multipath rules) and one backup to Border Router / Provider Edge.
  
  **Usecase**: large DC ECMP load ballancing

<table>
<thead>
<tr>
<th></th>
<th>IOS-XE</th>
<th>IOS-XR</th>
<th>NX-OS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add-n-path</td>
<td>![checkmark]</td>
<td>![checkmark]</td>
<td></td>
</tr>
<tr>
<td>Add-all-path</td>
<td>![checkmark]</td>
<td>![checkmark]</td>
<td>![x]</td>
</tr>
<tr>
<td>Add-all-multipath</td>
<td>![x]</td>
<td>![checkmark]</td>
<td>![x]</td>
</tr>
</tbody>
</table>
BGP Optimal Route Reflection Purpose

- Hot Potato Routing attempts to divert the traffic to the closest Egress Router
- Done in BGP using an IGP metric
- Route Reflector (RR) deployments will choose the Egress Router closest to the RR
- An RR selects best path based on IGP metric computed from its IGP database and announce this path to its client BGP Speakers
- RRs are typically located in the data path within a cluster
- Virtual RRs (VRR) are typically placed outside the clusters
- Hot Potato Routing is an issue for a VRR
- RRs can not select paths with optimal IGP metrics for their client BGP Speakers in different clusters
- ORR helps solve the Hot Potato Routing for VRRs
Hot Potato with full mesh

Prefix Z
Via NY
Via Paris

Prefix Z
Via NY
Via Paris

NY
Boston
Z
Paris
London
Hot Potato with route-reflection

Breaking Hot Potato

Prefix Z
Via NY

Prefix Z
Via NY

NY

Boston

RR

London

Paris

Z

RR

NY
Optimal Route Reflection Options

• 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)
Hot Potato with route-reflection + add-all-path

Prefix Z
Via NY
Via Paris

Prefix Z
Via NY
Via Paris

RISK: too many path
RR Based ORR (Option 2)

- RR runs SPF multiple times, one per cluster or each RR client BGP Speaker
- Resulting IGP metric values are stored in a Cluster/RR Client based rib table
- BGP best path mechanism modified to compute best path per Cluster/RR Client
- Best path computed using the appropriate rib table
- BGP route advertisement modified to announce best path computed for a given Cluster/RR Client

Pros:
- All changes are confined to RR. No change to RR Client BGP Speakers

Cons:
- Changes to BGP best path algorithm and BGP route announcement mechanism
- Addition of a new module to compute multiple SPFs
Hot Potato with RR Based ORR

neighbor x.x.x.x
  address-family ipv4 unicast
  optimal-route-reflection a.b.c.d

RISK: too many SPF to run

Prefix Z
  Via NY

Prefix Z
  Via Paris

IOS-XR demo code
IOS-XE no plan
NX-OS no plan
RR Client BGP Assisted ORR (Option 3)

- RR requests the IGP metric from the RR Client via BGP using:
  - NH SAFI (draft-varlashkin-bgp-nh-cost-00) or
  - BGP-LS (draft-ietf-idr-ls-distribution-03)

- RR stores IGP metric values in RR Client specific rib table

- BGP best path mechanism modified to compute best path per Cluster/RR Client

- Best path computed using the appropriate rib table

- BGP route advertisement modified to announce best path computed for a given Cluster/RR Client

- Pros:
  - RR does not have to run multiple SPFs

- Cons:
  - Changes required on RR Clients (upgrade needed)
  - Convergence impact on requesting metric values/changes on demand
  - Changes to BGP best path algorithm and route announcement mechanism
ORR Proof Of Concept code available

• Based on XRv (32-bit IOS XR)

• A new module called ORRSPF is written to:
  – Create multiple rib tables per ORR configuration
  – Create SPT for a particular context – IGP (ISIS/OSPF), area/level, rooted at a SPF root.
  – Store routes from multiple SPFIs in their respective tables.
  – Use IGP data from ISIS/OSPF or BGP-LS.

• Made the following changes to BGP:
  – CLI to configure ORR roots
  – Calculate best path per ORR root config (Cluster or RR Client IP)
  – Create update groups per ORR config
  – Consult appropriate RIB table for hot potato routing
  – Advertise best path per ORR config
  – Made use of addpath infrastructure on RR
BGP enabler for SDN
Control and Data Plane resides within Physical Device
Software defined networking (SDN) is an approach to building computer networks that separates and abstracts elements of these systems.
In other words...

In the SDN paradigm, not all processing happens inside the same device.
SDN Hybrid Approach

- 20+ Years investment in Distributed Control Planes—capex, skills and expertise—by both vendors and customers

- Distributed Control Planes designed to survive battlefield conditions with the possibility of multiple failures

- Leave the distributed control plane in place for “normal” traffic, use SDN for traffic that needs special handling (routing, bandwidth reservation etc.)

- In the event of an SDN Controller failure, you still have a network that works, maybe not as optimally
WAN BW optimization: 90%
- Distributed optimization
- Full Mesh Auto BW RSVP-TE tunnels
- HIGH OPEX (complex)
  - Cust A >50K tunnels
  - Cust B >100K tunnels
- Generate Network Oscillation (instability)

WAN BW, Latency, QoS optimization: 95%
- Hybrid optimization (Centralized/Distributed)
- SDN PCE controller driven WAN optimization
- Adequate Segment Routing TE tunnel
- Low OPEX (simple)
  - Cust A <50 tunnels
  - Cust B <100 tunnels
Gathering up-to-date WAN network state

- To do its job SDN WAN Controller requires up-to-date network visibility information, primarily about
  - **Topology**
    - IGP (OSPF/ISIS) information, direct link/passive, or better: **BGP**
  - **Load/Capacity**
    - SNMP, NetFlow, NETCONF/YANG
    - ISIS, OSPF, BGP LS extended TE attributes (future)
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 flowspec

- BGP (like any other routing protocol) influences destination-based routing
- BGP routing information can be injected from a central place (“SDN controller”)
- Why not use it for more than just giving a destination address to route packets to?

- “Flow Specification Rules”
  - Application aware Filtering/redirect/mirroring
  - Dynamic and adaptive technology
  - Simple to configure
Use case 1: Security DDoS mitigation

**Description:** The goal is to push policies to match on certain flows under DDoS attacks and drop/rate limit or redirect traffic to DDoS scrubber to protect peering / enterprise customers.

**Business:** SP to sale DDoS mitigation services to enterprise customers, generating add value to IP transit services.

- **BGP flowspec**
  - Match: DDOS flow
  - Action: redirect to DDOS scrubber

- **DDOS Controller**
  - **DDOS Analyser**

- **DDOS scrubber**

- **Flexible Netflow**

- **Scan Netflow data**
  - To detect DDOS signature
Use case 2: Redirection to DC/NfV

**Description:** The goal to redirect certain flows from IP NGN or Internet transit network to DC and NfV appliances

**Business:** SP to sale NfV appliance services to enterprise customers, generating add value to IP NGN and IP transit services
Cisco BGP flowspec is

**Standard supported**

- **BGP flowspec**: RFC5575
- **IPv6 support**: draft-ietf-idr-flow-spec-v6
- **IP Next Hop redirection options**: draft-ietf-idr-flowspec-redirect-ip
- **Origin check relax**: draft-ietf-idr-bgp-flowspec-oid
- Optimized flow based forwarding plane.
- **Controller, Route Reflection and Client.**

Tested with exaBGP (IPv4 controller), Arbor (IPv4 controller), Juniper (IPv4 client) and Alcatel (IPv4 & IPv6 client)
BGP flowspec infrastructure

Phase 1
Phase 2

BGP

BGP flowspec

BGP

Flowspec Manager

Policy Infrastructure (E-PBR)

Platform hardware

CLI

XR XML

YANG
Router acting as BGP flowspec client

BGP Flowspec
Match X
Action Y

BGP flowspec

Phase 2
Phase 1

CLI
XR XML
YANG

Flowspec Manager
Policy Infrastructure (E-PBR)
Platform hardware

BGP
Flowspec Manager

BGP
BGP

Phase 2
Phase 1
Router acting as BGP flowspec SDN controller

- **BGP Flowspec Manager**
  - CLI
  - XR XML
  - YANG
  - BGP Flowspec Match X Action Y

- **Policy Infrastructure** (E-PBR)
- **Platform hardware**

- **BGP flowspec**

- **BGP**

- **Flowspec Manager**
# BGP flowspec tuple support for IPv4/v6

<table>
<thead>
<tr>
<th>BGP Flowspec NLRI type</th>
<th>QoS match fields</th>
<th>Value input method</th>
<th>Controller</th>
<th>ASR9k As client</th>
<th>CRS As client</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>IPv4/v6 Destination address</td>
<td>Prefix length</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Type 2</td>
<td>IPv4/v6 Source address</td>
<td>Prefix length</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Type 3</td>
<td>IPv4/v6 protocol</td>
<td>Multi value range</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Type 4</td>
<td>IPv4/v6 source or destination port</td>
<td>Multi Value range</td>
<td></td>
<td>*</td>
<td>✓</td>
</tr>
<tr>
<td>Type 5</td>
<td>IPv4/v6 destination port</td>
<td>Multi Value range</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 6</td>
<td>IPv4/v6 Source port</td>
<td>Multi Value range</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Type 7</td>
<td>IPv4/v6 ICMP type</td>
<td>Multi value range</td>
<td>✓</td>
<td>Future</td>
<td>✓</td>
</tr>
<tr>
<td>Type 8</td>
<td>IPv4/v6 ICMP code</td>
<td>Multi value range</td>
<td>✓</td>
<td>Future</td>
<td>✓</td>
</tr>
<tr>
<td>Type 9</td>
<td>IPv4/v6 TCP flags (2 bytes include reserved bits)</td>
<td>Bit mask</td>
<td>✓</td>
<td>Only lower byte</td>
<td>not all bits</td>
</tr>
<tr>
<td>Type 10</td>
<td>IPv4/v6 Packet length</td>
<td>Multi value range</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Type 11</td>
<td>IPv4/v6 DSCP, Traffic Class</td>
<td>Multi value range</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Type 12</td>
<td>IPv4 fragmentation bits</td>
<td>Bit mask</td>
<td>✓</td>
<td>Only indication of fragment</td>
<td>✓</td>
</tr>
<tr>
<td>Type 13</td>
<td>IPv6 flow label optional header</td>
<td>Multi value range</td>
<td>✓</td>
<td>Future</td>
<td>future</td>
</tr>
</tbody>
</table>
# BGP flowspec extended community actions

<table>
<thead>
<tr>
<th>BGP ext-community value</th>
<th>PBR Action</th>
<th>XR PI</th>
<th>ASR9k</th>
<th>CRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0x8006 (RFC5575) Traffic Rate 0</td>
<td>drop</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>0x8006 (RFC5575) Traffic Rate &lt;rate&gt;</td>
<td>police</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>0x8008 (RFC5575) Redirect VRF</td>
<td>redirect vrf</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>0x8009 (RFC5575) Traffic Marking</td>
<td>Set dscp</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>0x800b (IP redirect draft) Redirect IP NH</td>
<td>nexthop IPv4/v6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
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

BMP client/collector

- BMP collector
  - BMP message
  - Adj-RIB-in (pre-inbound-filter) BGP Monitor Protocol update
- Loc-RIB (post-inbound-filter) iBGP update
- Inbound filtering policing
- BMP client
- BGP peer (internal)
- BGP peer's (external)

XR 5.2.2
XE 3.11
NX-OS no plan
BMP overview

1. BMP devices (e.g. routers) send BMP messages to a BMP collector.

2. Open Daylight (ODL) or NCS (TAIL-f) controller SQL plugin with SQL using Yang interfaces with the BMP receiver database. ODL or NCS in this fashion provides an abstract view of all BMP collector data.

3. Admins, Network Engineers, automated programs/scripts, etc. interact via ODL or NCS northbound interfaces to run various BMP analytics.

4. Admins, Network Engineers, automated programs/scripts, etc. can also go direct to the BMP database as needed.
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*
Egress Peering Traffic Engineering
BGP Traffic Engineering

BGP peering
- Follow BGP best path

BGP Egress Peering Traffic Engineering
- Egress Peering SLA and BW optimization
- Cisco leading architecture with BGP Segment Routing
  - BGP EPE SR
  - BGP Label Unicast with prefix-SID attribute
BGP LS EPE Segment Routing

<table>
<thead>
<tr>
<th>Interfaces</th>
<th>PeerAdj SID</th>
<th>PeerNode SID</th>
<th>PeerSet SID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eth0/1 (1.0)</td>
<td>101</td>
<td>111</td>
<td>121</td>
</tr>
<tr>
<td>Eth0/2 (2.0)</td>
<td>102</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>Eth0/3 (3.0)</td>
<td>103</td>
<td>113</td>
<td>122</td>
</tr>
<tr>
<td>Eth0/4 (3.4)</td>
<td>104</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## BGP / RPL – IOS XR

### XR 4.3.0
- BGP c-multicast signaling
- BGP Session/Prefix Scale

### XR 4.2.3
- XR BGP – show bgp (afi safi) neighbors (addr) advertised-routes enhancement
- BGP – Maintain list of dropped prefixes, updates for a duration – Per neighbor Knob

### XR 4.3.1
- BGP c-multicast signaling
- BGP Session/Prefix Scale
- L3VPN IAS option B and CSC label policing and label filtering (RPF check)
- L3VPN dynamic route leaking
- Mix of aggregate (per VRF) and specific (pe Prefix) labes
- BGP add-all-multipath
- Inbound/outbound policy for RT SAFI
- VPN route-limit

### XR 4.3.2
- PBB-EVPN for normalized hand-off
- BGP per neighbor link-bandwidth

### XR 4.3.3
- RPL – if best-external

### XR 4.5.0
- BGP mVPN PE-PE ingress replication
- BGP – mVPN PE-PE IR inter AS
- BGP Link State distribution
- BGP static network
- BGP mVPN multi-instance
- RPL set-label (vPE)
- RPL IPv6 bits address match
- RPL mVPN BGP sign/ad support
- XRVR – VRR (GA)

### XR 5.1.1
- BGP flowspec
- BGP VRF aware SAFI 2 and mVPN SAFI 129 multi-topology multicast routing
- RPL – set admin-distance
- BGP – CL to disable Next Hop Self for Multipath prefixes
- BGP – remove Private AS in inbound policy
- BGP – ignore AS-Path length for multipath
- RPL – match on atomic aggregate flag of a prefix

### XR 5.1.2
- BGP Multi-Segment pw (FEC 129)

### XR 5.2.0
- BMP
- BGP persistent
- NG mVPN musti-segment
- IBGP local-AS
<table>
<thead>
<tr>
<th>XE 3.8</th>
<th>AR1000, ASR90x, CSR1000v</th>
</tr>
</thead>
<tbody>
<tr>
<td>• VPLS BGP signaling</td>
<td></td>
</tr>
<tr>
<td>• mVPN BGP dampening</td>
<td></td>
</tr>
<tr>
<td>• BGP VPN distinguisher Attribute</td>
<td></td>
</tr>
<tr>
<td>• BGP multiclpuster id</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>XE 3.9</th>
<th>AR1000, ASR90x, CSR1000v</th>
</tr>
</thead>
<tbody>
<tr>
<td>• VRF aware conditionnal advertisement</td>
<td></td>
</tr>
<tr>
<td>• Wildcard for VPN distinguisher / Rt rewrite</td>
<td></td>
</tr>
<tr>
<td>• IPv6 NSR</td>
<td></td>
</tr>
<tr>
<td>• BGP local-AS allow-policy</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>XE 3.10</th>
<th>AR1000, ASR90x, CSR1000v</th>
</tr>
</thead>
<tbody>
<tr>
<td>• BGP per-CE-label</td>
<td></td>
</tr>
<tr>
<td>• BGP NSR for Inter-AS option B (IPv4 &amp; IPv6)</td>
<td></td>
</tr>
<tr>
<td>• L3VPN iBGP PE-CE</td>
<td></td>
</tr>
<tr>
<td>• VRF aware route-reflection</td>
<td></td>
</tr>
<tr>
<td>• iBGP multipath for native IPv4/IPv6</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>XE 3.11</th>
<th>AR1000, ASR90x, CSR1000v</th>
</tr>
</thead>
<tbody>
<tr>
<td>• VRF aware BGP translate-update</td>
<td></td>
</tr>
<tr>
<td>• EVVPN / PBB EVVPN route-reflection</td>
<td></td>
</tr>
<tr>
<td>• BMP (BGP monitoring protocol)</td>
<td></td>
</tr>
<tr>
<td>• RTC for legacy PE</td>
<td></td>
</tr>
<tr>
<td>• GSHUT enhancement</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>XE 3.12</th>
<th>AR1000, ASR90x, CSR1000v</th>
</tr>
</thead>
<tbody>
<tr>
<td>• BGP Accumulated IGP</td>
<td></td>
</tr>
<tr>
<td>• VPLS inter-AS option – BGP signaling</td>
<td></td>
</tr>
<tr>
<td>• BGP flowspec route-reflection</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>XE 3.13</th>
<th>AR1000, ASR90x, CSR1000v</th>
</tr>
</thead>
<tbody>
<tr>
<td>• BGP – set track xx into route-map</td>
<td></td>
</tr>
<tr>
<td>• BGP- Subcodes for BGP Cease notification (RFC 4486)</td>
<td></td>
</tr>
<tr>
<td>• BGP – NSR enhancement</td>
<td></td>
</tr>
<tr>
<td>• BGP – draft-ietf-idr-as4octet-extcomm-generic-subtype (RFC 5668)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>XE 3.14</th>
<th>AR1000, ASR90x, CSR1000v</th>
</tr>
</thead>
<tbody>
<tr>
<td>• BGP split horizon enhancement</td>
<td></td>
</tr>
<tr>
<td>• BGP flowspec route-reflection</td>
<td></td>
</tr>
</tbody>
</table>
### BGP – NX-OS

#### 6.2.6
- BGP – DFA Mc RR

#### 7.0
- BGP – DFA CP
- BGP remote-private-as

#### 6.2.8
- BGP+label (RFC3107)
- BGP - support for IPv6 next hop for IPv4 routes (RFC5549)
- cisco-BGP-MIBv2 (IPv6, 4bytes ASN)
- Update-delay Timer knob
- PIC Edge (IP only, no MPLS, multipath) – This is for the BGP PIC unipath for v4

#### 7.1
- BGP+label (RFC3107)
- BGP - support for IPv6 next hop for IPv4 routes (RFC5549)
- cisco-BGP-MIBv2 (IPv6, 4bytes ASN)
- Update-delay Timer knob
- PIC Edge (IP only, no MPLS, multipath) – This is for the BGP PIC unipath for v4
- Weighted ECMP (BGP, RIB, UFDM, FIB)
- MPLS-VPN Inter-AS Option B (full)
Many BGP books available
Call to Action

• Visit the World of Solutions for
  – Cisco
  – Walk in Labs
  – Technical Solution Clinics

• Meet the Engineer

• Lunch time Table Topics

• DevNet zone related labs and sessions

• Recommended Reading: for reading material and further resources for this session, please visit www.pearson-books.com/CLMilan2015
Complete Your **Online Session Evaluation**

- Please complete your online session evaluations after each session. Complete 4 session evaluations & the Overall Conference Evaluation (available from Thursday) to receive your Cisco Live T-shirt.

- All surveys can be completed via the Cisco Live Mobile App or the Communication Stations.
Thank you.