



BGP

BGP Overview

In This Section

- + What is BGP?
- + What is BGP Not?
- + IGP vs. BGP
- + Why use BGP?

What is BGP?

- + Border Gateway Protocol
 - + Open standards based path-vector exterior gateway protocol
- + Open Standards Based
 - + [RFC 4271 “A Border Gateway Protocol 4 \(BGP-4\)”](#)
- + Path-Vector
 - + Uses multiple “attributes” for routing decision
- + Exterior Gateway Protocol
 - + Used to route traffic between autonomous systems
 - + I.e. the Internet routing protocol

What Else is BGP?

- + The most scalable and extensible protocol (so far)
- + Highly Scalable
 - + Over 500,000 IPv4 prefixes and growing
 - + IPv6 space is growing but currently negligible
 - + See <http://bgp.potaroo.net> for table growth stats
- + Highly Extensible via AFI/SAFI
 - + Much more than just IPv4 Unicast routing
 - + IPv6 Unicast, IPv4 & IPv6 Multicast. IPv4 & IPv6 MPLS, MDT, VPLS...
many others

What is BGP Not?

- + BGP is not a *routing* protocol per-se
 - + BGP is an application
 - + I.e. BGP is a *reachability* protocol
- + Why the distinction?
 - + BGP (generally) cannot route the network alone
 - + BGP (generally) relies on IGP for transport and recursion
 - + More detail on this later...

IGP vs. BGP

- + IGP has visibility of the topology
 - + E.g. SPF and DUAL
- + IGP decides on link attributes
 - + I.e. cost belongs to link, not prefix
- + IGP policy application is hard
 - + TE hard to implement with IGP

IGP vs. BGP (cont.)

- + BGP (generally) does not have topology visibility
 - + BGP relies on IGP as a transport
- + BGP decides on prefix vector attributes
 - + I.e. attribute belongs to prefix, not link
- + BGP policy application is easy
 - + TE is simple per prefix

Why Use BGP?

- + Policy application
- + Easy to influence egress traffic
 - + I always choose how I route out
- + Easy to influence ingress traffic
 - + I (generally) choose how traffic returns
 - + Sometimes difficult on Internet scale
 - + More on this later...

When To Use BGP or Not

- + Who owns the address space?
- + Provider Aggregatable (PA) addresses
 - + ISP owns your addresses
 - + They (generally) dictate the policy
 - + BGP not needed
- + Provider Independent (PI) addresses
 - + You own your addresses and BGP ASN
 - + You (generally) dictate the policy
 - + BGP is needed

Other BGP Considerations

- + Do I really need a full view?
 - + 500,000+ IPv4 prefixes
 - + Does RIB have enough memory?
 - + Can FIB actually install it at the linecard?
- + Egress and Ingress policies are unrelated
 - + Default out is ok
 - + Influence traffic back in still allowed

Other BGP Use Cases

- + Scaling the Enterprise
 - + Islands of IGP, Core of BGP
- + Scaling DMVPN
 - + Hubs as BGP RRs
 - + BGP hierarchy is arbitrary
- + Scaling Data Center Fabric
 - + Use of BGP for routing in large-scale data centers
 - + [draft-lapukhov-bgp-routing-large-dc-07](#)

Recommended Reading

- + Recommended Books
 - + [Internet Routing Architectures](#)
 - + [Routing TCP/IP, Volume II](#)
 - + [Practical BGP](#)
 - + [Optimal Routing Design](#)





BGP

Basic BGP Operations



In This Section

- + BGP ASNs
- + Establishing BGP Peerings

BGP ASNs

- + Autonomous System (AS)
 - + “...a set of routers under a single technical administration, using an interior gateway protocol (IGP) and common metrics to determine how to route packets within the AS, and using an inter-AS routing protocol to determine how to route packets to other ASes.” (RFC 4271)
- + ASNs are allocated by Internet Assigned Numbers Authority (IANA)
 - + <http://www.iana.org/numbers/>

BGP ASN Values

- + Originally 2-byte field
 - + Values 0-65535
 - + Public ASNs 1 – 64511
 - + Private ASNs 64512 – 65535
- + Currently 4-byte field
 - + [RFC 4893 - BGP Support for Four-octet AS Number Space](#)
 - + IOS support as of 12.4(24)T

4-Byte BGP ASNs

- + 0.0 – 65535.65535 notation
 - + 0.[0-65535] denote original 2-byte ASNs
- + Requires backwards compatibility with old code
- + 4 Byte ASN support negotiated during capability exchange
 - + “Old” BGP speakers are sent ASdot numbers encoded as ASN “23456”
 - + Real AS-Path encoded with optional transitive attributes
AS4_AGGREGATOR and AS4_PATH

Establishing BGP Peerings

- + Like IGP, first step in BGP is to find neighbors to exchange information with
- + Unlike IGP...
 - + BGP does not have its own transport
 - + BGP has different types of neighbors
 - + BGP neighbors are not discovered by default
 - + BGP neighbors do not have to be directly connected

BGP Transport

- + BGP uses TCP port 179 for transport
 - + Implies that BGP needs IGP first
- + BGP neighbor statement tells process to...
 - + Listen for remote address via TCP 179
 - + Initiate a session to remote address via TCP 179
 - + If collision, higher router-id becomes TCP client

BGP Transport

- + TCP server must agree on where client's session is coming from
 - + If server does not expect session it will refuse
 - + **show control-plane host open-ports**
 - + May be version specific
- + Client's packet is sourced from outgoing interface in the routing table
 - + Can be modified with **update-source** per neighbor

BGP Peering Types

- + External BGP (EBGP) Peers
 - + Neighbors outside my Autonomous System
- + Internal BGP (iBGP) Peers
 - + Neighbors inside my Autonomous System
- + Update and path selection rules change depending on what type of peer a route is being sent to/received from
 - + More on this later...





BGP

iBGP vs. EBGPeering Rules



In This Section

- + iBGP vs. EBGP Transport Establishment
- + iBGP vs. EBGP Loop Prevention
- + BGP Next-Hop Significance
- + iBGP vs. EBGP Next-Hop Processing

iBGP Transport Establishment

- + iBGP packets default to TTL 255
 - + Implies neighbors do not have to be connected as long as IGP reachability exists
- + iBGP peers typically peer via Loopbacks
 - + Allows rerouting around failed paths via IGP
 - + Required for some application such as MPLS L3VPN

EBGP Transport Establishment

- + EBGP packets default to TTL 1
 - + Can be modified if neighbors are multiple hops away
 - + **neighbor ebgp-multihop [ttl]**
 - + **neighbor ttl-security hops [ttl]**
- + Non multi-hop peers must be directly connected by default
 - + Can be modified if connected neighbors peer via Loopbacks
 - + **neighbor disable-connected-check**

iBGP Loop Prevention

- + Loop prevention via route filtering
 - + iBGP learned routes cannot be advertised on to another iBGP neighbor
- + Implies iBGP requires either either...
 - + Fully meshed iBGP peerings
 - + Route reflectors
 - + Confederation
- + Pros and cons to all three designs
 - + More on this later

EBGP Loop Prevention

- + Loop prevention via AS-Path
 - + Local ASN is “prepended” to outbound updates
 - + Inbound updates containing local ASN are discarded
 - + Can be modified with **neighbor allowas-in** or **as-override**

BGP Next-Hop Significance

- + BGP is not a routing protocol
 - + BGP is an application used to exchange NLRI
 - + E.g. the route details but not the path details
- + IPv4 NLRI contains...
 - + Prefix/len
 - + Attributes
 - + Local-pref, AS-Path, MED, etc.
 - + Next-Hop

BGP Next-Hop Significance (cont.)

- + BGP Next-Hop controls IGP route recursion
 - + I.e. BGP knows the next-hop but not the outgoing interface
 - + IGP must be able to perform recursion otherwise the route cannot be used
- + Result of failed next-hop recursion is failed bestpath selection
 - + I.e. the route is not installed in RIB or advertised
- + Why do we care?
 - + Different peerings have different next-hop processing rules

iBGP Next-Hop Processing

- + By default...
 - + Outbound iBGP updates do not modify the next-hop attribute regardless of iBGP peer type
- + Can be modified
 - + **neighbor next-hop-self**
 - + route-map action set ip next-hop
 - + As of 15.1(1)SY, next-hop-self ALL for inserting RR in the data-path
 - + Used in Unified MPLS design

EBGP Next-Hop Processing

- + By default...
 - + Outbound EBGP updates have local update-source for neighbor set as next-hop
 - + E.g. if update-source is Loopback0, next-hop is Loopback0
- + Can be modified
 - + route-map action set ip next-hop
 - + **neighbor next-hop-unchanged**
 - + E.g. “Third Party” next-hop
 - + Limited corner case applications





BGP

iBGP Full Mesh



In This Section

- + Why is iBGP Full Mesh Needed?
- + iBGP Full Mesh Configuration Examples
- + iBGP Full Mesh Design Advantages
- + iBGP Full Mesh Design Disadvantages
- + Partial Mesh Designs

Why is iBGP Full Mesh Needed?

- + iBGP performs loop prevention via route filtering
 - + iBGP learned routes cannot be advertised on to another iBGP neighbor
- + Implies iBGP requires either...
 - + Fully meshed iBGP peerings
 - + Route reflectors
 - + Confederation





BGP

iBGP Full Mesh Configuration Examples



Full Mesh Design Advantages

- + Path Diversity
 - + All BGP peers learn all possible egress paths
- + Optimal Traffic Flows
 - + All BGP peers learn the closest egress path
 - + Path selection by default would be based on IGP metric to egress router
 - + I.e. Hot Potato Routing

Full Mesh Design Disadvantages

- + Control plane scaling is exponential
 - + Full mesh means $n*(n-1)/2$ peerings
 - + 10 routers = 45 peerings
 - + 100 routers = 4950 peerings
 - + 500 routers = 124750 peerings
- + Operationally hard to scale
 - + Adding or changing peering config is administratively prohibitive
 - + Could be automated, but few out of the box solutions

Partial Mesh Designs

- + You don't need to only choose one design
 - + E.g full mesh, RR, and Confed can interoperate
- + Why would you do this?
 - + Pockets of full mesh for path diversity
 - + Inter-cluster RR for scaling larger

Partial Mesh Design Example

- + [RFC 4456 - BGP Route Reflection: An Alternative to Full Mesh Internal BGP \(IBGP\)](#)
 - + Section 11 Configuration and Deployment Considerations
- + *“One of the key component of the route reflection approach in addressing the scaling issue is that the RR summarizes routing information and only reflects its best path.”*
- + *“Both Multi-Exit Discriminators (MEDs) and Interior Gateway Protocol (IGP) metrics may impact the BGP route selection... with certain route reflection topologies the route reflection approach may not yield the same route selection result as that of the full IBGP mesh approach. A way to make route selection the same... [is] by configuring the intra-cluster IGP metrics to be better than the inter-cluster IGP metrics, and maintaining full mesh within the cluster.”*





BGP

iBGP Scaling Techniques



In This Section

- + Route Reflectors
- + Route Reflector Clusters
- + Confederation

iBGP Route Reflection

- + Eliminates need for full mesh
 - + Only need peering(s) to the RR(s)
- + Like OSPF DR & IS-IS DIS, minimizes prefix replication
 - + Send one update to the RR
 - + RR sends the update to its “clients”
 - + Does not modify other attributes when reflecting routes

iBGP Route Reflection Loop Prevention

- + Loop prevention through Cluster-ID
 - + RR discards routes received with its own Cluster-ID
- + Sets Originator-ID attribute to the router-id of RR client on routes RECEIVED from the client
 - + Client uses Originator-ID for loop prevention

Route Reflector Peerings

- + Route reflector can have three types of peers
 - + EBGp peers
 - + iBGP Client Peers
 - + iBGP Non Client Peers

Route Reflector Update Processing

- + RR processes updates differently depending on what type of peer they came from
 - + EBGP learned routes...
 - + Pass to EBGP peers, Clients, & Non-Clients
 - + Client learned routes...
 - + Pass to EBGP peers, Clients, & Non-Clients
 - + Non-Client learned routes...
 - + Pass to EBGP peers & Clients
- + RR placement based upon these rules

Large Scale Route Reflection

- + Large scale BGP designs should not be serviced by a single RR
- + RR “clusters” allow redundancy and hierarchy
 - + Cluster is defined by the clients a RR serves
 - + RRs in the same cluster use the same Cluster-ID
- + Per address-family RR pairs
 - + Avoids “fate sharing” of services

Inter-Cluster Peerings

- + Inter-Cluster peerings between RRs can be client or non-client peerings
 - + Depends on redundancy design
 - + Duplicate routes, more processing
- + Cluster-ID is based on Router-ID
 - + By default all RR are in separate clusters
- + Same cluster-id on both clusters or not?
 - + Depends on redundancy design...

Virtual Route Reflectors

- + RR generally does not need to be in data-path
 - + E.g. I give you the route but you forward through someone else
- + High Memory/CPU VM can do the job
 - + IOS-XRv
 - + CSR1000v
 - + Quagga

Virtual Route Reflectors - Scale

- + No need to install 500k routes in the RIB/FIB if RR is not in the data-path
- + Selective RIB Download Feature
 - + Prevents BGP paths from being installed in RIB/FIB
 - + **table-map [route-map] filter**
 - + Scale to 20 + Million VPNv4 routes

BGP Confederation

- + Reduces full mesh iBGP requirement by splitting AS into smaller Sub-AS
 - + Inside Sub-AS full mesh or RR requirement remains
 - + Between Sub-AS acts like EBGp
- + Devices outside the confederation do not know about the internal structure
 - + Sub-AS numbers are stripped from advertisements to “true” EBGp peers

BGP Confederation (cont.)

- + Typically uses ASNs in private range
 - + 64512 – 65535
- + Can use different IGPs in each Sub-AS
 - + **Next-hop self** on border routers
- + Next-hop, Local-Pref, and MED are kept across Sub-AS eBGP peerings

BGP Confederation Loop Prevention

- + AS_CONFED_SEQUENCE
- + AS_CONFED_SET
- + These attributes never leave the confederation

Route Reflection vs. Confederation

- + Why both?
 - + Generally accomplish the same goal
- + Migrations paths are different
 - + Migrate to confederation is hard
 - + Greenfield confederation is easier
 - + Migrate to RR is easy, just add peers then remove old ones

For More Information

- + Google Talks - [BGP at 18: Lessons In Protocol Design](#)
 - + Dr. Yakov Rekhter, co-designer of BGP
- + Cisco Live 365
 - + [BRKRST-3321 Scaling BGP](#)





BGP

Configuration BGP Route Reflectors



In This Section

- + Configuring BGP Route Reflectors
- + Understanding BGP Cluster ID

Route Reflector Update Processing

- + RR processes updates differently depending on what type of peer they came from
 - + EBGP learned routes...
 - + Pass to EBGP peers, Clients, & Non-Clients
 - + Client learned routes...
 - + Pass to EBGP peers, Clients, & Non-Clients
 - + Non-Client learned routes...
 - + Pass to EBGP peers & Clients
- + RR placement based upon these rules

Route Reflection Configuration

- + Specify client peers
 - + **neighbor route-reflector-client**
- + Configuration is per-AF
 - + IPv4 RR doesn't not imply IPv6 RR
- + Client does not need to specify RR is a RR

BGP Cluster-ID

- + RR Loop prevention through Cluster-ID
 - + RR discards routes received with its own Cluster-ID
- + Sets Originator-ID attribute to the router-id of RR client on routes RECEIVED from the client
 - + Client uses Originator-ID for loop prevention





BGP

Configuring BGP Confederations



In This Section

- + Configuring BGP Confederation

BGP Confederation Configuration

- + Enable the BGP process
 - + **router bgp [sub-as]**
- + Specify the main AS number
 - + **bgp confederation-id [main-as]**
- + Specify other Sub-ASs that you peer with
 - + **bgp confederation-peers [sub-as1 sub-as_n]**
 - + Not all Sub-ASes, just those directly peered with





BGP

BGP Next-Hop Processing



In This Section

- + BGP Next-Hop Significance
- + Next-Hop-Self
- + Manual Next-Hop Modification
- + Third Party Next Hop
- + BGP Next Hop Tracking

BGP Next-Hop Significance

- + BGP is not a routing protocol
 - + BGP is an application used to exchange NLRI
 - + E.g. the route details but not the path details
- + IPv4 NLRI contains...
 - + Prefix/len
 - + Attributes
 - + Local-pref, AS-Path, MED, etc.
 - + Next-Hop

BGP Next-Hop Significance (cont.)

- + BGP Next-Hop controls IGP route recursion
 - + I.e. BGP knows the next-hop but not the outgoing interface
 - + IGP must be able to perform recursion otherwise the route cannot be used
- + Result of failed next-hop recursion is failed bestpath selection
 - + I.e. the route is not installed in RIB or advertised
- + Why do we care?
 - + Different peerings have different next-hop processing rules

Next-Hop-Self

- + Updates send to iBGP peer does not modify next-hop IP address by default
- + Can be modified with...
 - + **neighbor [address] next-hop-self**
 - + Update source IP address is used as next-hop

Manual Next-Hop Modification

- + Next-hop IP can be changed using a route-map
 - + For updates being or being received
- + Applied as...
 - + Neighbor [address] route-map [name] [in | out]
 - + Uses “set” clause in route-map

Third Party Next Hop

- + Used when multiple router are connected on a segment, same IP subnet
- + Not all routers form eBGP peerings
 - + Hub and Spoke design
- + Hub router does not change next hop IP address when sending eBGP updates on same segment
 - + Routers not directly peering end up send traffic directly to each other
- + CCIE Blog - [Understanding Third-Party Next-Hop](#)

Next-hop-Self At The Edge?

- + Should you set next hop self on an edge router?
- + Pros:
 - + Peer can use the same next-hop on outbound updates to iBGP peers
 - + Result is same dynamic update-group
 - + Don't need to include external links in IGP
- + Cons:
 - + Hinders fast convergence of external uplink failure
 - + If external links are unstable, can cause churn in IGP
 - + Type-5 LSAs are flooded to all non-stub areas





BGP

BGP NLRI Origination



In This Section

- + What is NLRI?
- + Methods to Originate NLRI
- + Conditional Advertisement
- + Conditional Route Injection

What is NLRI?

- + Network Layer Reachability Information
 - + E.g. a route
- + BGP uses UPDATE and WITHDRAW messages to exchange NLRI

Methods to Originate NLRI

- + BGP NLRI can be originated multiple ways...
- + **network** statement
 - + Requires exact match in the routing table first
- + **redistribute** statement
 - + Won't include OSPF External by default
- + **aggregate-address** statement
 - + Requires one subnet in BGP table first
- + **bgp inject-map** statement
 - + Opposite of aggregation

BGP Network Statement

- + Originates prefixes with ORIGIN of IGP (i)
- + Requires exact match in the routing table (RIB)
 - + Does not have to be a connected prefix, can be learned via IGP
- + Without the “mask” keyword , assumes classful mask
- + Sets Weight to 32768

BGP Redistribute Statement

- + Originates prefixes with ORIGIN INCOMPLETE (?)
- + Originates classful summary if auto-summary is enabled
- + Automatically copies IGP metric to BGP MED
- + Won't include OSPF External by default
 - + **redistribute ospf [pid] match internal external**
- + Sets Weight to 32768

BGP Redistribute Internal

- + By default only external BGP routes are redistributed into IGP with redistribution
- + **bgp redistribute internal** command lets internal BGP routes to be redistributed into IGP
- + Can result in a routing loop if you're not careful

BGP Conditional Advertisement

- + Neighbor advertise-map map1 [non-exist-map map2 | exist-map map2]
- + Advertise prefix matched in advertise-map;
 - + if prefix matched in non-exist-map does not exist
 - + or prefix matched in exist-map does exist
- + Typically used to track failure of a transit link
 - + E.g. Advertise to backup provider, only if primary provider is down

BGP Conditional Route Injection

- + Originates subnet(s) from aggregate for purpose of longest match traffic engineering

BGP Conditional Route Injection Configuration

- + Configured as `bgp inject-map inject-map exist-map exist-map [copy-attributes]`
 - + Inject Map
 - + Subnet to be advertised
 - + `set ip address prefix-list [list]`
 - + Exist Map
 - + Aggregate to be originated from
 - + `Match ip address prefix-list [list]`
 - + `Match ip route-source prefix-list [list]`





BGP

BGP Bestpath Selection



In This Section

- + BGP Best Path Selection Overview
- + Best Path Selection Order
- + Manipulating Best Path Selection

BGP Best Path Selection

- + Chooses a single BGP best-path which can be...
 - + Installed in the RIB/FIB
 - + Advertised to other BGP peers
- + In general path selection is standardized
 - + [RFC 4271](#) Section 9.1. “Decision Process”
 - + Cisco specific order at...
 - + [BGP Best Path Selection Algorithm](#)

Best Path Selection Prerequisites

- + Next-hop value must be in the routing table
 - + Prevents route recursion failure
- + Synchronization rule must be met or disabled
 - + Legacy black-hole prevention technique
- + AS-Path must not contain Local-AS
 - + Normal EBGp loop prevention
 - + Can be disabled with **allow-as in**
- + First ASN in path must be neighbor's ASN
 - + **bgp enforce-first-as** command

Best Path Selection Order

- + Pre-Bestpath community
 - + More on this later...
- + Weight
 - + Cisco Proprietary
 - + Locally significant
 - + Higher value is preferred
- + Local Preference
 - + Higher value is preferred
 - + Not advertised to eBGP peers
 - + Carried through confederation eBGP

Best Path Selection Order Continued

- + Locally Originated
 - + Locally originated gets Weight of 32768
- + AS-Path
 - + Smaller length is preferred
- + Origin
 - + IGP over EGP over Incomplete
 - + Status codes I > E > ?

Best Path Selection Order Continued

- + MED
 - + Smaller value is preferred
 - + Only compared between for peerings to same provider by default
 - + Lots of other exceptions
 - + More on this later...

Best Path Selection Order Continued

- + EBGP over iBGP
 - + If you learned it via EBGP it's not your prefix
- + IGP Metric to Next-Hop
 - + Can use multi-path if all equal after this step
 - + Hidden Command to allow multipath if AS_PATH is not the same (has to be same length)
 - + **bgp bestpath as-path multipath-relax**

Best Path Selection Order Continued

- + Tie breakers
 - + Oldest
 - + Lowest RID
 - + Shortest Cluster List
 - + Lowest Neighbor Address

Best Path Selection Exceptions

- + AS-Path
 - + **bgp bestpath as-path ignore**
- + MED
 - + **bgp always-compare-med**
 - + **bgp bestpath med-confed**
 - + Compares MED for routes locally originated in the Confederation
 - + **bgp bestpath med missing-as-worst**
 - + Assign MED of 4,294,967,294 to NULL MED
 - + **bgp deterministic med**
 - + Compare MED against all possible paths

Best Path Selection Exceptions Continued

- + IGP Metric
 - + **bgp bestpath igp-metric ignore**
 - + Introduced in IOS-XE 3.4S
- + Router-ID
 - + **No bgp bestpath compare-routerid**

Manipulating Best Path Selection

- + Outbound routing policy affects inbound traffic
- + Inbound routing policy affects outbound traffic
- + Longest Match Routing is above all
 - + Affects both directions

Attributes To Influence Outbound Path Selection

- + Weight & Local-Preference
 - + Set inbound
 - + Affects outbound traffic

Attributes To Influence Inbound Path Selection

- + AS-Path & MED
 - + Set outbound
 - + Affects inbound traffic

Multipath Load Balancing

- + Multipath load balancing for external links with unequal bandwidth capacity
 - + Enabled under IPv4, IPv6, VPNv4, VRF AF
 - + For iBGP, eBGP, eiBGP
- + Still only one best path advertised to peers





BGP

BGP Traffic Engineering with Weight & Local Preference

Attributes To Influence Outbound Path Selection

- + Weight & Local-Preference
 - + Set inbound
 - + Affects outbound traffic





BGP

BGP Traffic Engineering with AS-Path & MED



Attributes To Influence Inbound Path Selection

- + AS-Path & MED
 - + Set outbound
 - + Affects inbound traffic





BGP

BGP Traffic Engineering with Communities

Using BGP Communities To Influence Traffic

- + [RFC 1998](#) - An Application of the BGP Community Attribute in Multi-home Routing
- + Pre-stage an inbound policy to...
 - + Match on pre-defined communities
 - + Set local preference
 - + Set other attributes
- + Result is that outbound community signaling affects outbound path selection on other side





BGP

BGP Aggregation



In This Section

- + BGP Auto-Summary
- + Aggregate Address
- + Aggregation Arguments
- + Attribute and Advertise Map

BGP Auto-Summary

- + Disabled by default
- + Automatically summarizes prefixes
 - + Classful boundaries
 - + No mask required on **network** statement
 - + Requires subnet of network to be present in IGP

BGP Aggregation

- + BGP hierarchy is arbitrary
 - + Path vector is like distance vector, e.g. EIGRP & RIP
 - + You only learn what your neighbor tells you
 - + OSPF & IS-IS is strict leaf-spine-leaf
- + Implies summarization can apply anywhere
 - + Only Requirement is that subnet must exist in BGP table
- + Configured as...
 - + **aggregate-address [network] [mask] [args]**

BGP Aggregation Arguments

- + Arguments are...
 - + **summary-only**
 - + suppress more specific routes
 - + **suppress-map**
 - + suppress subset of more specific routes
 - + **as-set**
 - + un-order list of AS numbers and communities are associated with the aggregate; inherited from more specific routes

BGP Aggregation Arguments (cont.)

+ **advertise-map**

- + attributes of prefix or prefixes matched with advertised-map are inherited in aggregate address
- + Used in conjunction with as-set
 - + Prevents a specific AS from being included in the AS_SET of the new aggregate
 - + Circumvents loop prevention

+ **attribute-map | route-map**

- + Used to modify attributes associated with aggregate address
- + Used in conjunction with as-set
 - + Can set or remove attributes
 - + E.g. remove “no-export” community, add MED





BGP

BGP Traffic Engineering with Aggregation

In This Section

- + BGP Traffic Engineering with Aggregation
- + BGP Unsuppress-Map





BGP

BGP Communities



In This Section

- + What are BGP Communities?
- + Well Known Communities
- + Setting & Matching
- + Standard vs. Extended Communities
- + Other Community Applications

What Are BGP Communities?

- + BGP's implementation of a "route tag"
 - + No route-tag matching or setting support in BGP
 - + Defined in [RFC 1997 - BGP Communities Attribute](#)
- + Used to group prefixes together for...
 - + Advertisement policy
 - + Filtering policy
 - + Best Path Selection policy
 - + [RFC 1998 - An Application of the BGP Community Attribute in Multi-home Routing](#)

BGP Community Attribute

- + Community is an optional transitive attribute
 - + Not exchanged between peers by default
 - + **neighbor [address] send-community**

BGP Community Values

- + Standard Community is 4-byte value
- + Can be denoted as...
 - + Decimal (0 – 4294967296)
 - + AA:NN (0:0 – 65535:65535)
- + Not displayed in accordance to RFC-1997 by default
 - + **ip bgp new-format** or **ip bgp-community new-format**
 - + Displays communities in AA:NN format

BGP Well-Known Reserved Communities

- + No-Export (0xFFFFFFFF01)
 - + Don't advertise to EBGp peers
- + No-Advertise (0xFFFFFFFF02)
 - + Don't advertise to any peers
- + Local-AS (0xFFFFFFFF03)
 - + Don't advertise to Confederation EBGp peers
 - + RFC defines as NO_EXPORT_SUBCONFED

Setting Communities

- + Set occurs directly in route-map
 - + **set community {community-number [additive] [well-known-community] | none}**
 - + Not “additive” by default
- + Can Delete Community by setting to none on route-map

Matching Communities

- + Match occurs via community list
 - + Define list
 - + Standard list matches community name or number
 - + **ip community-list 1 standard permit no-export**
 - + Expanded matches regular expression
 - + **ip community-list expanded AS100 permit 100:[0-9]+**
 - + Reference from route-map
 - + **match community AS100**

Extended Communities

- + [RFC 4360 - BGP Extended Communities Attribute](#)
- + Used for extended applications such as...
 - + MPLS L3VPN Route Target
 - + MPLS L2VPN
 - + EIGRP Cost Community
 - + More on this in MPLS L3VPN

BGP Cost-Community

- + BGP Custom Decision Process
 - + <http://tools.ietf.org/html/draft-retana-bgp-custom-decision-02>
- + Only advertised within the AS or to confederation peers
- + Influences BGP path selection at the Point of Insertion
 - + “pre-bestpath” point of insertion can be used
 - + Compares cost-community value before Weight
- + EIGRP PE/CE
 - + Uses “pre-bestpath” cost-community to encode composite metric into BGP
- + More info at [BGP Cost Community](#)

GSHUT Community

- + Graceful BGP session shutdown
 - + <http://tools.ietf.org/html/draft-ietf-grow-bgp-gshut-05>
- + Used in conjunction with Graceful Restart
- + Takes the restarting peer out of the data-path by modifying local preference
 - + Similar to IS-IS Overload or OSPF Max Metric LSA
- + Introduced on Cisco IOS XE Release 3.6S





BGP

BGP Filtering



In This Section

- + BGP Filtering Methods
- + BGP Regular Expressions
- + Maximum Prefix
- + Outbound Route Filtering (ORF)

BGP Filtering Methods

- + BGP Update filtering occurs on a per peer basis with...
 - + **neighbor [address] distribute-list [standard acl | extended acl]**
 - + **neighbor [address] filter-list [as-path-acl]**
 - + **neighbor [address] prefix-list [prefix-list name]**
 - + **neighbor [address] route-map**
- + Using route-maps avoids order of operations issues

Regular Expressions

- + Used for string matching in...
 - + Show command outputs
 - + TCL/EEM scripting
 - + BGP AS-Path Access Lists
 - + BGP Expanded Community Lists
- + Documented in...
 - + Cisco IOS Terminal Services Configuration Guide
 - + Appendixes
 - + Regular Expressions

BGP Maximum Prefix

- + BGP default, full view allowed from all peers
- + Full view generally should not come from downstream customers
- + Simple solution, filter based on number of prefixes
 - + **neighbor maximum-prefix**
 - + Result can be log or shutdown peer

Outbound Route Filtering

- + In traditional filtering model upstream provider will send
 - + Full BGP table
 - + Default only
 - + Default plus local
 - + No complex view

Inbound Route Filtering

- + Inbound filtering on downstream neighbor
- + Inefficient because full view must still be processed by BGP inbound RIB
- + ORF used to allow downstream neighbor to control what upstream neighbor advertises downstream

BGP Outbound Route Filtering (ORF)

- + [RFC 5291 - Outbound Route Filtering Capability for BGP-4](#)
- + Downstream customer signals upstream provider what view they want
- + Result is outbound filter on provider with control on downstream customer
- + Negotiated during capabilities exchange
 - + **neighbor *ip-address* capability orf prefix-list [send | receive | both]**

Soft Reconfiguration Inbound

- + Saves unfiltered BGP routes received from peer in Adj-RIBs-In
 - + Adj-RIBs-In → Loc-RIB → Adj-RIBs-Out
- + Filtering policy run on local unfiltered copy stored in Adj-RIBs-In
- + **neighbor [address] soft-reconfiguration inbound**
- + Makes troubleshooting easier at the expense of more memory

Route Refresh Capability

- + Negotiated capability sent in OPEN message
- + Can ask peer to send an update without having to hard clear the session
 - + `clear bgp ipv4 unicast [neighbor] soft in`
 - + `clear bgp ipv4 unicast [neighbor] in`
 - + `clear bgp ipv4 unicast * in`
- + Does not consume memory like Soft Reconfiguration Inbound

