

BGP Techniques for Network Operators



Philip Smith
<philip@nsrc.org>
NANOG88, Seattle

Presentation Slides

- Will be available on:
 - <https://bgp4all.com/pfs/conferences/>
 - And on the NANOG88 website
- Feel free to ask questions any time

BGP Videos

- NSRC has made video recordings of excerpts of this presentation, as part of a library of BGP videos for the whole community to use:
 - https://learn.nsrc.org/bgp#intro_to_bgp

The screenshot shows the NSRC (Network Startup Resource Center) website. The navigation bar includes links for Home, About, BGP for All (highlighted), perfSONAR, ScienceDMZ, FedIdM, and Contact Us, along with a search bar. The main content area is divided into several sections:

- BGP for All:** A text-based introduction to BGP, explaining it as the primary routing protocol for the Internet and autonomous systems. It also mentions that understanding routing options can lead to efficiencies for institutions and research/education networks.
- Video Topics:** A list of video topics with 'BGP for All' highlighted in orange, and other topics like perfSONAR, ScienceDMZ, and FedIdM in different colors.
- Introduction to Routing:** A list of video topics including Internet Routing, Routing Protocols, Introduction to IS-IS (UPDATED), IS-IS Levels, IS-IS Adjacencies, Best Configuration Practices for IS-IS on Cisco IOS, IS-IS Authentication, Default Routes and IPv6, Introduction to OSPF, OSPF Areas, OSPF Adjacencies, Best Configuration Practices for OSPF on Cisco IOS, OSPF Authentication, Default Routes and IPv6, Comparing OSPF and IS-IS, Choosing between OSPF and IS-IS, Migrating from OSPF to IS-IS, Migration Plan, and Finalizing Migration.
- Introduction to BGP:** A list of video topics including Introduction to Border Gateway Protocol, Transit and Peering, Autonomous Systems (UPDATED), How BGP works, Supporting Multiple Protocols, IBGP versus EBGP, Setting up EBGP, and Setting up IBGP.
- BGP Case Studies:** A list of video topics including Peering Priorities (NEW), Transit Provider Peering at an IXP (NEW), Customer Multihomed between two IXP members (NEW), Traffic Engineering for an ISP connected to two IXes (NEW), Traffic Engineering for an ISP with two interfaces on one IX LAN (NEW), and Traffic Engineering and CDNs (NEW).
- Communities:** A list of video topics including Communities: RFC 1998 Traffic Engineering, Communities: Simplifying Traffic Engineering, How to Apply Communities to Originated Routes, and How to Use Communities for Service Identification.

A video player is embedded in the center, showing a video titled 'BGP for All' from the 'Internet Routing' series. The video player includes controls for watch later, share, and a progress bar. The video thumbnail features the NSRC logo and the text 'BGP for All' and 'Internet Routing'.

BGP Techniques for Network Operators

- **BGP Basics**
- Scaling BGP
- Using Communities
- Deploying BGP in a Service Provider Network

BGP Basics

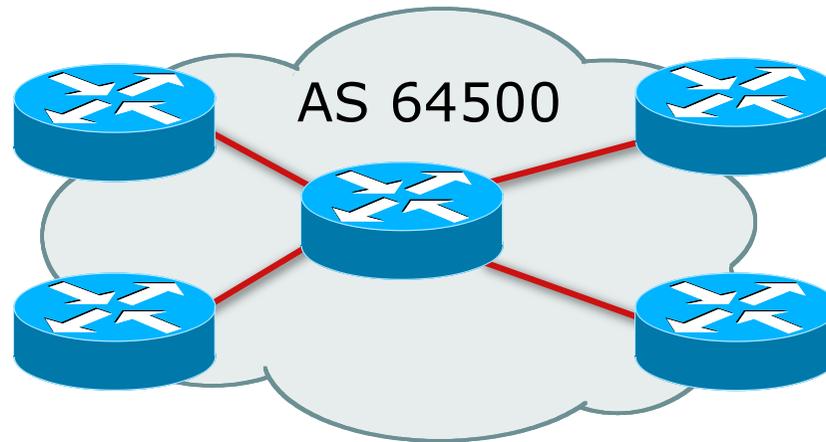


What is BGP?

Border Gateway Protocol

- A Routing Protocol used to exchange routing information between different networks
 - Exterior gateway protocol
- Described in RFC4271
 - RFC4276 gives an implementation report on BGP
 - RFC4277 describes operational experiences using BGP
- The Autonomous System is the cornerstone of BGP
 - It is used to uniquely identify networks with a common routing policy

Autonomous System (AS)



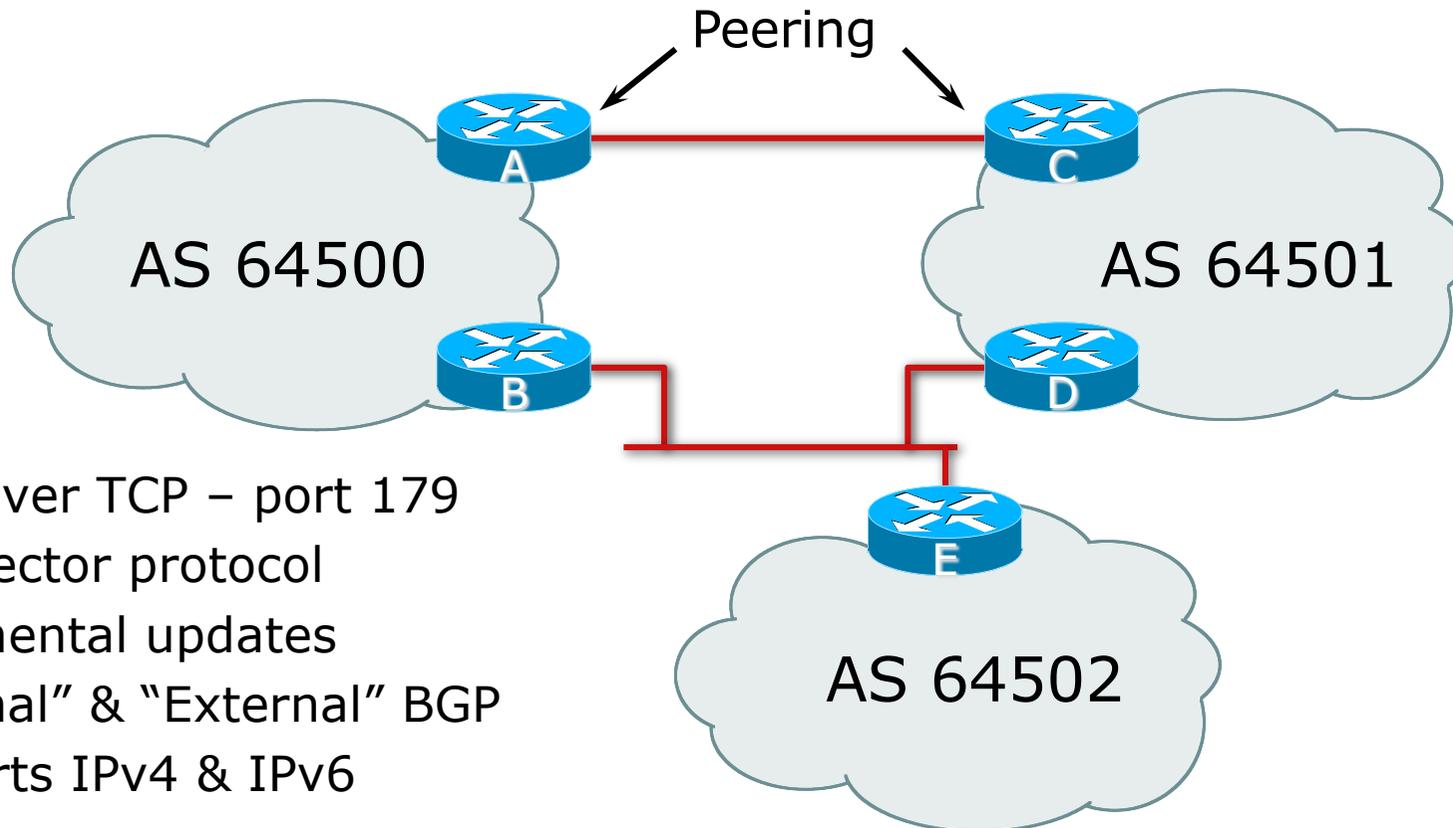
- ❑ Collection of networks with same routing policy
- ❑ Single routing protocol
- ❑ Usually under single ownership, trust and administrative control
- ❑ Identified by a unique 32-bit integer (ASN)

Autonomous System Number

Range:	
0-4294967295	(32-bit range – RFC6793)
	(0-65535 was original 16-bit range)
Usage:	
0 and 65535	(IANA Reserved)
1-64495	(public Internet)
64496-64511	(documentation – RFC5398)
64512-65534	(private use only)
23456	(represent 32-bit range in 16-bit world)
65536-65551	(documentation – RFC5398)
65552-131071	(IANA Reserved)
131072-458751	(public Internet)
458752-4199999999	(IANA Reserved/Unallocated)
4200000000-4294967294	(private use only – RFC6996)
4294967295	(IANA Reserved – RFC7300)

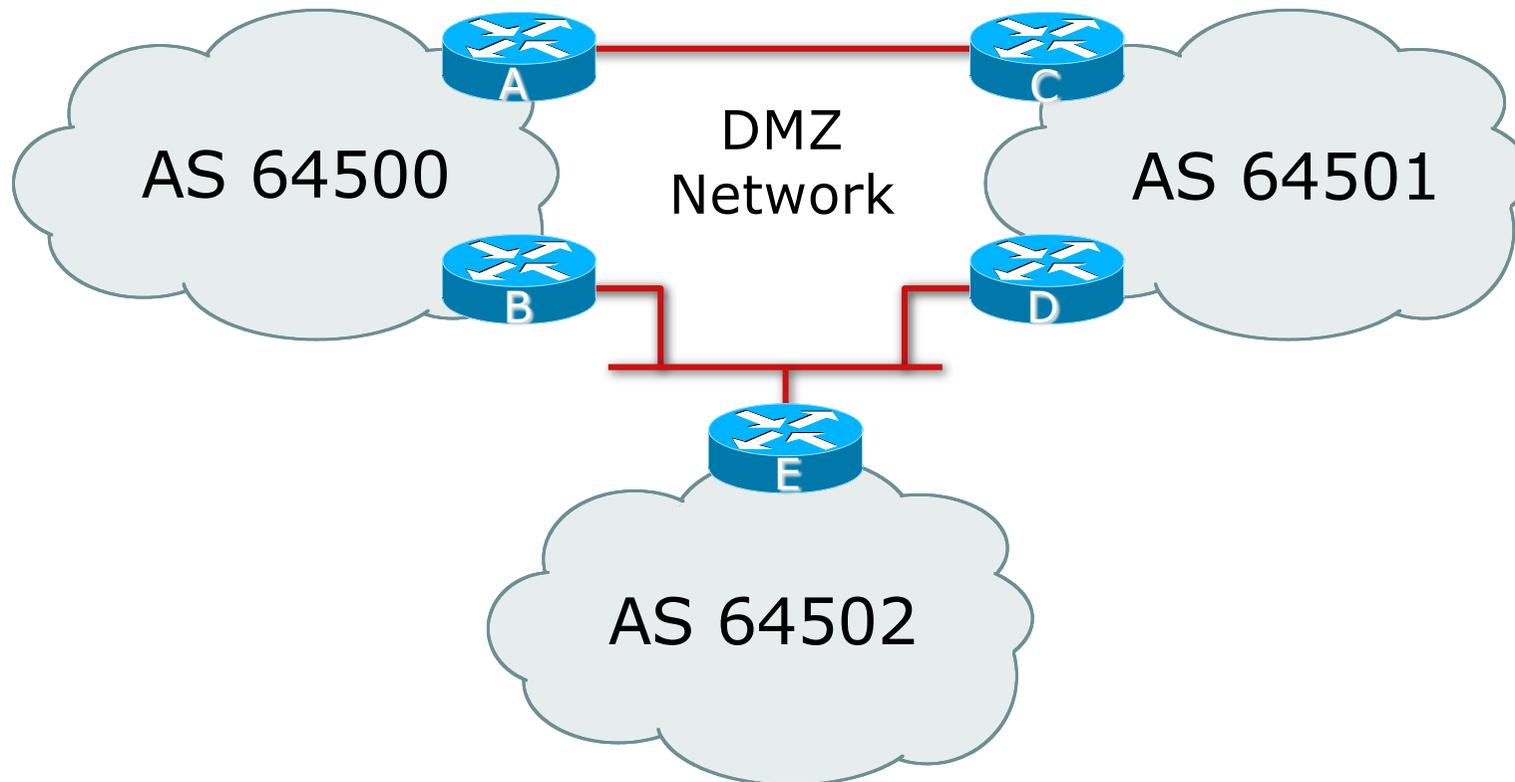
- 32-bit range representation specified in RFC5396
 - Defines “asplain” (traditional format) as standard notation

BGP Basics



- ❑ Runs over TCP – port 179
- ❑ Path vector protocol
- ❑ Incremental updates
- ❑ “Internal” & “External” BGP
- ❑ Supports IPv4 & IPv6

Demarcation Zone (DMZ)



- DMZ is the link or network shared between ASes

BGP General Operation

- ❑ Learns multiple paths via internal and external BGP speakers
- ❑ Picks the best path and installs it in the routing table (RIB)
- ❑ Best path is sent to external BGP neighbours
- ❑ Policies are applied by influencing the best path selection

Supporting Multiple Protocols

□ RFC4760

- Defines Multi-protocol Extensions for BGP4
- Enables BGP to carry routing information of protocols other than IPv4
 - e.g. MPLS, IPv6, Multicast etc
- Exchange of multiprotocol NLRI must be negotiated at session startup

□ RFC2545

- Use of BGP Multiprotocol Extensions for IPv6 Inter-Domain Routing
- Address family for IPv6

Supporting Multiple Protocols

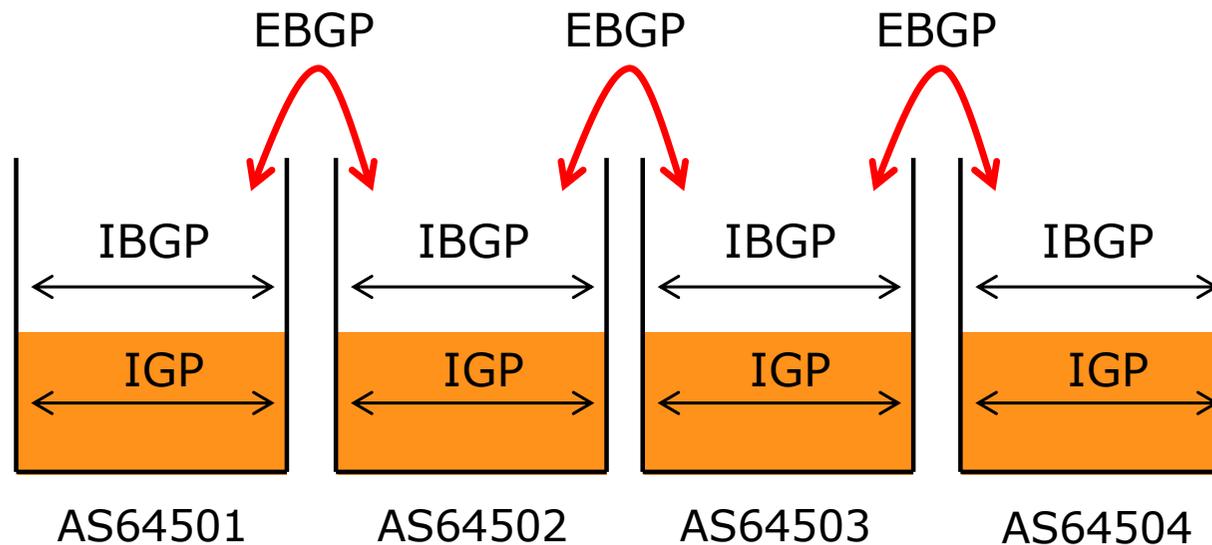
- Independent operation
 - One RIB per protocol
 - IPv6 routes in BGP's IPv6 RIB
 - IPv4 routes in BGP's IPv4 RIB
 - Each protocol can have its own policies
- NEXTHOP
 - The IP address of the next router must belong to the same address family as that of the local router

EBGP & IBGP

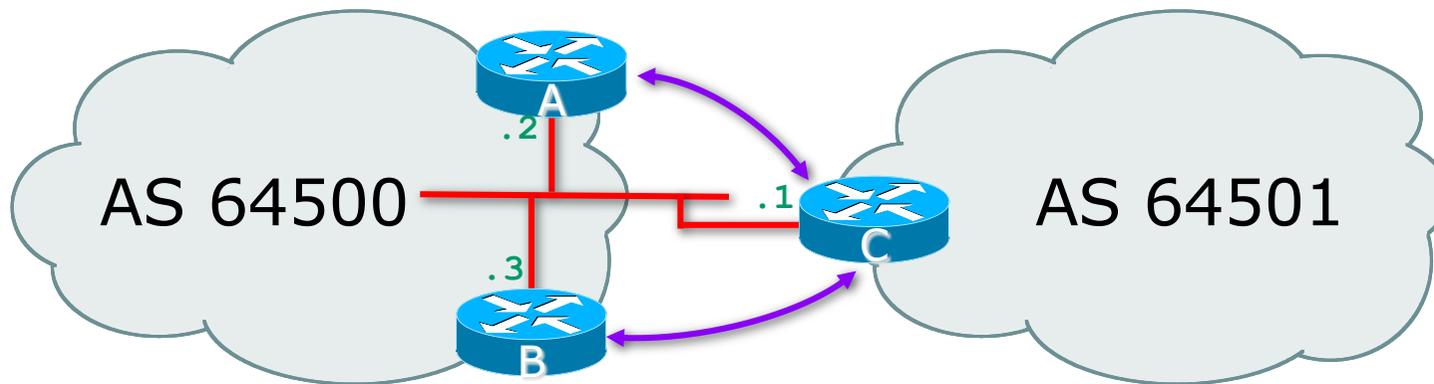
- BGP is used
 - Internally (IBGP)
 - Externally (EBGP)
- IBGP used to carry
 - Some/all Internet prefixes across Service Provider backbone
 - Service Provider's customer prefixes
- EBGP used to
 - Exchange prefixes with other ASes
 - Implement routing policy

BGP/IGP model used in Service Provider networks

□ Model representation



External BGP Peering (EBGP)

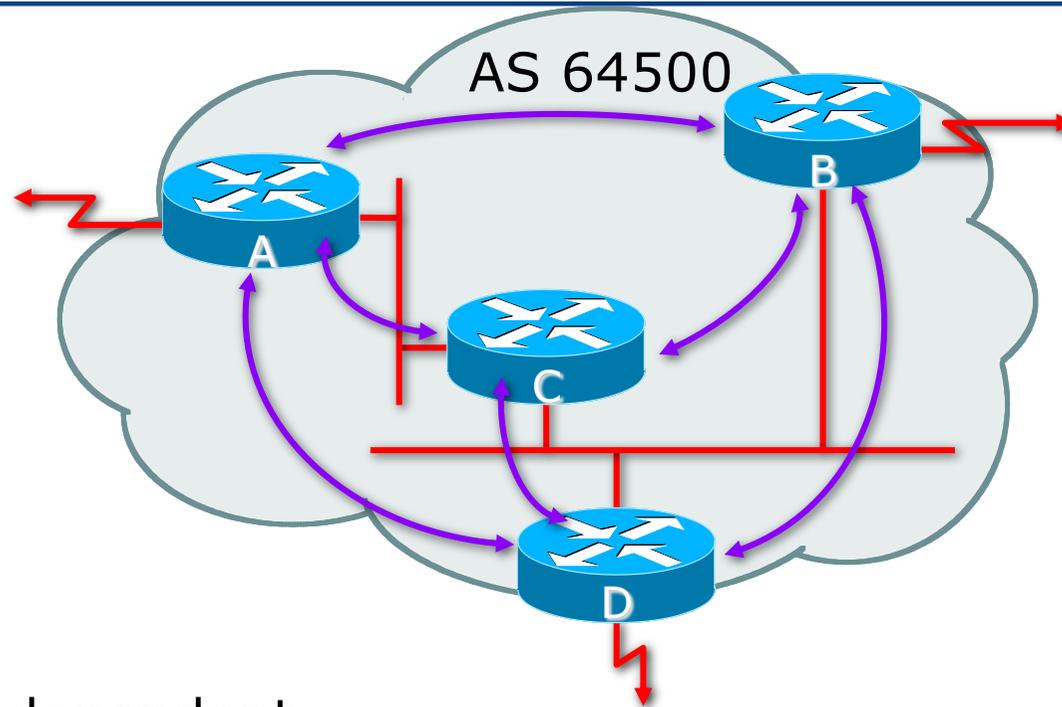


- ❑ Between BGP speakers in different AS
- ❑ Should be directly connected
- ❑ **Never** run an IGP between EBGP peers

Internal BGP (IBGP)

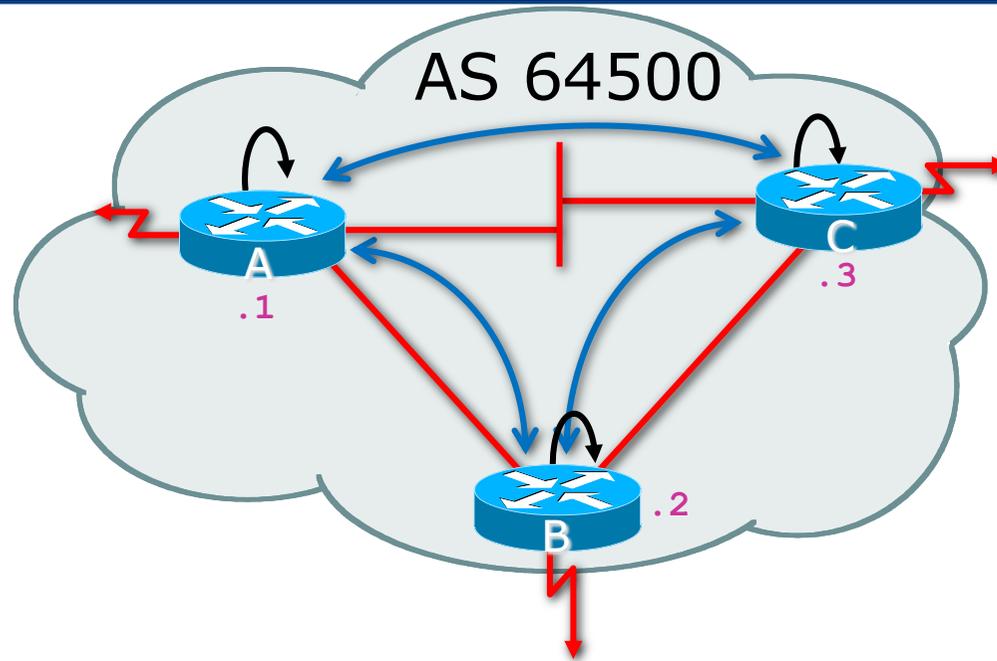
- BGP peer within the same AS
- Not required to be directly connected
 - IGP takes care of inter-BGP speaker connectivity
- IBGP speakers must be fully meshed:
 - They originate connected networks
 - They pass on prefixes learned from outside the AS
 - **They do not pass on prefixes learned from other IBGP speakers**

Internal BGP Peering (IBGP)



- ❑ Topology independent
- ❑ Each IBGP speaker must peer with every other IBGP speaker in the AS as per \longleftrightarrow

Peering between Loopback Interfaces



- ❑ Peer with loop-back interface
 - Loop-back interface does not go down – ever!
- ❑ Do not want IBGP session to depend on state of a single interface or the physical topology

Summary

BGP neighbour status (Cisco IOS IPv4)

```
Router6>show ip bgp summary
BGP router identifier 10.0.15.246, local AS number 10
BGP table version is 16, main routing table version 16
7 network entries using 819 bytes of memory
14 path entries using 728 bytes of memory
2/1 BGP path/bestpath attribute entries using 248 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
BGP using 1795 total bytes of memory
BGP activity 7/0 prefixes, 14/0 paths, scan interval 60 secs
```

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
10.0.15.241	4	10	9	8	16	0	0	00:04:47	2
10.0.15.242	4	10	6	5	16	0	0	00:01:43	2
10.0.15.243	4	10	9	8	16	0	0	00:04:49	2
...									

BGP Version

Updates sent
and received

Updates waiting

Summary

BGP neighbour status (Cisco IOS IPv6)

```
Router1>sh bgp ipv6 unicast summary
BGP router identifier 10.10.15.224, local AS number 10
BGP table version is 28, main routing table version 28
18 network entries using 2880 bytes of memory
38 path entries using 3040 bytes of memory
9/6 BGP path/bestpath attribute entries using 1152 bytes of memory
4 BGP AS-PATH entries using 96 bytes of memory
0 BGP route-map cache entries using 0 bytes of memory
0 BGP filter-list cache entries using 0 bytes of memory
BGP using 7168 total bytes of memory
BGP activity 37/1 prefixes, 95/19 paths, scan interval 60 secs
```

Neighbor	V	AS	MsgRcvd	MsgSent	TblVer	InQ	OutQ	Up/Down	State/PfxRcd
2001:DB8::2	4	10	185	182	28	0	0	02:36:11	16
2001:DB8::3	4	10	180	181	28	0	0	02:36:08	11
2001:DB8:0:4::1	4	40	153	152	28	0	0	02:05:39	9



Neighbour Information



BGP Messages Activity

Summary

BGP neighbour status (JunOS)

```
philip@R6> show bgp summary
Groups: 1 Peers: 14 Down peers: 0
Table          Tot Paths  Act Paths  Suppressed    History  Damp State    Pending
inet.0         20         20         0             0        0     0         0
inet6.0        20         20         0             0        0     0         0

Peer          AS        InPkt      OutPkt      OutQ      Flaps  Last Up/Dwn  State|#Active/Received/Accepted/Damped..
10.0.15.241   10        1067980    202487      0         0 9w1d 4:32:05  Establ  inet.0: 10/10/10/0
10.0.15.242   10        204577     1001705     0         0 9w1d 4:32:09  Establ  inet.0: 3/3/3/0
10.0.15.243   10        277630     1886656     0         0 9w1d 4:32:06  Establ  inet.0: 4/4/4/0
...
2001:DB8::1   10        416832     202568      0         0 9w1d 4:30:46  Establ  inet6.0: 10/10/10/0
2001:DB8::2   10        204605     411166      0         0 9w1d 4:34:47  Establ  inet6.0: 3/3/3/0
2001:DB8::3   10        277568     729073      0         0 9w1d 1:03:31  Establ  inet6.0: 2/2/2/0
...
```

AS Number

Updates sent
and received

Updates waiting

Address Family

Summary

BGP Table (Cisco IOS IPv4)

```
Router6>sh ip bgp
BGP table version is 18, local router ID is 10.0.15.246
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

	Network	Next Hop	Metric	LocPrf	Weight	Path
*>i	10.0.0.0/26	10.0.15.241	0	100	0	i
*>i	10.0.0.64/26	10.0.15.242	0	100	0	i
*>i	10.0.0.128/26	10.0.15.243	0	100	0	i
*>i	10.0.0.192/26	10.0.15.244	0	100	0	i
*>i	10.0.1.0/26	10.0.15.245	0	100	0	i
*>	10.0.1.64/26	0.0.0.0	0		32768	i
*>i	10.0.1.128/26	10.0.15.247	0	100	0	i
*>i	10.0.1.192/26	10.0.15.248	0	100	0	i
*>i	10.0.2.0/26	10.0.15.249	0	100	0	i
*>i	10.0.2.64/26	10.0.15.250	0	100	0	i
*>i	10.0.2.128/26	10.0.15.251	0	100	0	i
*>i	10.0.2.192/26	10.0.15.252	0	100	0	i
*>i	10.0.3.0/26	10.0.15.253	0	100	0	i
*>i	10.0.3.64/26	10.0.15.254	0	100	0	i

Summary

BGP Table (Cisco IOS IPv6)

```
Router6>sh bgp ipv6 unicast
BGP table version is 18, local router ID is 10.0.15.246
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
               r RIB-failure, S Stale, m multipath, b backup-path, f RT-Filter,
               x best-external, a additional-path, c RIB-compressed,
Origin codes: i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found
```

Network	Next Hop	Metric	LocPrf	Weight	Path
*>i 2001:DB8:1::/48	2001:DB8::1	0	100	0	i
*>i 2001:DB8:2::/48	2001:DB8::2	0	100	0	i
*>i 2001:DB8:3::/48	2001:DB8::3	0	100	0	i
*>i 2001:DB8:4::/48	2001:DB8::4	0	100	0	i
*>i 2001:DB8:5::/48	2001:DB8::5	0	100	0	i
*> 2001:DB8:6::/48	::	0		32768	i
*>i 2001:DB8:7::/48	2001:DB8::7	0	100	0	i
*>i 2001:DB8:8::/48	2001:DB8::8	0	100	0	i
*>i 2001:DB8:9::/48	2001:DB8::9	0	100	0	i
*>i 2001:DB8:A::/48	2001:DB8::A	0	100	0	i
*>i 2001:DB8:B::/48	2001:DB8::B	0	100	0	i
*>i 2001:DB8:C::/48	2001:DB8::C	0	100	0	i
*>i 2001:DB8:D::/48	2001:DB8::D	0	100	0	i
*>i 2001:DB8:E::/48	2001:DB8::E	0	100	0	i

Summary

BGP Table (JunOS)

```
philip@R6> show route protocol bgp terse
```

```
inet.0: 14 destinations, 14 routes (14 active, 0 holddown, 0 hidden)
```

```
+ = Active Route, - = Last Active, * = Both
```

A	V	Destination	P	Prf	Metric 1	Metric 2	Next hop	AS path
	?	10.0.0.0/26	B	100			>10.0.15.241	I
		unverified						
	?	10.0.0.64/26	B	100			>10.0.15.241	I
		unverified						
	...							
	?	10.1.0.0/24	B	100			>10.0.15.242	20 I
		unverified						
	?	10.4.0.0/24	B	100			>10.0.15.241	20 I
		unverified						

```
...
```

```
inet6.0: 14 destinations, 14 routes (14 active, 0 holddown, 0 hidden)
```

```
+ = Active Route, - = Last Active, * = Both
```

A	V	Destination	P	Prf	Metric 1	Metric 2	Next hop	AS path
	?	2001:DB8:1::/48	B	100			>fe80::82ac:acff:fed2:ea88	I
		unverified						
	?	2001:DB8:2::/48	B	100			>fe80::82ac:acff:fed2:ea88	I
		unverified						
	...							
	?	2001:DB9::/32	B	100			>fe80::224e:71ff:fe90:2500	20 I
		unverified						
	?	2001:DB9::/32	B	100			>fe80::82ac:acff:fed2:ea88	20 I
		unverified						

```
...
```

BGP Attributes



BGP's policy tool kit

What Is an Attribute?

...	Origin	AS Path	Next Hop	MED	...
-----	--------	---------	----------	-----	-----

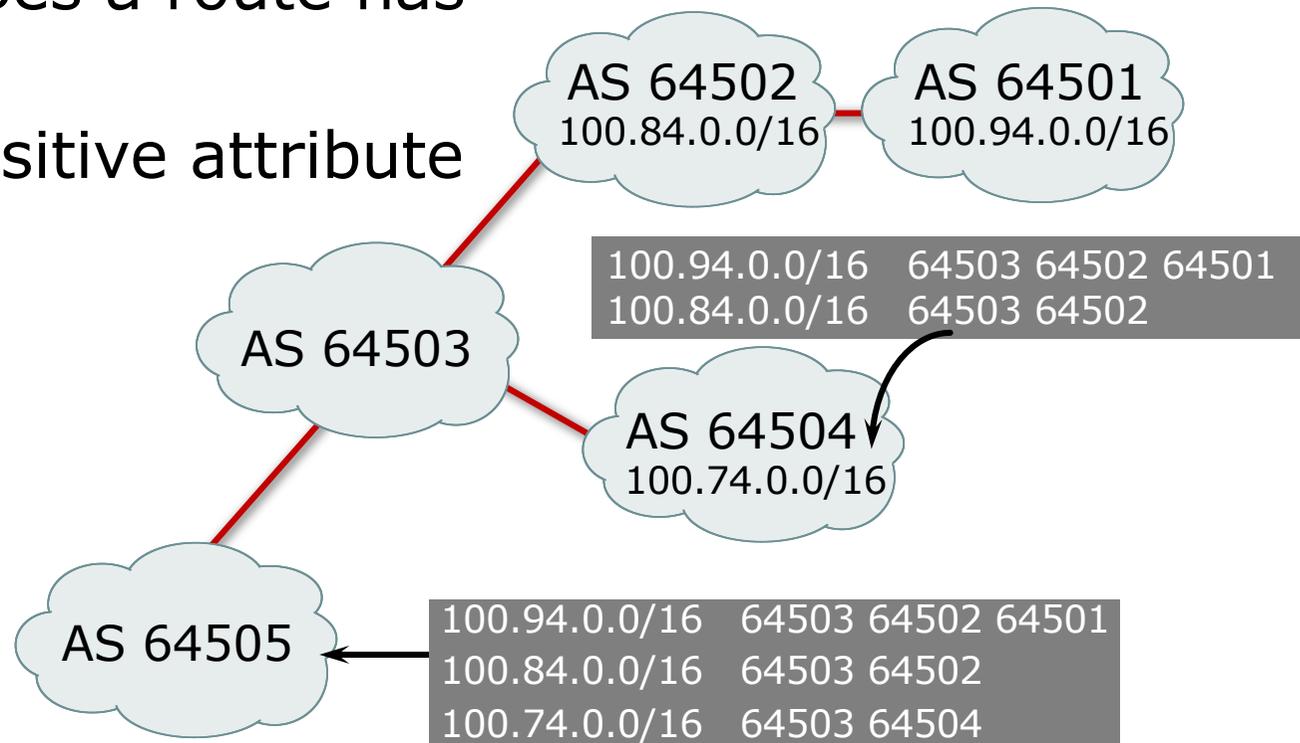
- ❑ Part of a BGP Update
- ❑ Describes the characteristics of prefix
- ❑ Can either be transitive or non-transitive
- ❑ Some are mandatory

BGP Attributes

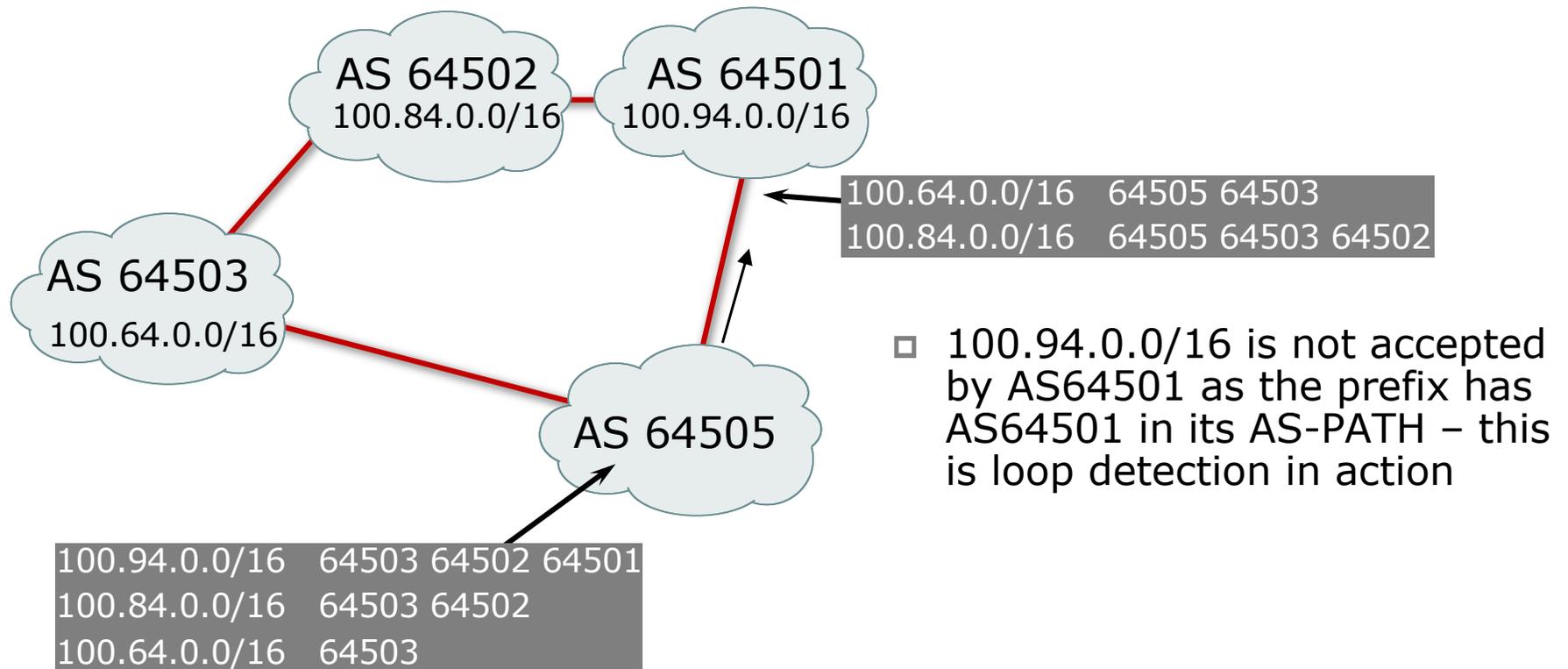
- Carry various information about or characteristics of the prefix being propagated
 - AS-PATH
 - NEXT-HOP
 - ORIGIN
 - AGGREGATOR
 - LOCAL_PREFERENCE
 - Multi-Exit Discriminator
 - (Weight)
 - COMMUNITY

AS-Path

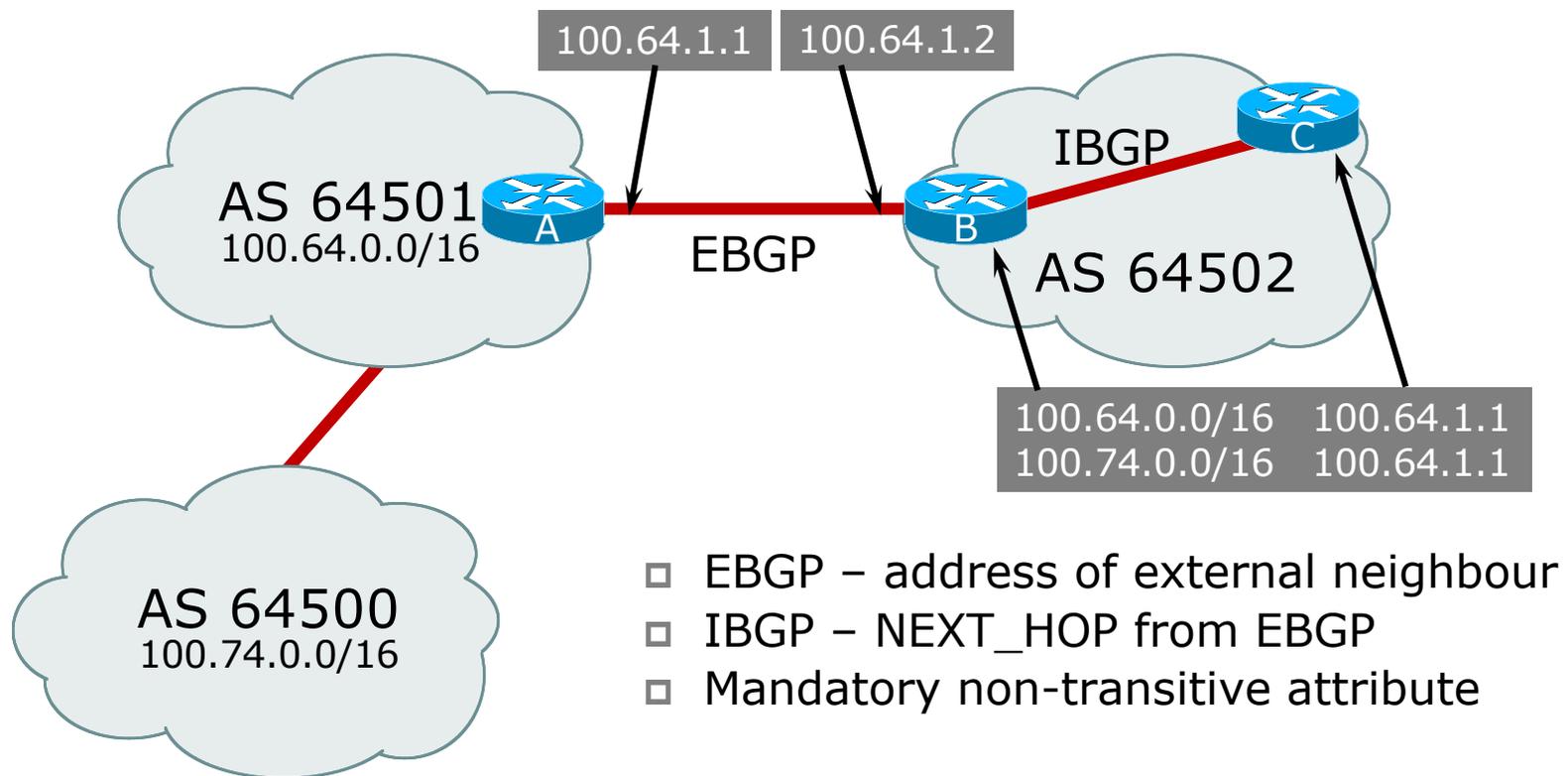
- ❑ Sequence of ASes a route has traversed
- ❑ Mandatory transitive attribute
- ❑ Used for:
 - Loop detection
 - Applying policy



AS-Path loop detection



Next Hop



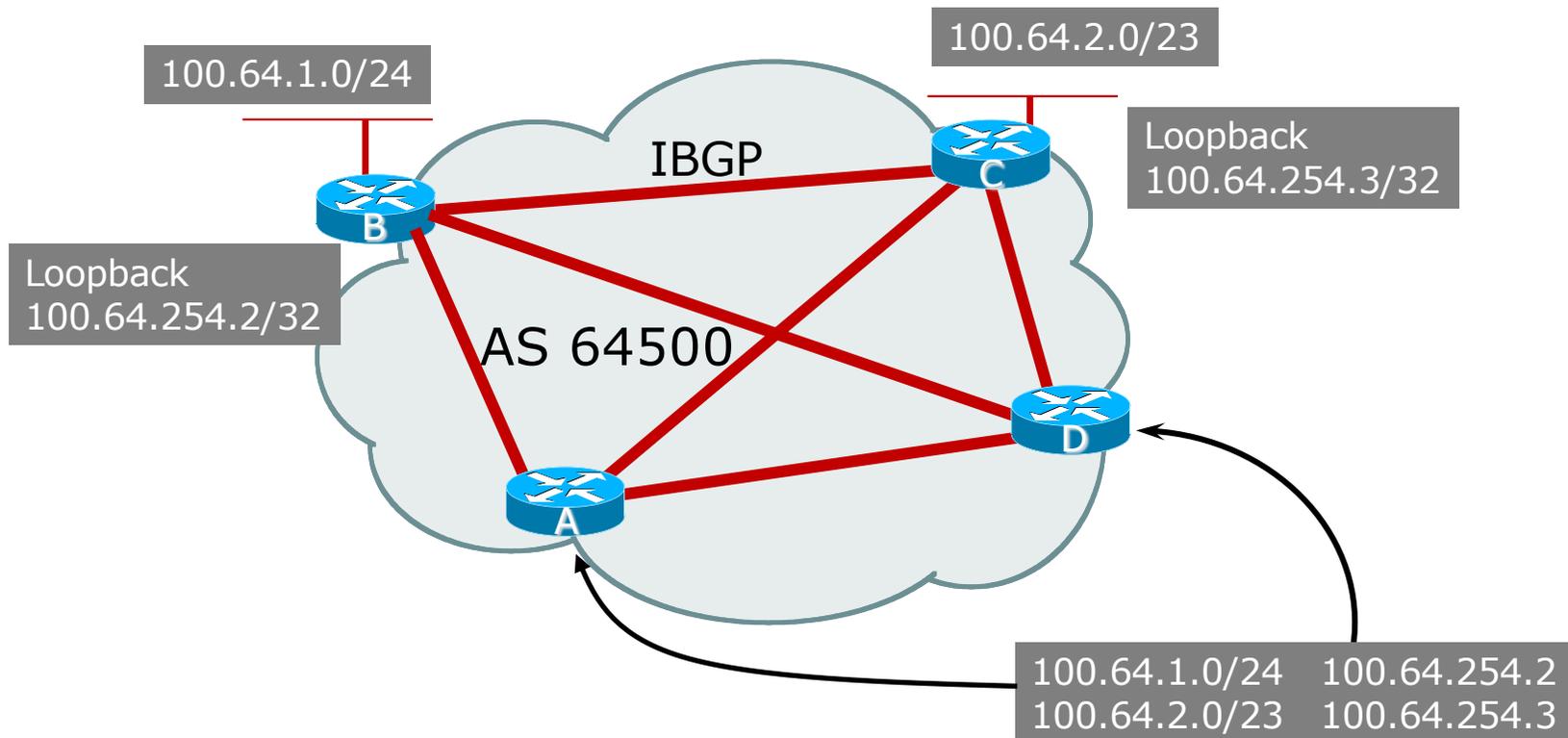
Next Hop Best Practice

- The default behaviour is for external next-hop to be propagated unchanged to IBGP peers
 - This means that IGP has to carry external next-hops
 - Forgetting means external network is invisible
 - With many EBGP peers, it is unnecessary extra load on IGP
- Network operator Best Practice is to change external next-hop to be that of the local router

- Cisco IOS: `neighbor x.x.x.x next-hop-self`

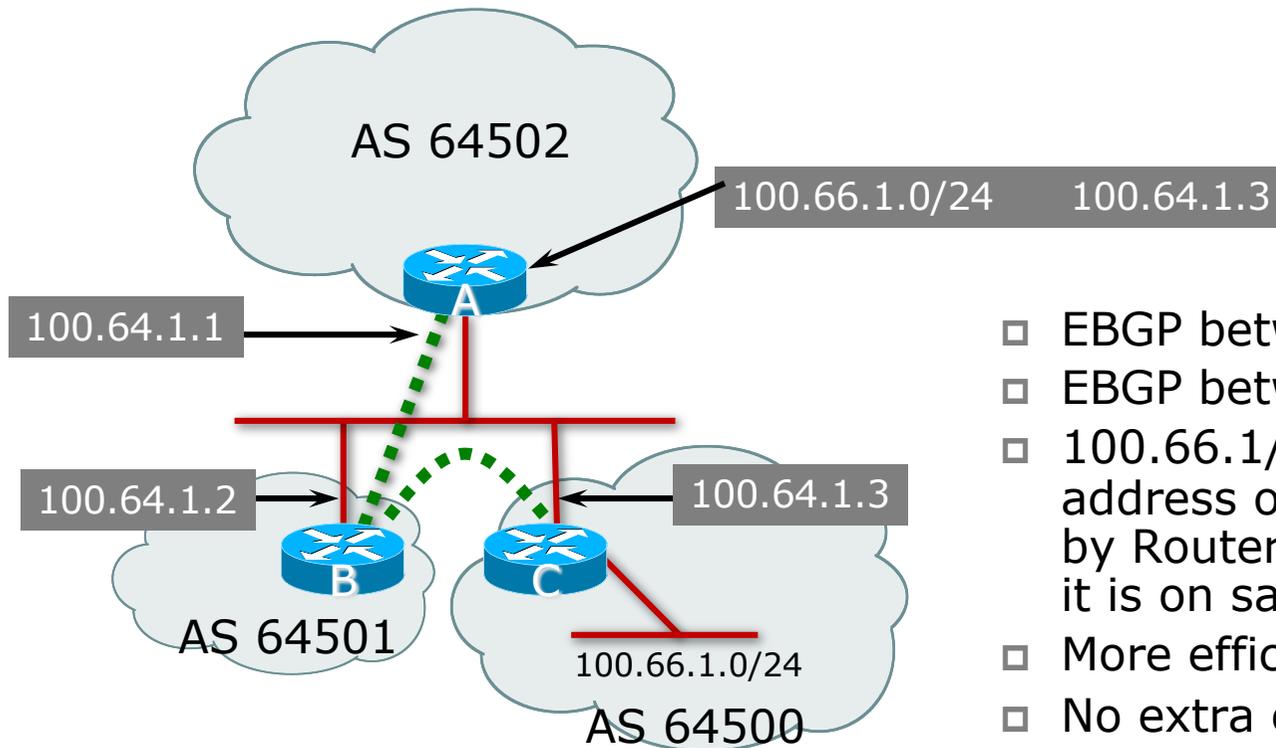
- JunOS: `set policy-options
policy-statement <name> term <name> then next-hop self`

IBGP Next Hop



- ❑ Next hop is IBGP router loopback address
- ❑ Recursive route look-up

Third Party Next Hop



- ❑ EBGp between Router A and Router B
- ❑ EBGp between Router B and Router C
- ❑ 100.66.1/24 prefix has next hop address of 100.64.1.3 – this is used by Router A instead of 100.64.1.2 as it is on same subnet as Router B
- ❑ More efficient
- ❑ No extra configuration needed

Next Hop (Summary)

- ❑ IGP should carry route to next hops
- ❑ Recursive route look-up
- ❑ Unlinks BGP from actual physical topology
- ❑ Use “next-hop-self” for external next hops
- ❑ Allows IGP to make intelligent forwarding decision

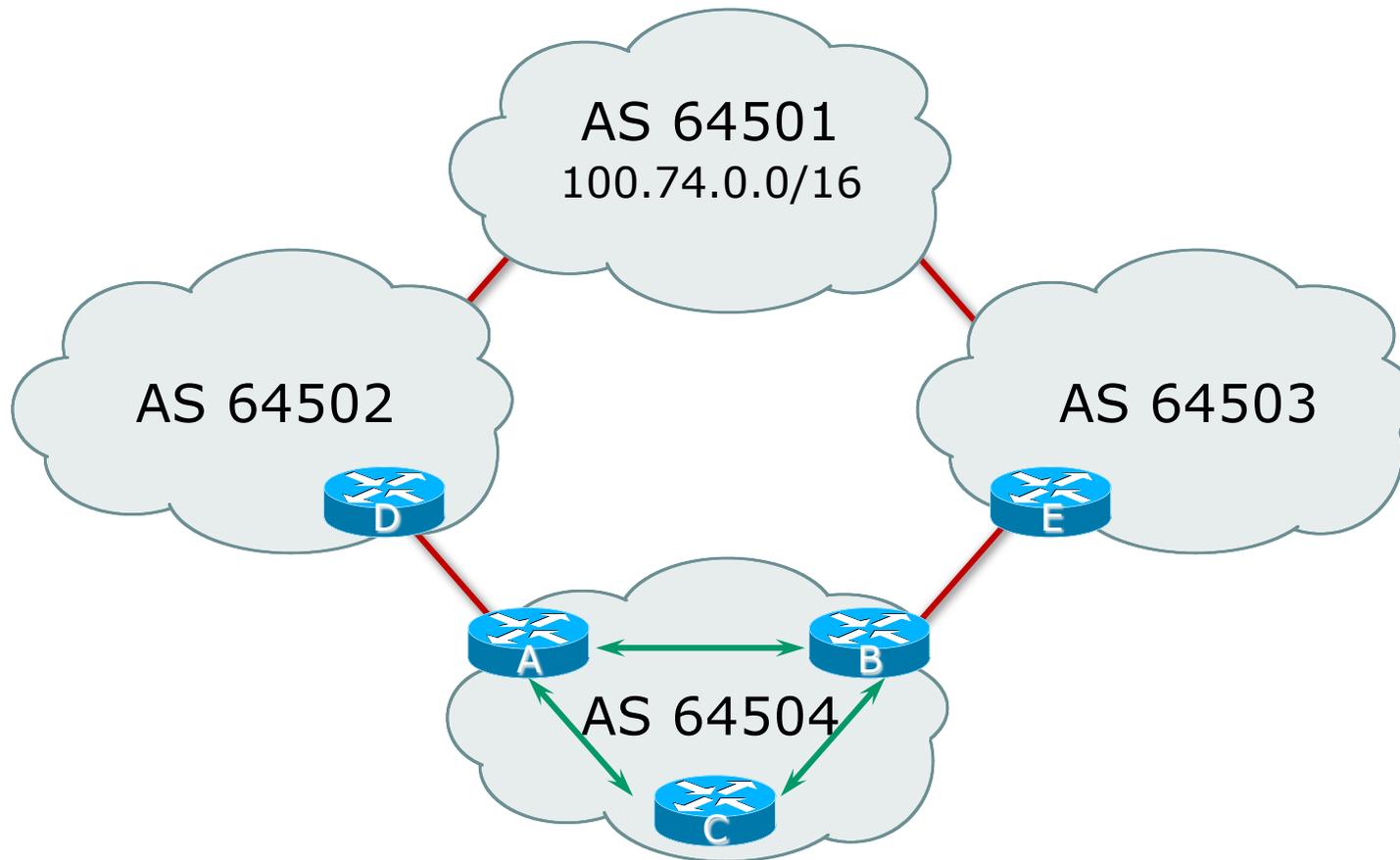
Origin

- ❑ Conveys the origin of the prefix
- ❑ **Historical** attribute
 - Used in transition from EGP to BGP
- ❑ Transitive and Mandatory Attribute
- ❑ Influences best path selection
- ❑ Three values: IGP, EGP, incomplete
 - IGP – generated by BGP network statement
 - EGP – generated by EGP
 - incomplete – redistributed from another routing protocol

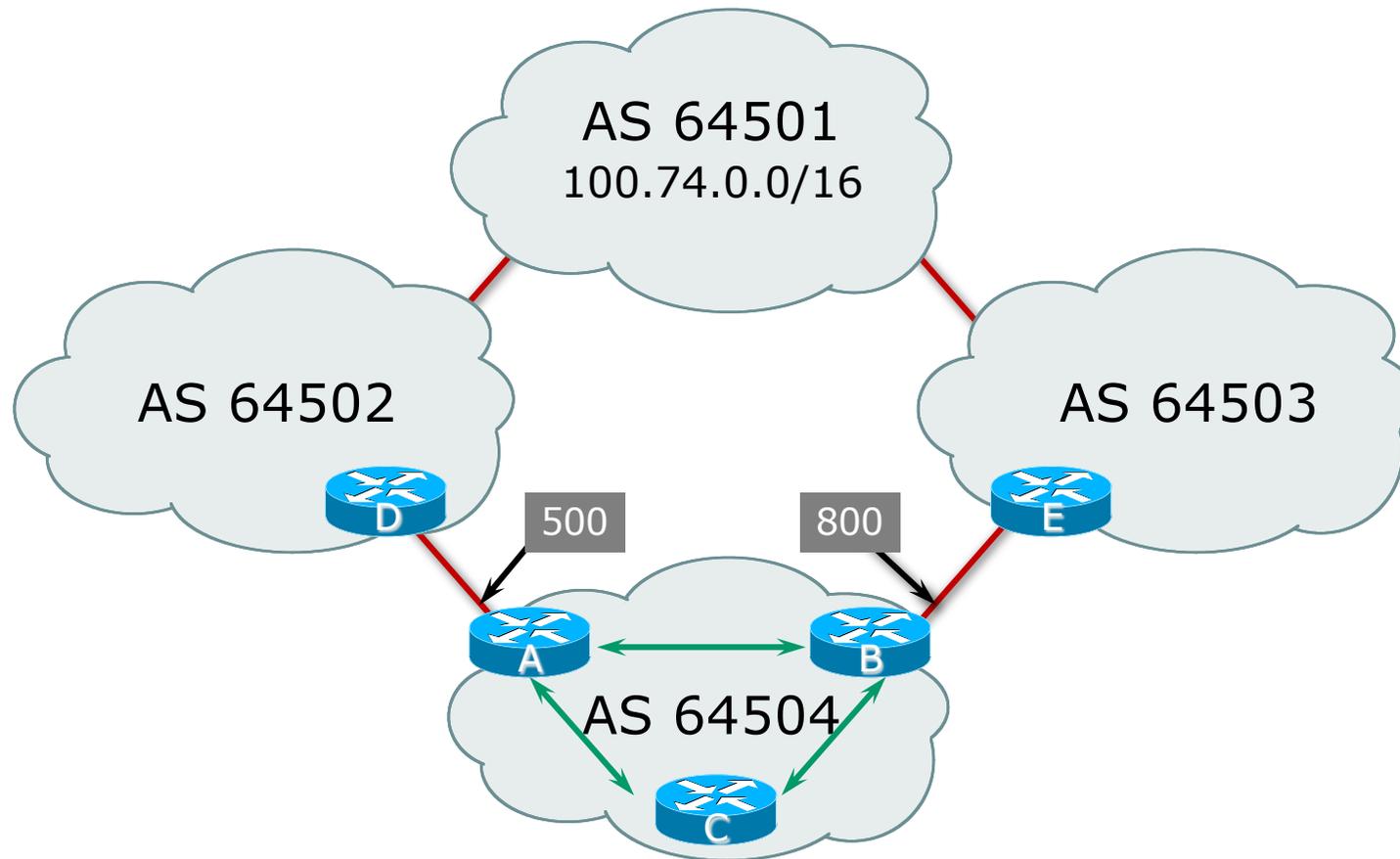
Aggregator

- ❑ Conveys the IP address of the router or BGP speaker generating the aggregate route
- ❑ Optional & transitive attribute
- ❑ Useful for debugging purposes
- ❑ Does not influence best path selection

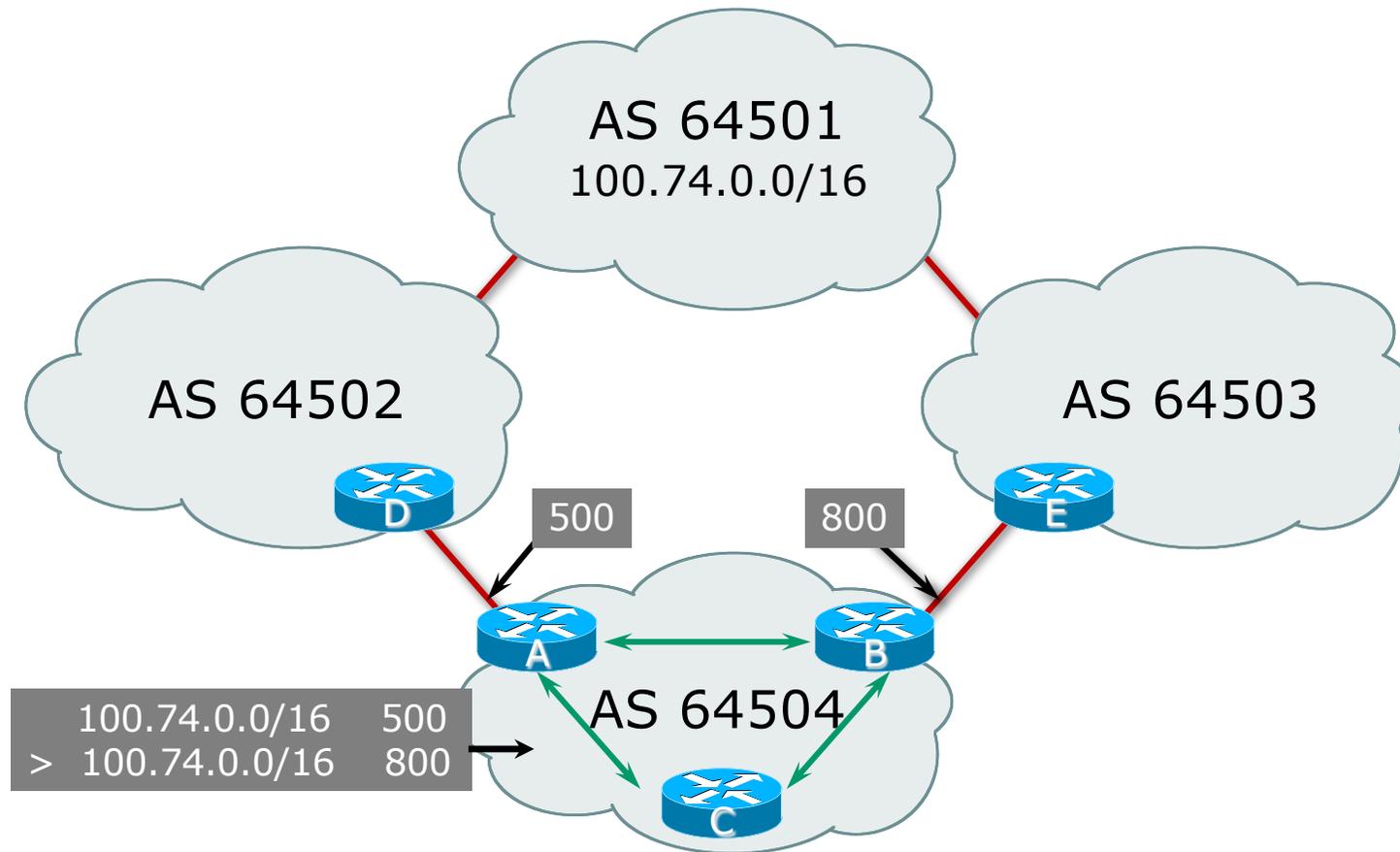
Local Preference



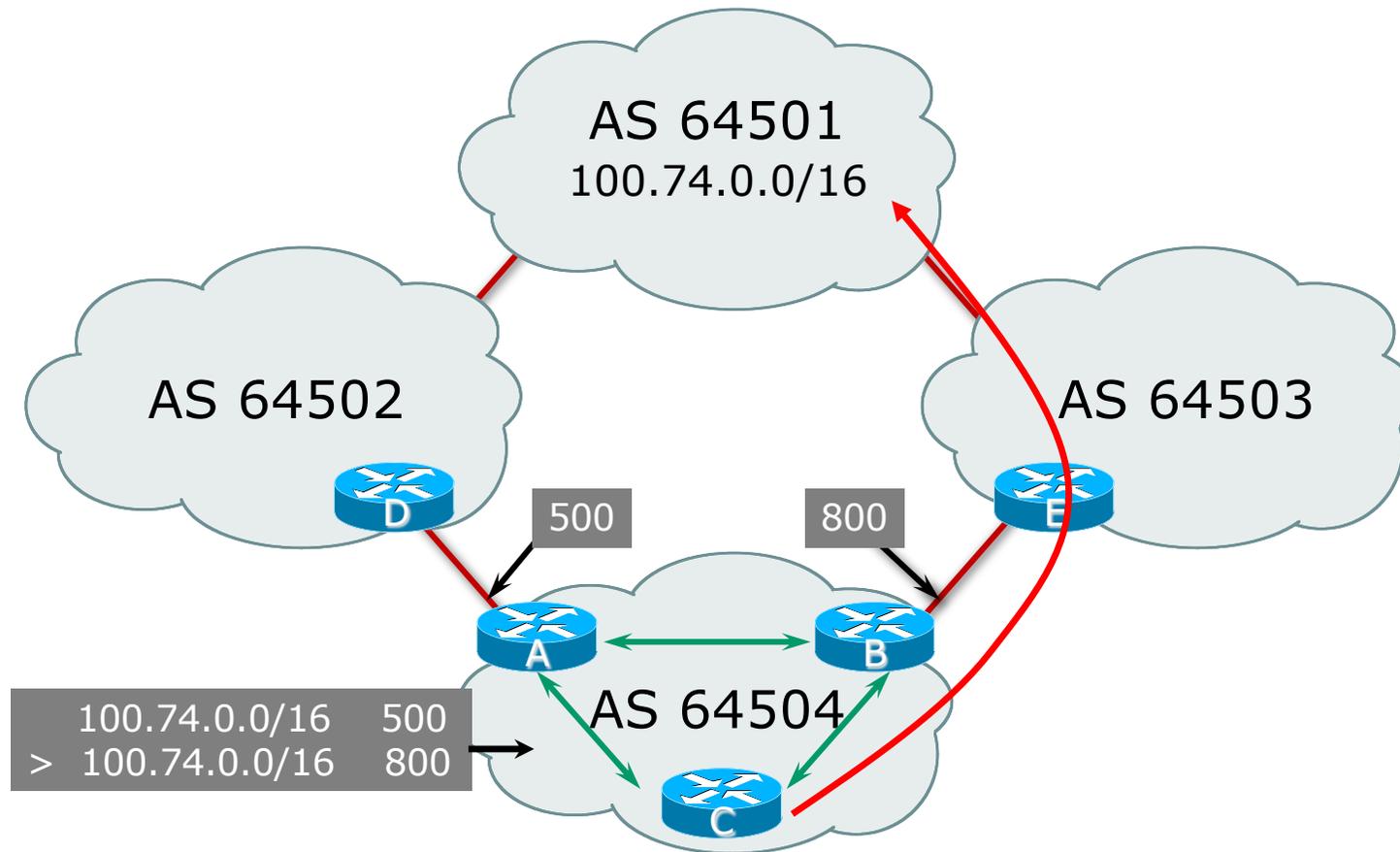
Local Preference



Local Preference



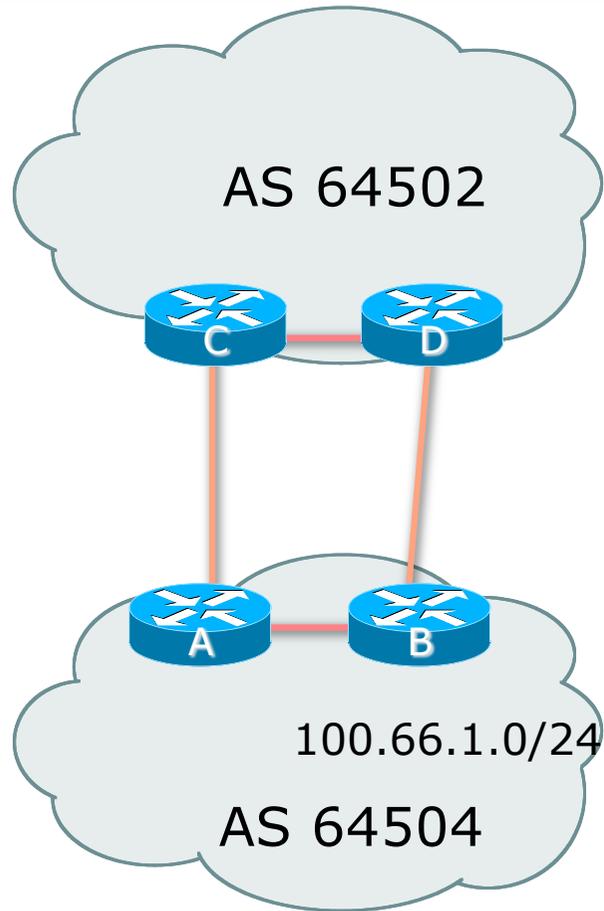
Local Preference



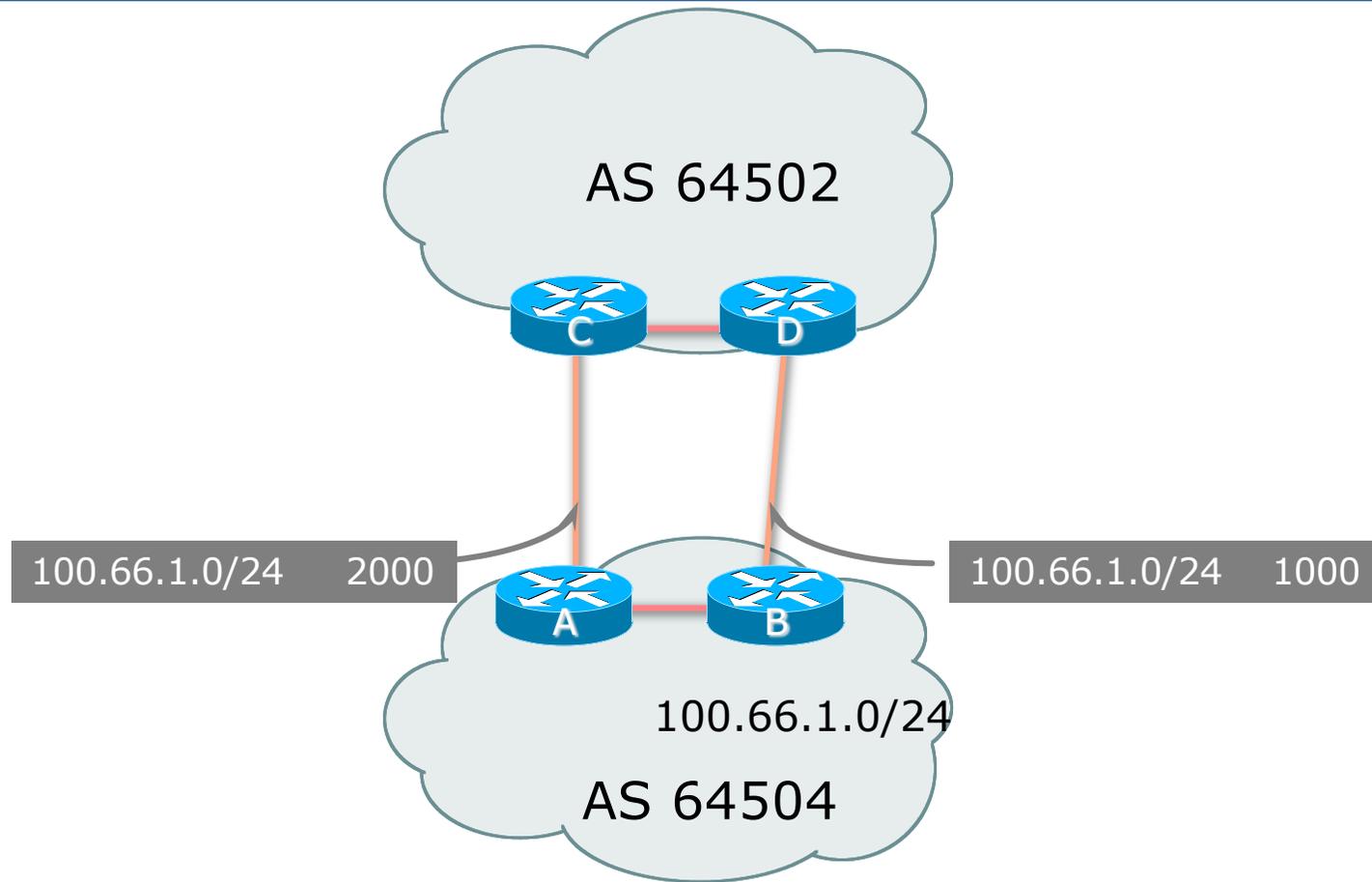
Local Preference

- ❑ Non-transitive and optional attribute
- ❑ Local to an AS only
 - Default local preference is 100 (most implementations)
- ❑ Used to influence BGP path selection
 - Determines best path for *outbound* traffic
- ❑ Path with highest local preference wins

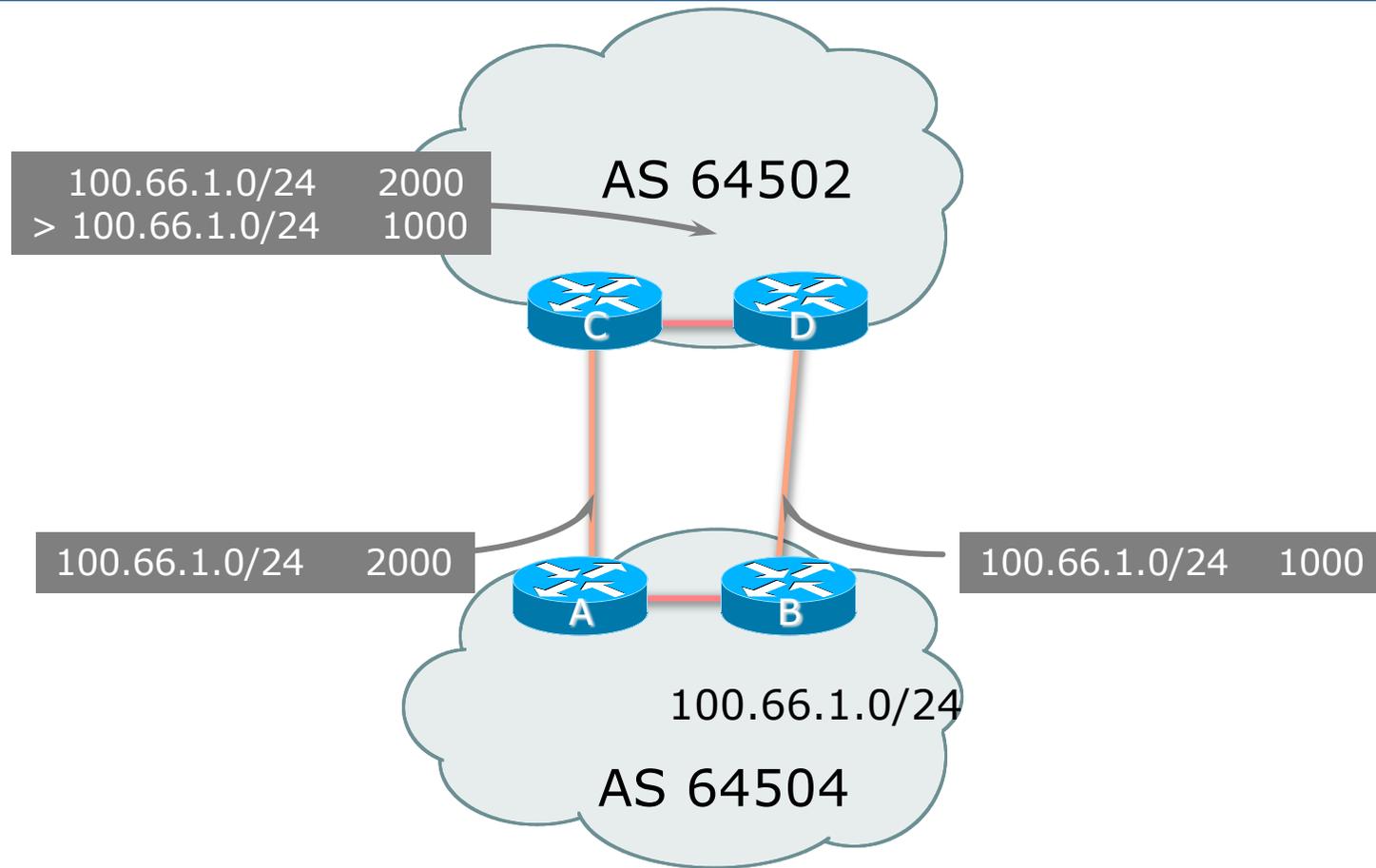
Multi-Exit Discriminator (MED)



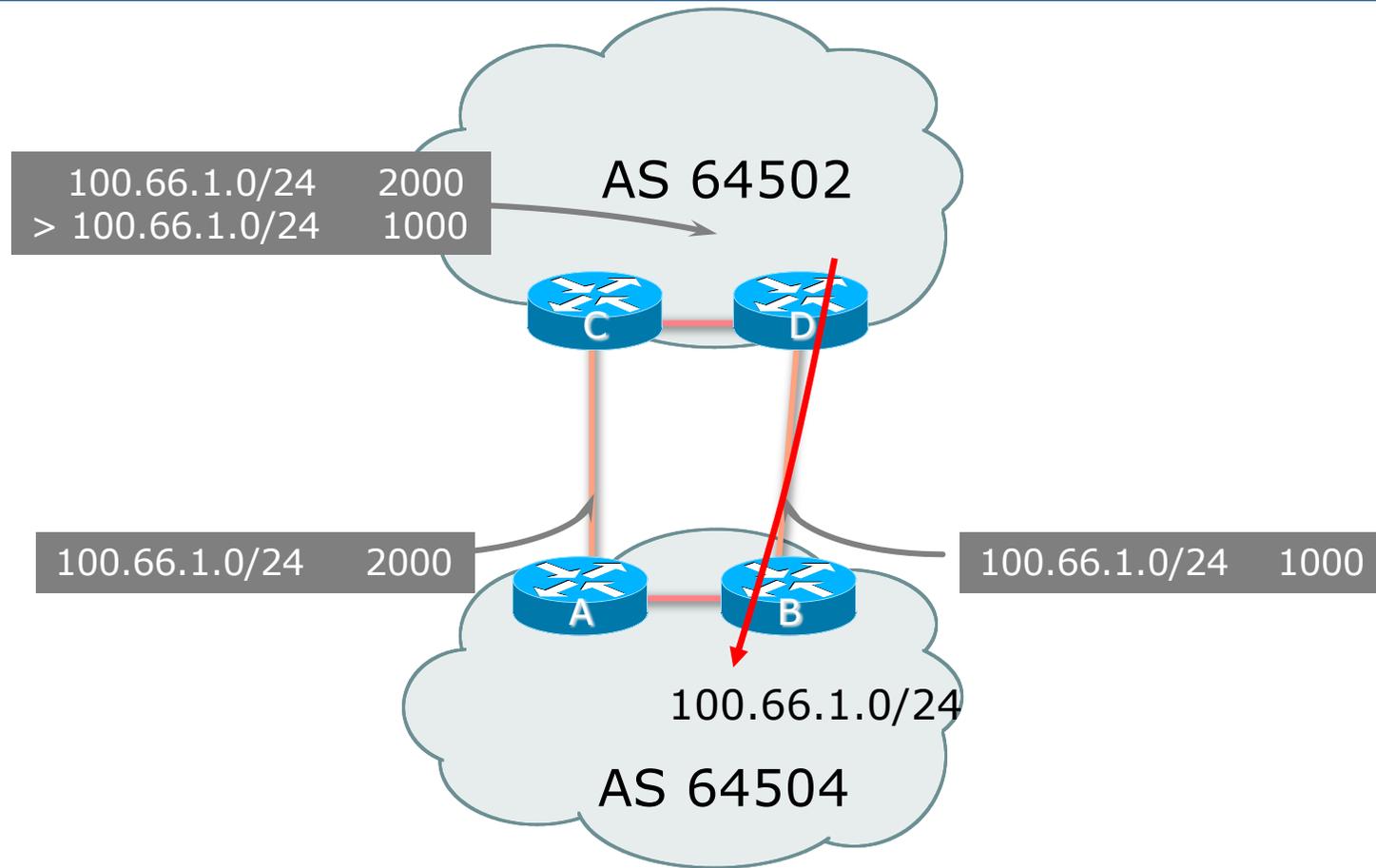
Multi-Exit Discriminator (MED)



Multi-Exit Discriminator (MED)



Multi-Exit Discriminator (MED)



Multi-Exit Discriminator

- ❑ Inter-AS – non-transitive & optional attribute
- ❑ Used to convey the relative preference of entry points
 - Determines best path for inbound traffic
- ❑ Comparable if paths are from same AS
 - Cisco's `bgp always-compare-med` allows comparisons of MEDs from different ASes
 - Also available in JunOS:

```
set protocols bgp path-selection always-compare-med
```

- ❑ Path with lowest MED wins
- ❑ Absence of MED attribute implies MED value of **zero** (RFC4271)

Deterministic MED

- ❑ Cisco IOS compares paths in the order they were received
 - Leads to inconsistent decisions when comparing MED
- ❑ Deterministic MED
 - Configure on all BGP speaking routers in AS
 - Orders paths according to their neighbouring ASN
 - Best path for each neighbour ASN group is selected
 - Overall bestpath selected from the winners of each group

```
router bgp 10
  bgp deterministic-med
```

- ❑ Deterministic MED is default in JunOS
 - Non-deterministic behaviour enabled with

```
set protocols bgp path-selection cisco-non-deterministic
```

MED & IGP Metric

□ IGP metric can be conveyed as MED

- `set metric-type internal` in route-map
 - Enables BGP to advertise a MED which corresponds to the IGP metric values
 - Changes are monitored (and re-advertised if needed) every 600s
 - Monitoring period can be changed using:

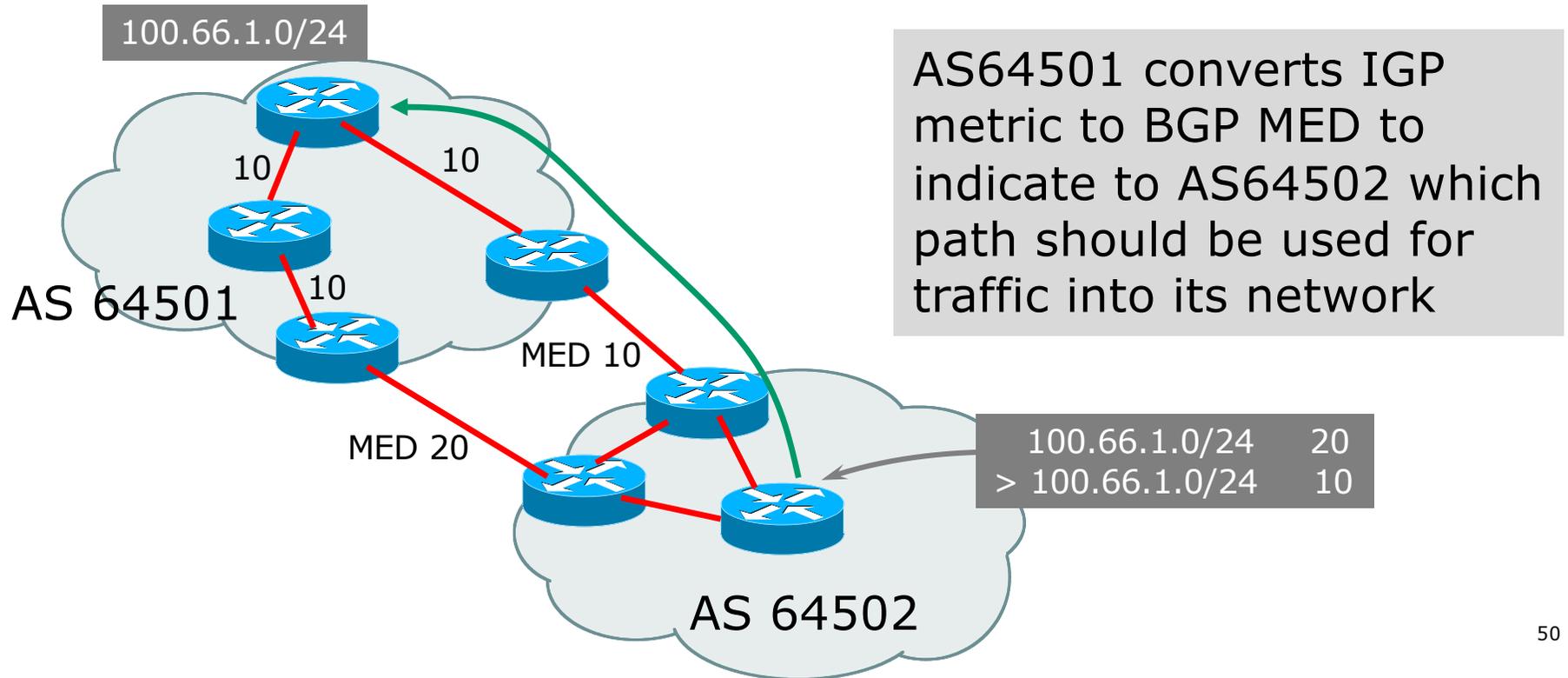
```
bgp dynamic-med-interval <secs>
```

- Also available in JunOS:

```
set protocols bgp path-selection med-plus-igp
```

MED & IGP Metric

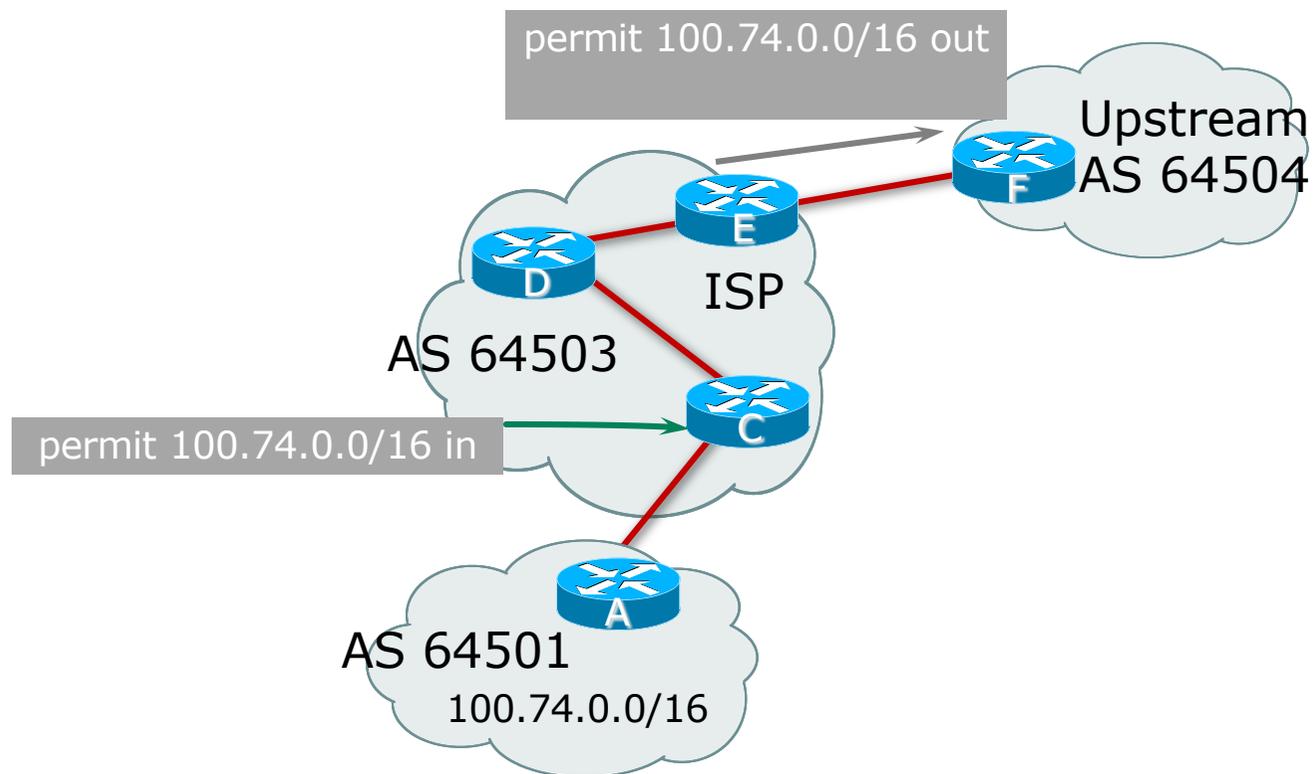
- Example: IGP metric conveyed as MED



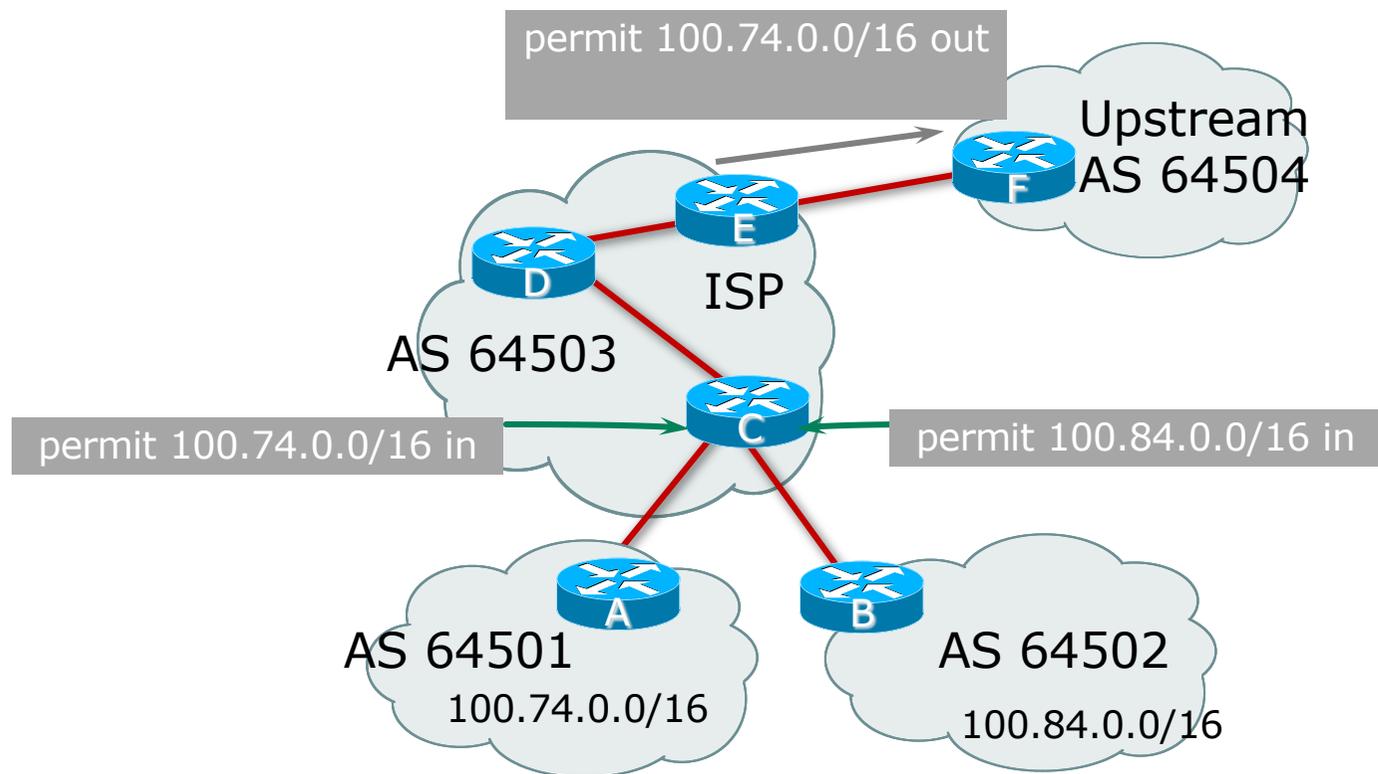
Community

- Communities are described in RFC1997
 - Transitive and Optional Attribute
- 32-bit integer
 - Represented as two 16-bit integers (RFC1998)
 - Common format is <local-ASN>:xx
 - 0:0 to 0:65535 and 65535:0 to 65535:65535 are reserved
- Used to group destinations
 - Each destination could be member of multiple communities
- Very useful in applying policies within and between ASes

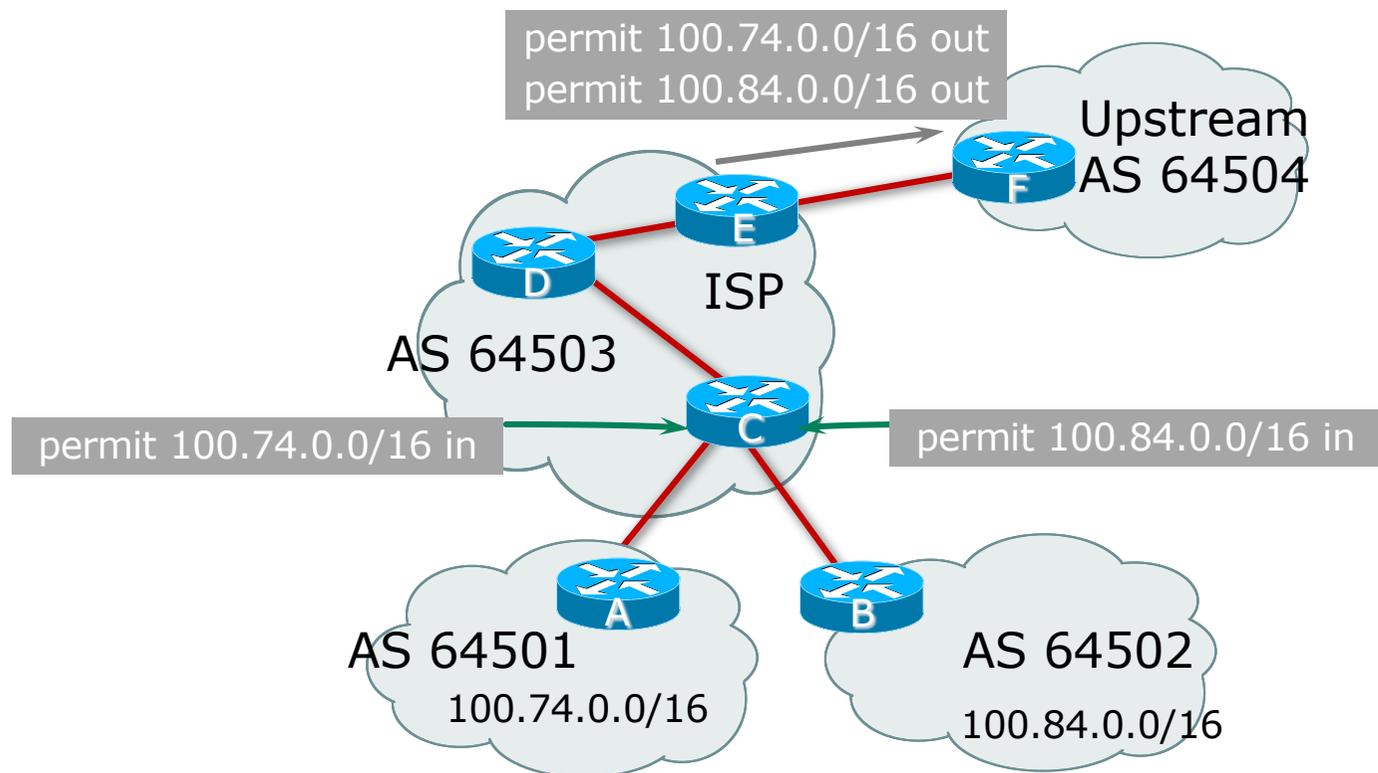
Community Example (before)



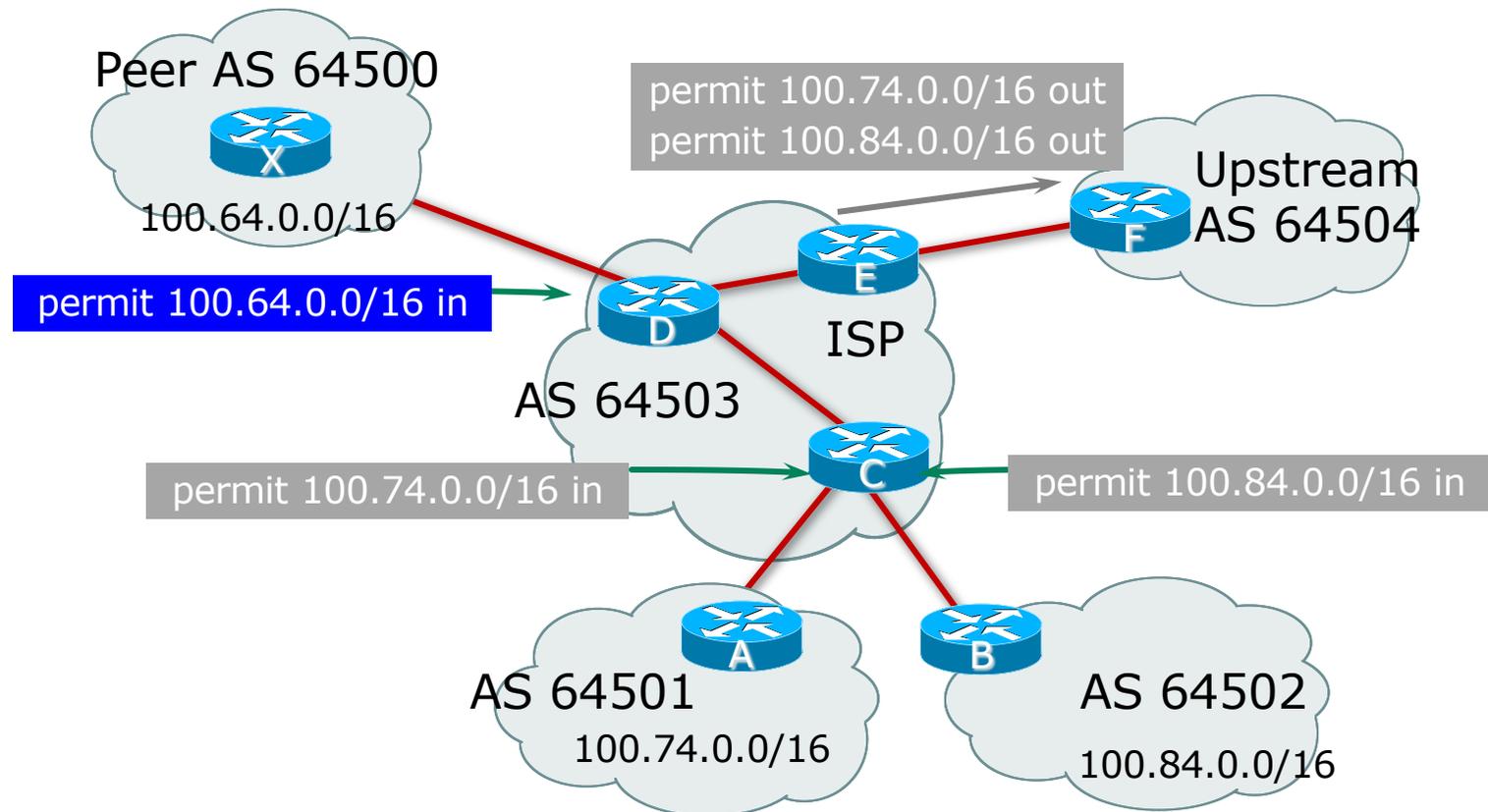
Community Example (before)



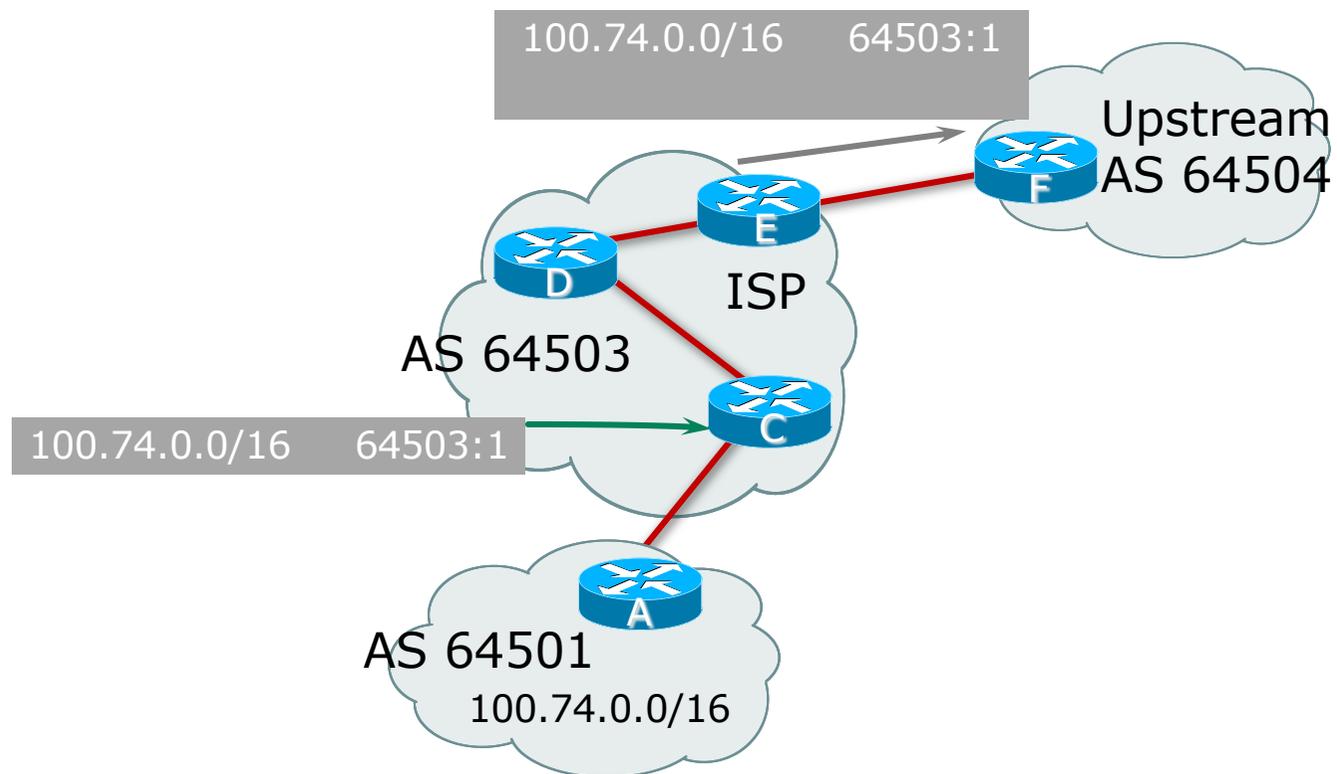
Community Example (before)



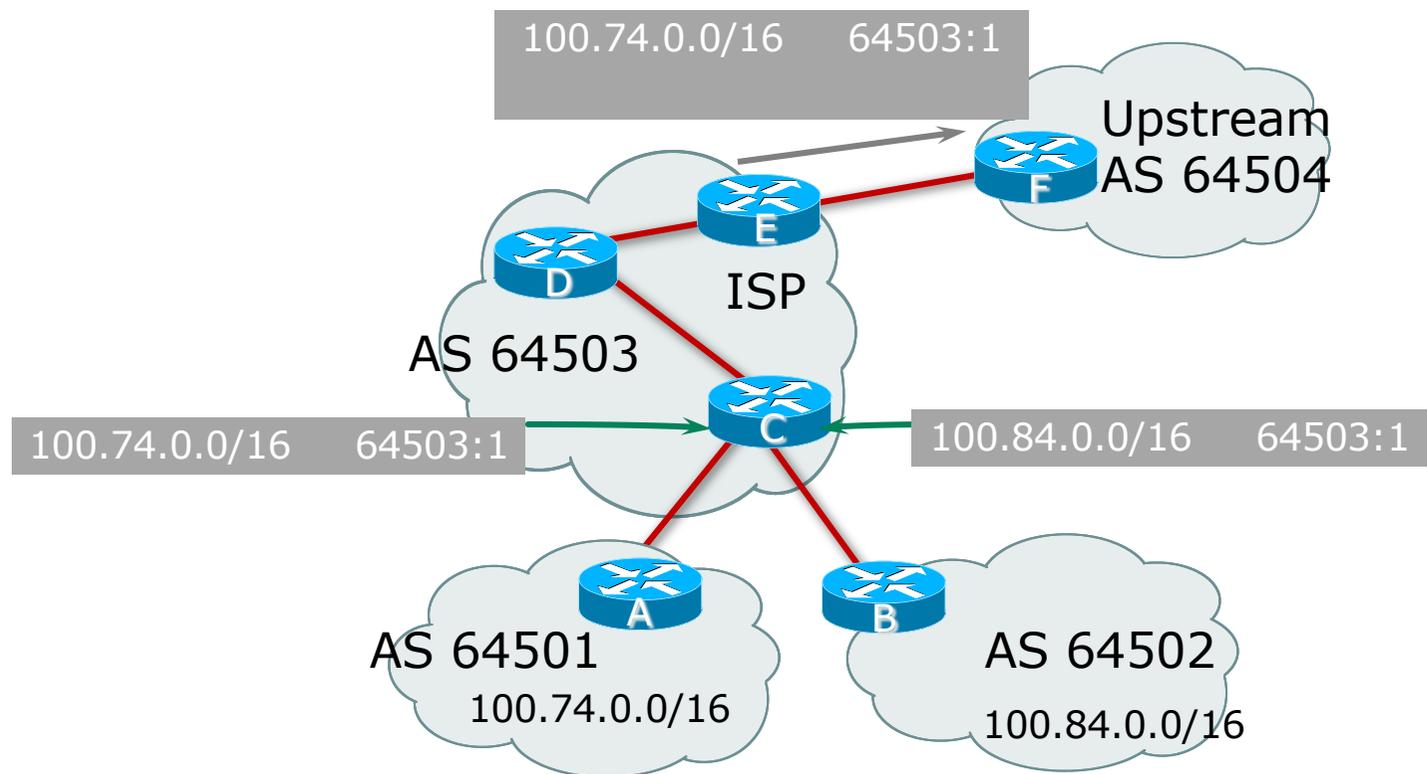
Community Example (before)



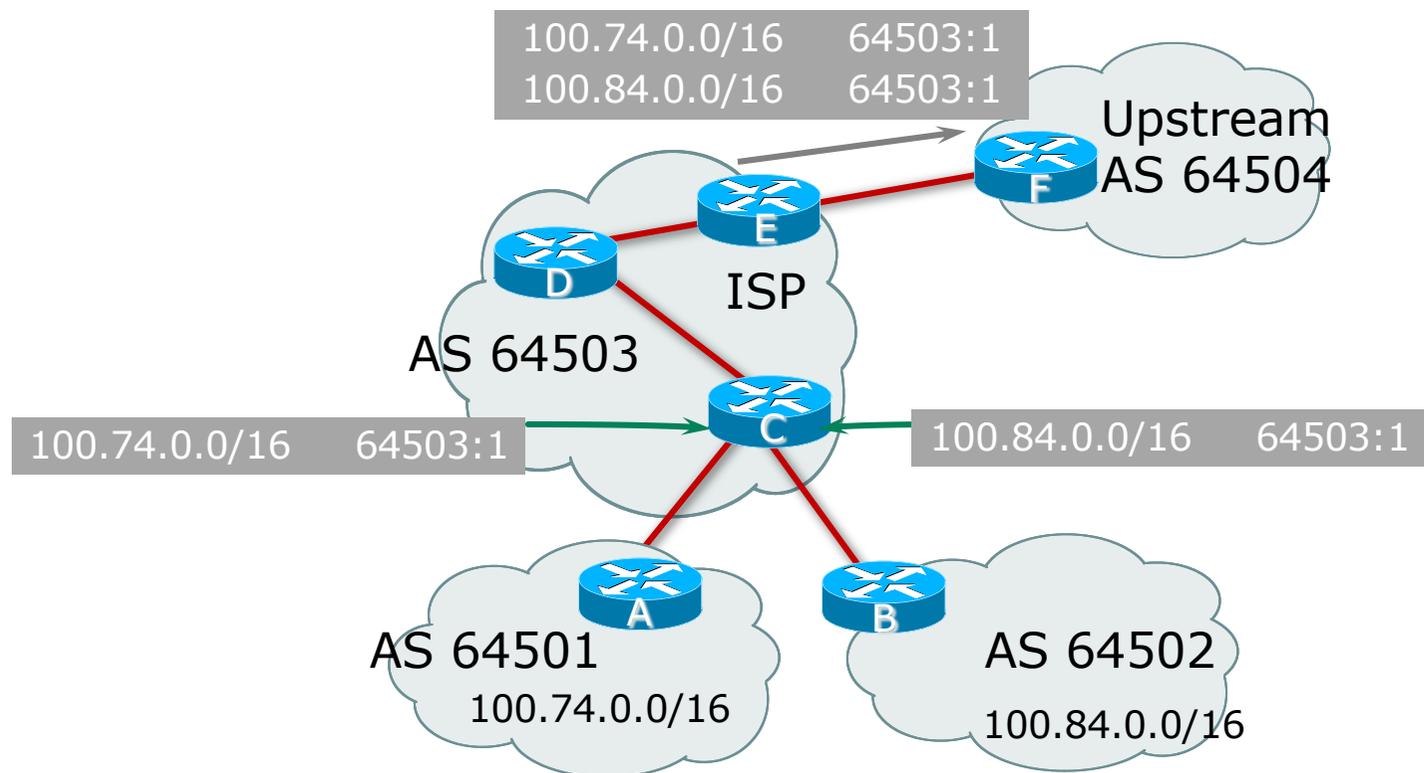
Community Example (after)



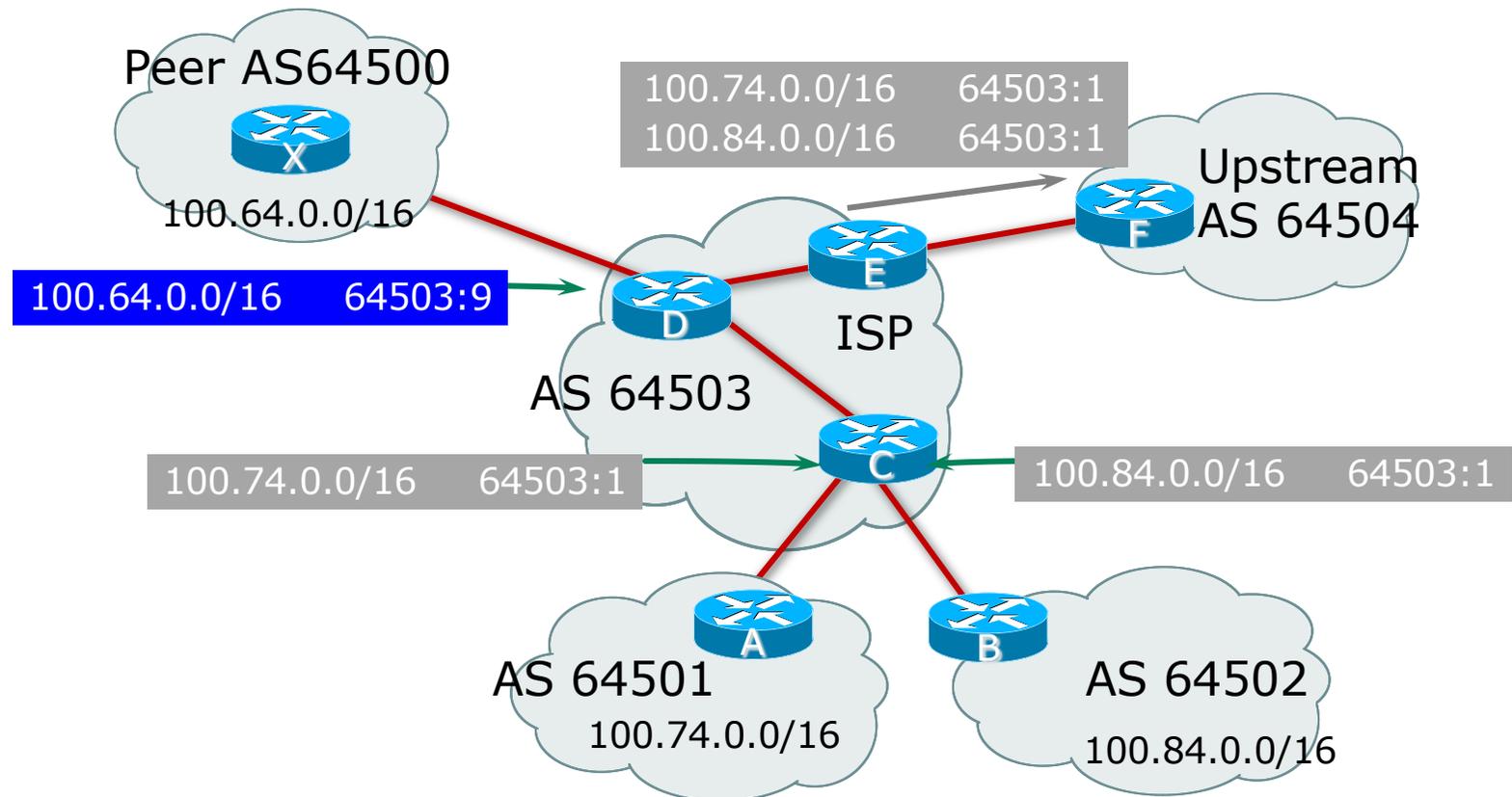
Community Example (after)



Community Example (after)



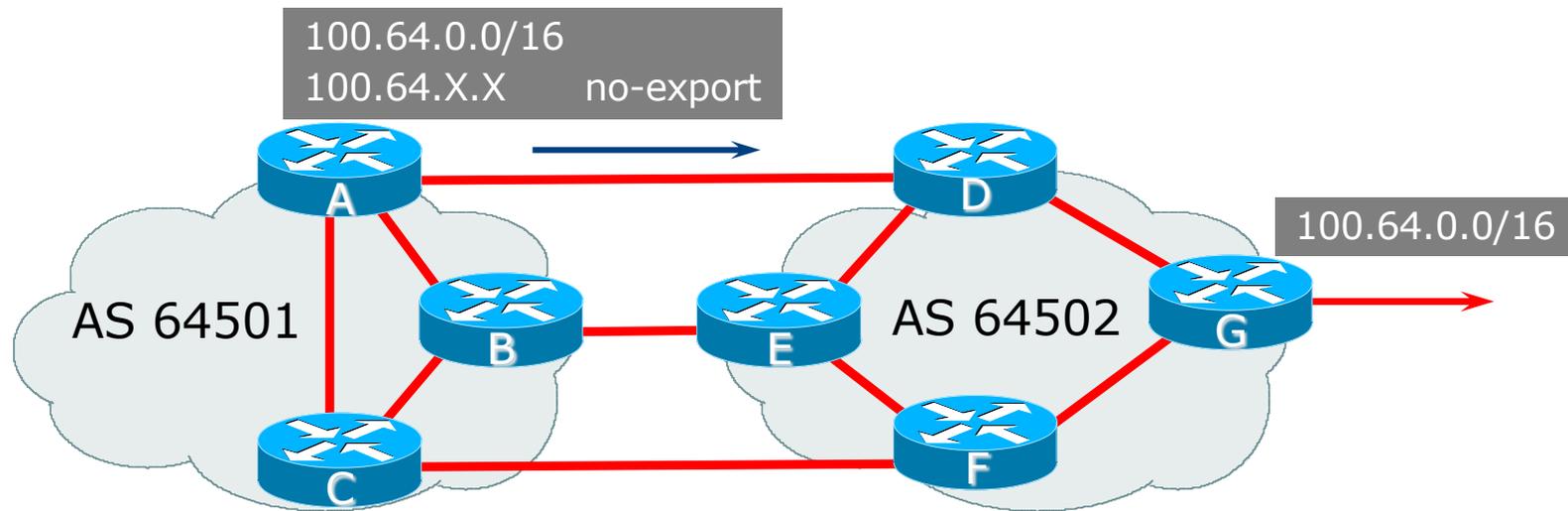
Community Example (after)



Well-Known Communities

- Several well-known communities
 - www.iana.org/assignments/bgp-well-known-communities
- Five most common:
 - *no-export* 65535:65281
 - Do not advertise to any EBGp peers
 - *no-advertise* 65535:65282
 - Do not advertise to any BGP peer
 - *no-peer* 65535:65284
 - Do not advertise to bi-lateral peers (RFC3765)
 - *blackhole* 65535:666
 - Null route the prefix (RFC7999)
 - *graceful-shutdown* 65535:0
 - Indicate imminent graceful shutdown (RFC8326)

No-Export Community



- ❑ AS64501 announces aggregate and subprefixes
 - Intention is to improve loadsharing by leaking subprefixes to upstream AS64502 only
- ❑ Subprefixes marked with **no-export** community
- ❑ Router G in AS64502 does not announce prefixes with **no-export** community set

Vendor Policy implementation

- Be aware that each vendor has differing policy language behaviours for:
 - Treatment of well known communities
 - Setting communities
 - Removing communities
 - Replacing communities
- Consult:
 - Vendor documentation
 - <https://www.rfc-editor.org/rfc/rfc8642.txt> for discussion of some of the issues operators need to be aware of

What about 4-byte ASNs?

- Communities are widely used for encoding routing policy
 - 32-bit attribute
- RFC1998 format is now “standard” practice
 - ASN:number
- Fine for 2-byte ASNs, but 4-byte ASNs cannot be encoded
- Solutions:
 - Use “private ASN” for the first 16 bits
 - **RFC8092 – “BGP Large Communities”**

BGP 'Large Community' Attribute

- New attribute designed to accommodate:
 - Local 32-bit ASN
 - Local Operator Defined Action (32-bits)
 - Remote Operator Defined Action (32-bits)
- This allows operators using 32-bit ASNs to peer with others using 32-bit ASNs and define policy actions
 - Compare with standard Communities which only accommodated 16-bit ASNs and 16-bits of action

BGP 'Large Community' Examples

- Some examples using common community conventions
 - (see BGP Community section for more detailed examples of typical network operator BGP Community policy)
 - **131072:3:131074**
 - AS 131072 requests AS 131074 to do a **three** times prepend of this prefix on AS 131074's peerings
 - **131072:0:131074**
 - AS 131072 requests AS 131074 not to announce this prefix

BGP Path Selection Algorithm



Why is this the best path?

BGP Path Selection Algorithm: Part One

1. Do not consider path if no route to next hop
2. Do not consider IBGP path if not synchronised (historical)
3. Highest weight (local to router)
4. Highest local preference (global within AS)
5. Prefer locally originated route
6. Shortest AS path
7. Lowest origin code
 - IGP < EGP < incomplete

BGP Path Selection Algorithm: Part Two

8. Lowest Multi-Exit Discriminator (MED)
 - Cisco IOS: if **bgp deterministic-med**, order the paths by AS number before comparing
 - Cisco IOS: if **bgp always-compare-med**, then compare for all paths
 - Otherwise only consider MEDs if paths are from the same neighbouring AS
9. Prefer EBGP path over IBGP path
10. Path with lowest IGP metric to next-hop

BGP Path Selection Algorithm: Part Three

11. For EBGP paths:

- Cisco IOS: if multipath is enabled, install N parallel paths in forwarding table
- If router-id is the same, go to next step (as per RFC)
- If router-id is not the same, select the oldest path (non-RFC)
 - ▣ To turn off @ Cisco: `bgp bestpath compare-routerid`
 - ▣ To turn off @ Juniper: `path-selection external-router-id`

12. Lowest router-id (originator-id for reflected routes)

13. Shortest cluster-list

- Client must be aware of Route Reflector attributes!

14. Lowest neighbour address

BGP Path Selection Algorithm

- In multi-vendor environments:
 - Make sure the path selection processes are understood for each brand of equipment
 - All must follow the RFC, but because of “customer demand”, each vendor has:
 - Slightly different implementations
 - Extra steps
 - Extra features
 - Watch out for possible MED confusion

Applying Policy with BGP



Controlling Traffic Flow & Traffic
Engineering

Applying Policy in BGP: Why?

- Network operators rarely “plug in routers and go”
- External relationships:
 - Control who they peer with
 - Control who they give transit to
 - Control who they get transit from
- Traffic flow control:
 - Efficiently use the scarce infrastructure resources (external link load balancing)
 - Congestion avoidance
 - Terminology: Traffic Engineering

Applying Policy in BGP: How?

- Policies are applied by:
 - Setting BGP attributes (local-pref, MED, AS-PATH, community), thereby influencing the path selection process
 - Advertising or Filtering prefixes
 - Advertising or Filtering prefixes according to ASN and AS-PATHs
 - Advertising or Filtering prefixes according to Community membership

Applying Policy with BGP: Tools

- Most implementations have tools to apply policies to BGP:
 - Prefix manipulation/filtering
 - AS-PATH manipulation/filtering
 - Community Attribute setting and matching
- Implementations also have policy language which can do various match/set constructs on the attributes of chosen BGP routes

BGP Capabilities



Extending BGP

BGP Capabilities

- Documented in RFC2842
- Capabilities parameters passed in BGP open message
- Unknown or unsupported capabilities will result in NOTIFICATION message
- Codes:
 - 0 to 63 are assigned by IANA by IETF consensus
 - 64 to 238 are assigned by IANA “first come first served”
 - 239 to 254 are “Experimental Use”

BGP Capabilities

- ❑ Current capabilities are listed opposite
- ❑ Most implementations support:
 - Multiprotocol extensions
 - Route Refresh
 - BGP ORF
 - Graceful Restart
 - 4-byte ASNs

0	Reserved	[RFC3392]
1	Multiprotocol Extensions for BGP-4	[RFC4760]
2	Route Refresh Capability for BGP-4	[RFC2918]
3	Outbound Route Filtering Capability	[RFC5291]
4	Multiple routes to a destination capability	[RFC3107]
5	Extended Next Hop Encoding	[RFC5549]
6	BGP Extended Message	[RFC8654]
7	BGPsec Capability	[RFC8205]
8	Multiple Labels Capability	[RFC8277]
9	BGP Role	[RFC9234]
64	Graceful Restart Capability	[RFC4724]
65	Support for 4 octet ASNs	[RFC6793]
66	Deprecated	
67	Support for Dynamic Capability	[ID]
68	Multisession BGP	[ID]
69	Add Path Capability	[RFC7911]
70	Enhanced Route Refresh Capability	[RFC7313]
71	Long Lived Graceful Restart	[ID]
72	Routing Policy Distribution	[ID]
73	FQDN Capability	[ID]
74	BFD Capability	[ID]
75	Software Version Capability	[ID]
128-131 & 184-185	Deprecated	[RFC8810]

<https://www.iana.org/assignments/capability-codes>

BGP Techniques for Network Operators

- BGP Basics
- **Scaling BGP**
- Using Communities
- Deploying BGP in a Service Provider Network

BGP Scaling Techniques



Scaling BGP

BGP Scaling Techniques

- Original BGP specification and implementation was fine for the Internet of the early 1990s
 - But didn't scale
- Issues as the Internet grew included:
 - Scaling the IBGP mesh beyond a few peers?
 - Implement new policy without causing flaps and route churning?
 - Keep the network stable, scalable, as well as simple?

BGP Scaling Techniques

- BGP Configuration Scaling
 - Grouping BGP peers

- Industry Best Practice Scaling Techniques
 - Route Refresh
 - Route Reflectors

- Historical Scaling Techniques
 - Soft Reconfiguration
 - Confederations
 - Route Flap Damping

BGP Configuration Scaling



Cisco's peer-groups
&
Juniper's BGP groups

Grouping similar BGP peers

- What are they for?
 - Allows operators to group peers with the same outbound policy
 - Makes configuration easier
 - Makes configuration less prone to error
 - Makes configuration more readable
 - Members can have different inbound policy
 - Can be used for EBGP neighbours too!

Grouping similar BGP peers

□ Cisco:

■ peer-groups

- Originally designed to speed IBGP convergence – now for scaling BGP configuration management

■ Internal code optimisation called *update-groups*

- Speeds IBGP convergence; update only calculated once for neighbours with the same outbound policy

□ Juniper:

■ BGP groups

Configuring a Peer Group in IOS

```
router bgp 64500
  address-family ipv4
    neighbor IBGP peer-group
    neighbor IBGP remote-as 64500
    neighbor IBGP update-source loopback 0
    neighbor IBGP send-community
    neighbor IBGP route-map outfilter out
    neighbor 100.64.0.1 peer-group IBGP
    neighbor 100.64.0.2 peer-group IBGP
    neighbor 100.64.0.2 route-map infilter in
    neighbor 100.64.0.3 peer-group IBGP
!
```

- Note how 100.64.0.2 has an additional inbound filter over the peer-group

Configuring a Peer Group in IOS

```
router bgp 64500
  address-family ipv4
    neighbor EBGP peer-group
    neighbor EBGP send-community
    neighbor EBGP route-map set-metric out
    neighbor 100.89.1.2 remote-as 64502
    neighbor 100.89.1.2 peer-group EBGP
    neighbor 100.89.1.4 remote-as 64503
    neighbor 100.89.1.4 peer-group EBGP
    neighbor 100.89.1.6 remote-as 64504
    neighbor 100.89.1.6 peer-group EBGP
    neighbor 100.89.1.6 filter-list infiltrer in
  !
```

- Can be used for EBGP as well

Juniper BGP groups

- JunOS has very similar configuration concept
 - Simply known as bgp groups, for example:

```
protocols {
  bgp {
    group ibgp {
      type internal;
      local-address 10.0.15.241;
      family inet {
        unicast;
      }
      export export-ibgp;
      peer-as 10;
      neighbor 10.0.15.242 {
        description "Router 2";
      }
      neighbor 10.0.15.243 {
        description "Router 3";
      }
      ...etc...
    }
  }
}
```

Grouping similar BGP peers

- Always configure peer-groups or BGP groups for IBGP
 - Even if there are only a few IBGP peers
 - Easier to scale network in the future
 - Makes configuration easier to read
- Consider using peer-groups for EBGP
 - Especially useful for multiple BGP customers using same AS (RFC2270)
 - Also useful at Exchange Points:
 - Where network operator policy is generally the same to each peer
 - For Route Server where all peers receive the same routing updates

Dynamic Reconfiguration



Non-destructive policy changes

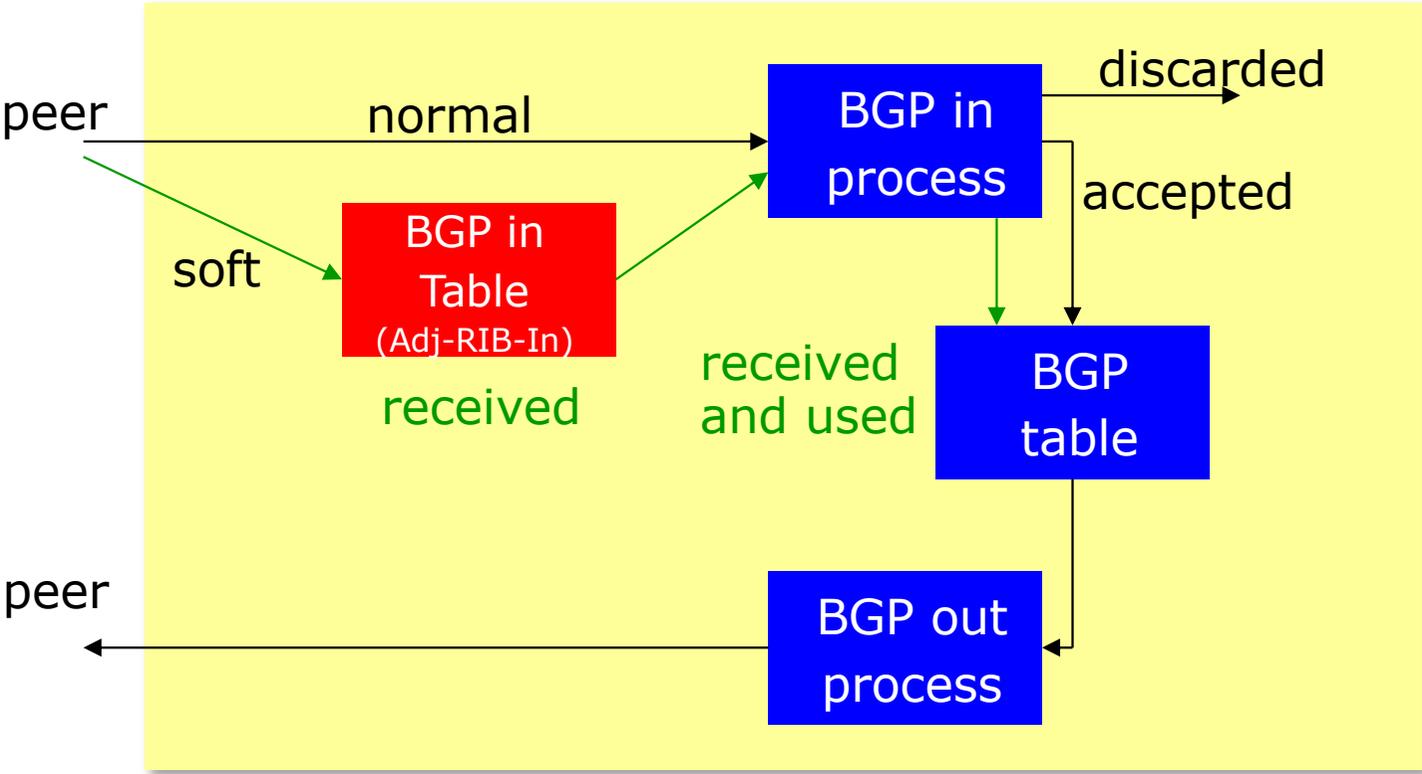
Route Refresh: History

- Historically, routers only stored prefixes which were accepted by incoming policy
 - Those rejected by policy were discarded
 - No storage of discarded prefixes
- If a change of incoming policy was required:
 - The EBGP session had to be shutdown, and then brought up again
 - Destructive change: EBGP session down means lost connectivity to that peer, and potentially the rest of the Internet (outage of many minutes!)
- Changes in BGP policy usually had to be carried out during published scheduled maintenance timeslots
 - To minimise impact on end-users

Route Refresh: Step One

- First step at solving this problem was by Cisco with the “soft reconfiguration” concept
 - Router keeps a record of all prefixes received **before** any policy applied (known as Adj-RIB-In)
 - Needed extra memory (highly problematic in early routers and modern routers with limited memory)
 - Full BGP table with policy change could require double the control plane memory for BGP
 - Policy changes applied to the stored received prefixes
 - No shutdown and restart of the BGP session needed when implementing policy changes

Cisco's Soft Reconfiguration



Route Refresh: Step Two

- Second step at solving this problem was the introduction of “route refresh”
 - A BGP Capability: RFC2918
 - Peering remains active
 - Impacts only those prefixes affected by the policy change
 - No configuration needed
 - Automatically negotiated at peer establishment
 - No extra memory needed (no need for Adj-Rib-In)
- Today most implementations do an automatic route-refresh after BGP Policy changes
 - Beware: not all vendor implementations do an automatic route-refresh – know your software implementation!

Route Refresh

- Use Route Refresh capability, *not* hard reset
 - Supported on virtually all BGP implementations
 - Find out from the detailed BGP neighbour status
 - Non-disruptive, “Good For the Internet”
- Only hard-reset a BGP peering as a last resort

Consider the impact of a hard-reset of BGP to be equivalent to a router reboot

Route Refresh: Route Origin Validation

- Route Origin Validation means checking if the prefix received has a valid ROA
 - Route Origination Authorisation – digital object indicating the origin AS for the prefix (and subnet size) using RPKI
 - Valid ROA means that the prefix (and subnet) is being originated from the correct origin AS
 - See the “BGP Origin Validation” presentation for more in-depth content
- Routers implementing ROV apply the validation results via the existing policy language & process
 - Valid – allow; Invalid – drop; NotFound – allow (at lower preference?)
- **Problem**: how is incoming policy applied on routers today?

Route Refresh: Route Origin Validation

- Routers which maintain the Adj-RIB-In:
 - Apply the ROV policy to the stored received BGP table
 - Updates are applied “automatically” to the BGP table and therefore the FIB
 - No impact on any BGP peers (Route Refresh not needed)

Route Refresh: Route Origin Validation

- Routers which do NOT maintain the Adj-RIB-In:
 - Apply the ROV policy by sending a Route Refresh to peers
 - When there are a large number of ROAs (June 2023 sees over 360k IPv4 and 78k IPv6), and frequent changes or updates of ROAs:
 - Routers are sending frequent Route Refresh requests to peers (typically every few minutes)
 - Peers are being “bombed” by Route Refresh requests: significant resource burden when they send the full or a large portion of the BGP table
 - Severe control plane CPU impact on the peer router (effectively a Denial of Service on the peer router)
 - As more and more ROAs are created and altered globally, this problem becomes significantly more serious!

Route Refresh: Route Origin Validation

- JunOS implements Adj-RIB-In by default
 - ROA updates do not cause a problem when operating ROV

- Cisco does not implement Adj-RIB-In by default:
 - Applies to all Cisco IOS/IOS-XE/IOS-XR apart from the most recent releases
 - **MUST turn on soft-reconfiguration if running ROV on the router**
 - Soft-reconfiguration is similar concept to Adj-RIB-In

Route Reflectors

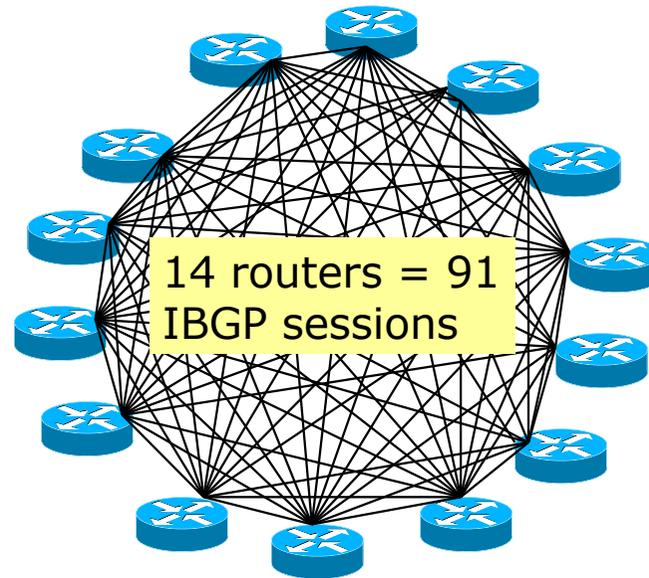


Scaling the IBGP mesh

Scaling the IBGP mesh

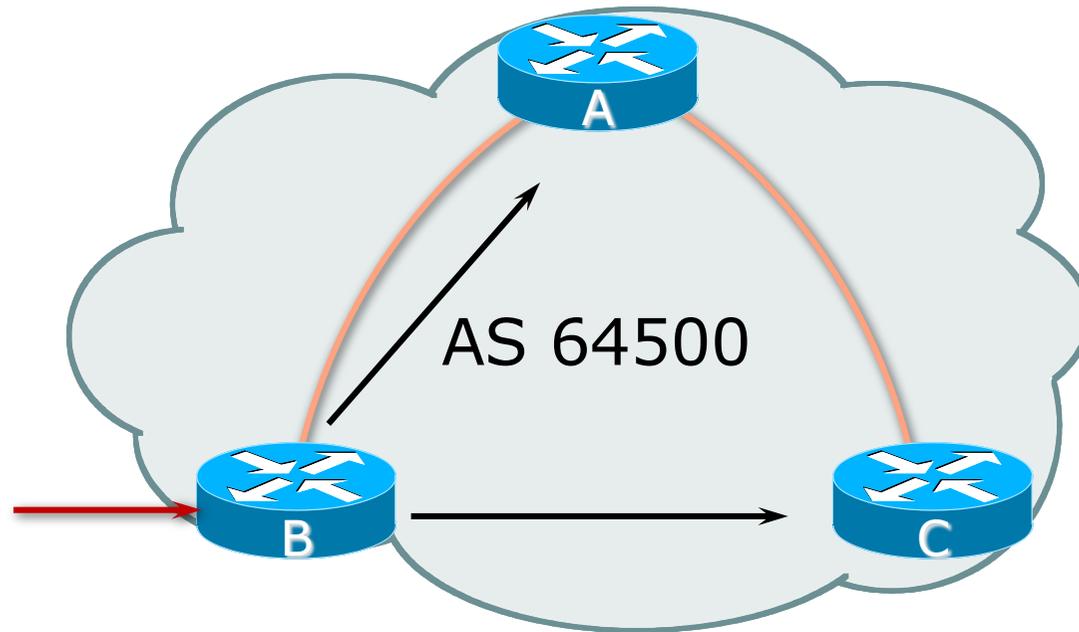
- Avoid $\frac{1}{2}n(n-1)$ IBGP mesh

**$n=1000 \Rightarrow$ nearly
half a million
IBGP sessions!**

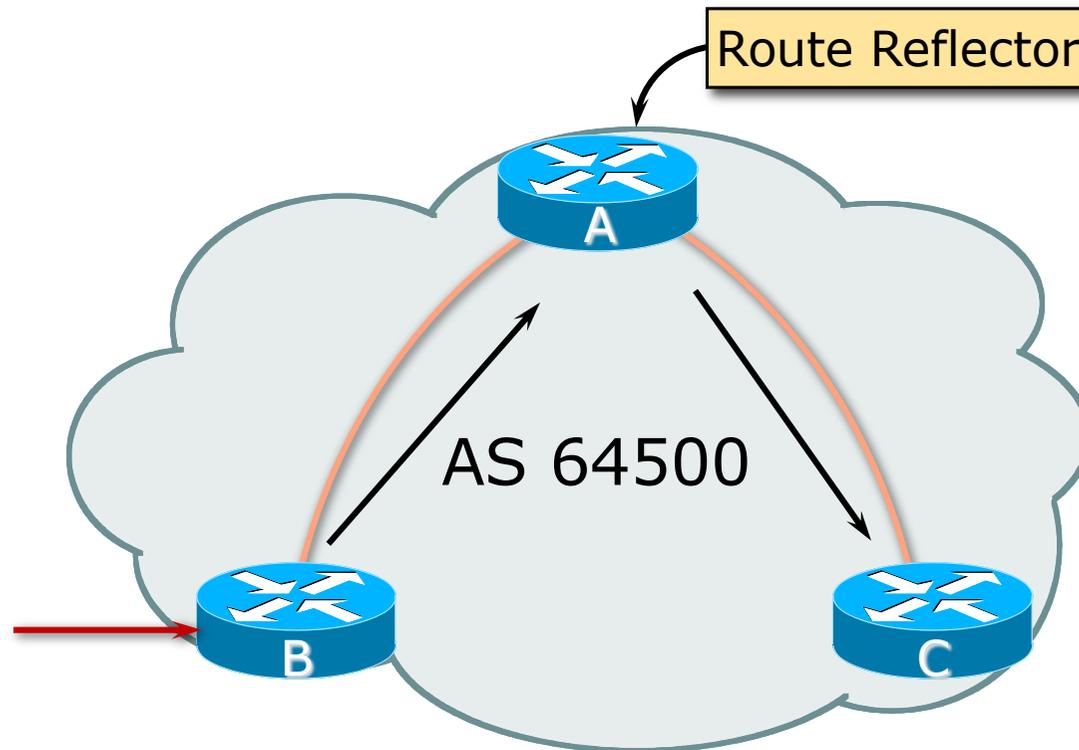


- Two solutions
 - Route reflector: simpler to deploy and run
 - BGP Confederation: more complex, has corner case advantages

Route Reflector: Principle

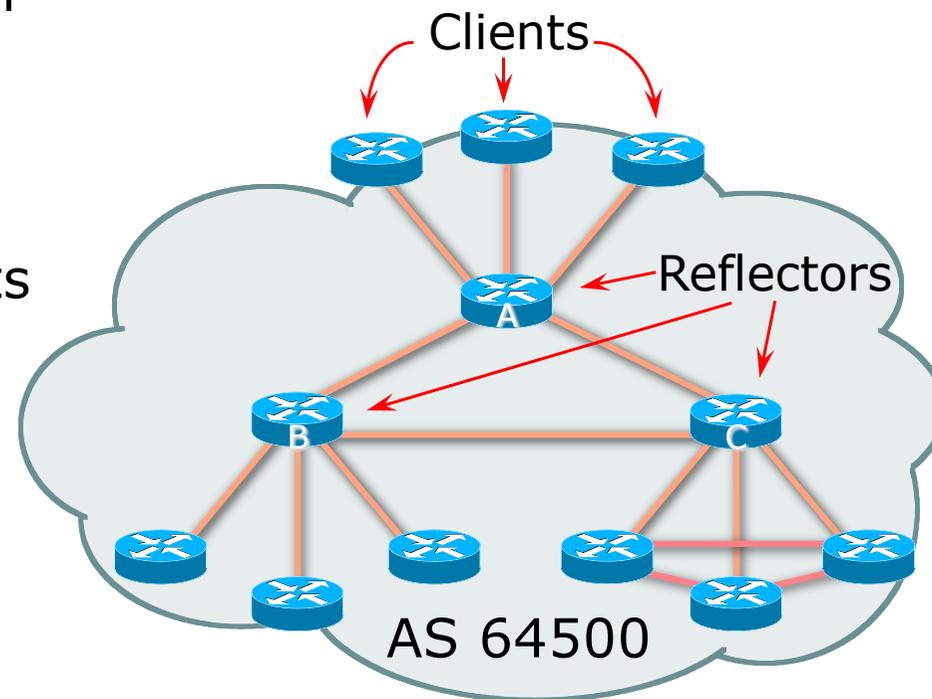


Route Reflector: Principle



Route Reflector: Rules

- ❑ Reflector receives path from clients and non-clients
- ❑ Selects best path
- ❑ If best path is from client, reflect to other clients and non-clients
- ❑ If best path is from non-client, reflect to clients only
- ❑ Non-meshed clients
- ❑ Described in RFC4456



Route Reflector: Topology

- ❑ Divide the backbone into multiple clusters
- ❑ Provision at least one Route Reflector (RR) and few clients per cluster
- ❑ Route reflectors are fully meshed
- ❑ Clients in a cluster could be fully meshed
- ❑ Single IGP still carries next-hop and any local routes

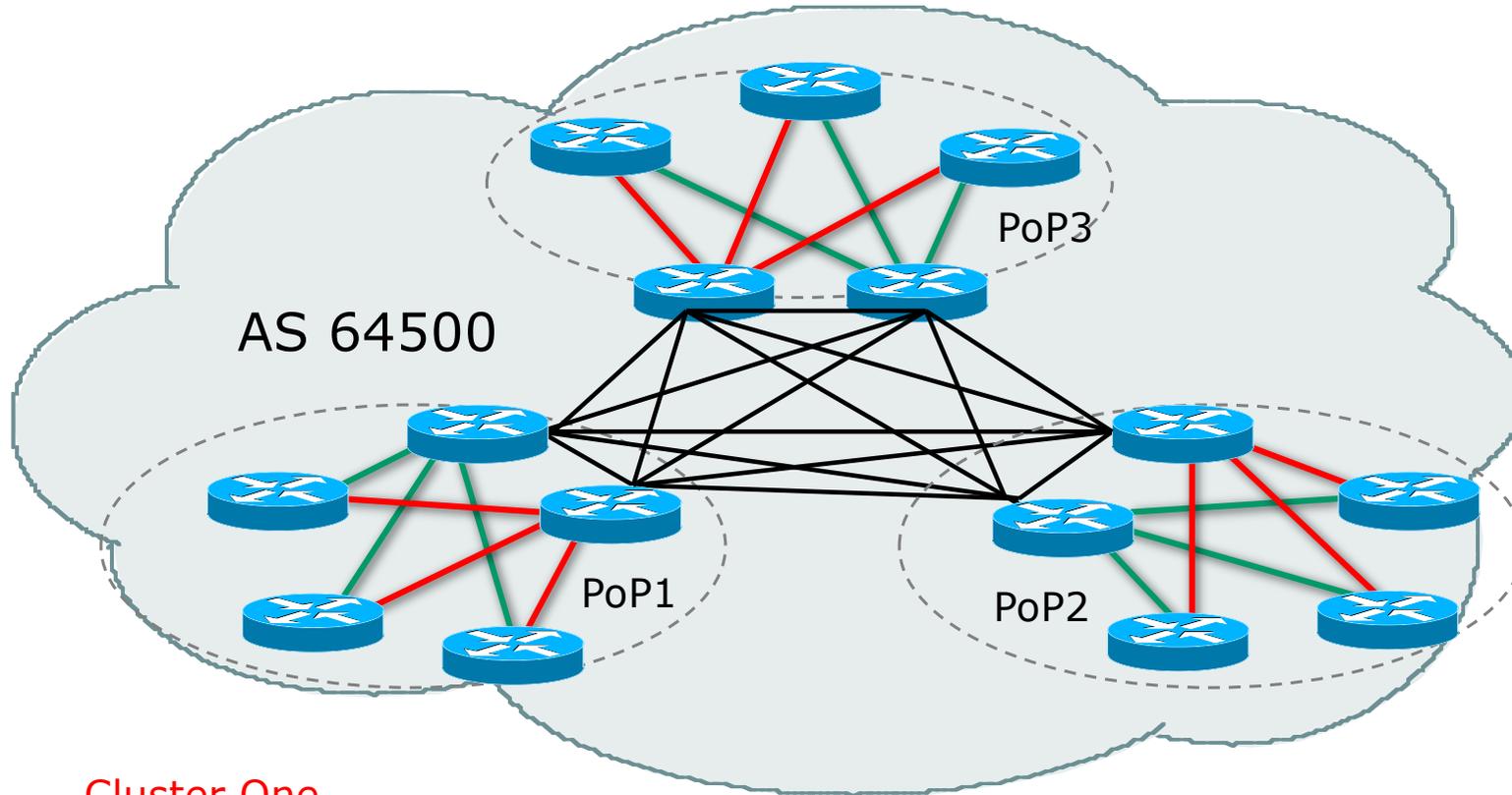
Route Reflector: Loop Avoidance

- Originator_ID attribute
 - Carries the RID of the originator of the route in the local AS (created by the RR)
- Cluster_list attribute
 - The local cluster-id is added when the update is sent by the RR
 - Best to set cluster-id from router-id by (address of loopback interface)
 - (Some Network Operators use their own cluster-id assignment strategy – but needs to be well documented!)

Route Reflector: Redundancy

- Multiple RRs can be configured in the same cluster – not advised!
 - All RRs in the cluster must have the same cluster-id (otherwise it is a different cluster)
- A router may be a client of RRs in different clusters
 - Common today in service provider networks to overlay two clusters – redundancy achieved that way
 - → Each client has two RRs = redundancy

Route Reflector: Redundancy



Cluster One

Cluster Two

Route Reflector: Benefits

- ❑ Solves IBGP mesh problem
- ❑ Packet forwarding is not affected
- ❑ Normal BGP speakers co-exist
- ❑ Multiple reflectors for redundancy
- ❑ Easy migration
- ❑ Multiple levels of route reflectors

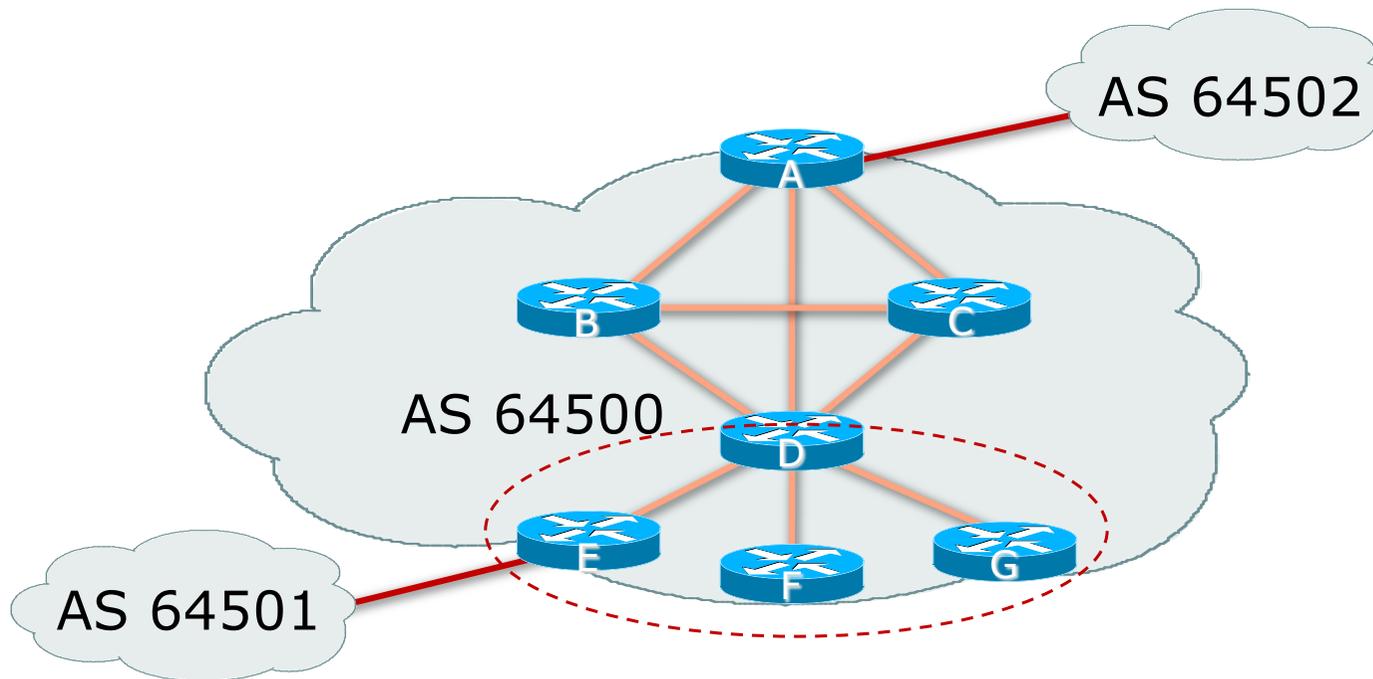
Route Reflector: Deployment

- Where to place the route reflectors?
 - Always follow the physical topology!
 - This will guarantee that the packet forwarding won't be affected
- Typical Service Provider network:
 - PoP has two core routers
 - Core routers are RR for the PoP
 - Two overlaid clusters

Route Reflector: Migration

- Typical Service Provider network:
 - Core routers have fully meshed IBGP
 - Create further hierarchy if core mesh too big
 - Split backbone into regions
- Configure one cluster pair at a time
 - Eliminate redundant IBGP sessions
 - Place maximum one RR per cluster
 - Easy migration, multiple levels

Route Reflector: Migration



- ❑ Migrate small parts of the network, one part at a time.

BGP Scaling Techniques

- These two standards-based techniques must be designed in from the beginning for all network operator infrastructure
 1. Route Refresh
 2. Route Reflectors

BGP Confederations



(Historical IBGP Scaling)

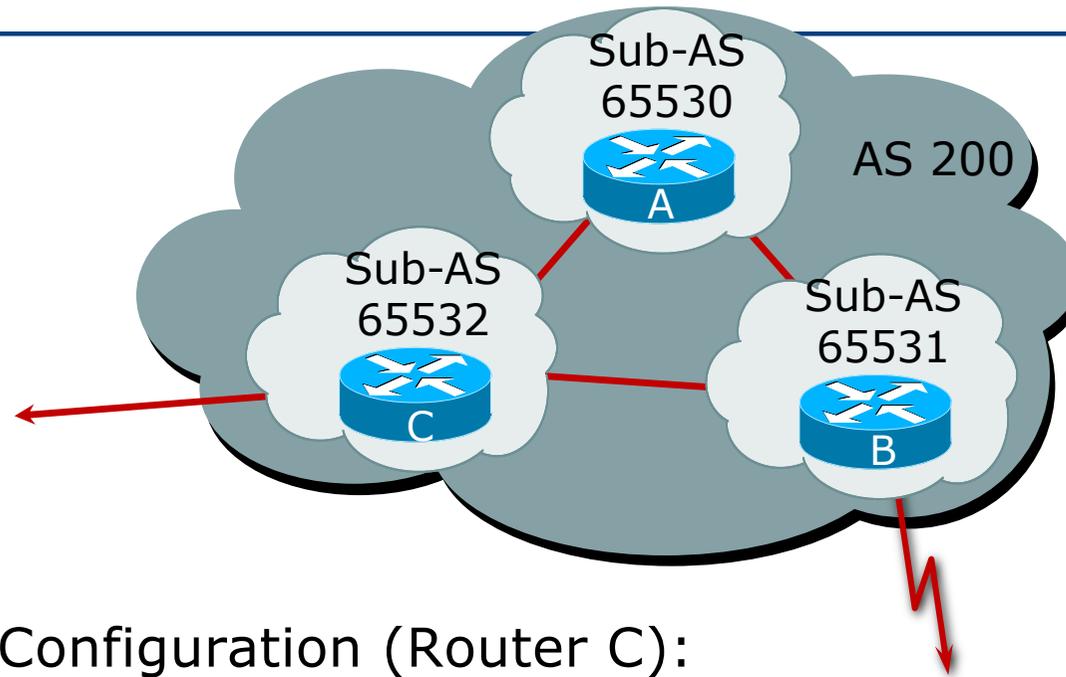
Confederations

- Divide the AS into sub-AS
 - EBGp between sub-AS, but some IBGP information is kept
 - Preserve NEXT_HOP across the sub-AS (IGP carries this information)
 - Preserve LOCAL_PREF and MED
- Usually a single IGP
- Described in RFC5065

Confederations

- Visible to outside world as single AS – “Confederation Identifier”
 - Each sub-AS uses a number from the private space (64512-65534)
- IBGP speakers in sub-AS are fully meshed
 - The total number of neighbors is reduced by limiting the full mesh requirement to only the peers in the sub-AS
 - Can also use Route-Reflector within sub-AS

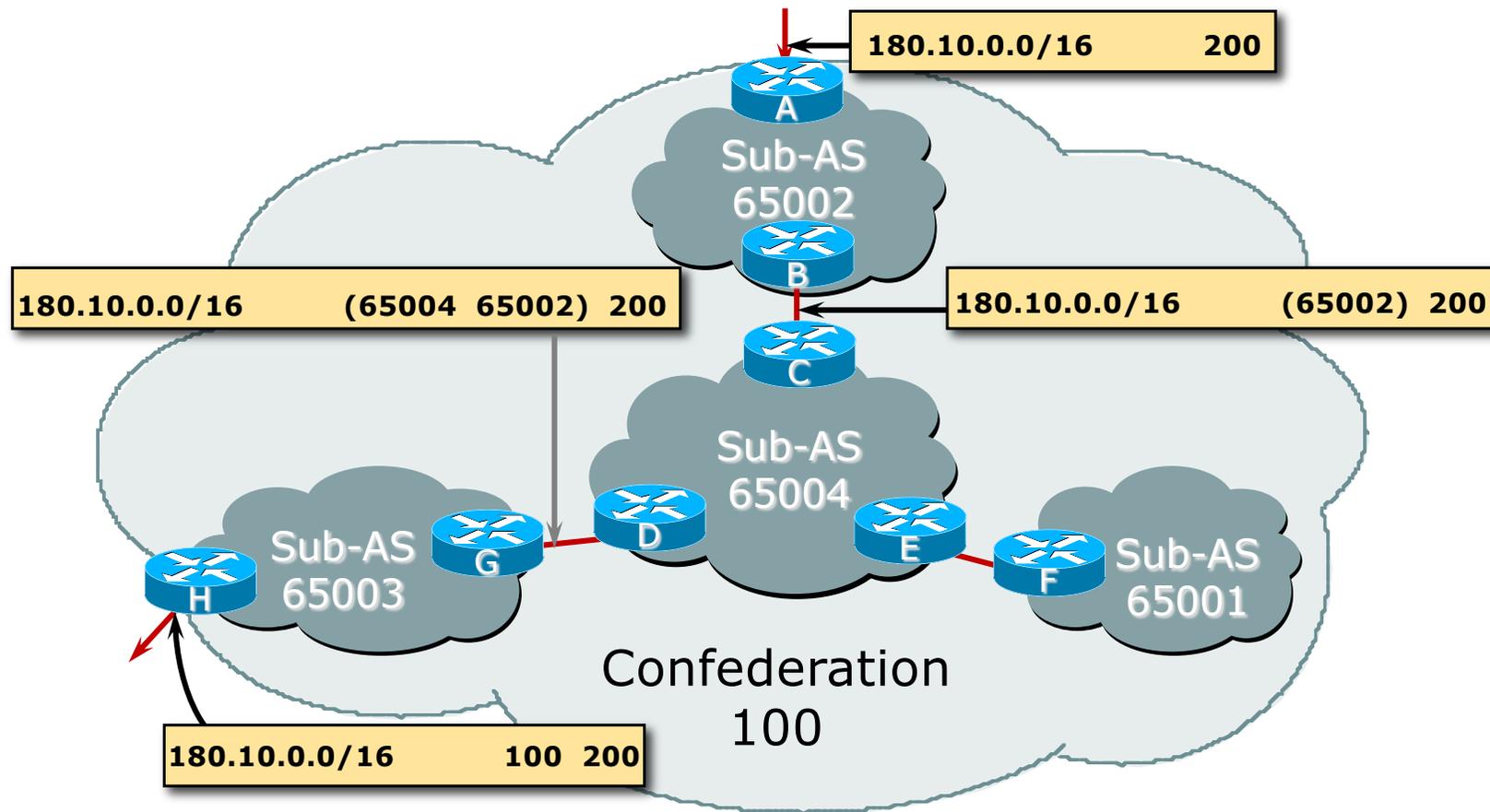
Confederations



- Configuration (Router C):

```
router bgp 65532
  bgp confederation identifier 200
  bgp confederation peers 65530 65531
  neighbor 141.153.12.1 remote-as 65530
  neighbor 141.153.17.2 remote-as 65531
```

Confederations: AS-Sequence



Route Propagation Decisions

- Same as with “normal” BGP:
 - From peer in same sub-AS → only to external peers
 - From external peers → to all neighbors
- “External peers” refers to
 - Peers outside the confederation
 - Peers in a different sub-AS
 - Preserve LOCAL_PREF, MED and NEXT_HOP

RRs or Confederations ?

	Internet Connectivity	Multi-Level Hierarchy	Policy Control	Scalability	Migration Complexity
Confederations	Anywhere in the network	Yes	Yes	Medium	Medium to High
Route Reflectors	Anywhere in the network	Yes	Yes	Very High	Very Low

New network operators deploy Route Reflectors from Day One

More points about Confederations

- Can ease “absorbing” other Service Providers into your provider – e.g., if one provider buys another
 - Or can use AS masquerading feature available in some implementations to do a similar thing
- Can use route-reflectors with confederation sub-AS to reduce the sub-AS iBGP mesh

Route Flap Damping



Network Stability for the 1990s

Network Instability for the 21st Century!

Route Flap Damping

- ❑ In the late 1990s, Route Flap Damping was a strongly recommended practice
- ❑ Now it is **strongly discouraged** as it causes far greater network instability than it cures
- ❑ But first, the theory...

Route Flap Damping

- Route flap
 - Going up and down of path or change in attribute
 - BGP WITHDRAW followed by UPDATE = 1 flap
 - EBGP neighbour going down/up is NOT a flap
 - Ripples through the entire Internet
 - Wastes CPU
- Damping aims to reduce scope of route flap propagation

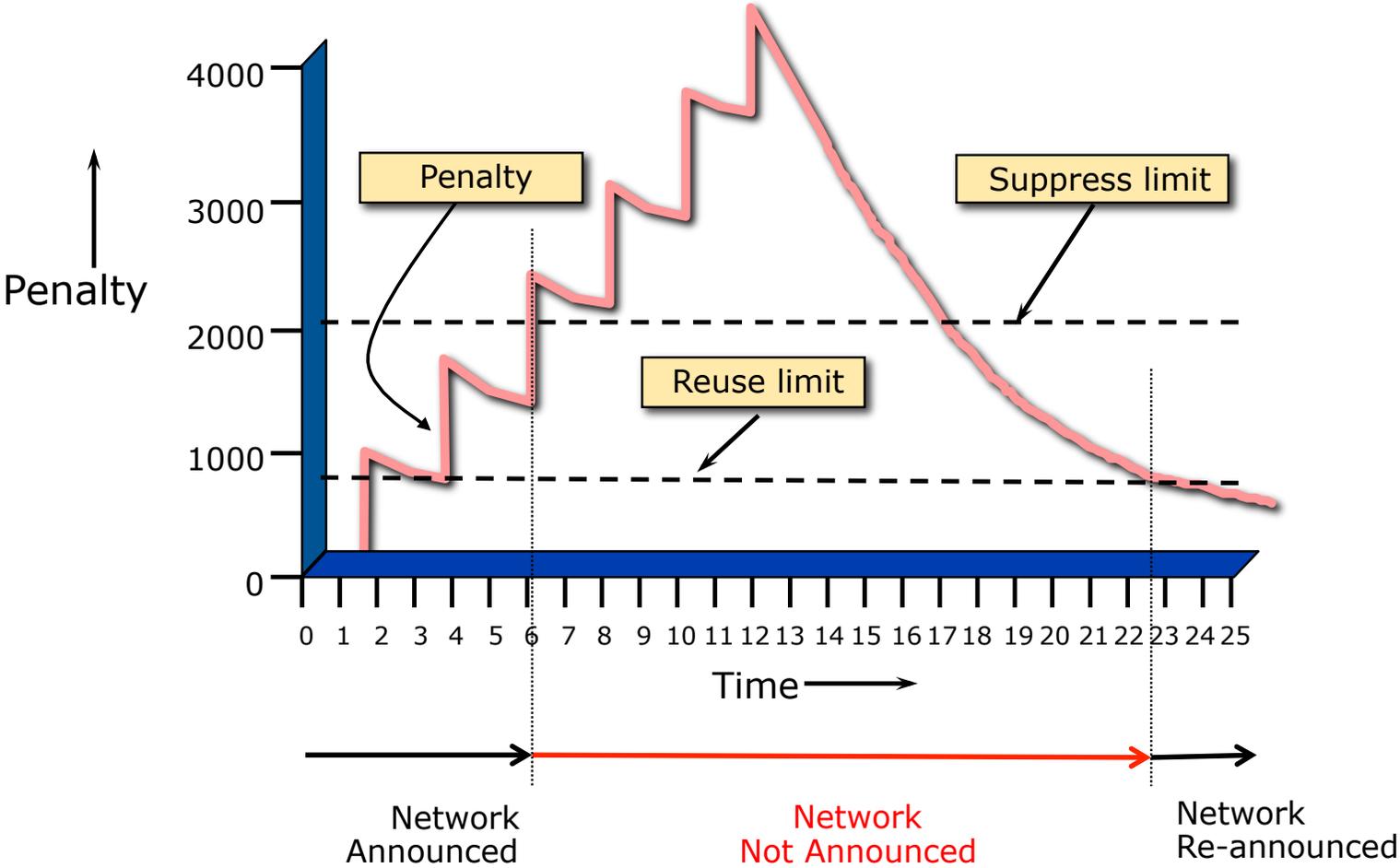
Route Flap Damping (continued)

- Requirements
 - Fast convergence for normal route changes
 - History predicts future behaviour
 - Suppress oscillating routes
 - Advertise stable routes
- Implementation described in RFC 2439

Operation

- Add penalty (1000) for each flap
 - Change in attribute gets penalty of 500
- Exponentially decay penalty
 - Half-life determines decay rate
- Penalty above suppress-limit
 - Do not advertise route to BGP peers
- Penalty decayed below reuse-limit
 - Re-advertise route to BGP peers
 - Penalty reset to zero when it is half of reuse-limit

Operation



Operation

- ❑ Only applied to inbound announcements from EBGP peers
- ❑ Alternate paths still usable
- ❑ Controllable by at least:
 - Half-life
 - reuse-limit
 - suppress-limit
 - maximum suppress time

Configuration

- Implementations allow various policy control with flap damping
 - Fixed damping, same rate applied to all prefixes
 - Variable damping, different rates applied to different ranges of prefixes and prefix lengths

Route Flap Damping History

- First implementations on the Internet by 1995
- Vendor defaults too severe
 - RIPE Routing Working Group recommendations in ripe-178, ripe-210, and ripe-229
 - <http://www.ripe.net/ripe/docs>
 - But many network operators simply switched on the vendors' default values without thinking

Serious Problems:

- “Route Flap Damping Exacerbates Internet Routing Convergence”
 - Zhuoqing Morley Mao, Ramesh Govindan, George Varghese & Randy H. Katz, August 2002
- “What is the sound of one route flapping?”
 - Tim Griffin, June 2002
- Various work on routing convergence by Craig Labovitz and Abha Ahuja a few years ago
- “Happy Packets”
 - Closely related work by Randy Bush et al

Problem 1:

- One path flaps:
 - BGP speakers pick next best path, announce to all peers, flap counter incremented
 - Those peers see change in best path, flap counter incremented
 - After a few hops, peers see multiple changes simply caused by a single flap → prefix is suppressed

Problem 2:

- Different BGP implementations have different transit time for prefixes
 - Some hold onto prefix for some time before advertising
 - Others advertise immediately
- Race to the finish line causes appearance of flapping, caused by a simple announcement or path change → prefix is suppressed

Solution:

- Misconfigured Route Flap Damping will seriously impact access to:
 - Your network *and*
 - The Internet
- More background contained in RIPE Routing Working Group document:
 - www.ripe.net/ripe/docs/ripe-378
- Recommendations now in:
 - www.rfc-editor.org/rfc/rfc7196.txt and www.ripe.net/ripe/docs/ripe-580

BGP Techniques for Network Operators

- BGP Basics
- Scaling BGP
- **Using Communities**
- Deploying BGP in a Service Provider Network

BGP Communities



Scaling BGP Policies

Multihoming and Communities

- The BGP community attribute is a very powerful tool for assisting and scaling BGP Policies and BGP Multihoming
 - BGP Communities were introduced earlier – now a more detailed look at this power BGP tool is required
- Most major Network Operators make extensive use of BGP communities:
 - Internal policies (IBGP)
 - Inter-provider relationships (MED replacement)
 - Customer traffic engineering

Well-known BGP Communities



How the “well-known” BGP communities
are used

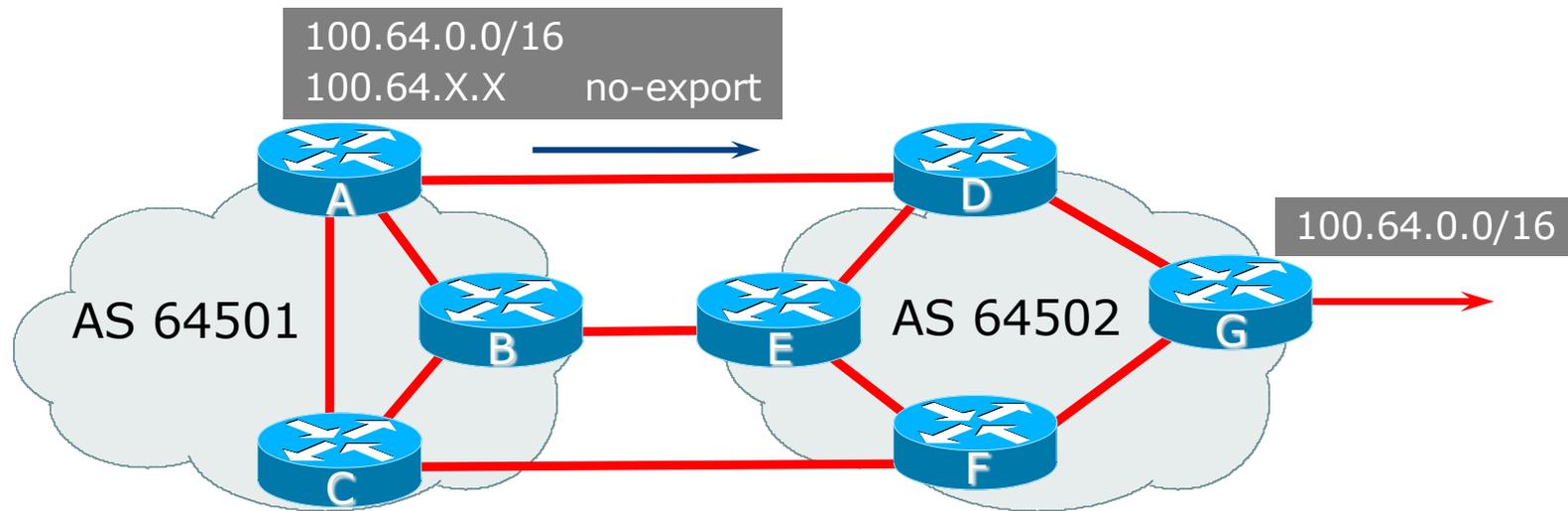
Well-Known Communities

- Several well-known communities
 - www.iana.org/assignments/bgp-well-known-communities
- Five most common:
 - *no-export* 65535:65281
 - Do not advertise to any EBGp peers
 - *no-advertise* 65535:65282
 - Do not advertise to any BGP peer
 - *no-peer* 65535:65284
 - Do not advertise to bi-lateral peers (RFC3765)
 - *blackhole* 65535:666
 - Null route the prefix (RFC7999)
 - *graceful-shutdown* 65535:0
 - Indicate imminent graceful shutdown (RFC8326)

Well-Known Communities: Notes

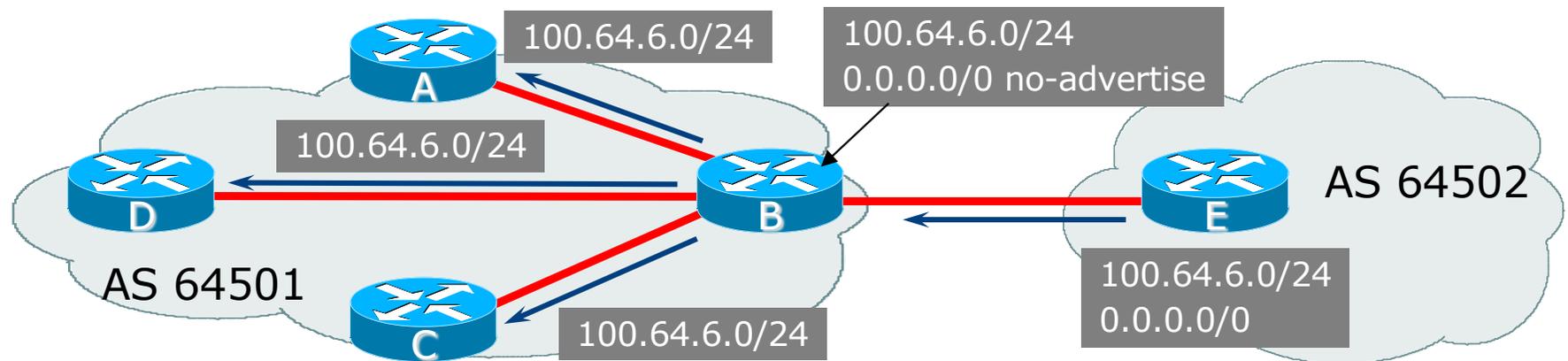
- Even though there are several well-known communities there are variations in implementation support
 - Not all vendors will create configuration key-words to support them
 - Not all vendors will automatically implement their behaviours
 - Not all vendors will allow them to be overwritten
 - *And so on*
- Check vendor documentation for implementation details
 - RFC8642 will give some idea as to the issues to be aware of
- Advice:
 - If the key-word does not exist, create a community declaration that implements the key-word (for configuration clarity & simplicity)

No-Export Community



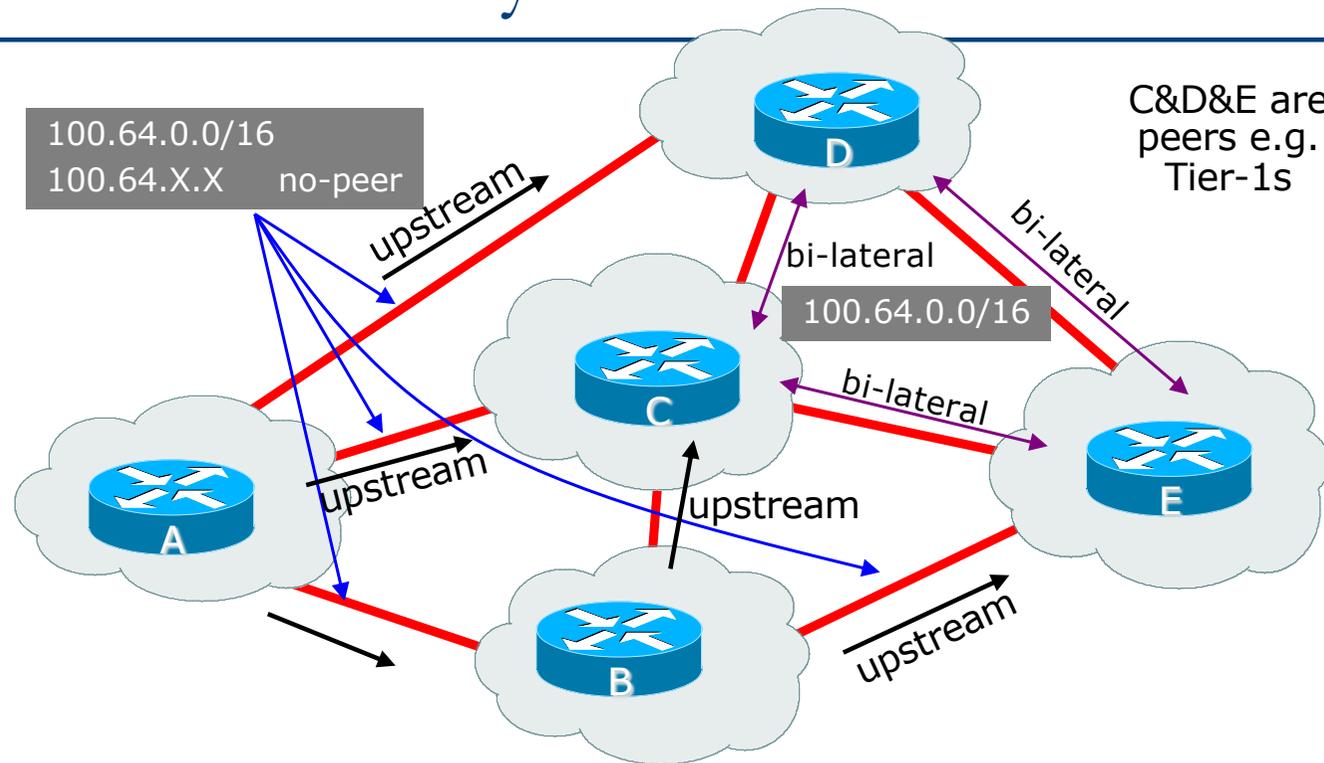
- ❑ AS64501 announces aggregate and subprefixes
 - Intention is to improve loadsharing by leaking subprefixes to upstream AS64502 only
- ❑ Subprefixes marked with **no-export** community
- ❑ Router G in AS64502 does not announce prefixes with **no-export** community set

No-Advertise Community



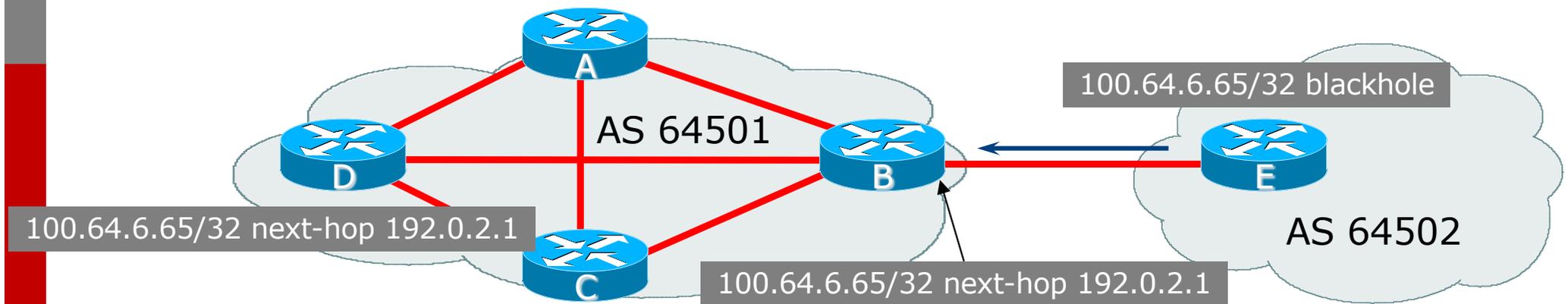
- Used to not advertise a prefix in IBGP
 - B hears 0.0.0.0/0 from EBGP peer E
 - Tags 0.0.0.0/0 as *no-advertise*
 - B will (automatically) not announce prefix to A, C or D
 - Easier/more scalable than using a prefix filter

No-Peer Community



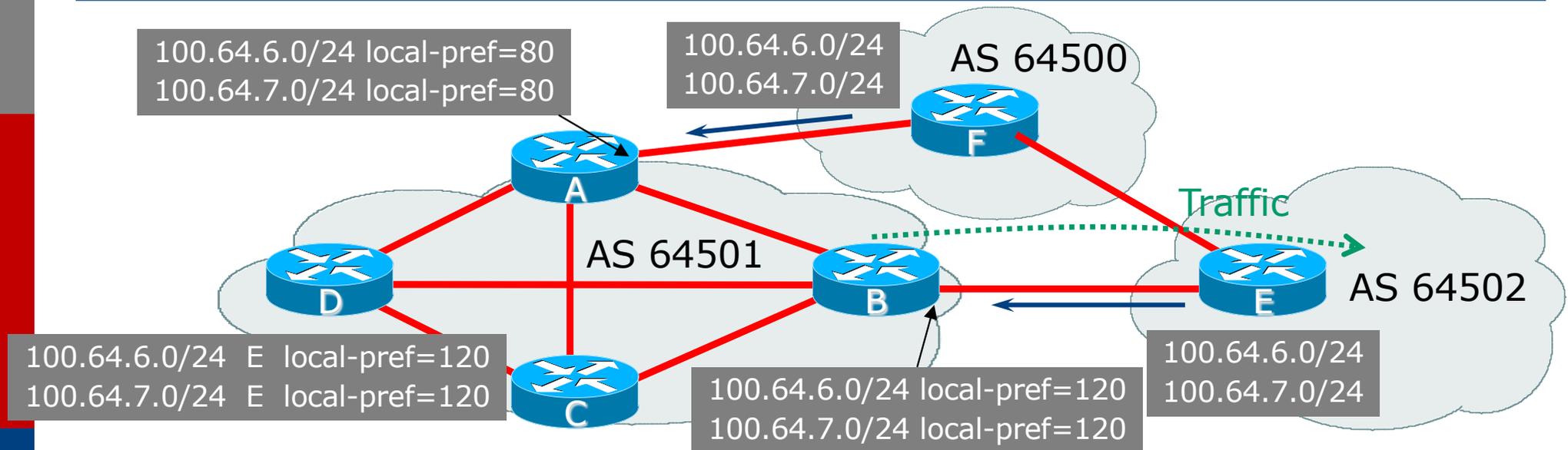
- Sub-prefixes marked with **no-peer** community are not sent to bi-lateral peers
 - They are only sent to upstream providers

Blackhole Community



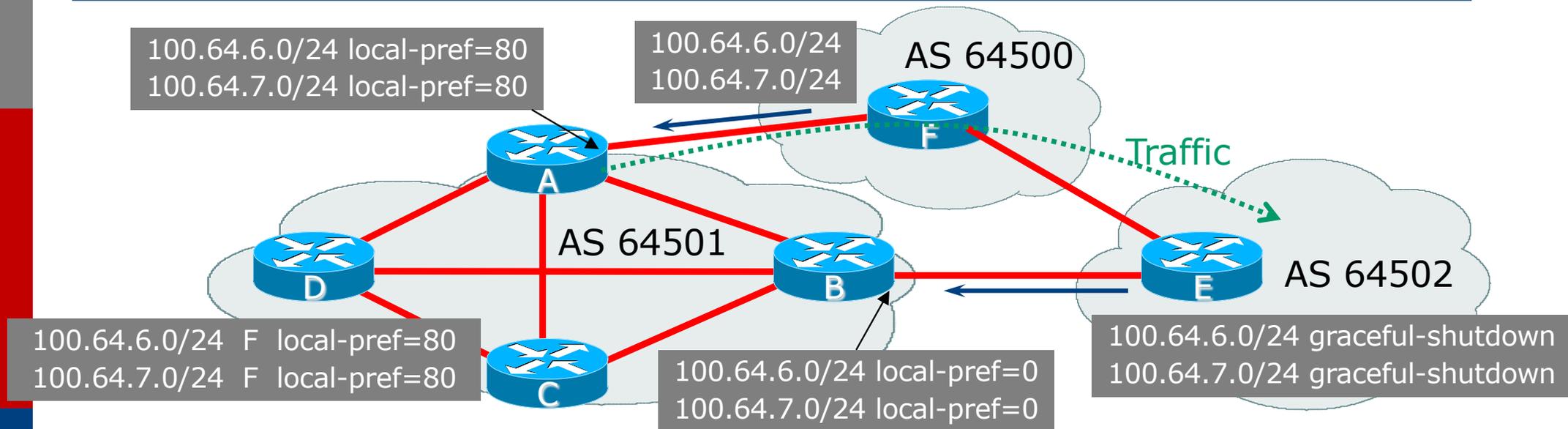
- Used to signal to a BGP neighbour to null route traffic
 - Router E sets *blackhole* community
 - Router B detects *blackhole* community on incoming EBGP announcements and sets *next-hop* to 192.0.2.1
 - 192.0.2.1 is routed to Null interface on all routers within the Autonomous System
 - All traffic to 100.64.6.65 is Null routed

Graceful-Shutdown Community (before)



- Used to inform an EBGP peer that the peering will be going down soon
 - Steady state is primary path between AS64502 and 64501 is via routers E & B
 - AS64502 wants to shutdown direct link, which means traffic will use path via AS64500
 - Graceful-Shutdown ensures that this can be achieved without traffic loss by informing AS64501 that the link is going away

Graceful-Shutdown Community (after)



- Used to inform an EBGP peer that the peering will be going down soon
 - Router E sets *graceful-shutdown* community
 - Router B detects *graceful-shutdown* community on incoming EBGP announcements and sets *local-preference* to 0
 - Best path to 100.64.6.0/24 and 100.64.7.0/24 is now via Router F
 - Allows graceful transition of external best path from Router E to Router F

Service Provider use of Communities



Some examples of how Network Operators
make life easier for themselves

BGP Communities

- Communities are generally set at the edge of the service provider network
 - **Customer edge:** customer prefixes belong to different communities depending on the services they have purchased
 - **Internet edge:** transit provider prefixes belong to different communities, depending on the loadsharing or traffic engineering requirements of the local Service Provider, or what the demands from its BGP customers might be
- Two simple examples follow to explain the concept

Community Example: Customer Edge

- Service Providers tag prefixes learned from their BGP and static customers with communities
 - To identify services the customer may have purchased
 - To identify prefixes which are part of the Provider's own address space
 - To identify customer independent address space
 - To control prefix distribution in IBGP
 - To control prefix announcements to customers and upstreams
 - (amongst several other reasons)

Community Example: Customer Edge

- No need to alter filters at the network border when adding a new customer
- New customer simply is added to the appropriate community
 - Border filters already in place take care of announcements
 - ⇒ Ease of operation!

Community Example: Customer Edge

Community Value	Description
X:1000	Aggregates
X:1001	Subprefixes of AS X aggregates
X:1005	Static (non-BGP) customer Provider Independent addresses
X:2000	BGP customers who get Transit
X:2100	BGP customers announced to Private Peers
X:2200	BGP customers announced at IXPs
X:2500	BGP customers announced to other BGP customers
X:3100	Routes received from Private Peers
X:3200	Routes received from IXP peers

Community Example: Internet Edge

- This demonstrates how communities might be used at the peering edge of a service provider network
- Service Provider has four types of BGP peers:
 - Customer
 - IXP peer
 - Private peer
 - Transit provider
- The prefixes received from each can be classified using communities
- Customers can opt to receive any or all of the above

Community Example: Internet Edge

- Referring to Customer Edge assignment table:
 - BGP customer who buys local connectivity gets X:2500
 - BGP customer who buys local and IXP connectivity receives community X:2500 and X:3200
 - BGP customer who buys full peer connectivity receives community X:2500, X:3100, and X:3200
- Customer who wants “the Internet” gets everything
 - Gets default route originated by aggregation router
 - Or pays money to get the full BGP table! 😊

Community Example: Internet Edge

- No need to create customised filters when adding customers
 - Border router already sets communities
 - Installation engineers pick the appropriate community set when establishing the customer BGP session
 - ⇒ Ease of operation!
- Communities also available for customers to do traffic engineering with Network Operator's peers and upstreams
 - Common examples in the following table

Communities for EBGP

Community Value	Action	Description
X:80	<code>set local-preference 80</code>	Backup path
X:120	<code>set local-preference 120</code>	Primary path (over-ride BGP path selection default)
X:1	<code>set as-path prepend X</code>	Single prepend when announced to X's upstreams
X:2	<code>set as-path prepend X X</code>	Double prepend when announced to X's upstreams
X:3	<code>set as-path prepend X X X</code>	Triple prepend when announced to X's upstreams
X:666	<code>set ip next-hop 192.0.2.1</code>	Blackhole route – very useful for DoS attack mitigation (RFC7999)

Community Example – Summary

- ❑ Two examples of customer edge and internet edge can be combined to form a simple community solution for network operator prefix policy control
- ❑ More experienced operators tend to have more sophisticated options available
 - Advice is to start with the easy examples given, and then proceed onwards as experience is gained

Network Operator BGP Communities

- There are no recommended Network Operator BGP communities apart from
 - RFC1998
 - The well-known communities
 - www.iana.org/assignments/bgp-well-known-communities
- Efforts have been made to document from time to time
 - Collection of Network Operator communities at www.onesc.net/communities
 - NANOG Tutorial:
 - www.nanog.org/meetings/nanog40/presentations/BGPcommunities.pdf
- Network Operator policy is usually published
 - On the Operator's website
 - Referenced in the AS Object in the IRR

WHAT YOU CAN CONTROL

AS-PATH PREPENDS

Sprint allows customers to use AS-path prepending to adjust route preference on the network. Such prepending will be received and passed on properly without notifying Sprint of your change in announcements.

Additionally, Sprint will prepend AS1239 to eBGP sessions with certain autonomous systems depending on a received community. Currently, the following ASes are supported: 1668, 209, 2914, 3300, 3356, 3549, 3561, 4635, 701, 7018, 702 and 8220.

String Resulting AS Path to ASXXX

65000:XXX	Do not advertise to ASXXX
65001:XXX	1239 (default) ...
65002:XXX	1239 1239 ...
65003:XXX	1239 1239 1239 ...
65004:XXX	1239 1239 1239 1239 ...

Example: Sprint

String Resulting AS Path to ASXXX in Asia

65070:XXX	Do not advertise to ASXXX
65071:XXX	1239 (default) ...
65072:XXX	1239 1239 ...
65073:XXX	1239 1239 1239 ...
65074:XXX	1239 1239 1239 1239 ...

String Resulting AS Path to ASXXX in Europe

65050:XXX	Do not advertise to ASXXX
65051:XXX	1239 (default) ...
65052:XXX	1239 1239 ...
65053:XXX	1239 1239 1239 ...
65054:XXX	1239 1239 1239 1239 ...

More info at
https://www.sprint.net/index.php?p=policy_bgp

BGP customer communities

Customers wanting to alter local preference on their routes.

NTT BGP customers may choose to affect our local preference on their routes by marking their routes with the following communities. Our regions are listed [here](#).

Community	Local-pref	Description
(default)	120	customer
65520:nnnn	50	only within country origin <nnnn> (see country origin list below)
65530:nnnn	50	only within region <nnnn> (see region list below)
2914:435	50	only beyond the connected country
2914:436	50	only beyond the connected region
2914:450	96	customer fallback
2914:460	98	peer backup
2914:470	100	peer
2914:480	110	customer backup
2914:490	120	customer default
2914:666		blackhole

Example: NTT

Customers wanting to alter their route announcements to other customers.

NTT BGP customers may choose to prepend to all other NTT BGP customers with the following communities:

Community	Description
2914:411	prepends o/b to customer 1x
2914:412	prepends o/b to customer 2x
2914:413	prepends o/b to customer 3x

Customers wanting to alter their route announcements to peers.

NTT BGP customers may choose to prepend to all NTT peers with the following communities:

Community	Description
2914:421	prepends o/b to peer 1x
2914:422	prepends o/b to peer 2x
2914:423	prepends o/b to peer 3x
2914:429	do not advertise to any peer
2914:439	do not advertise to any peer outside region

Note: 2914 is the ASN prepend in all cases. If used, 654xx:nnn overrides 655xx:nnn and 2914:429, 655xx:nnn overrides the 2914:42x communities.

Create Routing Registry Objects

Using PGP Authentication

Frequently Asked Questions

UTC Conversion Chart

Get More Information

 [Product Collateral](#)

 [Case Studies](#)

 [White Papers](#)

 [Audio & Video](#)

Get Started

To find out which solutions will best benefit your business, contact one of our Account Managers.

 [Click Here to Get Connected](#)

 Call us at **+1 425-250-7051**

Stay Connected

 [Follow Us on Twitter](#)

 [Friend Us on Facebook](#)

 [Join Us on LinkedIn](#)

More info at

<http://www.us.ntt.net/support/policy/routing.cfm>

Example: Verizon Europe

```
aut-num:      AS702
descr:        Verizon Business EMEA - Commercial IP service provider in Europe
<snip>
remarks:      -----
              Verizon Business filters out inbound prefixes longer than /24.
              We also filter any networks within AS702:RS-INBOUND-FILTER.
              -----
              VzBi uses the following communities with its customers:
              702:80   Set Local Pref 80 within AS702
              702:120 Set Local Pref 120 within AS702
              702:20   Announce only to VzBi AS'es and VzBi customers
              702:30   Keep within Europe, don't announce to other VzBi AS's
              702:1    Prepend AS702 once at edges of VzBi to Peers
              702:2    Prepend AS702 twice at edges of VzBi to Peers
              702:3    Prepend AS702 thrice at edges of VzBi to Peers
              -----
              Advanced communities for customers
              702:7020 Do not announce to AS702 peers with a scope of
              National but advertise to Global Peers, European
              Peers and VzBi customers.
              702:7001 Prepend AS702 once at edges of VzBi to AS702
              peers with a scope of National.
              702:7002 Prepend AS702 twice at edges of VzBi to AS702
              peers with a scope of National.
              -----
              Additional details of the VzBi communities are located at:
              http://www.verizonbusiness.com/uk/customer/bgp/
              -----
<snip>
```

← And many more!

Example: Arelion

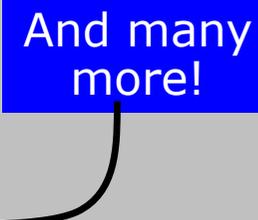
```
aut-num: AS1299
as-name: TWELVE99
descr: Arelion, f/k/a Telia Carrier
<snip>
remarks: BGP COMMUNITY SUPPORT FOR AS1299 TRANSIT CUSTOMERS:
remarks:
remarks: Community Action (default local pref 200)
remarks: -----
remarks: 1299:50 Set local pref 50 within AS1299 (lowest possible)
remarks: 1299:150 Set local pref 150 within AS1299 (equal to peer, backup)
remarks: 1299:1y050 Set local pref 50 in region y
remarks: 1299:1y150 Set local pref 150 in region y
remarks: Where y is:
remarks: 0= outside own continent
remarks: 2= Europe
remarks: 5= North America
remarks: 7= Asia Pacific
remarks:
remarks: European peers
remarks: Community Action
remarks: -----
remarks: 1299:200x All peers Europe incl:
remarks:
remarks: 1299:252x NTT/2914
remarks: 1299:253x Zayo/6461
remarks: 1299:254x Orange/5511
remarks: 1299:256x Lumen/3356
remarks: 1299:257x Verizon/702
remarks: 1299:258x AT&T/2686
remarks: 1299:259x Telxius/12956
<snip>
remarks: Where x is number of prepends (x=0,1,2,3) or do NOT announce (x=9)
```

And many
many more!

Example: BT Ignite

```
aut-num:      AS5400
descr:        BT Ignite European Backbone
<snip>
remarks:      The following BGP communities can be set by BT
remarks:      BGP customers to affect announcements to major peers.
remarks:
remarks:      5400:NXXX
remarks:      N=1      not announce
remarks:      N=2      prepend an extra "5400 5400" on announcement
remarks:      Valid values for XXX:
remarks:      000      All peers and transits
remarks:      500      All transits
remarks:      503      Level3 AS3356
remarks:      509      Telia AS1299
remarks:      510      NTT Verio AS2914
remarks:      002      Sprint AS1239
remarks:      003      Savvis AS3561
remarks:      004      C&W AS1273
remarks:      005      Verizon EMEA AS702
remarks:      014      DTAG AS3320
remarks:      016      Opentransit AS5511
remarks:      018      GlobeInternet Tata AS6453
remarks:      023      Tinet AS3257
remarks:      027      Telia AS1299
remarks:      045      Telecom Italia AS6762
remarks:      073      Eurorings AS286
remarks:      169      Cogent AS174
<snip>
```

And many
more!



Example: Level3

```
aut-num:      AS3356
descr:       Level 3 Communications
<snip>
remarks:     -----
remarks:     customer traffic engineering communities - Suppression
remarks:     -----
remarks:     64960:XXX - announce to AS XXX if 65000:0
remarks:     65000:0  - announce to customers but not to peers
remarks:     65000:XXX - do not announce at peerings to AS XXX
remarks:     -----
remarks:     customer traffic engineering communities - Prepending
remarks:     -----
remarks:     65001:0   - prepend once   to all peers
remarks:     65001:XXX - prepend once   at peerings to AS XXX
remarks:     65002:0   - prepend twice  to all peers
remarks:     65002:XXX - prepend twice  at peerings to AS XXX
<snip>
remarks:     -----
remarks:     customer traffic engineering communities - LocalPref
remarks:     -----
remarks:     3356:70   - set local preference to 70
remarks:     3356:80   - set local preference to 80
remarks:     3356:90   - set local preference to 90
remarks:     -----
remarks:     customer traffic engineering communities - Blackhole
remarks:     -----
remarks:     3356:9999 - blackhole (discard) traffic
<snip>
```

And many
more!



Creating your own community policy

- Consider creating communities to give policy control to customers
 - Reduces technical support burden
 - Reduces the amount of router reconfiguration, and the chance of mistakes
 - Use previous Network Operator and configuration examples as a guideline

BGP Communities

- There are no “standard communities” for Network Operators
- Best practices today consider that Network Operators should use BGP communities extensively for:
 - Scaling IBGP
 - Multihoming support of traffic engineering
- Look in the Network Operator AS Object in the IRR or on their website for documented community support

BGP Techniques for Network Operators

- ❑ BGP Basics
- ❑ Scaling BGP
- ❑ Using Communities
- ❑ **Deploying BGP in a Service Provider Network**

Deploying BGP in a Service Provider Network

Okay, so we've learned all about BGP
now; how do we use it on our network??



Deploying BGP

- ❑ The role of IGPs and iBGP
- ❑ Aggregation
- ❑ Receiving Prefixes
- ❑ Configuration Tips

The role of IBGP and IGP

Ships in the night?

Or

Good foundations?

BGP versus OSPF/ISIS

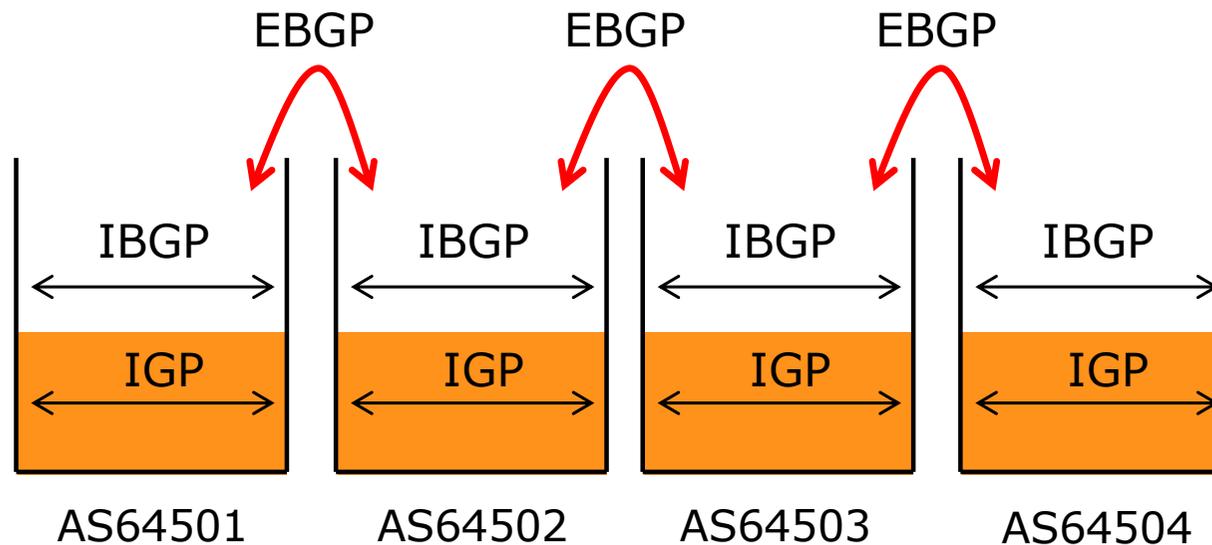
- Internal Routing Protocols (IGPs)
 - Examples are IS-IS and OSPF
 - Used for carrying **infrastructure** addresses
 - NOT used for carrying Internet prefixes or customer prefixes
 - Design goal is to **minimise** number of prefixes in IGP to aid **scalability** and **rapid convergence**

BGP versus OSPF/IS-IS

- BGP is used
 - Internally (IBGP)
 - Externally (EBGP)
- IBGP is used to carry:
 - Some/all Internet prefixes across backbone
 - Customer prefixes
- EBGP is used to:
 - Exchange prefixes with other ASes
 - Implement routing policy

BGP/IGP model used in Service Provider networks

□ Model representation



BGP versus OSPF/IS-IS

- DO NOT:
 - Distribute BGP prefixes into an IGP
 - Distribute IGP routes into BGP
 - Use an IGP to carry customer prefixes
- **YOUR NETWORK WILL NOT SCALE**

Injecting prefixes into IBGP

- Use IBGP to carry customer prefixes
 - Don't ever use IGP
- Point static route to customer interface
- Enter network into BGP process
 - Ensure that implementation options are used so that the prefix always remains in IBGP, regardless of state of interface
 - i.e. avoid IBGP flaps caused by interface flaps

Aggregation



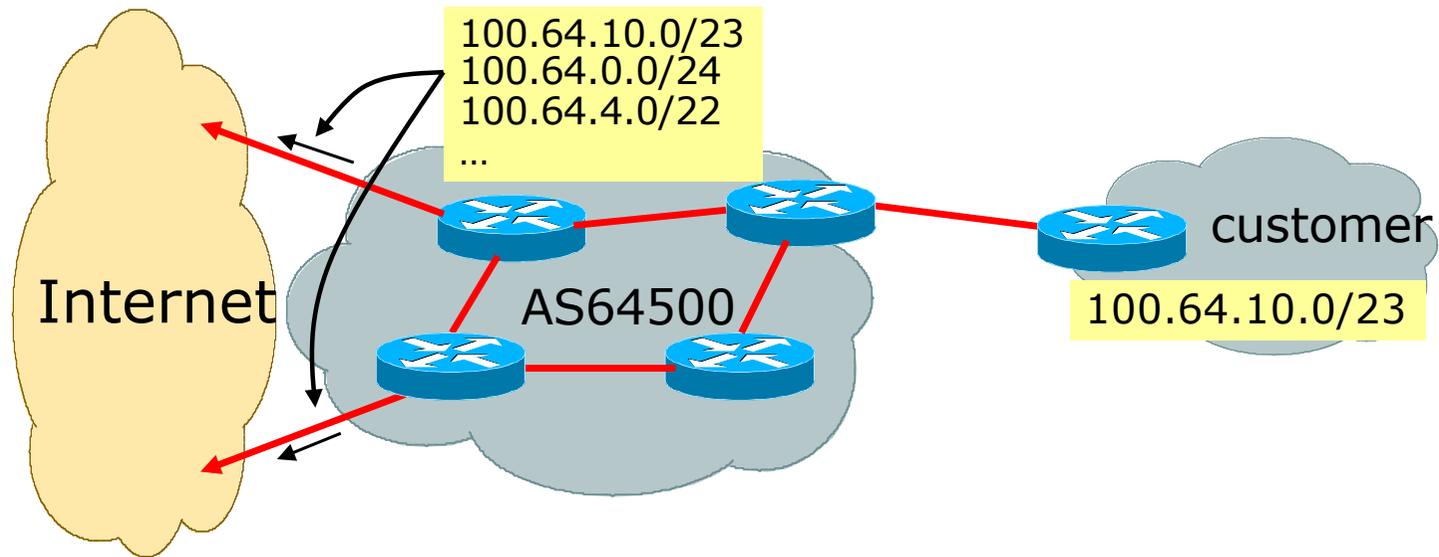
Aggregation

- ❑ Aggregation means announcing the address block received from the RIR to the other ASes connected to your network
- ❑ Subprefixes of this aggregate may be:
 - Used internally in the provider network
 - Announced to other ASes to aid with multihoming
- ❑ Too many operators are still thinking about class Cs, resulting in a proliferation of /24s in the Internet routing table
 - June 2023: 555542 /24s in IPv4 table of 925195 prefixes
- ❑ **The same is happening for /48s with IPv6**
 - June 2023: 85690 /48s in IPv6 table of 174464 prefixes

Announcing an Aggregate

- Network Operators who don't and won't aggregate are held in poor regard by community
- Registries publish their minimum allocation size
 - For IPv4:
 - /24
 - For IPv6:
 - /48 for assignment, /32 for allocation
- Until 2010, there was no real reason to see anything longer than a /22 IPv4 prefix on the Internet. But now?
 - IPv4 run-out is having an impact
 - It is expected that eventually the global IPv4 table will be mostly /24s

Aggregation – Example

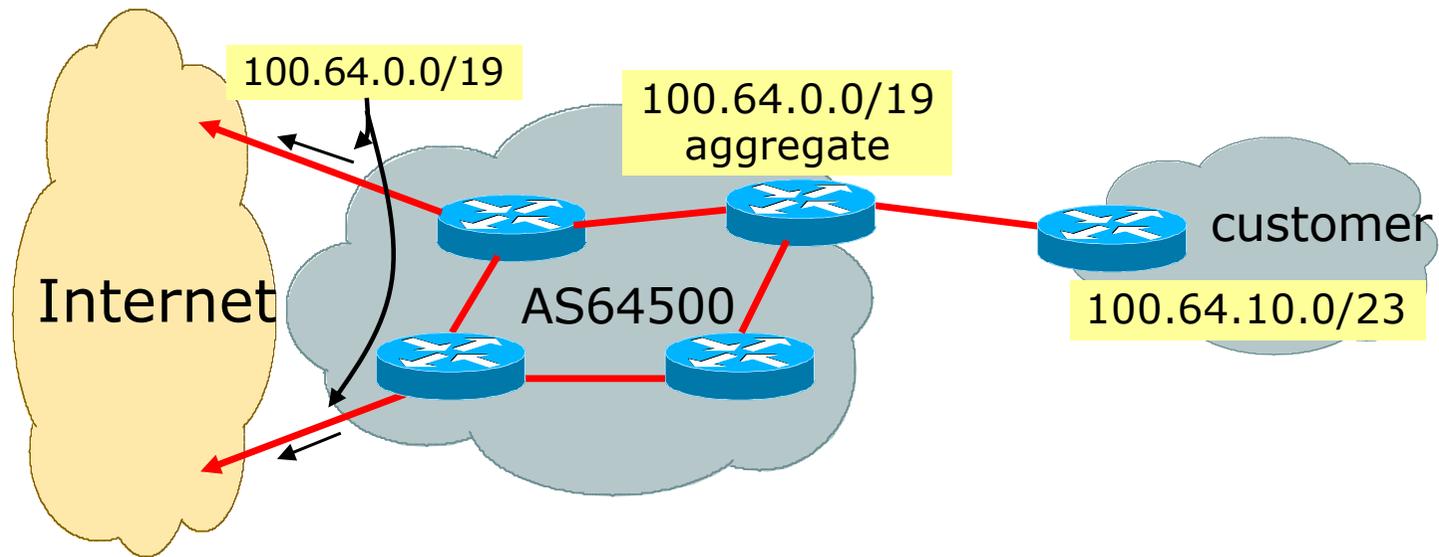


- ❑ Customer has /23 network assigned from AS64500's /19 address block
- ❑ AS64500 announces customers' individual networks to the Internet

Aggregation – Bad Example

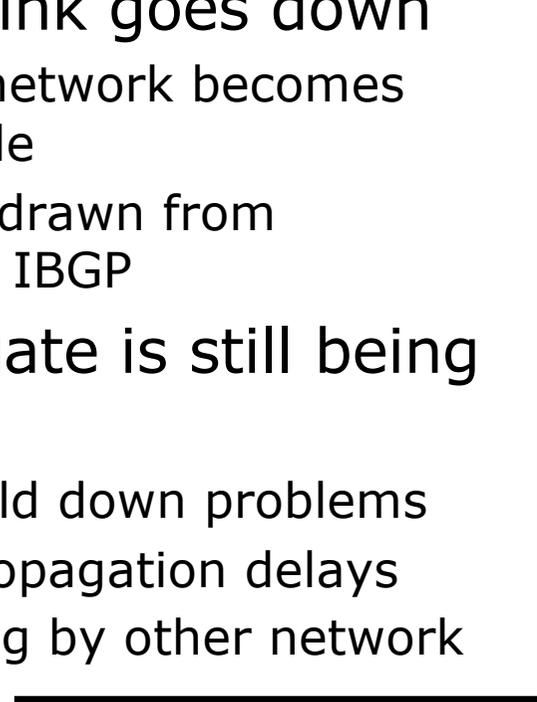
- Customer link goes down
 - Their /23 network becomes unreachable
 - /23 is withdrawn from AS64500's IBGP
- Their Service Provider doesn't aggregate its /19 network block
 - /23 network withdrawal announced to peers
 - Starts rippling through the Internet
 - Added load on all Internet backbone routers as network is removed from routing table
- Customer link returns
 - Their /23 network is now visible to their Service Provider
 - Their /23 network is re-advertised to peers
 - Starts rippling through Internet
 - Load on Internet backbone routers as network is reinserted into routing table
 - Some Network Operators suppress the flaps
 - Internet may take 10-20 min or longer to be visible
 - Where is the Quality of Service???

Aggregation – Example



- ❑ Customer has /23 network assigned from AS64500's /19 address block
- ❑ AS64500 announced /19 aggregate to the Internet

Aggregation – Good Example

- Customer link goes down
 - Their /23 network becomes unreachable
 - /23 is withdrawn from AS64500's IBGP
 - /19 aggregate is still being announced
 - No BGP hold down problems
 - No BGP propagation delays
 - No damping by other network operators
- 
- Customer link returns
 - Their /23 network is visible again
 - The /23 is re-injected into AS64500's IBGP
 - The whole Internet becomes visible immediately
 - Customer has Quality of Service perception

Aggregation – Summary

- Good example is what everyone should do!
 - Adds to Internet stability
 - Reduces size of routing table
 - Reduces routing churn
 - Improves Internet QoS for **everyone**
- Bad example is what too many still do!
 - Why? Lack of knowledge?
 - Laziness?

Separation of IBGP and EBGP

- Many Network Operators do not understand the importance of separating IBGP and EBGP
 - IBGP is where all customer prefixes are carried
 - EBGP is used for announcing aggregate to Internet and for Traffic Engineering
- Do **NOT** do traffic engineering with customer originated IBGP prefixes
 - Leads to instability similar to that mentioned in the earlier bad example
 - Even though aggregate is announced, a flapping subprefix will lead to instability for the customer concerned
- **Generate traffic engineering prefixes on the Border Router**

The Internet Today

(June 2023)

□ Current IPv4 Internet Routing Table Statistics

BGP Routing Table Entries	925195
Prefixes after maximum aggregation	350380
Unique prefixes in Internet	450399
/24s announced	555542
ASNs in use	74484

- (maximum aggregation is calculated by Origin AS)
- (unique prefixes > max aggregation means that operators are announcing prefixes from their blocks without a covering aggregate)

The Internet Today

(June 2023)

□ Current IPv6 Internet Routing Table Statistics

BGP Routing Table Entries	174464
/48s announced	85690
ASNs in use	30968

Efforts to improve aggregation

□ The CIDR Report

- Initiated and operated for many years by Tony Bates
- Now combined with Geoff Huston's routing analysis
 - www.cidr-report.org
 - (covers both IPv4 and IPv6 BGP tables)
- Results e-mailed on a weekly basis to most operations lists around the world
- Lists the top 30 service providers who could do better at aggregating

□ RIPE Routing WG aggregation recommendations

- IPv4: RIPE-399 — www.ripe.net/ripe/docs/ripe-399.html
- IPv6: RIPE-532 — www.ripe.net/ripe/docs/ripe-532.html

Efforts to Improve Aggregation

The CIDR Report

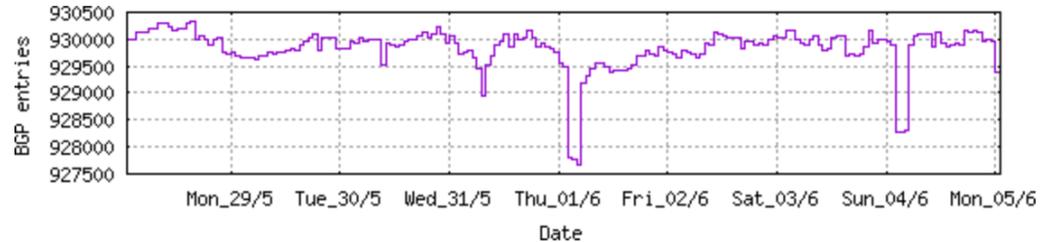
- Also computes the size of the routing table assuming network operators performed optimal aggregation
- Website allows searches and computations of aggregation to be made on a per AS basis
 - Flexible and powerful tool to aid Network Operators
 - Intended to show how greater efficiency in terms of BGP table size can be obtained without loss of routing and policy information
 - Shows what forms of origin AS aggregation could be performed and the potential benefit of such actions to the total table size
 - Very effectively challenges the traffic engineering excuse

Status Summary

Table History

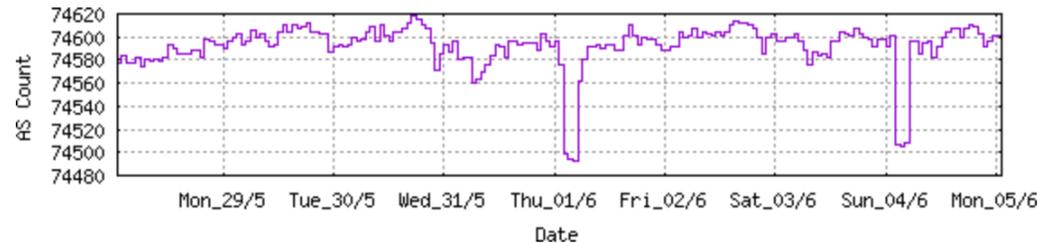
Date	Prefixes	CIDR Aggregated
29-05-23	929723	515436
30-05-23	929816	515170
31-05-23	929921	515773
01-06-23	929771	515579
02-06-23	929783	516013
03-06-23	930046	515775
04-06-23	929978	516038
05-06-23	929975	515775

Plot: [BGP Table Size](#)



AS Summary

74591	Number of ASes in routing system
26450	Number of ASes announcing only one prefix
11368	Largest number of prefixes announced by an AS AS8151 : Uninet S.A. de C.V., MX
228722688	Largest address span announced by an AS (/32s) AS749 : DNIC-AS-00749, US



Plot: [AS count](#)

Plot: [Average announcements per origin AS](#)

Report: [ASes ordered by originating address span](#)

Report: [ASes ordered by transit address span](#)

Report: [Autonomous System number-to-name mapping \(from Registry WHOIS data\)](#)

Announced Prefixes

Rank	AS	Type	Originate	Addr Space (pfx)	Transit	Addr space (pfx)	Description
84	AS6389		ORG+TRN Originate:	8674816 /8.95	Transit:	54528 /16.27	BELLSOUTH-NET-BLK, US

Long term deaggregator
– BellSouth in the US

Aggregation Suggestions

Filter: [Aggregates](#), [Specifics](#)

This report does not take into account conditions local to each origin AS in terms of policy or traffic engineering requirements, so this is an approximate guideline as to aggregation

Rank	AS	AS Name	Current	Wthdw	Aggte	Annce	Redctn	%
92	AS6389	BELLSOUTH-NET-BLK, US	662	558	7	111	551	83.23%

Prefix	AS Path	Aggregation Suggestion
12.81.120.0/24	4777 2516 1299 7018 6389	
12.130.209.0/24	4777 2516 1299 7018 6389 6389 6389 6389	
65.5.0.0/16	4777 2516 1299 7018 6389	
65.5.64.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.5.118.0/23	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.5.160.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.5.164.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.5.172.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.5.200.0/21	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.5.228.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.5.232.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.5.236.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.5.240.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.5.244.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.5.248.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.5.252.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.5.0.0/16 4777 2516 1299 7018 6389
65.6.0.0/15	4777 2516 1299 7018 6389	
65.6.192.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.6.0.0/15 4777 2516 1299 7018 6389
65.6.196.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.6.0.0/15 4777 2516 1299 7018 6389
65.7.64.0/18	4777 2516 1299 7018 6389	
65.12.0.0/14	4777 2516 1299 7018 6389	
65.12.32.0/20	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.12.0.0/14 4777 2516 1299 7018 6389
65.13.84.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.12.0.0/14 4777 2516 1299 7018 6389
65.13.92.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.12.0.0/14 4777 2516 1299 7018 6389
65.13.120.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.12.0.0/14 4777 2516 1299 7018 6389
65.13.124.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.12.0.0/14 4777 2516 1299 7018 6389
65.13.136.0/22	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.12.0.0/14 4777 2516 1299 7018 6389
65.13.176.0/21	4777 2516 1299 7018 6389	- Withdrawn - matching aggregate 65.12.0.0/14 4777 2516 1299 7018 6389

Announced Prefixes

Rank	AS	Type	Originate	Addr Space (pfx)	Transit	Addr space (pfx)	Description
303	AS18566	ORIGIN	Originate:	1710592	/11.29	Transit:	0 /0.00 MEGAPATH5-, US

Aggregation Suggestions

Filter: [Aggregates](#), [Specifics](#)

Long term deaggregator
– Megapath in the US

This report does not take into account conditions local to each origin AS in terms of policy or traffic engineering requirements, so this is an approximate guideline as to aggregation

Rank	AS	AS Name	Current	Wthdw	Aggte	Annce	Redctn	%
34	AS18566	MEGAPATH5-, US	1852	1467	267	652	1200	64.79%

Prefix	AS Path	Aggregation Suggestion
64.6.160.0/23	4777 2516 3257 18566	
64.6.164.0/22	4777 2516 3257 18566	+ Announce - aggregate of 64.6.164.0/23 (4777 2516 3257 18566) and 64.6.166.0/23 (4777 2516 3257 18566)
64.6.164.0/23	4777 2516 3257 18566	- Withdrawn - aggregated with 64.6.166.0/23 (4777 2516 3257 18566)
64.6.166.0/24	4777 2516 3257 18566	- Withdrawn - aggregated with 64.6.167.0/24 (4777 2516 3257 18566)
64.6.167.0/24	4777 2516 3257 18566	- Withdrawn - aggregated with 64.6.166.0/24 (4777 2516 3257 18566)
64.50.206.0/23	4777 2516 3257 18566	
64.51.126.0/23	4777 2516 3257 18566	
64.81.16.0/20	4777 2516 3257 18566	+ Announce - aggregate of 64.81.16.0/21 (4777 2516 3257 18566) and 64.81.24.0/21 (4777 2516 3257 18566)
64.81.16.0/22	4777 2516 3257 18566	- Withdrawn - aggregated with 64.81.20.0/22 (4777 2516 3257 18566)
64.81.20.0/22	4777 2516 3257 18566	- Withdrawn - aggregated with 64.81.16.0/22 (4777 2516 3257 18566)
64.81.22.0/24	4777 2516 3257 18566	- Withdrawn - matching aggregate 64.81.20.0/22 4777 2516 3257 18566
64.81.24.0/22	4777 2516 3257 18566	- Withdrawn - aggregated with 64.81.28.0/22 (4777 2516 3257 18566)
64.81.28.0/22	4777 2516 3257 18566	- Withdrawn - aggregated with 64.81.24.0/22 (4777 2516 3257 18566)
64.81.32.0/19	4777 2516 3257 18566	+ Announce - aggregate of 64.81.32.0/20 (4777 2516 3257 18566) and 64.81.48.0/20 (4777 2516 3257 18566)
64.81.32.0/20	4777 2516 3257 18566	- Withdrawn - aggregated with 64.81.48.0/20 (4777 2516 3257 18566)
64.81.32.0/24	4777 2516 3257 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 3257 18566
64.81.33.0/24	4777 2516 3257 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 3257 18566
64.81.34.0/24	4777 2516 3257 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 3257 18566
64.81.39.0/24	4777 2516 3257 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 3257 18566
64.81.40.0/24	4777 2516 3257 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 3257 18566
64.81.44.0/24	4777 2516 3257 18566	- Withdrawn - matching aggregate 64.81.32.0/20 4777 2516 3257 18566
64.81.48.0/20	4777 2516 3257 18566	- Withdrawn - aggregated with 64.81.32.0/20 (4777 2516 3257 18566)
64.81.48.0/24	4777 2516 3257 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3257 18566
64.81.50.0/24	4777 2516 3257 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3257 18566
64.81.53.0/24	4777 2516 3257 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3257 18566
64.81.57.0/24	4777 2516 3257 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3257 18566
64.81.60.0/24	4777 2516 3257 18566	- Withdrawn - matching aggregate 64.81.48.0/20 4777 2516 3257 18566
64.81.64.0/19	4777 2516 3257 18566	+ Announce - aggregate of 64.81.64.0/20 (4777 2516 3257 18566) and 64.81.80.0/20 (4777 2516 3257 18566)

Announced Prefixes

Rank	AS	Type	Originate	Addr Space (pfx)	Transit	Addr space (pfx)	Description
145	AS7545		ORG+TRN Originate:	5190656 /9.69	Transit:	7385344 /9.18	TPG-INTERNET-AP TPG Telecom Limited, AU

Aggregation Suggestions

Filter: [Aggregates](#), [Specifics](#)

Long term deaggregator
– TPG in Australia

This report does not take into account conditions local to each origin AS in terms of policy or traffic engineering requirements, so this is an approximate guideline as to aggregation p

Rank	AS	AS Name	Current	Wthdw	Aggte	Annce	Redctn	%
4	AS7545	TPG-INTERNET-AP TPG Telecom Limited, AU	5723	5174	162	711	5012	87.58%

Prefix	AS Path	Aggregation Suggestion
14.2.0.0/17	4608 7575 7545	+ Announce - aggregate of 14.2.0.0/18 (4608 7575 7545) and 14.2.64.0/18 (4608 7575 7545)
14.2.0.0/19	4608 7575 7545	- Withdrawn - aggregated with 14.2.32.0/19 (4608 7575 7545)
14.2.32.0/19	4608 7575 7545	- Withdrawn - aggregated with 14.2.0.0/19 (4608 7575 7545)
14.2.32.0/21	4608 7575 7545	- Withdrawn - matching aggregate 14.2.32.0/19 4608 7575 7545
14.2.40.0/21	4608 7575 7545	- Withdrawn - matching aggregate 14.2.32.0/19 4608 7575 7545
14.2.48.0/21	4608 7575 7545	- Withdrawn - matching aggregate 14.2.32.0/19 4608 7575 7545
14.2.56.0/21	4608 7575 7545	- Withdrawn - matching aggregate 14.2.32.0/19 4608 7575 7545
14.2.64.0/19	4608 7575 7545	- Withdrawn - aggregated with 14.2.96.0/19 (4608 7575 7545)
14.2.96.0/19	4608 7575 7545	- Withdrawn - aggregated with 14.2.64.0/19 (4608 7575 7545)
14.2.128.0/18	4608 7575 7545	
14.2.192.0/20	4608 7575 7545	
14.200.0.0/14	4608 7575 7545	
14.200.1.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.2.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.4.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.5.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.6.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.7.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.8.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.10.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.11.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.12.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.13.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.14.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.15.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.16.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.17.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545
14.200.18.0/24	4608 7575 7545	- Withdrawn - matching aggregate 14.200.0.0/14 4608 7575 7545

Announced Prefixes

```
Rank AS      Type   Originate Addr Space (pfx)  Transit Addr space (pfx)  Description
58   AS12479   ORG+TRN Originate: 13940992 /8.27   Transit: 340480 /13.62  UNI2-AS, ES
```

Long term deaggregator
– Orange in Spain

Aggregation Suggestions

Filter: [Aggregates](#), [Specifics](#)

This report does not take into account conditions local to each origin AS in terms of policy or traffic engineering requirements, so this is an approximate guideline as to aggregation p

```
Rank AS      AS Name      Current Wthdw Aggte Annce Redctn  %
12  AS12479   UNI2-AS, ES  7301  4029  936  4208  3093  42.36%
```

Prefix	AS Path	Aggregation Suggestion
1.178.224.0/19	4608 1221 4637 5511 12479	
1.178.224.0/21	4777 2516 1299 5511 12479	
1.178.232.0/21	4608 1221 4637 5511 12479	- Withdrawn - matching aggregate 1.178.224.0/19 4608 1221 4637 5511 12479
1.178.240.0/20	4777 2516 1299 5511 12479	+ Announce - aggregate of 1.178.240.0/21 (4777 2516 1299 5511 12479) and 1.178.248.0/21 (4777 2516 1
1.178.240.0/21	4777 2516 1299 5511 12479	- Withdrawn - aggregated with 1.178.248.0/21 (4777 2516 1299 5511 12479)
1.178.248.0/21	4777 2516 1299 5511 12479	- Withdrawn - aggregated with 1.178.240.0/21 (4777 2516 1299 5511 12479)
37.11.0.0/16	4608 7575 6461 5511 12479	
37.11.0.0/21	4777 2516 1299 5511 12479	+ Announce - aggregate of 37.11.0.0/22 (4777 2516 1299 5511 12479) and 37.11.4.0/22 (4777 2516 1299
37.11.0.0/22	4777 2516 1299 5511 12479	- Withdrawn - aggregated with 37.11.4.0/22 (4777 2516 1299 5511 12479)
37.11.4.0/22	4777 2516 1299 5511 12479	- Withdrawn - aggregated with 37.11.0.0/22 (4777 2516 1299 5511 12479)
37.11.8.0/22	4608 7575 6461 5511 12479	- Withdrawn - matching aggregate 37.11.0.0/16 4608 7575 6461 5511 12479
37.11.12.0/22	4777 2516 1299 5511 12479	
37.11.16.0/22	4608 1221 4637 5511 12479	
37.11.20.0/22	4608 7575 6461 5511 12479	- Withdrawn - matching aggregate 37.11.0.0/16 4608 7575 6461 5511 12479
37.11.24.0/21	4777 2516 1299 5511 12479	+ Announce - aggregate of 37.11.24.0/22 (4777 2516 1299 5511 12479) and 37.11.28.0/22 (4777 2516 129
37.11.24.0/22	4777 2516 1299 5511 12479	- Withdrawn - aggregated with 37.11.28.0/22 (4777 2516 1299 5511 12479)
37.11.28.0/22	4777 2516 1299 5511 12479	- Withdrawn - aggregated with 37.11.24.0/22 (4777 2516 1299 5511 12479)
37.11.32.0/22	4608 7575 6461 5511 12479	- Withdrawn - matching aggregate 37.11.0.0/16 4608 7575 6461 5511 12479
37.11.36.0/22	4608 7575 6461 5511 12479	- Withdrawn - matching aggregate 37.11.0.0/16 4608 7575 6461 5511 12479
37.11.40.0/21	4777 2516 1299 5511 12479	+ Announce - aggregate of 37.11.40.0/22 (4777 2516 1299 5511 12479) and 37.11.44.0/22 (4777 2516 129
37.11.40.0/22	4777 2516 1299 5511 12479	- Withdrawn - aggregated with 37.11.44.0/22 (4777 2516 1299 5511 12479)
37.11.44.0/22	4777 2516 1299 5511 12479	- Withdrawn - aggregated with 37.11.40.0/22 (4777 2516 1299 5511 12479)
37.11.48.0/22	4777 2516 1299 5511 12479	
37.11.52.0/22	4608 7575 6461 5511 12479	- Withdrawn - matching aggregate 37.11.0.0/16 4608 7575 6461 5511 12479
37.11.56.0/23	4608 7575 6461 5511 12479	- Withdrawn - matching aggregate 37.11.0.0/16 4608 7575 6461 5511 12479
37.11.58.0/23	4608 7575 6461 5511 12479	- Withdrawn - matching aggregate 37.11.0.0/16 4608 7575 6461 5511 12479
37.11.60.0/22	4608 7575 6461 5511 12479	- Withdrawn - matching aggregate 37.11.0.0/16 4608 7575 6461 5511 12479
37.11.64.0/22	4777 2516 1299 5511 12479	

Importance of Aggregation

- Size of routing table
 - Router Memory is not so much of a problem as it was in the 1990s
 - Routers routinely carry over 2 million prefixes
- Convergence of the Routing System
 - This is a problem
 - Bigger table takes longer for CPU to process
 - BGP updates take longer to deal with
 - BGP Instability Report tracks routing system update activity
 - bgpupdates.potaroo.net/instability/bgpupd.html

The BGP Instability Report

The BGP Instability Report is updated daily. This report was generated on 04 June 2023 06:20 (UTC+1000)

50 Most active ASes for the past 14 days

RANK	ASN	UPDs	% Prefixes	UPDs/Prefix	AS NAME	
1	9381	330481	3.72%	535	617.72	HKBNES-AS-AP HKBN Enterprise Solutions HK Limited, HK
2	16509	245727	2.77%	8586	28.62	AMAZON-02, US
3	8151	143443	1.61%	11448	12.53	Uninet S.A. de C.V., MX
4	399077	102133	1.15%	539	189.49	TERAEXCH, US
5	9829	87139	0.98%	1924	45.29	BSNL-NIB National Internet Backbone, IN
6	11830	67801	0.76%	2273	29.83	Instituto Costarricense de Electricidad y Telecom., CR
7	7552	63105	0.71%	3618	17.44	VIETEL-AS-AP Viettel Group, VN
8	55577	61026	0.69%	91	670.62	BEAMTELE-AS-AP Atria Convergence Technologies pvt ltd, IN
9	647	60136	0.68%	370	162.53	DNIC-ASBLK-00616-00665, US
10	7713	56448	0.64%	3574	15.79	TELKOMNET-AS-AP PT Telekomunikasi Indonesia, ID
11	18403	52314	0.59%	4252	12.30	FPT-AS-AP FPT Telecom Company, VN
12	36903	51710	0.58%	1242	41.63	MT-MPLS, MA
13	13977	51117	0.58%	887	57.63	CTELCO, US
14	58224	50221	0.57%	2150	23.36	TCI, IR
15	23666	50004	0.56%	49	1020.49	SISTELINDO-AS-ID PT Sistelindo Mitralintas, ID
16	21575	48509	0.55%	169	287.04	ENTEL PERU S.A., PE
17	10620	45595	0.51%	3524	12.94	Telmex Colombia S.A., CO
18	203214	38504	0.43%	997	38.62	HULUMTELE, IQ
19	18209	36561	0.41%	129	283.42	BEAMTELE-AS-AP Atria Convergence Technologies pvt ltd, IN
20	38322	34913	0.39%	86	405.97	TTSSB-MY TM TECHNOLOGY SERVICES SDN. BHD., MY
21	53135	34168	0.38%	89	383.91	BRASIL TECNOLOGIA E PARTICIPACOES SA, BR
22	265566	33106	0.37%	4	8276.50	TELESISTEMAS PENINSULARES SA DE CV, MX
23	266407	31818	0.36%	1	31818.00	ASSEMBLEIA LEGISLATIVA DO ESTADO DE GOIAS, BR
24	6364	31405	0.35%	48	654.27	ATLANTIC-NET-1, US
25	8007	30434	0.34%	32	951.06	Servicios Administrados Mexis, S.A. de C.V., MX
26	28135	30385	0.34%	66	460.38	ASSOCIACAO NACIONAL PARA INCLUSAO DIGITAL - ANID, BR

50 Most active Prefixes for the past 14 days

RANK	PREFIX	UPDs	%	Origin AS -- AS NAME
1	45.172.92.0/22	33003	0.36%	265566 -- TELESISTEMAS PENINSULARES SA DE CV, MX
2	170.238.225.0/24	31818	0.34%	266407 -- ASSEMBLEIA LEGISLATIVA DO ESTADO DE GOIAS, BR
3	69.28.87.0/24	31137	0.34%	6364 -- ATLANTIC-NET-1, US
4	207.249.70.0/24	30374	0.33%	8007 -- Servicios Administrados Mexis, S.A. de C.V., MX
5	177.46.7.0/24	27701	0.30%	28135 -- ASSOCIACAO NACIONAL PARA INCLUSAO DIGITAL - ANID, BR
6	67.211.53.0/24	24031	0.26%	26405 -- HDCS, US
7	142.67.31.0/24	23118	0.25%	22636 -- NOVA-SCOTIA-POWER, CA
8	12.129.82.0/24	22679	0.25%	40631 -- MASTEC-PRIVATE, US
9	185.96.215.0/24	22115	0.24%	200770 -- REQUISIS-AS requisis.com, DE
10	103.248.132.0/22	21689	0.23%	132829 -- NCPL-AS-AP Navigate Communications S Pte Ltd, SG
11	186.219.224.0/22	21686	0.23%	270598 -- FUSION INTERNET WIFI LTDA, BR
12	45.183.178.0/24	18345	0.20%	270142 -- Castor Networks Mexico s de rl, MX
13	195.13.50.0/24	15496	0.17%	58319 -- KAZAKOV-AS, RU
14	103.223.2.0/24	13983	0.15%	135445 -- IDNIC-AIRPAY-AS-ID PT. Airpay International Indonesia, ID
15	209.22.66.0/24	13503	0.15%	2046 -- DNIC-AS-02046, US
16	209.22.67.0/24	13495	0.15%	2046 -- DNIC-AS-02046, US
17	41.209.0.0/18	12628	0.14%	9129 -- KE-NET2000, ZA
18	193.232.56.0/24	12468	0.13%	50276 -- MUCTR, RU
19	193.232.57.0/24	12447	0.13%	50276 -- MUCTR, RU
20	27.131.0.0/24	12407	0.13%	38753 -- SOLUSINDO-AS-ID Solusindo Bintang Pratama, PT, ID
21	187.251.64.0/19	12213	0.13%	18592 -- Corporacion Universitaria para el Desarrollo de Internet, A.C., MX
22	200.23.60.0/24	12186	0.13%	18592 -- Corporacion Universitaria para el Desarrollo de Internet, A.C., MX
23	181.225.48.0/24	11952	0.13%	8053 -- IFX Networks Venezuela C.A., VE
24	181.225.43.0/24	11688	0.13%	8053 -- IFX Networks Venezuela C.A., VE
25	85.187.16.0/24	11629	0.13%	203182 -- DIGITURK-, BG
26	135.84.68.0/24	10297	0.11%	17289 -- USG-PROD-WEST01, US 397133 -- USG-NPE-WEST01, US
27	185.18.201.0/24	9559	0.10%	47855 -- PRIME Moscow branch, RU
28	207.244.192.0/22	9467	0.10%	6428 -- CDM, US
29	130.137.108.0/24	8692	0.09%	16509 -- AMAZON-02, US
30	130.137.227.0/24	8461	0.09%	16509 -- AMAZON-02, US
31	130.137.111.0/24	8454	0.09%	16509 -- AMAZON-02, US

The BGP IPv6 Instability Report

This report is updated daily. The current report was generated on 5 June 2023 01:10 (UTC+1000)

50 Most active ASes for the past 14 days

RANK	ASN	UPDs	%	Prefixes	UPDs/Prefix	AS NAME
1	202256	265579	7.08%	2886	92.02	LAWLIETNET, CN
2	7602	190326	5.07%	12	15860.50	SPT-AS-VN Sai gon Postel Corporation, VN
3	38266	186776	4.98%	1301	143.56	VIL-AS-AP Vodafone Idea Ltd, IN
4	18403	183905	4.90%	4314	42.63	FPT-AS-AP FPT Telecom Company, VN
5	11172	129918	3.46%	5517	23.55	Alestra, S. de R.L. de C.V., MX
6	20473	102960	2.75%	857	120.14	AS-CHOOA, US
7	272605	77974	2.08%	18	4331.89	G4 FIBRA, BR
8	53667	75847	2.02%	1929	39.32	PONYNET, US
9	28573	63515	1.69%	1899	33.45	Claro NXT Telecomunicacoes Ltda, BR
10	134326	58679	1.56%	16	3667.44	AIRDESIGNBROADCAST-AS Airdesign Broadcast Media Pvt Ltd, IN
11	32098	56566	1.51%	1361	41.56	TRANSTELCO-INC, US
12	9829	45274	1.21%	516	87.74	BSNL-NIB National Internet Backbone, IN
13	263499	39620	1.06%	3	13206.67	Telecom Conectividade, BR
14	212034	37377	1.00%	10	3737.70	sclrnet.com research net, CA
15	264525	36492	0.97%	23	1586.61	Coelho Tecnologia, BR
16	19180	35499	0.95%	1	35499.00	AMERICATEL PERU S.A., PE
17	205610	31697	0.85%	8	3962.12	JAMESITS Jamesits Networks, CN
18	61637	31312	0.83%	31	1010.06	RADIOBRAS TELECOM LTDA ME, BR
19	210384	29912	0.80%	9	3323.56	KUDOS-AS, TW
20	13780	28201	0.75%	10	2820.10	MILEHIGHWALRUS-NETWORK, US
21	48695	27985	0.75%	427	65.54	ATHEEB-AS, SA
22	7552	27452	0.73%	2202	12.47	VIETEL-AS-AP Viettel Group, VN
23	5385	26817	0.71%	4	6704.25	RUSSMEDIA-IT, AT
24	7545	26330	0.70%	1974	13.34	TPG-INTERNET-AP TPG Telecom Limited, AU
25	270326	22239	0.59%	9	2471.00	LUKAS S. FREITAS ME, BR
26	200952	21803	0.58%	7	3114.71	BARATA, PT
27	16509	20675	0.55%	3564	5.80	AMAZON-02, US

50 Most active Prefixes for the past 14 days

RANK	PREFIX	UPDs	%	Origin AS -- AS NAME
1	2001:1380:2000::/36	35499	0.87%	19180 -- AMERICATEL PERU S.A., PE
2	2a0a:2f40::/32	26806	0.65%	5385 -- RUSSMEDIA-IT, AT
3	2804:20fc:1b00::/48	19938	0.49%	264525 -- Coelho Tecnologia, BR
4	2804:12cc::/33	19798	0.48%	263499 -- Telecom Conectividade, BR
5	2804:12cc:8000::/33	19742	0.48%	263499 -- Telecom Conectividade, BR
6	2402:f800::/32	16709	0.41%	7602 -- SPT-AS-VN Sai gon Postel Corporation, VN
7	2a0b:4340:99::/48	15848	0.39%	205610 -- JAMESITS Jamesits Networks, CN
8	2a0b:4340:90::/48	15822	0.39%	205610 -- JAMESITS Jamesits Networks, CN
9	2402:f800:ff00::/40	15796	0.39%	7602 -- SPT-AS-VN Sai gon Postel Corporation, VN
10	2402:f800:cf00::/48	15796	0.39%	7602 -- SPT-AS-VN Sai gon Postel Corporation, VN
11	2402:f800:ef00::/40	15794	0.39%	7602 -- SPT-AS-VN Sai gon Postel Corporation, VN
12	2402:f800:cf00::/40	15788	0.39%	7602 -- SPT-AS-VN Sai gon Postel Corporation, VN
13	2402:f800::/40	15787	0.39%	7602 -- SPT-AS-VN Sai gon Postel Corporation, VN
14	2402:f800:df00::/40	15787	0.39%	7602 -- SPT-AS-VN Sai gon Postel Corporation, VN
15	2402:f800::/48	15782	0.39%	7602 -- SPT-AS-VN Sai gon Postel Corporation, VN
16	2402:f800:fff::/48	15781	0.39%	7602 -- SPT-AS-VN Sai gon Postel Corporation, VN
17	2402:f800:df00::/48	15779	0.39%	7602 -- SPT-AS-VN Sai gon Postel Corporation, VN
18	2402:f800:ef00::/48	15772	0.39%	7602 -- SPT-AS-VN Sai gon Postel Corporation, VN
19	2402:f800:ff00::/48	15755	0.38%	7602 -- SPT-AS-VN Sai gon Postel Corporation, VN
20	2400:58a0:1800::/44	15201	0.37%	150404 -- THUNDER-AS-AP Thunder Network Limited, HK
21	2a0a:6040:e6ff::/48	14959	0.37%	200882 -- SAHIL-S-NETWORK, US
22	2a06:a001:a108::/48	14936	0.36%	210384 -- KUDOS-AS, TW
23	2404:f4c0:8868::/48	14914	0.36%	210384 -- KUDOS-AS, TW
24	2804:2000:1fff::/48	14654	0.36%	264459 -- NET ALTERNATIVA PROVEDOR DE INTERNET LTDA - ME, BR
25	2a0a:6040:c742::/48	14402	0.35%	212164 -- HE-AS Kakan Network, CN
26	2a0a:afc7::/48	13882	0.34%	201196 -- BERMOND-AS, VG
27	2602:feda:d64::/48	13765	0.34%	212034 -- sclrnet.com research net, CA
28	2406:840:f895::/48	13394	0.33%	212034 -- sclrnet.com research net, CA
29	2a03:5a00:d::/48	12725	0.31%	58319 -- KAZAKOV-AS, RU
30	2804:7160::/32	12500	0.31%	270785 -- PRNET PROVEDOR DE INTERNET LTDA, BR
31	2a0e:fd45:bee::/48	12352	0.30%	139921 -- UNIXFY-AS Unixfy, SG

Aggregation: Summary

- Aggregation on the Internet could be **MUCH** better
 - 50% saving on Internet routing table size is quite feasible
 - Tools **are** available
 - Commands on the routers are not hard
 - CIDR-Report webpage

Receiving Prefixes



Receiving Prefixes

- There are three scenarios for receiving prefixes from other ASes
 - Customer talking BGP
 - Peer talking BGP
 - Upstream/Transit talking BGP
- Each has different filtering requirements and need to be considered separately

Receiving Prefixes: From Customers

- Network Operators must only accept prefixes which have been assigned or allocated to their downstream customer
- If Network Operator has assigned address space to its customer, then the customer IS entitled to announce it back to their Network Operator
- If the Network Operator has NOT assigned address space to its customer, then:
 - Check in the five RIR databases to see if this address space really has been assigned to the customer
 - The tool: `whois -h jwhois.apnic.net x.x.x.0/24`
 - (jwhois is "joint whois" and queries the 5 RIR databases)

Receiving Prefixes: From Customers

- Example use of whois to check if customer is entitled to announce address space:

```
$ whois -h jwhois.apnic.net 202.12.29.0
```

```
inetnum:      202.12.29.0 - 202.12.29.255
netname:      APNIC-SERVICES-AU
descr:        Asia Pacific Network Information Centre
descr:        Regional Internet Registry for the Asia-Pacific Region
descr:        6 Cordelia Street
descr:        South Brisbane
geoloc:       27.4731138 153.0141194
country:      AU
admin-c:      AIC1-AP
tech-c:       AIC1-AP
mnt-by:       APNIC-HM
mnt-irt:      IRT-APNIC-IS-AP
status:       ASSIGNED PORTABLE
changed:      hm-changed@apnic.net 20170327
changed:      hm-changed@apnic.net 20170331
source:       APNIC
```

inetnum – means it is an address delegation to an entity

Portable – means its an assignment to the customer, the customer can announce it to you

Receiving Prefixes: From Customers

- Example use of whois to check if customer is entitled to announce address space:

```
$ whois -h jwhois.apnic.net 194.15.141.0

inetnum:      194.15.141.0 - 194.15.141.255
netname:      INETTECH
country:      SE
org:          ORG-ITAS2-RIPE
admin-c:      KEL5-RIPE
tech-c:       KEL5-RIPE
status:       ASSIGNED PI
mnt-by:       RIPE-NCC-END-MNT
mnt-by:       KURTIS-PP-MNT
mnt-routes:   KURTIS-PP-MNT
mnt-domains:  KURTIS-PP-MNT
created:      2003-12-04T09:33:09Z
last-modified: 2016-04-14T08:21:55Z
source:       RIPE
sponsoring-org: ORG-NIE1-RIPE
```

inetnum – means it is an address delegation to an entity

Assigned PI – means its an assignment to the customer, the customer can announce it to you

Receiving Prefixes: From Customers

- Example use of whois to check if customer is entitled to announce address space:

```
$ whois -h jwhois.apnic.net 193.128.0.0/22
```

```
inetnum:      193.128.0.0 - 193.128.6.255
netname:      UK-PIPEX-19931014
country:      GB
org:          ORG-UA24-RIPE
admin-c:      WERT1-RIPE
tech-c:       UPHM1-RIPE
status:       ALLOCATED PA
remarks:      Please send abuse notification to abuse@uk.uu.net
mnt-by:       RIPE-NCC-HM-MNT
mnt-by:       AS1849-MNT
mnt-routes:   AS1849-MNT
mnt-routes:   WCOM-EMEA-RICE-MNT
mnt-irt:      IRT-MCI-GB
created:      2018-07-30T09:42:04Z
last-modified: 2018-07-30T09:42:04Z
source:       RIPE # Filtered
```

inetnum – means it is an address delegation to an entity

ALLOCATED – means that this is Provider Aggregatable address space and can only be announced by the provider holding the allocation (in this case Verizon UK)

Receiving Prefixes: From Peers

- A peer is a Network Operator with whom you agree to exchange prefixes you originate into the Internet routing table
 - Prefixes you accept from a peer are only those they have indicated they will announce
 - Prefixes you announce to your peer are only those you have indicated you will announce

Receiving Prefixes: From Peers

- Agreeing what each will announce to the other:
 - Exchange of e-mail documentation as part of the peering agreement, and then ongoing updates

OR

- Use of the Internet Routing Registry and configuration tools such as:
 - IRRToolSet:
<https://github.com/irrtoolset/irrtoolset>
 - bgpq4:
<https://github.com/bgp/bgpq4>

Receiving Prefixes: From Upstream/Transit Provider

- Upstream/Transit Provider is a Network Operator who you pay to give you transit to the **WHOLE** Internet
- Receiving prefixes from them is not desirable unless really necessary
 - Traffic Engineering – see BGP Multihoming presentations
- Ask upstream/transit provider to either:
 - originate a default-route
 - OR
 - announce one prefix you can use as default

Receiving Prefixes: From Upstream/Transit Provider

- If it is necessary to receive prefixes from any provider, care is required.
 - Don't accept default (unless you need it)
 - Don't accept your own prefixes
- Special use prefixes for IPv4 and IPv6:
 - <http://www.rfc-editor.org/rfc/rfc6890.txt>
- For IPv4:
 - Don't accept prefixes longer than /24 (?)
 - /24 was the historical class C
- For IPv6:
 - Don't accept prefixes longer than /48 (?)
 - /48 is the design minimum delegated to a site

Receiving Prefixes: From Upstream/Transit Provider

- Check Team Cymru's list of "bogons"
 - <http://www.team-cymru.com/bogon-reference-http>
- For IPv4 also consult:
 - <https://www.rfc-editor.org/rfc/rfc6441.txt> (BCP171)
- Bogon Route Server:
 - <https://www.team-cymru.com/bogon-reference-bgp>
 - Supplies a BGP feed (IPv4 and/or IPv6) of address blocks which should not appear in the BGP table

Receiving IPv4 Prefixes

```
deny 0.0.0.0/0                ! Default
deny 0.0.0.0/8 to /32         ! RFC1122 local host
deny 10.0.0.0/8 to /32       ! RFC1918
deny 100.64.0.0/10 to /32    ! RFC6598 shared address
deny 101.10.0.0/19 to /32    ! Local prefix
deny 127.0.0.0/8 to /32     ! Loopback
deny 169.254.0.0/16 to /32   ! Auto-config
deny 172.16.0.0/12 to /32    ! RFC1918
deny 192.0.0.0/24 to /32    ! RFC6598 IETF protocol
deny 192.0.2.0/24 to /32    ! TEST1
deny 192.168.0.0/16 to /32   ! RFC1918
deny 198.18.0.0/15 to /32    ! Benchmarking
deny 198.51.100.0/24 to /32  ! TEST2
deny 203.0.113.0/24 to /32  ! TEST3
deny 224.0.0.0/3 to /32     ! Multicast & Experimental
deny 0.0.0.0/0 from /25 to /32 ! Prefixes >/24
deny subnets of your own address space
permit everything else
```

Receiving IPv6 Prefixes

```
permit 64:ff9b::/96          ! RFC6052 v4v6trans
deny 2001::/23 to /128      ! RFC2928 IETF prot
deny 2001:2::/48 to /128    ! Benchmarking
deny 2001:10::/28 to /128   ! ORCHID
deny 2001:db8::/32 to /128  ! Documentation
deny 2002::/16 to /128      ! Deny all 6to4
deny 2020:3030::/32 to /128 ! Local Prefix
deny 3ffe::/16 to /128      ! Old 6bone
deny subnets of your own address space
permit 2000::/3 to 48        ! Global Unicast
deny ::/0 to /128           ! Deny everything else
```

Note: These filters block Teredo (serious security risk) and 6to4 (deprecated by RFC7526)

Receiving Prefixes

- Paying attention to prefixes received from customers, peers and transit providers assists with:
 - The integrity of the local network
 - The integrity of the Internet
- Responsibility of all Network Operators to be good Internet citizens

Receiving BGP attributes



Receiving BGP attributes

- BGP attributes are sent as part of the BGP updates for each prefix
- Common attributes operators need to be aware of, for routing best practice, are:
 - MED
 - AS numbers (only public ASNs are routable)
 - BGP Communities

Receiving Prefixes: MEDs?

- MEDs are used by EBGP neighbours to indicate preferred entry point into their network over two or more links with their neighbour
 - Allows the operator to determine entry path into their network
 - Might have unintended consequences within their peer's network
 - Many operators will override MEDs attached to BGP announcements by setting their own local-preference values

Receiving Prefixes: Bogon ASNs?

- What about prefixes originated by bogon AS numbers?
 - Public ranges are 1-64495 (excluding 23456) and 131072-458751
 - IANA is distributing AS blocks to the RIRs from the latter range
 - All other ASNs are either for documentation, or for private use, or are unassigned
 - And any prefixes originating from those need to be dropped
 - Configuration error? Malicious intent?
- What would the AS_PATH filter look like?
 - Challenging with regular expression (as per IOS)
 - Easier with AS ranges (as per Bird or JunOS)

Filtering bogon ASNs – BIRD

- Here is a function showing how to filter bogon ASNs, as described previously:

```
function as_path_contains_bogons()
int set invalid_asns;
{
    invalid_asns = [
        0,                # Reserved
        23456,            # Transition AS
        64496..64511,    # Documentation ASNs
        64512..65534,    # Private ASNs
        65535,           # Reserved
        65536..65551,    # Documentation ASNs
        65552..131071,   # Reserved
        458752..419999999, # IANA Reserved
        4200000000..4294967294, # Private ASNs
        4294967295       # Reserved
    ];
    return bgp_path ~ invalid_asns;
}
```

Filtering bogon ASNs – FRR

- Here is an AS-PATH regexp showing how to filter bogon ASNs:

```
bgp as-path access-list Bogon_ASNs deny _0_  
bgp as-path access-list Bogon_ASNs deny _23456_  
bgp as-path access-list Bogon_ASNs deny _6449[6-9]_  
bgp as-path access-list Bogon_ASNs deny _64[5-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _6[5-9][0-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _[7-9][0-9][0-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _1[0-2][0-9][0-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _130[0-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _1310[0-6][0-9]_  
bgp as-path access-list Bogon_ASNs deny _13107[0-1]_  
bgp as-path access-list Bogon_ASNs deny _45875[2-9]_  
bgp as-path access-list Bogon_ASNs deny _4587[6-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _458[8-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _459[0-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _4[6-9][0-9][0-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _[5-9][0-9][0-9][0-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _[0-9][0-9][0-9][0-9][0-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _[0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _[0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs deny _[0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9]_  
bgp as-path access-list Bogon_ASNs permit .*
```

Receiving Prefixes: BGP Communities?

- BGP communities are attached to BGP announcements to indicate:
 - Internal policy within an AS
 - External policy supported by a peer, for:
 - Onward routing policy/traffic engineering
 - Filtering (eg Remotely Triggered Blackhole Filtering)
 - Traffic engineering between the two networks
- Different BGP implementations have different default BGP community behaviours – consult:
 - Vendor documentation
 - <https://www.rfc-editor.org/rfc/rfc8642.txt> for discussion of some of the issues operators need to be aware of

Receiving Prefixes: BGP Communities

- Don't accept community values that are not expected
 - Match expected values
 - Overwrite received community values with your own default value

```
ip community-list standard lp-250 permit 65534:250
!  
route-map ebgp-import permit 5  
  description Set high preference  
  match community lp-250  
  set local-preference 250  
  set community 65534:100  
!  
route-map ebgp-import permit 10  
  description Set our default community  
  set community 65534:100  
!
```

We only expect this community from our EBGp neighbour

Cisco IOS: this overwrites all incoming community values

Receiving Prefixes: BGP Communities

- Don't send community values that are not needed by peer
 - This avoids propagating your internal communities to other networks

```
route-map ebgp-export permit 5
description Tell upstream to set local-pref 250
set community 65534:250
!
```

← Cisco IOS: this overwrites all other community values

- Propagate all communities within the AS (by IBGP)
 - This may need changes to your equipment's default!

Receiving BGP attributes

- Care is needed when receiving prefixes, to be aware of some of the optional BGP attributes that may be attached
 - BGP communities are only intended for policy decisions within an AS or between two peering ASes
 - MEDs may have unexpected consequences for traffic flows on the peer's network
 - Bogon ASNs, like bogon address space, must never be used or announced to the global Internet

Preparing the network



Before we begin...

Preparing the Network

- We will deploy BGP across the network before we try and multihome
- BGP will be used therefore an ASN is required
- If multihoming to different providers, a public ASN will be needed:
 - Either go to upstream provider who is a registry member, or
 - Apply to the RIR yourself for a one-off assignment, or
 - Ask a network operator who is a registry member, or
 - **Join the RIR and get your own IP address allocation too**
 - (this option strongly recommended)!
 - Gives independence of any upstream provider
 - Provides minimal IPv4 (depending on RIR region), IPv6 block (/32) and an ASN

Preparing the Network

Initial Assumptions

- The network is not running any BGP at the moment
 - Single statically routed connection to upstream provider
- The network is not running any IGP at all
 - Static default and routes through the network to do “routing”

Preparing the Network

First Step: IGP

- ❑ Decide on an IGP: OSPF or ISIS 😊
- ❑ Assign loopback interfaces and /32 address to each router which will run the IGP
 - Loopback is used for OSPF and BGP router id anchor
 - Used for IBGP and route origination
- ❑ Deploy IGP (e.g. ISIS)
 - IGP can be deployed with NO IMPACT on the existing static routing
 - e.g. ISIS distance might be 115, static distance is 1
 - Lowest distance wins

Preparing the Network

IGP (cont)

- Be prudent deploying IGP – keep the Link State Database Lean!
 - Router loopbacks go in IGP
 - WAN point to point links go in IGP
 - (In fact, any link where IGP dynamic routing will be run should go into IGP)
 - Summarise on area/level boundaries (if possible) – i.e. think about your IGP address plan

Preparing the Network

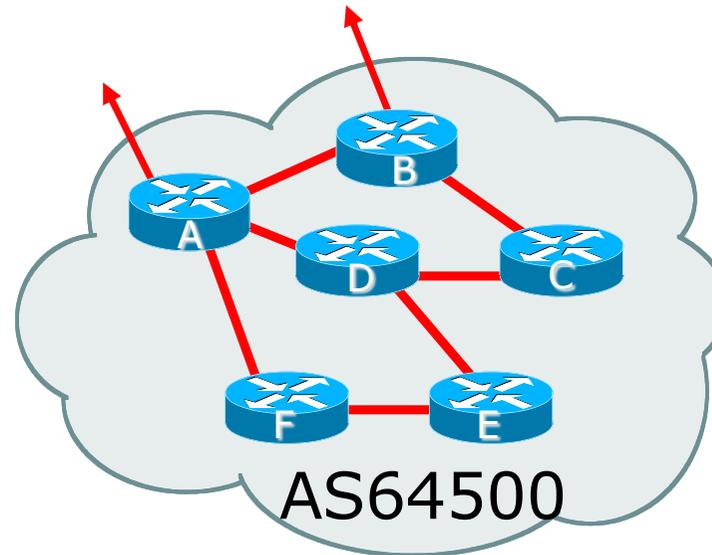
IGP (cont)

- Routes which don't go into the IGP include:
 - Dynamic assignment pools (DSL/Cable/Dial)
 - Customer point to point link addressing
 - (using next-hop-self in IBGP ensures that these do NOT need to be in IGP)
 - Static/Hosting LANs
 - Customer assigned address space
 - Anything else not listed in the previous slide

Preparing the Network

Second Step: IBGP

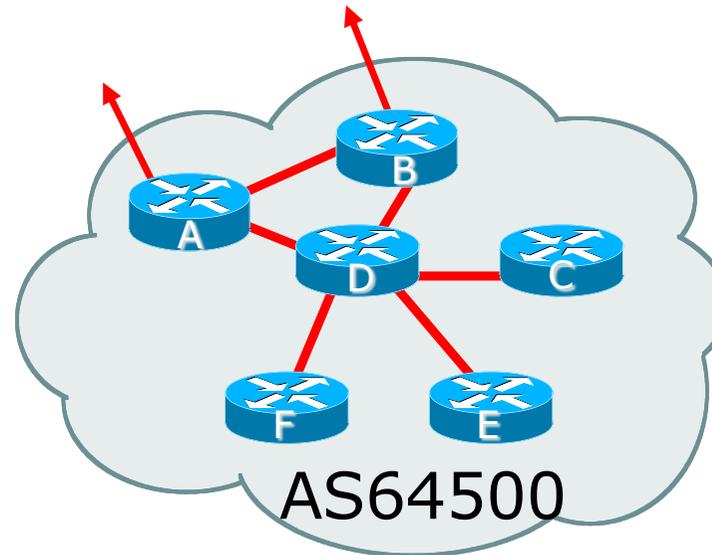
- ❑ Second step is to configure the local network to use IBGP
- ❑ IBGP can run on
 - All routers, or
 - A subset of routers, or
 - Just on the upstream edge
- ❑ **IBGP must run on all routers which are in the transit path between external connections**



Preparing the Network

Second Step: IBGP (Transit Path)

- ❑ IBGP must run on all routers which are in the transit path between external connections
- ❑ Routers C, E and F are not in the transit path
 - Static routes or IGP will suffice
- ❑ Router D is in the transit path
 - Will need to be in IBGP mesh, otherwise routing loops will result



Preparing the Network Layers

- Typical SP networks have three layers:
 - Core – the backbone, usually the transit path
 - Distribution – the middle, PoP aggregation layer
 - Aggregation – the edge, the devices connecting customers

Preparing the Network

Aggregation Layer

- IBGP is optional
 - Many service providers run IBGP here, either partial routing (more common) or full routing (less common)
 - Full routing is not needed unless customers want full table
 - Partial routing is cheaper/easier, might usually consist of internal prefixes and, optionally, external prefixes to aid external load balancing
 - Communities and peer-groups make this administratively easy
- Many aggregation devices can't run IBGP
 - Static routes from distribution devices for address pools
 - IGP for best exit

Preparing the Network Distribution Layer

- Usually runs IBGP
 - Partial or full routing (as with aggregation layer)
- But does not have to run IBGP
 - IGP is then used to carry customer prefixes (does not scale)
 - IGP is used to determine nearest exit
- Networks which plan to grow large should deploy IBGP from day one
 - Migration at a later date is extra work
 - No extra overhead in deploying IBGP, indeed IGP benefits

Preparing the Network

Core Layer

- Core of network is usually the transit path
- IBGP necessary between core devices
 - Full routes or partial routes:
 - Transit providers carry full routes in core
 - Edge providers carry partial routes only
- Core layer includes AS border routers

Preparing the Network

IBGP Implementation

Decide on:

- Best IBGP policy
 - Will it be full routes everywhere, or partial, or some mix?
- IBGP scaling technique
 - Community policy?
 - Route-reflectors?
 - Techniques such as BGP peer groups?

Preparing the Network

IBGP Implementation

□ Then deploy IBGP:

- Step 1: Introduce IBGP mesh on chosen routers
 - make sure that iBGP distance is greater than IGP distance (it usually is)
- Step 2: Install “customer” prefixes into IBGP

Check! Does the network still work?
- Step 3: Carefully remove the static routing for the prefixes now in IGP and IBGP

Check! Does the network still work?
- Step 4: Deployment of EBGP follows

Preparing the Network

IBGP Implementation

- *Install “customer” prefixes into IBGP?*
- Customer assigned address space
 - Network statement/static route combination
 - Use unique community to identify customer assignments
- Customer facing point-to-point links
 - Redistribute connected through filters which only permit point-to-point link addresses to enter IBGP
 - Use a unique community to identify point-to-point link addresses (these are only required for your monitoring system)
- Dynamic assignment pools & local LANs
 - Simple network statement will do this
 - Use unique community to identify these networks

Preparing the Network

IBGP Implementation

- *Carefully remove static routes?*
- Work on one router at a time:
 - Check that static route for a particular destination is also learned by the IBGP
 - If so, remove it
 - If not, establish why and fix the problem
 - (Remember to look in the RIB, not the FIB!)
- Then the next router, until the whole PoP is done
- Then the next PoP, and so on until the network is now dependent on the IGP and IBGP you have deployed

Preparing the Network Completion

- Previous steps are NOT flag day steps
 - Each can be carried out during different maintenance periods, for example:
 - Step One on Week One
 - Step Two on Week Two
 - Step Three on Week Three
 - And so on
 - And with proper planning will have NO customer visible impact at all

Preparing the Network Configuration Summary

- ❑ IGP essential networks are in IGP
- ❑ Customer networks are now in IBGP
 - IBGP deployed over the backbone
 - Full or Partial or Upstream Edge only
- ❑ BGP distance is greater than any IGP
- ❑ Now ready to deploy EBGP

Configuration Tips



Of passwords, tricks and templates

IBGP and IGP

Reminder!

- Make sure loopback is configured on router
 - IBGP between loopbacks, NOT physical interfaces
- Make sure IGP carries loopback IPv4 /32 and IPv6 /128 address
- Consider the DMZ nets:
 - Use unnumbered interfaces?
 - Use next-hop-self on IBGP neighbours
 - Or carry the DMZ IPv4 /30s and IPv6 /127s in the IBGP
 - Basically, keep the DMZ nets out of the IGP!

IBGP: Next-hop-self

- BGP speaker announces external network to IBGP peers using router's local address (loopback) as next-hop
- Used by many service providers on edge routers
 - Preferable to carrying DMZ point-to-point link addresses in the IGP
 - Reduces size of IGP to just core infrastructure
 - Alternative to using unnumbered interfaces
 - Helps scale network
 - Many network operators consider this "best practice"

Limiting AS Path Length

- Some BGP implementations have problems with long AS_PATHS
 - Memory corruption
 - Memory fragmentation
- Even using AS_PATH prepends, it is not normal to see more than 20 ASNs in a typical AS_PATH in the Internet Routing Table today
 - The Internet is around 5 ASes deep on average
 - Largest AS_PATH is usually 16-20 ASNs

Limiting AS Path Length

- Some announcements have ridiculous lengths of AS-paths
 - This example is an error in one IPv6 implementation

```
*> 3FFE:1600::/24      22 11537 145 12199 10318 10566 13193 1930 2200 3425 293 5609 5430
13285 6939 14277 1849 33 15589 25336 6830 8002 2042 7610 i
```

- This example shows 100 prepends (for no obvious reason)

```
*>i193.105.15.0      2516 3257 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404
50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404
50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404
50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404
50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404
50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404 50404
50404 i
```

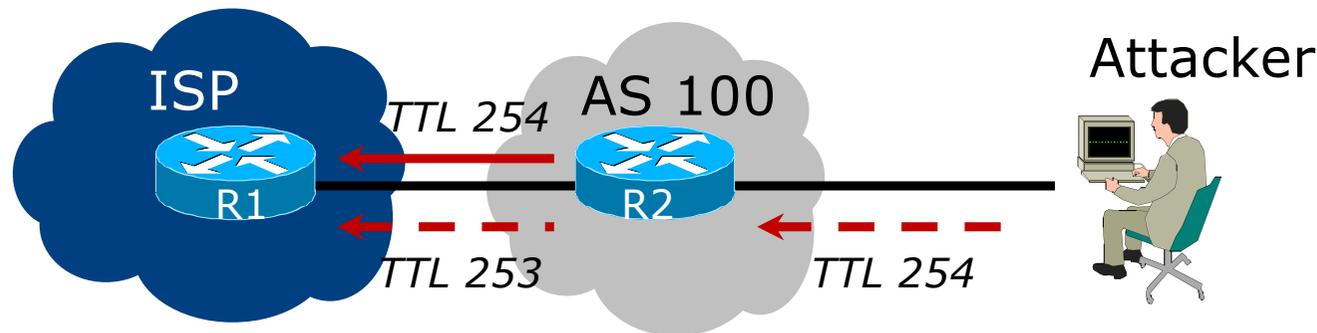
- If your implementation supports it, consider limiting the maximum AS-path length you will accept

BGP Maximum Prefix Tracking

- Allow configuration of the maximum number of prefixes a BGP router will receive from a peer
 - Supported by good BGP implementations
- Usually have two level control for prefix count:
 - Reaches warning threshold: log a warning message
 - Threshold is configurable
 - Reaches maximum:
 - Only send warnings
 - Tear down BGP, manual intervention required to restart
 - Tear down BGP and automatically restart after a delay (configurable)

BGP TTL “hack”

- Implement RFC5082 on BGP peerings
 - (Generalised TTL Security Mechanism)
 - Neighbour sets TTL to 255
 - Local router expects TTL of incoming BGP packets to be 254
 - No one apart from directly attached devices can send BGP packets which arrive with TTL of 254, so any possible attack by a remote miscreant is dropped due to TTL mismatch



BGP TTL “hack”

- TTL Hack:
 - Both neighbours must agree to use the feature
 - TTL check is much easier to perform than MD5
 - (Called BTSH – BGP TTL Security Hack)
- Provides “security” for BGP sessions
 - In addition to packet filters of course
 - MD5 should still be used for messages which slip through the TTL hack
 - See <https://www.nanog.org/meetings/nanog27/presentations/meyer.pdf> for more details

Templates

- Good practice to configure templates for everything
 - Vendor defaults tend not to be optimal or even very useful for network operators
 - Network operators create their own defaults by using configuration templates
- EBGP and IBGP examples follow
 - Also see Team Cymru's BGP templates
 - <http://www.team-cymru.com/community-services>

IBGP Template

Example

- ❑ IBGP between loopbacks!
- ❑ Next-hop-self
 - Keep DMZ and external point-to-point out of IGP
- ❑ Always send communities in IBGP
 - Otherwise BGP policy accidents will happen
 - (Default on some vendor implementations, optional on others)
- ❑ Hardwire BGP to version 4
 - Yes, this is being paranoid!
 - Prevents accidental configuration of BGP version 3 which is still supported in some implementations

IBGP Template

Example continued

- Use passwords on IBGP session
 - Not being paranoid, **VERY** necessary
 - It's a secret shared between you and your peer
 - If arriving packets don't have the correct MD5 hash, they are ignored
 - Helps defeat miscreants who wish to attack BGP sessions
- Powerful preventative tool, especially when combined with filters and the TTL "hack"

EBGP Template

Example

- BGP damping
 - Do **NOT** use it unless you understand the impact
 - Do **NOT** use the vendor defaults without thinking
- Cisco's Soft Reconfiguration
 - Do **NOT** use unless troubleshooting or doing Route Origin Validation – it will consume considerable amounts of extra memory for BGP
- Remove private ASNs from announcements
 - Common omission today
- Use extensive filters, with “backup”
 - Use AS-path filters to backup prefix filters
 - Keep policy language for implementing policy, rather than basic filtering

EBGP Template

Example continued

- ❑ Use password agreed between you and peer on EBGP session
- ❑ Use maximum-prefix tracking
 - Router will warn you if there are sudden increases in BGP table size, bringing down EBGP if desired
- ❑ Limit maximum as-path length inbound
- ❑ Log changes of neighbour state
 - ...and monitor those logs!
- ❑ Make BGP admin distance higher than that of any IGP
 - Otherwise, prefixes heard from outside your network could override your IGP!!

Mutually Agreed Norms for Routing Security

Industry Best Practices to ensure Security
of the Routing System



MANRS

Routing Security

□ Implement the recommendations in

<https://www.manrs.org>

1. Prevent propagation of incorrect routing information
 - Filter BGP peers, in & out!
2. Prevent traffic with spoofed source addresses
 - BCP38 – Unicast Reverse Path Forwarding
3. Facilitate communication between network operators
 - NOC to NOC Communication
 - Up-to-date details in Route and AS Objects, and PeeringDB
4. Facilitate validation of routing information
 - Route Origin Authorisation using RPKI



MANRS

MANRS 1)

- Filtering prefixes inbound and outbound
 - RFC8212 requires all EBGP implementations to reject prefixes received and announced in the absence of any policy

- Advice: **Never** set up an EBGP session without inbound and outbound prefix filters
 - If full table required, block at least the bogons (see earlier)

MANRS 2)

- Implementing BCP 38
 - Unicast Reverse Path Forwarding
 - (Deny outbound traffic from customers which has spoofed source addresses)
- Advice: implement uRPF on ***all*** single-homed customer facing interfaces
 - Cheaper (CPU & RAM) than implementing packet filters

MANRS 3)

- Facilitate NOC to NOC communication
 - Know the **direct** NOC contacts for your customer Network Operators, your peer Network Operators, and your upstream Network Operators
 - This is not calling their “customer support line”
 - Make sure NOC contact info is part of any service contract
 - Up to date info in Route and AS Objects
 - Up to date AS info in PeeringDB

- Advice: NOC contact info for all connected Autonomous Networks is known to your NOC

MANRS 4)

- Facilitate validation of Routing Information
 - RPKI and Route Origin Authorisation (ROA)
 - All routes originated need to be signed to indicate that your AS is authorised to originate these routes
 - Helps secure the global routing system

- Advice: Sign ROAs for all originated routes using RPKI
 - And make sure all customer originated routes are also signed
 - Validate received routes from all peers
 - High priority for validated routes
 - Discard invalid routes
 - Low priority for unsigned routes

MANRS summary

- If your organisation supports and implements all 4 techniques in your network
 - Then join MANRS
 - <https://www.manrs.org/join/>
 - MANRS for Operators
 - MANRS for IXPs
 - MANRS for CDN & Cloud Providers



MANRS

Summary

- ❑ Use configuration templates
- ❑ Standardise the configuration
- ❑ Be aware of standard “tricks” to avoid compromise of the BGP session
- ❑ Anything to make your life easier, network less prone to errors, network more likely to scale
- ❑ Implement the four fundamentals of MANRS
- ❑ It’s all about scaling – if your network won’t scale, then it won’t be successful

BGP Techniques for Network Operators



The End!