MPLS

BGP Tunneling – the MPLS Use Case

In This Section

- Tunneling BGP Over the Core
- Where MPLS Fits In

The BGP Problem Statement

- Global BGP table is huge, and growing
  - Over 500,000 IPv4 prefixes and growing
  - IPv6 space is growing but currently negligible
  - See http://bgp.potaroo.net/ for table growth stats
- Why is this a problem?
  - IP routing is destination based
  - All devices in the transit path must know the destination
  - E.g. all transit routers must have full BGP feed(s)

Routing Through vs. To the Core

- Transit providers sell transit, not applications
  - E.g. ISP is not the same as an ASP
- Traffic routes through the SP, not to the SP
  - E.g. end client needs to ping end application, not core link
- How does this affect core routing?
  - To SP core, only the ingress point and egress point matter
  - Original source and final destination are arbitrary

Tunnels – The Ultimate Band-Aid

- Simple transit solution is to tunnel traffic over core from ingress to egress
  - Only the ingress and egress devices need full end-to-end information
  - Core only needs info about ingress and egress devices
- How can we tunnel?
  - QinQ, GRE, IPinIP, MPLS, etc.
  - MPLS is de-facto standard

Example Case - BGP over GRE over Core

- Form a GRE tunnel from ingress to egress
  - Tunnel subnet is link-local and arbitrary
- Peer BGP from ingress to egress
- Recurse BGP next-hop to tunnel
  - Either peer through the tunnel, or modify next-hop to the tunnel
- What is the core’s data plane result?
  - Core routes ingress PE to egress PE
  - Core does not need end-to-end information

Where MPLS Fits In

- MPLS is the core's tunnel encapsulation
  - Same exact logic as GRE
- MPLS is more flexible
  - Arbitrary transport
  - Arbitrary payload
  - Extensible applications
  - Much more on this later…

Example Case – BGP over MPLS over Core

- Form an MPLS tunnel from ingress to egress
  - Typically IGP + LDP is used for this
  - Could be BGP or RSVP (MPLS TE)
  - More on this later…
- Peer BGP from ingress to egress
- Recurse BGP next-hop to MPLS label
- What is the core’s data plane result?
  - Core label switches ingress PE to egress PE
  - Core does not need end-to-end information

MPLS

MPLS Overview

In This Section

- What is MPLS?
- Why Use MPLS?
- MPLS Device Roles
- MPLS Device Operations
- MPLS Label Distribution Protocols

What is MPLS?

- Multiprotocol Label Switching
- Originally Cisco proprietary
  - Previously called “tag switching”
- Now an open standard
  - RFC 3031 - Multiprotocol Label Switching Architecture

What is MPLS? (cont.)

- Name implies two important parts
  - Multiprotocol
  - Label switching

What is MPLS? (cont.)

- Multiprotocol
  - Can transport different payloads
- Layer 2 payloads
  - Ethernet, Frame Relay, ATM, PPP, HDLC, etc.
- Layer 3 payloads
  - IPv4, IPv6, etc.
- Extensible for new future payloads

What is MPLS? (cont.)

- Label Switching
  - Switches traffic between interfaces based on locally significant label values
- Similar to a legacy virtual circuit switching
  - Frame Relay input/output DLCI
  - ATM input/output VPI/VCI

Why Use MPLS?

- Transparent tunneling over SP network
- BGP free core
  - Saves routing table space on Provider (P) routers
- Offer L2/L3 VPN service to customers
  - No need for overlay VPN model
  - Much more on this later…

Why Use MPLS? (cont.)

- Traffic engineering
  - Distribute load over underutilized links
  - Give bandwidth guarantees
  - Route based on service type
  - Detect and repair failures quickly
    - i.e. Fast Reroute (FRR)

MPLS Label Format

- [RFC 3032 - MPLS Label Stack Encoding]
- 4 byte header used to “switch” packets
  - 20 bit Label = Locally significant to router
  - 3 bit EXP = Class of Service
  - S bit = Defines last label in the label stack
  - 8 bit TTL = Time to Live

How Labels Work

- MPLS Labels are bound to FECs
  - Forwarding Equivalency Class
  - IPv4 prefix for our purposes
- Router uses MPLS LFIB instead of IP routing table to switch traffic
- Switching logic
  - If traffic comes in if1 with label X send it out if2 with label Y

MPLS Device Roles

- MPLS network consists of three types of devices
  - Customer Edge (CE)
  - Provider Edge (PE)
  - Provider (P)

CE Devices

- Customer Edge (CE)
- Last hop device in customer’s network
  - Connects to provider’s network
- Can be layer 2 only or layer 3 aware
- Typically not MPLS aware

PE Devices

- Provider Edge (PE)
  - Previously called Label Edge Routers (LER)
- Last hop device in provider’s network
  - Connects to CE and Provider (P) core devices
- PE performs both IP routing & MPLS lookups

PE Devices (cont.)

- For traffic from customer to core…
  - Receives unlabeled packets (e.g. IPv4)
  - Adds one or more MPLS labels
  - Forwards labeled packet to core

PE Devices (cont.)

- For traffic from core to customer…
  - Receives MPLS labeled packets
  - Removes one or more MPLS labels
  - Forwards packet to customer

P Devices

- Provider (P)
  - Previously called Label Switch Router (LSR)
- Core devices in provider's network
  - Connects to PEs and/or other P routers
- Switches traffic based only on MPLS labels

MPLS Device Operations

- PE & P routers perform three major MPLS operations
- Label push
  - Add a label to an incoming packet
  - AKA label imposition
- Label swap
  - Replace the label on an incoming packet
- Label pop
  - Remove the label from an outgoing packet
  - AKA label disposition

Label Distribution

- Labels are advertised via a Label Distribution Protocol
- Label Distribution Protocol (LDP)
  - Advertises labels for IGP learned routes
  - RFC 5036 - LDP Specification
- MP-BGP
  - Advertises labels for BGP learned routes
  - RFC 3107 - Carrying Label Information in BGP-4
- RSVP
  - Used for MPLS Traffic Engineering (MPLS TE)
  - RFC 3209 - RSVP-TE: Extensions to RSVP for LSP Tunnels

MPLS

Label Distribution Protocol (LDP)

In This Section

- Label Distribution Protocol (LDP) Operations
  - Discovering LDP Neighbors
  - Forming LDP Adjacencies
  - Advertising Labels
- LDP Configuration
- LDP Verification

Discovering LDP Neighbors

- Like IGPs, LDP automatically discovers neighbors with a Hello protocol
- Hellos use multicast UDP
  - 224.0.0.2 – “All Routers Multicast”
  - UDP src & dst port 646
- Hello include “IPv4 Transport Address”
  - Address to use for the TCP session
  - Defaults to the LDP Router-ID

Forming LDP Adjacencies

- LDP sessions are formed reliably over TCP
  - Unicast between transport addresses
  - TCP port 646
- Implies peers must have routes to each other’s transport addresses
  - E.g. their Loopbacks
- Transport addresses could be modified, but not normally needed
  - Like IGP, hardcoding Router-ID is recommended

Advertising Labels

- Once LDP session is established, Label is advertised for FEC
  - I.e. Label to IPv4 prefix mapping
- Label distribution can be implicit or explicit
  - Unsolicited Downstream vs. Downstream on Demand
  - Depends on implementation & config options
- Labels could be advertised for some or all routes
  - Cisco default is all IGP routes
  - Really only /32 Loopback matters
    - More on this later…

LDP Configuration

- Enable CEF
  - Should already be on by default
- Agree on label protocol
  - **mpls label protocol**
  - Should already be LDP by default
- Recommended to define Router-ID
  - **mpls ldp router-id**
- Enable LDP
  - Interface **mpls ip**
  - IGP process **mpls ldp autoconfig**

LDP Verification

- Verify LDP is enabled
  - **show mpls interfaces**
- Verify LDP sessions
  - **show mpls ldp neighbor**
- Verify LFIB
  - **show mpls forwarding-table**
- Troubleshooting LDP Adjacencies
  - **debug mpls ldp transport events**

MPLS

Virtual Routing & Forwarding Instances (VRFs)

In This Section

+ VRFs Overview
+ VRF Aware Routing Protocols
+ VRF Lite Configuration & Verification

What Are VRFs?

- Virtual Routing and Forwarding Instances (VRF)
- VRFs create “virtual routers” inside IOS
  - VRF defines a new instance of the routing table
  - Interfaces assigned to the VRF belong to that routing table
  - Interfaces not in a VRF are in the “global” or “default” vrf
- Result is essentially a VPN
  - Separates control plane instances
  - Separates data plane based on routing
    - E.g. I can’t reach you if I have no route to you
  - Addressing can overlap in different VRFs

VRF Aware Routing

- Routing inside a VRF can be through…
  - VRF aware static routes
  - VRF aware dynamic routing
    - RIP
    - EIGRP
    - OSPF
    - IS-IS
    - MP-BGP
  - Policy Routing

Creating VRFs

- Specify locally significant VRF name
  - **ip vrf [name]**
    - IPv4 only
  - **vrf definition [name]**
    - Supports both IPv4 and IPv6
- Specify “route distinguisher”
  - **rd [ASN:nn | IP-address:nn]**
  - More on this in MP-BGP

Applying VRFs

- Apply VRF to interface
  - **ip vrf forwarding [name] | vrf forwarding [name]**
  - Removes IP address from interface
- This minimum configuration is called “VRF Lite”
  - I.e. VRFs without any MPLS config
  - VRFs do not always mean MPLS
  - MPLS does not always mean VRFs

VRF Verification

- Normal commands become “VRF Aware”
  - **show ip route vrf [vrf]**
  - **ping vrf [vrf]**
  - **traceroute vrf [vrf]**
- Features must become “VRF Aware”
  - E.g. VRF Aware NAT, VRF Aware IPsec, etc.
  - Not all features are supported
- Not all vrf parser syntax is intuitive
  - ? is your friend

MPLS

MPLS Layer 3 VPNs Overview

In This Section

- MPLS Layer 3 VPN Overview
- MPLS L3VPN Components
  - VPNv4 BGP
  - Route Distinguishers vs. Route Targets
  - Transport Labels vs. VPN Labels

How MPLS Layer 3 VPNs Work

- MPLS L3VPNs have two basic components…
- Separation of customer routing information
  - VRF – Virtual Routing and Forwarding Instance
  - Different customers have different “virtual” routing tables
  - IGP/BGP run inside the VRF between the customer and SP
- Exchange of customer’s routing info inside SP
  - MP-BGP through the SP network
  - Traffic is label switched towards BGP next-hops

VRF Lite vs. MPLS VPNs

- In VRF lite all devices in transit path must carry all routes in all VRF tables
- In MPLS VPNs only PE routers need customer routes
- Accomplished through…
  - VPNv4 BGP
    - RD + Prefix makes VPN routes globally unique
  - MPLS VPN tag/label
    - P routers only need to know how to reach BGP next-hop
    - Uses “BGP free core” logic

MPLS L3VPN High Level Steps

- Establish an LSP between PEs
  - E.g. IGP + LDP
- Exchange routes with customer
  - PE-CE IGP or BGP
- Exchange customer routes between PEs
  - iBGP + MPLS VPN Label
- Label switch between PEs
  - Data follows the IGP + LDP transport label

Multiprotocol BGP

- How do PE routers exchange VRF info?
  - RFC 4364 - BGP/MPLS IP Virtual Private Networks (VPNs)
- MP-BGP defines AFI 1 & SAFI 128 as VPN-IPv4 or “VPNv4”
  - 8 byte Route Distinguisher (RD)
    - Unique per VPN or per VPN site
    - ASN:nn or IP-address:nn
  - 4 byte IPv4 address
    - Unique per VPN
  - Implies globally unique routes
- VPNv4 includes MPLS VPN label

VPNv4 NLRI Format

- VPNv4 NLRI main attributes include…
  - 8 byte Route Distinguisher (RD)
    - Unique per VPN or per VPN site
    - ASN:nn or IP-address:nn
  - IPv4 prefix & len
    - Unique per VPN because of RD
  - Next Hop
  - MPLS VPN label
- Regular BGP attributes stay the same

Controlling VPNv4 Routes

- Route distinguisher used solely to make route unique
  - Allows for overlapping IPv4 addresses between customers
- New BGP extended community “route-target” used to control what enters/exits VRF table
  - “export” route-target
    - What routes will be go from VRF into BGP
  - “import” route-target
    - What routes will go from BGP into VRF
- Allows granular control over what sites have what routes
  - “import map” and “export map” allow control on a per prefix basis

Route Distinguisher vs. Route Target

- Route Distinguisher
  - Makes the route unique
  - Only one RD per VPNv4 route
- Route Target
  - Controls the route’s VPN membership(s)
  - Can be multiple RTs per VPNv4 route

VPNv4 Route Targets

- 8 byte field
  - [RFC 4360 - BGP Extended Communities Attribute]
- Format similar to route distinguisher
  - ASN:nn or IP-address:nn
- VPNv4 speakers only accept VPNv4 routes with a route-target matching a local VRF
  - Some exceptions, e.g. route reflectors

VPNv4 Route Targets (cont.)

- VPNv4 routes can have more than one RT
- Allows complex VPN topologies
  - Full mesh
  - Hub and Spoke
  - Central services

Transport Label vs. VPN Label

- L3VPN needs at least 2 labels to deliver traffic
  - Can be more with applications like MPLS TE, FRR, etc.
- Transport Label
  - Tells the SP core routers which PE traffic is destined to
    - I.e. who is the exit point
  - Typically derived from LDP
    - Sometimes called the IGP label
- VPN Label
  - Tells the PE router which CE traffic is destined to
  - Derived from VPNv4 advertisements of PEs

MPLS

MPLS Layer 3 VPN Configuration

In This Section

+ MPLS Layer 3 VPN Configuration

MPLS L3VPN Components

- Customer (CE) devices
- PE-CE Routing
  - No MPLS required
  - Normal IPv4 and IPv6 routing
  - All IPv4 protocols supported
  - Some IPv6 protocols supported

MPLS L3VPN Components (cont.)

- MPLS Core (P & PE) devices
- IGP + LDP
  - Goal is to establish LSP between PE /32 Loopbacks
  - Traceroute between loopbacks for verification
- Other labeling methods supported, but outside this scope
  - E.g. BGP + Label, RSVP-TE, etc.

MPLS L3VPN Components (cont.)

- MPLS Edge (PE) devices
- VRF
  - VRF Aware PE-CE Routing
  - Used to locally separate customer routes & traffic
- VPNv4 BGP
  - iBGP peering to remove PE /32 Loopbacks
  - Separates customer control & data plane over MPLS core
  - Other designs supported outside this scope
    - E.g. VPNv4 RR, Multihop EBGP VPNv4, etc.
- Redistribution
  - VRF to BGP import and export policy

MPLS

MPLS Layer 3 VPN Verification

In This Section

- MPLS Layer 3 VPN Verification
  - MPLS L3VPN Components
  - How to Verify Them Individually

MPLS L3VPN Components Review

+ Core IGP + LDP on PEs and Ps
+ VRF PE to CE
+ PE to CE Routing
+ VPNv4 BGP between PEs
+ VRF to VPNv4 Import/Export
+ Data Plane

Core IGP + LDP Verification

- Did LDP properly form?
  - **show mpls interfaces**
  - **show mpls ldp neighbor**
  - **show mpls ldp binding**
  - **show mpls forwarding-table**

Core IGP + LDP Verification (cont.)

- Is there an LSP between PEs /32 Loopbacks?
  - **show mpls forwarding-table**
  - **show ip cef**
  - **traceroute**

VRF PE to CE Verification

- Was the VRF properly allocated?
  - **show [ip] vrf**
  - **show [ip] vrf detail**
  - **show run vrf**
- Were interfaces properly allocated to the VRF?
  - **show ip route vrf ***
  - **show ipv6 route vrf ***

PE to CE Routing Verification

- CE verifications are normal global routing
  - **show ip rip database**
  - **show ip eigrp neighbor**
  - **show ip ospf neighbor**
  - **show ip bgp summary**
- PE verifications are VRF aware

VPNv4 BGP between PEs Verification

- Did VPNv4 BGP establish?
  - **show bgp vpnv4 unicast all summary**
- Are extended communities being sent?
  - **debug bgp vpnv4 unicast updates**

VRF to VPNv4 Import/Export Verification

- Is the RT import/export policy correct?
  - **show [ip] vrf detail**
  - **show run vrf**

VRF to VPNv4 Import/Export Verification (cont.)

+ Did IGP to BGP redistribution occur?
  + **show bgp vpnv4 unicast all**
  + **show ip ospf database**
  + **show ip eigrp topology**
  + **show ip rip database**

VRF to VPNv4 Import/Export Verification (cont.)

- Are VPNv4 routes being sent/received?
  - **show bgp vpnv4 unicast all**
  - **show bgp vpnv4 unicast all neighbor advertised-routes**
  - **debug bgp vpnv4 unicast updates**
  - **clear bgp vpnv4 unicast * [in|out]**

Data Plane Verification

- CE commands are normal global
  - Ensure verification comes from correct source
  - E.g. PE-CE link might not be exported into VPNv4
- PE commands are VRF aware
  - **ping vrf**
  - **traceroute vrf**
  - **telnet vrf**

MPLS

MPLS Layer 3 VPN Troubleshooting

In This Section

- MPLS Layer 3 VPN Troubleshooting
  - What can go wrong?
  - How do I find the problem?
  - How do I fix the problem?

Troubleshooting LDP Adjacency

- Is LDP actually enabled?
  - **show mpls interfaces**
- Is LDP transport working?
  - **debug mpls ldp transport events**
- Is LDP session authenticated?
  - Like BGP, LDP uses TCP auth (option 19)

Troubleshooting LDP Advertisements

- Are labels actually bound?
  - **show mpls ldp binding**
  - **show mpls forwarding-table**
  - **debug mpls ldp binding**
- Is allocation being filtered?
  - Advertise filter vs. allocate filter
  - Typically only /32 needs an allocation

Troubleshooting PE-CE Routing

- Is loop prevention being violated?
  - OSPF down-bit & domain-tag
  - BGP AS-path & Site-of-Origin (SoO)
  - EIGRP Site-of-Origin
  - More on this later…
- Is the VRF just hung?
  - **clear ip route vrf ***
  - Forces re-import/export

Troubleshooting VPNv4 BGP

- Is the route target being filtered?
  - Wrong import/export policy
  - Default route target filter
  - VPNv4 Route Reflection

Troubleshooting MPLS Data Plane

- Is the VPNv4 peering to a /32?
  - Problems in PHP
  - Problems in route summarization
- Are LDP and IGP Synced?
  - LFIB can only install label for RIB/LRIB intersection

MPLS

MPLS PE-CE Routing with EIGRP

MPLS

MPLS PE-CE Routing with OSPF

MPLS

MPLS PE-CE Routing with BGP

In This Section

+ BGP as PE-CE for MPLS L3VPN
+ AS Override vs. Allow AS In
+ BGP Site of Origin (SoO)

MPLS

BGP Multipath for MPLS Layer 3 VPNs

In This Section

- eiBGP Multipath
- VPNv4 Route Reflectors

MPLS

BGP Prefix Independent Convergence (BGP PIC)

In This Section

- Convergence Review
- BGP PIC Edge

What Factors Affect Convergence Time?

+ Generally four factors affect convergence time
  + Failure Detection Time
  + Event Propagation Time
  + Recalculation Time
  + Forwarding Table Update Time

BGP PIC Edge

- BGP PIC Edge is FRR for BGP
  - Install next-best route as a repair path in FIB
  - Once failure is detected, immediately switchover
- Configured as…
  - bgp additional-paths install
- Like IGP FRR, can selective protect certain prefixes