BGP Case Studies

ISP Workshops

Last updated 28th September 2020

Acknowledgements

- This material was developed by Philip Smith with the support of the Network Startup Resource Center

- Use of these materials is encouraged as long as the source is fully acknowledged and this notice remains in place

- Bug fixes and improvements are welcomed
  - Please email *workshop (at) bgp4all.com*

Philip Smith

Topics Covered

- Peering Priorities
- Transit Provider Peering at an IXP
- Customer Multihomed between two IXP members
- Traffic Engineering for an ISP connected to two IXes
- Traffic Engineering for an ISP with two interfaces on one IX LAN
- Traffic Engineering and CDNs
- Multihoming with Partial Routes: propagating the default using the IGP

Peering Priorities for a Network Operator

Peering Priorities

□ As network operators move from having a single upstream to deploying BGP with multiple external connections, they need to:

- Establish priorities for BGP customers
- Prioritise different peering partners
- Establish cost/benefits for participating at different IXPs
- Establish cost/benefits for different transit connections

Peering Policy

- Typical prioritisation:
  - Most preferred – BGP customers
    - We would like traffic from us to our BGP customers to go directly, not via our peers or transits
  - Next preference – private peers
    - Connect by direct cross-connection
  - Next preference – local IXP
    - Keep local traffic local
  - Next preference – regional IXP
    - Keep regional traffic regional
    - Will cost money for physical connectivity to regional IXP
  - Last preference – paid transit
    - Will cost money for physical connectivity and for traffic

Peering Policy – Local Preference

☐ Example Local Preference Table

Peering Policy	Local Preference
BGP Customer	250
Private Peer	200
Local IXP	170
Regional IXP	140
(default)	100
Primary Paid Transit	50
Backup Paid Transit	40

Agenda

- Peering Priorities
- Transit Provider Peering at an IXP
- Customer Multihomed between two IXP members
- Traffic Engineering for an ISP connected to two IXes
- Traffic Engineering for an ISP with two interfaces on one IX LAN
- Traffic Engineering and CDNs
- Multihoming with Partial Routes: propagating the default using the IGP

IXP Peering: When a Transit Provider is Also a Peer

IXP Peering, when Transit Provider is also a Peer

- Relatively common situation
  - Several local ISPs providing access to the local market
  - One or two licensed transit providers
  - Licensed transits also wish to peer at the IXP
- Desired outcome:
  - Transit provider wants to:
    - Peer domestic traffic at the IX
    - Sell transit access for all other destinations
- How to ensure that:
  - Transit traffic only goes on transit link
 - Peering traffic only goes on peering link

10

IXP Peering, when Transit Provider is also a Peer
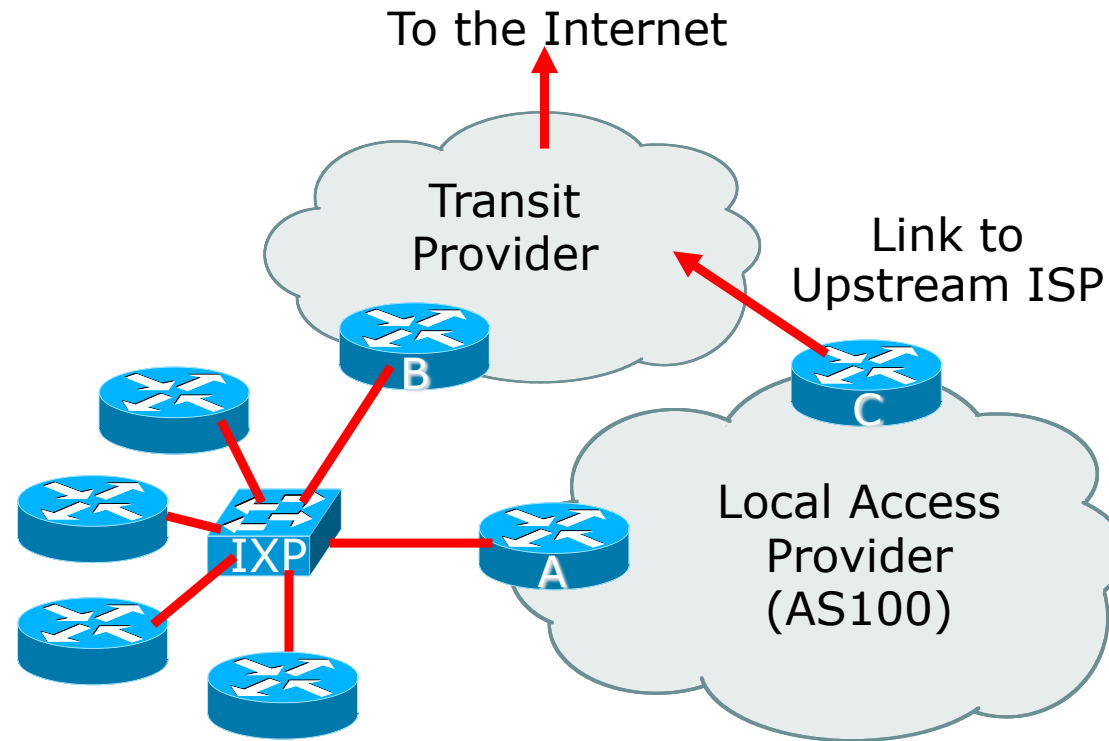
IXP Peering, when Transit Provider is also a Peer

- Outbound traffic from AS100:
  - Upstream sends full BGP table to AS100 on direct peering link
  - Upstream sends domestic routes to IXP peers
  - AS100 uses IXP for domestic traffic
  - AS100 uses Upstream link for international traffic
- Inbound traffic to AS100:
  - AS100 sends address block to IXP peers
  - AS100 sends address block to upstream
  - Best path from upstream to AS100 preferred via the IXP (see previous scenario)
- Problem: how to separate international and domestic traffic towards AS100?

Solution: AS Separation

To the Internet

Transit
AS160

Domestic
AS150

Link to
Upstream ISP

D

B

C

A

IXP

AS 100

13
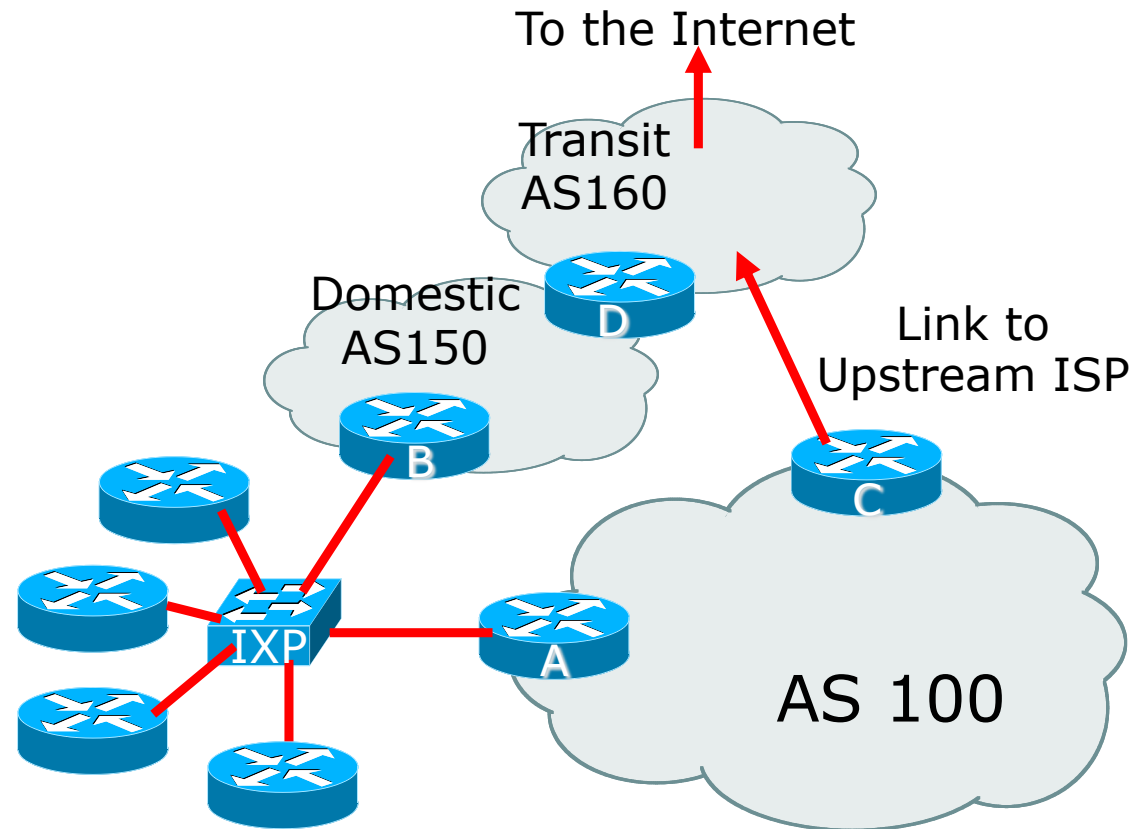
Solution: AS Separation

- The transit provider needs to separate their network:
  - Domestic (AS150: local routes)
  - Transit (AS160: non-local routes)
- Transit customers connect to transit AS (AS160)
  - Receive default route (or full BGP if desires)
  - Send just their address blocks
- Domestic AS (AS150) peers at the IX
  - Receives local routes from other IX peers
  - Sends AS150 originated routes to IX peers

Solution: AS Separation Outcome

□ Inbound traffic to AS100 now:
- AS100 sends address block to IXP peers (including AS150)
- AS100 sends address block to upstream (AS160)
- Best path from upstream to AS100 preferred via the transit link

□ Important notes:
- AS150 must NOT pass prefixes learned from IX peers to AS160

IXP Peering, when Transit Provider is also a Peer

- Transit providers who peer with their customers at an IX for local routes need to split their ASNs into two:
  - One AS for domestic routes
  - One AS for transit routes

- Two ASNs are justifiable because the two ASNs have completely different routing policies
  - Domestic AS peers at IXP
  - Transit AS connects transit customers and upstreams

IXP Peering, when Transit Provider is also a Peer

- This solution is scalable
- This solution is much easier to implement than other solutions such as complex source address policy routing

- Remember:
  - An Autonomous System is used for representing a distinct routing policy
  - An Autonomous System doesn't necessarily map onto an organisation
  - A transit business WILL have different routing policy from an access business or a hosting business, and therefore will quite likely need a different ASN

Agenda

- Peering Priorities
- Transit Provider Peering at an IXP
- Customer Multihomed between two IXP members
- Traffic Engineering for an ISP over two routers on one IX LAN
- Traffic Engineering for an ISP connected to two IXes
- Traffic Engineering and CDNs
- Multihoming with Partial Routes: propagating the default using the IGP

Customer multihomed between two IXP members

Customer multihomed between two IXP members

- It is quite a common for an end user network to be multihomed between two ISPs who are also members of the same IXP
  - What are the traffic engineering considerations here?
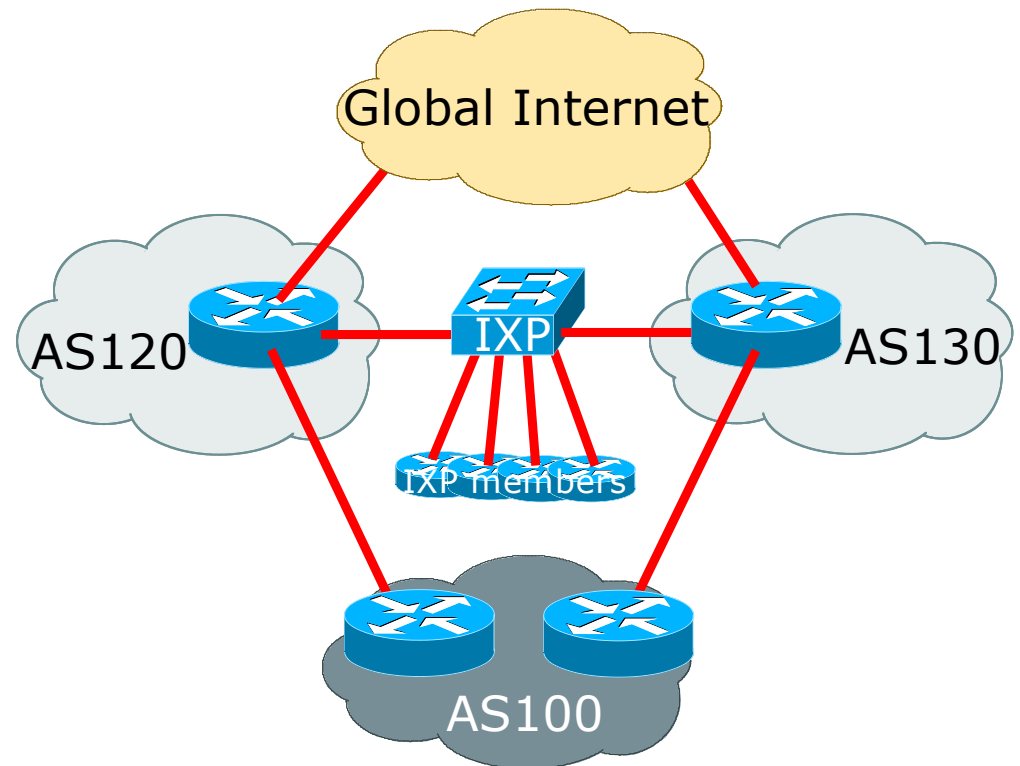
Customer multihomed between two IXP members

- Issues to address:
  - Transit:
    - Which path does incoming transit traffic to the end-site take?
    - Which path does outgoing transit traffic from the end-site take?
  - Peering:
    - What about peering traffic from other IXP members to the end-site?
    - What about traffic from the end-site to other IXP members?
- Each ISP would consider the end-site to be their customer, and for their customer's traffic to transit their link

Route announcements (1)

- AS100 announces its address block to both upstreams
  - AS120 and AS130 will both tag high local preference on the announcement from AS100
  - Therefore traffic from AS120 and AS130 will go over their respective links
- AS120 and AS130 will announce AS100's address block to their upstreams and onwards to the rest of the Internet
  - The global Internet will choose the path to AS100 according to best path rules

Route announcements (2)

- AS120 and AS130 will also announce AS100's routes to the IXP Route Server and the IXP members
- Following the peering priorities discussed earlier:
  - AS120 will hear AS100's routes from direct connection
    - Local preference 250
  - AS120 will hear AS100's routes across the IXP
    - Local preference 170
  - (And the same is true for AS130)
- In steady state, traffic from AS120 or AS130 to their customer will traverse the direct transit link
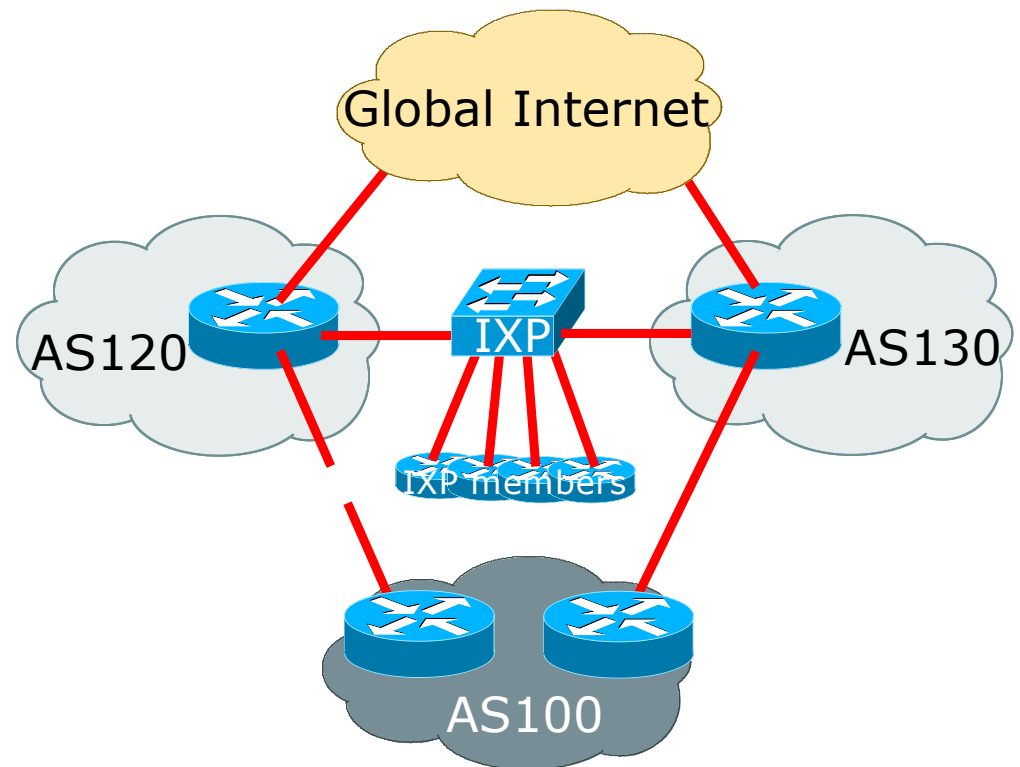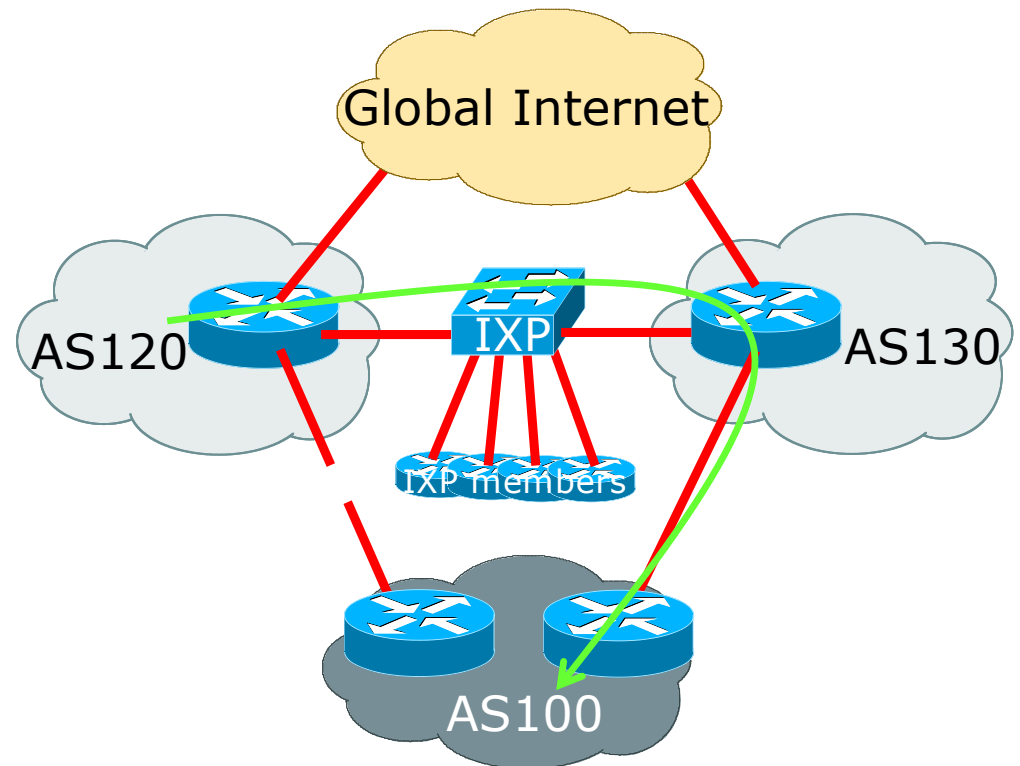
Traffic flow

Incoming Traffic

Outgoing Traffic

Global Internet

AS120

IXP

AS130

IXP members
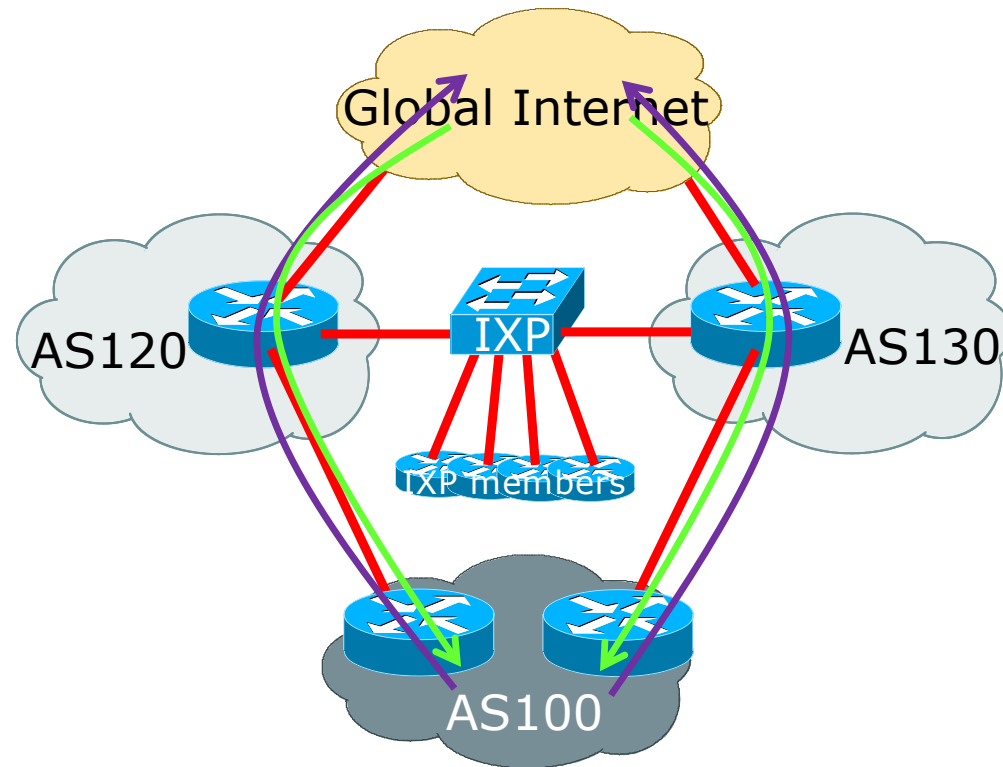
AS100

Example: Failure mode

□ What happens when the link
   from AS120 to its customer
   AS100 goes down?

Example: Failure mode

- What happens when the link from AS120 to its customer AS100 goes down?
- AS120 will still see a path to AS100
  - Will be across the IXP to AS130 and using AS130's link
  - Which means traffic from AS120 to AS100 will go via the IXP and AS130

Global Internet

AS120

IXP

AS130

IXP members

AS100

26

Example: Failure mode

- ☐ What happens when the link from AS120 to its customer AS100 goes down?
- ☐ AS120 will still see a path to AS100
  - ■ Will be across the IXP to AS130 and using AS130's link
  - ■ Which means traffic from AS120 to AS100 will go via the IXP and AS130
  - ■ Traffic from AS100 will go to AS120 via AS130 and the IXP

Global Internet

AS120
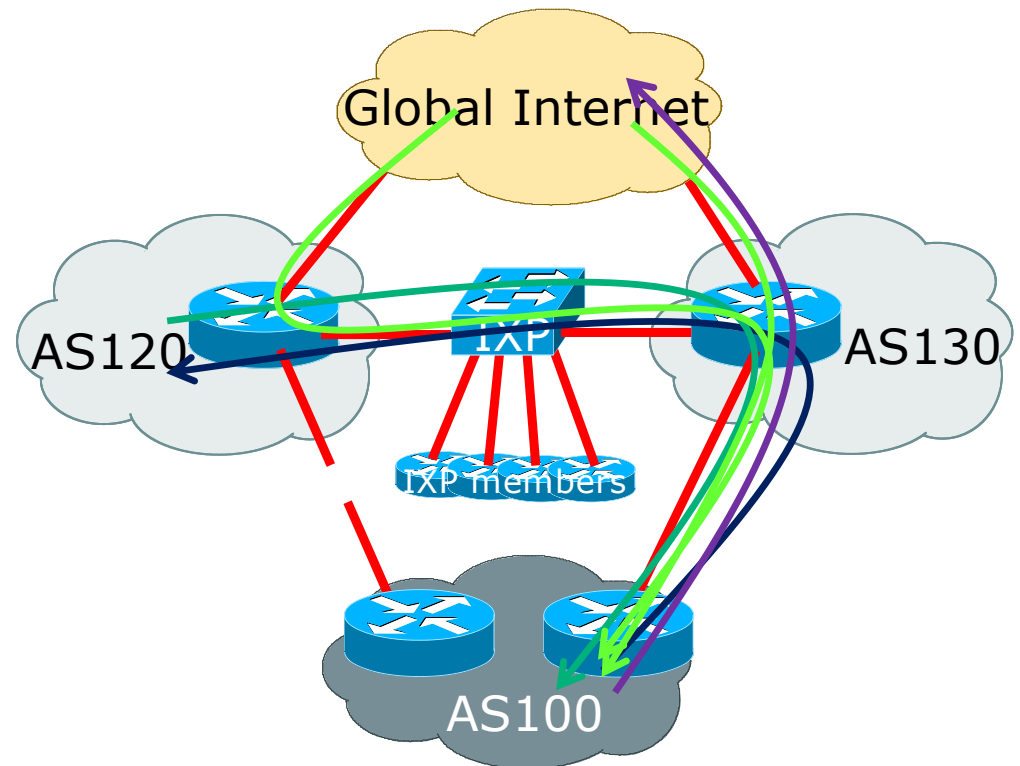
AS130

IXP

IXP members

AS100

27
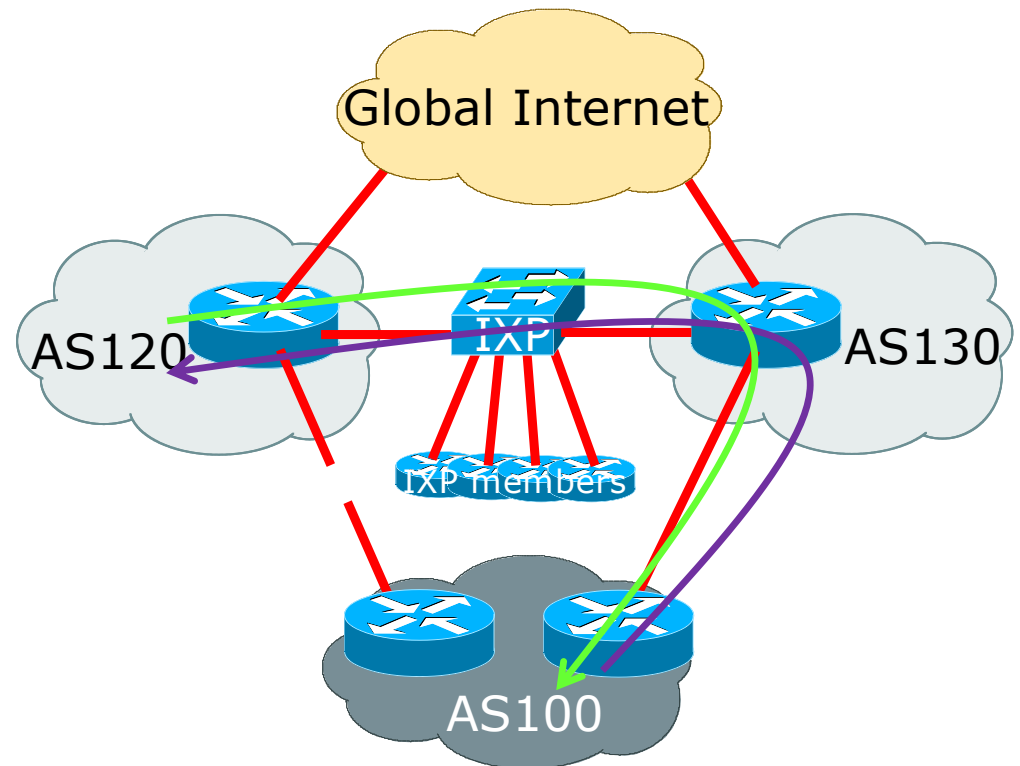
Failure mode: Traffic flow

- Traffic to the Internet:
  - Exits via AS130 upstream link

  → (purple arrow)

- Traffic from the Internet:
  - Will enter via AS120 and AS130 (latter will likely be higher volume due to shorter AS-PATH length)

  → (light green arrow)

- Traffic to AS120
  - Crosses the IXP via AS130

  → (dark blue arrow)

- Traffic from AS120
  - Crosses the IXP via AS130

  → (green arrow)

Global Internet

AS120   IXP   AS130
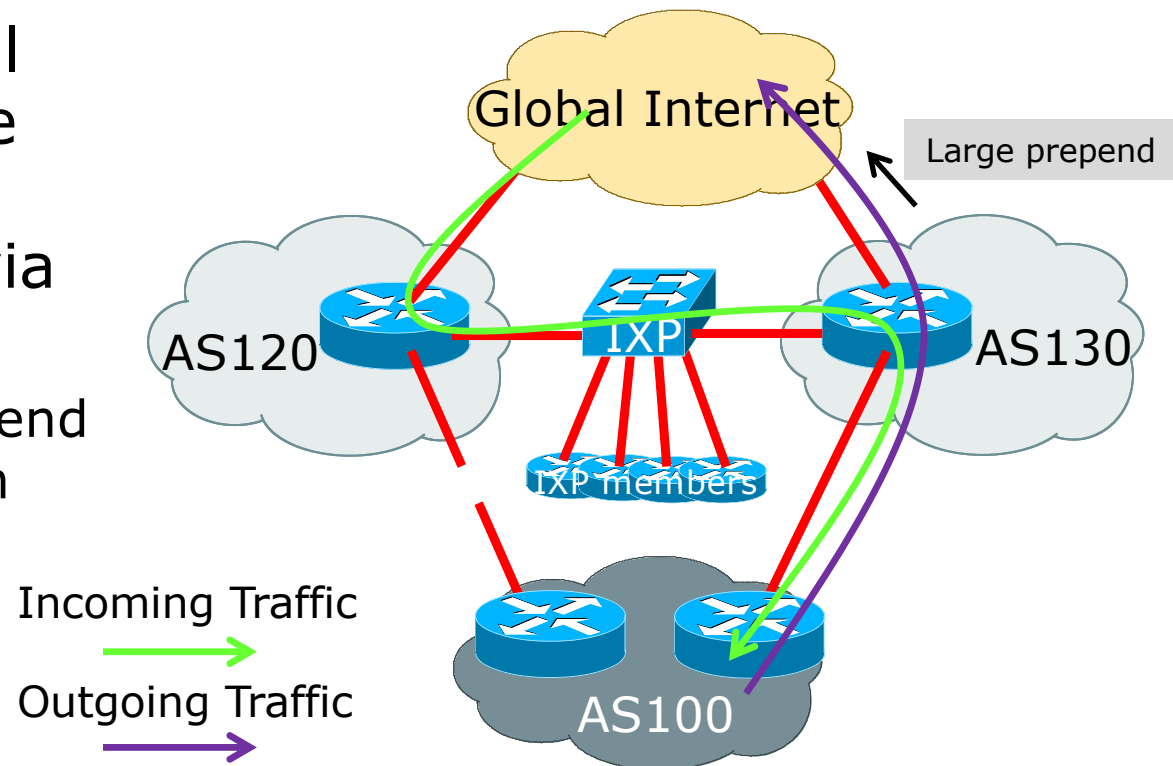
IXP members

AS100

28

Failure mode observations

□ If AS130 implements packet filtering by source address on the peering link:

■ Incoming traffic from the Internet via AS120 and the IXP to AS100 will be dropped, affecting connectivity for AS100

□ Multihomed customers of IXP members will result in this "unusual" traffic flow

■ This is a side-effect of the customer having redundant connections – nothing wrong with it

■ And one reason why it is very important for a network operator to set higher local preference on routes heard directly from a BGP customer versus the same routes heard via peering or transit links
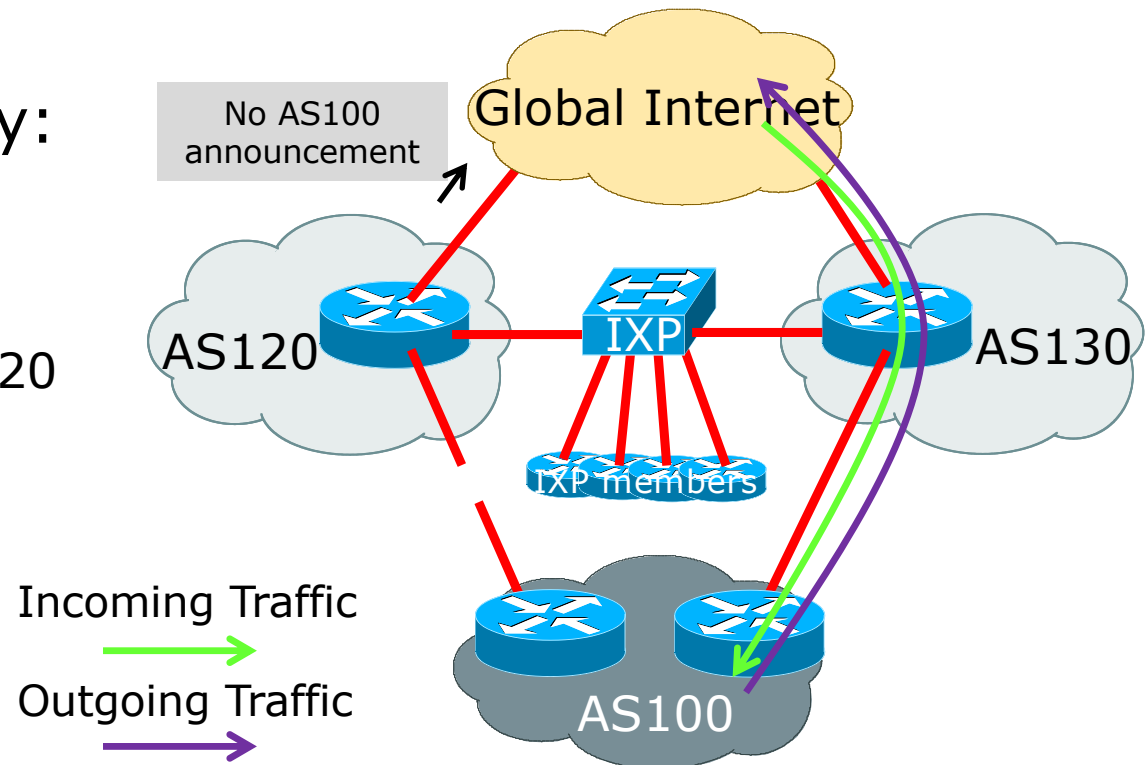
Failure mode observations

□ Customer network traffic engineering could cause all transit traffic during failure case to transit to AS120, across the IXP, and then via AS130 to the customer

▪ For example: AS-PATH prepend on their announcement from AS130 to upstream

Global Internet

Large prepend

AS120

IXP

AS130

IXP members

Incoming Traffic

Outgoing Traffic

AS100

Failure mode observations

□ **If AS120 tags routes it hears from IXP peers with the "no-export" community:**

  ▪ When the link to its AS100 customer fails, AS100's prefix will be withdrawn from BGP announcements made by AS120 to the Internet

    □ Due to best path via the IXP being "no-export"

No AS100 announcement

Global Internet

AS120

IXP

AS130

IXP members

Incoming Traffic

Outgoing Traffic

AS100

Summary

- This is a common scenario for when a customer is multihomed on to members of an IXP

- Steady state works as expected

- If either customer transit link fails, then incoming transit traffic will cross the IXP
  - This is not unusual, nor is it a problem

- To avoid having transit traffic crossing the IXP
  - Tag routes heard from the IXP with "no-export"
  - Direct customer routes only announced to transits when direct link to customer is active

Agenda

- Peering Priorities
- Transit Provider Peering at an IXP
- Customer Multihomed between two IXP members
- Traffic Engineering for an ISP over two routers on one IX LAN
- Traffic Engineering for an ISP connected to two IXes
- Traffic Engineering and CDNs
- Multihoming with Partial Routes: propagating the default using the IGP

Traffic Engineering over two routers on one IX LAN

Two connections to one IXP

- In early stages of IX development:
  - IX has one ethernet switch
  - Members have a single ethernet connection to IX switch
- As IX grows:
  - It becomes critical infrastructure for local Internet economy
  - More members join
  - IX adds second switch for extra capacity and to provide redundancy for members
  - Second switch is on same L2 infrastructure as original
- How to configure BGP & Traffic engineering for two connections to the IX?

Two connections to one IXP

IXP
IXP Member
(AS100)
A
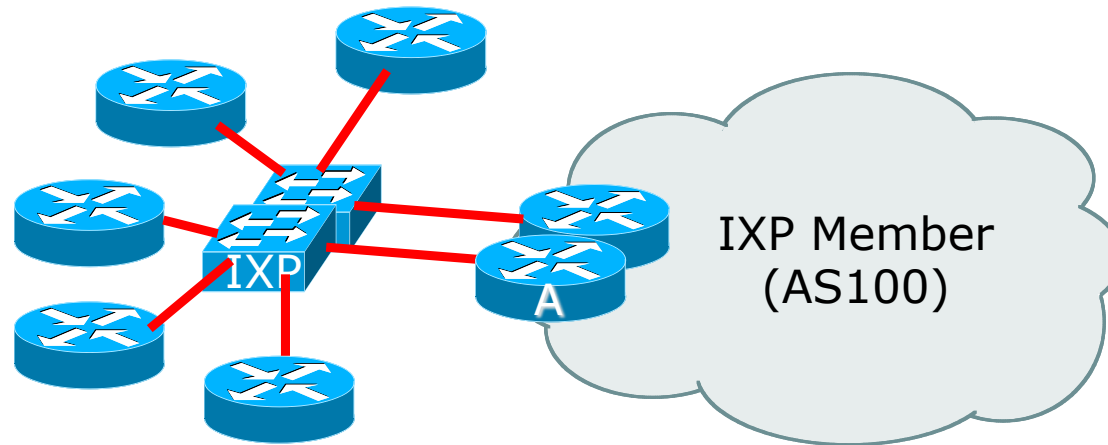
□ Diagram shows two ethernet links from separate switches to two routers

Two connections to one IXP

- IXP LAN configuration:
  - Second connection is on same subnet on IXP
  - Member receives another IP address from the same subnet
- BGP configuration:
  - Second EBGP session is established
    - With the IXP Route Server (if present)
    - With the other IXP members (with their second router, if they have one)
    - With IXP services infrastructure (if applicable)

Two connections to one IXP

- Outbound Traffic Engineering configuration:
  - By default, the link chosen will follow BGP best path rules
    - In the absence of any other member policy (e.g. MEDs), best path will be lowest neighbour IP address
    - Which most likely means that one link carries all the traffic; the other link remains relatively empty
  - AS100 could load balance over the two physical links by:
    - Setting local preferences on particular announcements from peers
    - Using any BGP community policy implemented by other members

Two connections to one IXP

- Inbound Traffic Engineering configuration:
  - By default, the link chosen will follow BGP best path rules
    - In the absence of any local policy (e.g. MEDs), best path will be lowest IP address on the IX LAN
  - AS100 could load balance over the two physical links by:
    - Setting MEDs on particular announcements to peers
      - Half the peers could have announcements of MED 10 on one link and MED 20 on the other link
      - And the other half of the peers have the MED values reversed
      - Which assumes that peers even respect MEDs
    - Implementing a BGP community policy available for other members to use
      - Sometimes IXPs recommend what a community policy might be
    - Using AS-PATH prepends (care needed so the IX path doesn't have longer AS path than via paid transit links)

Two connections to one IXP

- Bonding two ethernet connections
  - In some circumstances, the IXP may offer the facility of creating an aggregated link (LAG – Link Aggregation Group)
  - This provides redundancy at L2
    - For example, two GigabitEthernet links will effectively present as 2Gbps on a single connection on the router
    - The BGP session is established over the LAG rather than on individual links
    - Load balancing is at L2, contained within the LAG itself
- Note: this is only possible if the member only provisions one router for the IXP connection
  - And not desirable if the IXP provisions the two links on separate switches (assuming the switch vendor supports creating a LAG shared over two switches)

Agenda

- Peering Priorities
- Transit Provider Peering at an IXP
- Customer Multihomed between two IXP members
- Traffic Engineering for an ISP with two interfaces on one IX LAN
- Traffic Engineering for an ISP connected to two IXes
- Traffic Engineering and CDNs
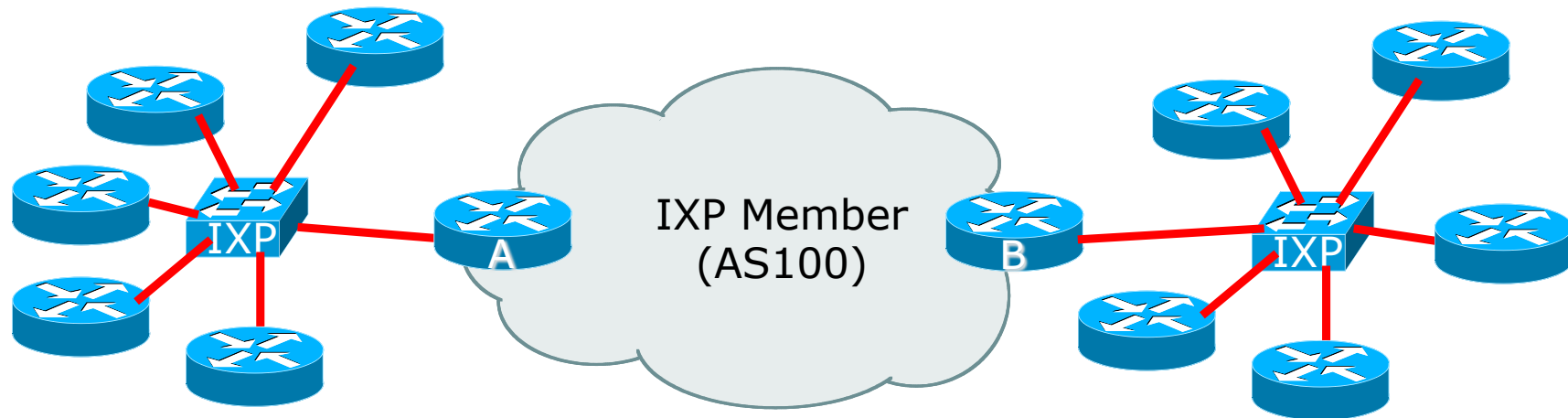- Multihoming with Partial Routes: propagating the default using the IGP

Traffic Engineering when connected to two IXPs

Traffic Engineering when connected to two IXPs

- **Several variations possible on this theme**
  - Peering at two local IXPs
    - Shouldn't really happen as an IXP is intended to be a collaborative effort between members/participants to peer local traffic
    - Two IXPs serving the same local market doubles the costs for all operators and makes the traffic engineering more challenging
  - Peering at local IXP and regional IXP
    - Very common where an ISP participates in the local IXP and also turns up at one or more regional IXPs for greater peering opportunities
  - Peering at two regional IXPs
    - Occurs in the absence of a local IXP

Peering at two local IXPs

- Diagram shows ISP connecting to two different IXPs
  - Could also be the case where one IXP operates two independent sites

Peering at two local IXPs

- Second IXP LAN configuration:
  - Connection to the second IXP set up in the same way as the connection to the first IXP
  - Member has access to same facilities (Route Server, IX services, etc)
- BGP configuration:
  - EBGP sessions established
    - With the IXP Route Server (if present)
    - With the other IXP members
    - With IXP services infrastructure (if applicable)
- Traffic Engineering
  - Load balancing across IXP links needed when members are present at both IXPs
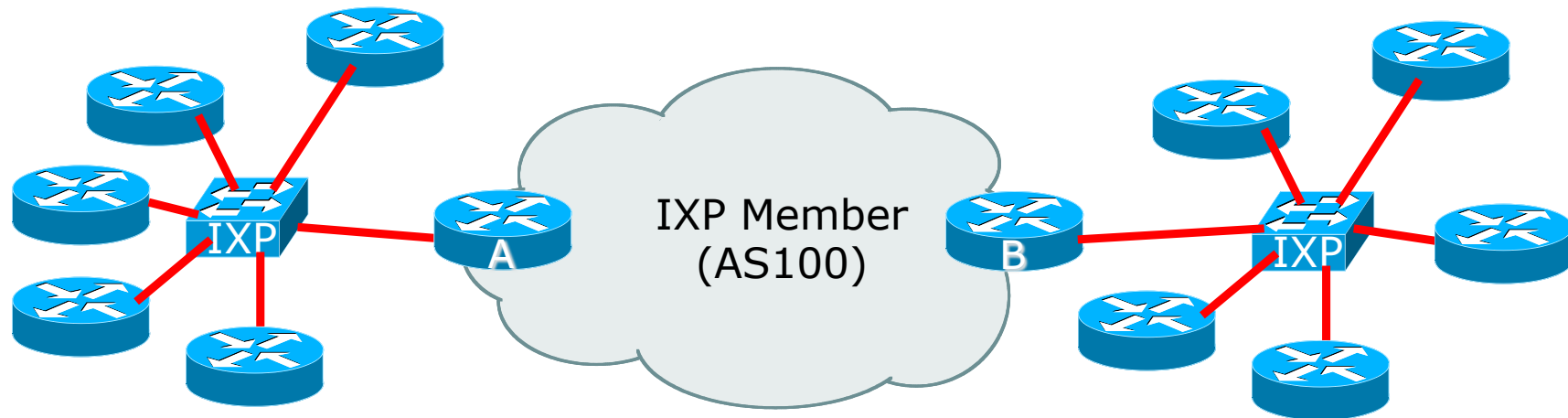
Peering at two local IXPs

- Outbound Traffic Engineering configuration:
  - By default, the link chosen will follow BGP best path rules
    - In the absence of any other member policy (e.g. MEDs), best path will be lowest neighbour IP address
    - Which most likely means that the link to one IXP carries all the traffic; the other link remains relatively empty
    - Could end up with situation with outbound traffic going through one IXP, and return traffic coming through the other IXP
  - AS100 could load balance over the two IXPs by:
    - Setting local preferences on particular announcements from peers
    - Using any BGP community policy implemented by other members

Peering at two local IXPs

- Inbound Traffic Engineering configuration:
  - By default, the link chosen will follow BGP best path rules
    - In the absence of any local policy (e.g. MEDs), best path will be lowest neighbour IP address (i.e. entirely dependent on the address block the IX has received from the RIR)
  - AS100 could load balance over the two IXP links to other members by:
    - Setting MEDs on particular announcements to peers
      - Half the peers could have announcements of MED 10 on one link and MED 20 on the other link
      - And the other half of the peers have the MED values reversed
      - Which assumes that peers even respect MEDs
    - Implementing a BGP community policy available for other members to use
      - Sometimes IXPs recommend what a community policy might be
    - Using AS-PATH prepends (care needed so the IX path doesn't have longer AS path than via paid transit links)

Peering at one local IXP and one regional IXP

□ Diagram shows ISP connecting to one local and one regional IXP

Peering at one local IXP and one regional IXP

- ☐ Regional IXP LAN configuration:
  - ■ Connection to the Regional IXP set up in the same way as the connection to the Local IXP
  - ■ Member has access to same facilities (Route Server, IX services, etc)
- ☐ BGP configuration:
  - ■ EBGP sessions established
    - ☐ With the IXP Route Server (if present)
    - ☐ With the other IXP members
    - ☐ With IXP services infrastructure (if applicable)
- ☐ Traffic Engineering
  - ■ Load balancing across IXP links needed when members are present at both IXPs
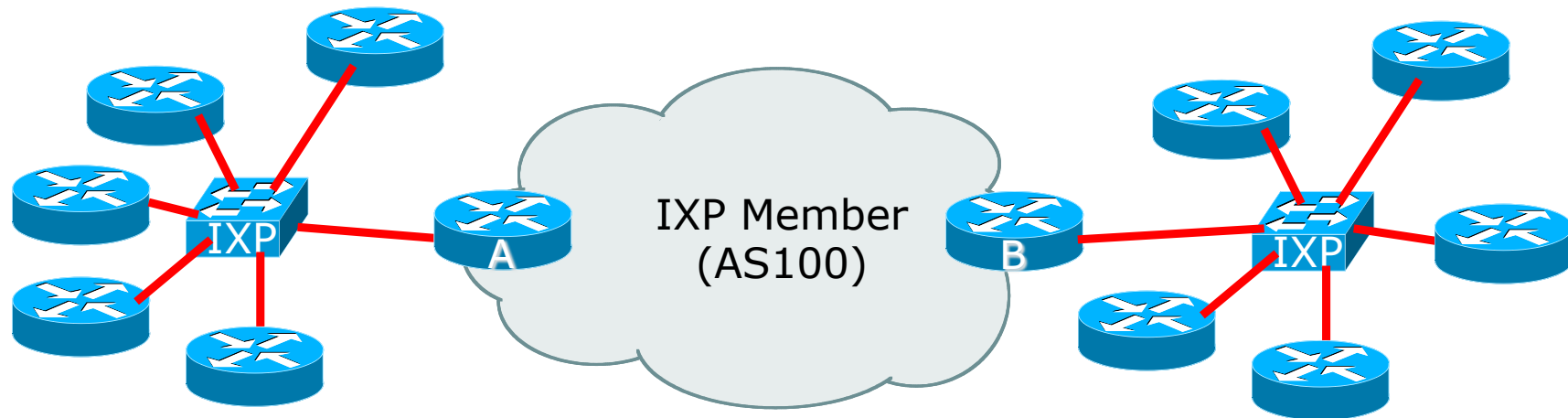
Peering at one local IXP and one regional IXP

- ❑ Outbound Traffic Engineering configuration:
  - ■ By default, the link chosen will follow BGP best path rules
    - ❑ In the absence of any other member policy (e.g. MEDs), best path will be lowest neighbour IP address
    - ❑ Setting local preference on BGP routes learned from different classes of BGP neighbours becomes very important
  - ■ AS100 could prioritise between the IXPs by:
    - ❑ Setting local preferences (see earlier table)
    - ❑ Using any BGP community policy implemented by other members

Peering at one local IXP and one regional IXP

- Inbound Traffic Engineering configuration:
  - By default, the link chosen will follow BGP best path rules
    - In the absence of any local policy (e.g. MEDs), best path will be lowest neighbour IP address (i.e. entirely dependent on the address block the IX has received from the RIR)
  - AS100 needs to prioritise incoming traffic over the local IXP rather than the regional IXP (considered backup)
    - Outbound traffic follows the local preference table in earlier slides
    - Prioritisation can be done by
      - Using AS-PATH prepend (carefully – don't want path to be longer than through transit provider)
      - Subdividing address blocks (de-aggregating) for private peer and local IXP connections, and not subdividing for regional IXP and Transit

Peering at two regional IXPs

- Diagram shows ISP connecting to two different IXPs
  - Could also be the case where one IXP operates two independent sites

Peering at two regional IXPs

- Second IXP LAN configuration:
  - Connection to the second IXP set up in the same way as the connection to the first IXP
  - Member has access to same facilities (Route Server, IX services, etc)
- BGP configuration:
  - EBGP sessions established
    - With the IXP Route Server (if present)
    - With the other IXP members
    - With IXP services infrastructure (if applicable)
- Traffic Engineering
  - Load balancing across IXP links needed when members are present at both IXPs

Peering at two regional IXPs

- Outbound Traffic Engineering configuration:
  - By default, the link chosen will follow BGP best path rules
    - In the absence of any other member policy (e.g. MEDs), best path will be lowest neighbour IP address
    - Which most likely means that the link to one IXP carries all the traffic; the other links remains relatively empty
    - Could end up with situation with outbound traffic going through one IXP, and return traffic coming through the other IXP
    - Not good if the two IXPs have a significant geographical separation
  - AS100 could load balance over the two IXPs by:
    - Setting local preferences on particular announcements from peers, paying close attention to geographical or regional interconnect issues
    - Using any BGP community policy implemented by other members

Peering at two local IXPs

- Inbound Traffic Engineering configuration:
  - By default, the link chosen will follow BGP best path rules
    - In the absence of any local policy (e.g. MEDs), best path will be lowest neighbour IP address (i.e. entirely dependent on the address block the IX has received from the RIR)
  - AS100 needs to prioritise incoming traffic between the two regional IXPs according to geographical needs/issues
    - Outbound traffic afterall follows the local preference table in earlier slides
    - Prioritisation can be done by
      - Using AS-PATH prepend (carefully – don't want path to be longer than through transit provider)
      - Subdividing address blocks (de-aggregating) for private peer and regional IXP connections, and not subdividing for Transit

Agenda

- Peering Priorities
- Transit Provider Peering at an IXP
- Customer Multihomed between two IXP members
- Traffic Engineering for an ISP with two interfaces on one IX LAN
- Traffic Engineering for an ISP connected to two IXes
- Traffic Engineering and CDNs
- Multihoming with Partial Routes: propagating the default using the IGP

Traffic Engineering and CDNs

Traffic Engineering and CDNs

- Each CDN has its own configuration recommendations
  - These slides are only a guideline – it is best to consult directly with the CDN in question about their operational and traffic engineering policies
- CDN implementations:
  - Present at IXP via the IXP Services Infrastructure
    - Transit (backhaul/cache-fill) via one of the IX members or a transit provider or their own infrastructure
  - Peering directly at the IXP
  - Hosted at IX member, and made available to other IX members

CDN at an IXP – on Services LAN
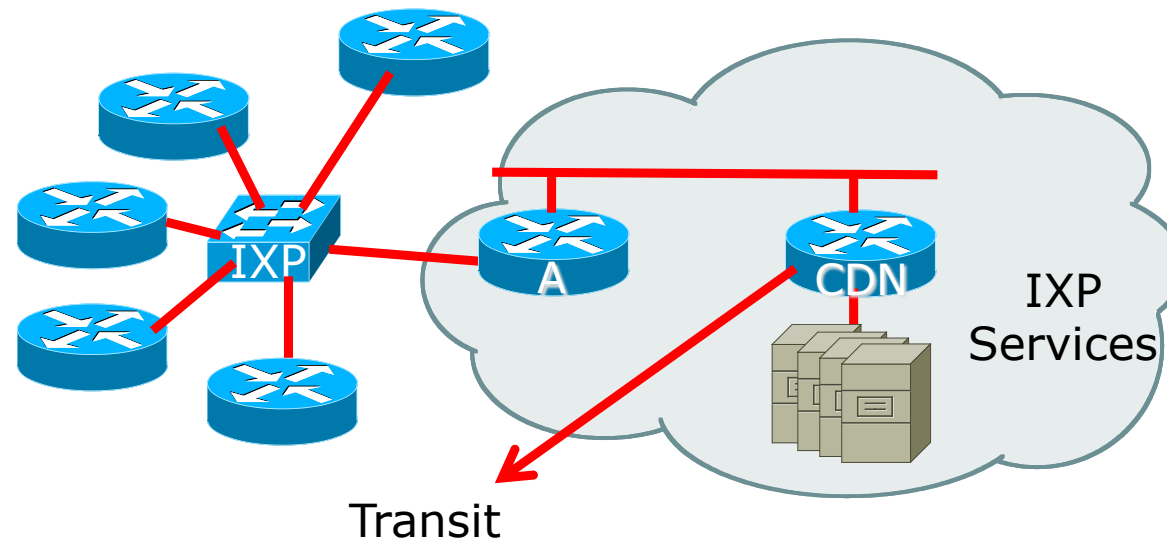
□ Diagram shows content provider hosted on IXP Services LAN
  ▪ Transit connection for Cache Fill

CDN at an IXP – on Services LAN

- BGP configuration:
  - IXP members peer with IXP Services Router (Router A)
  - Receive the routes originated by the CDN
  - IXP Services announces routes to be served to the CDN
  - CDN has its own transit arrangements
    - Either via IXP member or separate infrastructure

CDN at an IXP – on Services LAN

- CDNs usually serve content to operators based on a combination of:
  - Lowest round trip time (latency)
    - End users expect "instant access"
  - BGP announcements of the peer
    - Following most specific announcements
    - AS-path length
    - BGP MED
- Operators need to:
  - Talk to CDN operator about BGP policy!
  - Watch the bandwidth to the CDN
  - Pay attention to BGP announcements

CDN at an IXP – direct peering
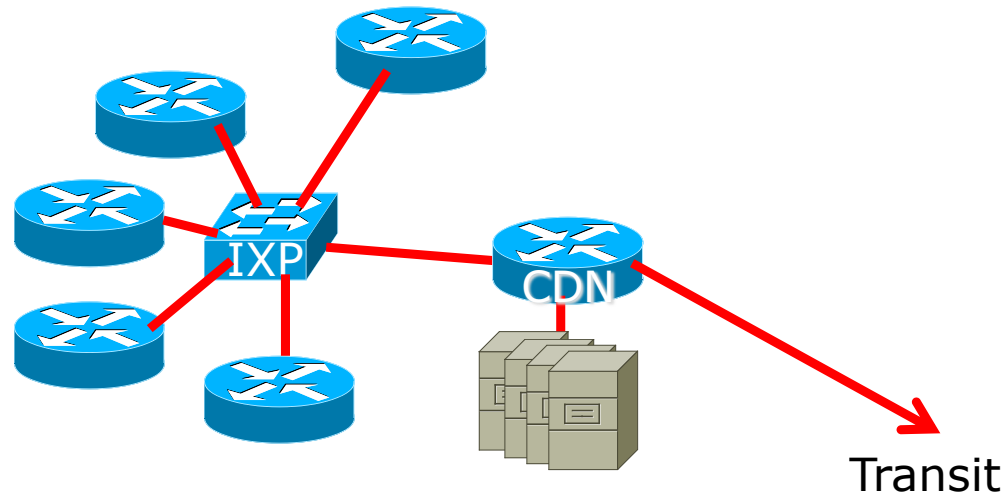
- ☐ Diagram shows content provider peering directly at the IXP
  - ▪ Transit connection for Cache Fill

CDN at an IXP – direct peering

- BGP configuration:
  - IXP members peer with CDN Router
  - IXP members receive the routes originated by the CDN
  - CDN has its own transit arrangements
    - Either via IXP member or separate infrastructure
- Operations:
  - Same as for the previous example

CDN at an IXP – hosted by a member
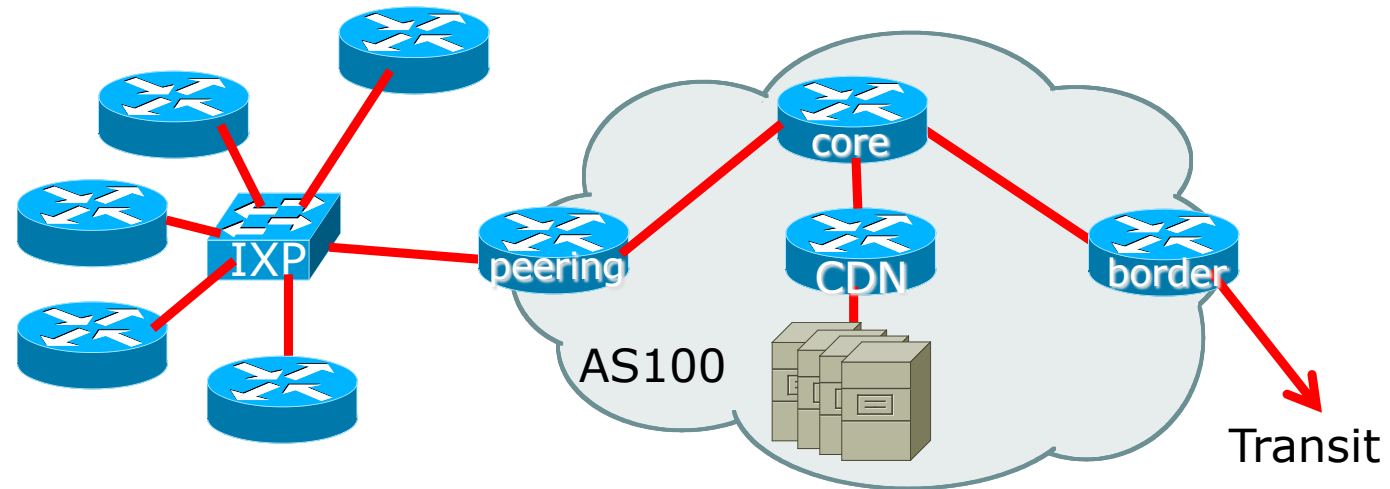
- ❑ Diagram shows content provider hosted by IXP Member
  - ▪ Transit connection for Cache Fill

CDN at an IXP – hosted by a member

- BGP configuration:
  - IXP members peer with AS100 (Peering Router A)
  - IXP members receive the routes originated by the CDN (as well as those originated by AS100)
  - AS100 announces routes to be served to the CDN
    - This could depend on AS100's agreement with each of its peering partners
      - AS100 may charge for access to the CDN content (they have to pay for the backhaul)
      - AS100 may limit access to the CDN content to certain peering partners

CDN at an IXP – hosted by a member

- In addition to the previous advice:
  - Pay attention to the AS path length – CDNs may pay attention to BGP attributes
    - Make sure shortest path to the CDN is via the IXP member, rather than your own transit links (similar case to when the IXP hosts the CDN)
  - Stay in touch with the member who is giving you access to the cache/CDN
    - Especially for any change in policy
    - Especially for any bandwidth or latency issues

Finally: Connection to a CDN in two locations

- Circumstance happens to many operators
  - See the CDN via the local IXP (or local IXP member)
  - See the same CDN through their transit provider
  - How do they ensure that their end-users access the local CDN, and not the one hosted via the transit provider??
- CDNs normally:
  - Pay attention to BGP announcements
    - But will they accept traffic engineering?
  - Pay attention to RTTs
- Solution:
  - Talk to the CDN and discuss the situation
  - They want the best for their "eyeballs" – like the operator wants the best of end-users

Agenda

- Peering Priorities
- Transit Provider Peering at an IXP
- Customer Multihomed between two IXP members
- Traffic Engineering for an ISP with two interfaces on one IX LAN
- Traffic Engineering for an ISP connected to two IXes
- Traffic Engineering and CDNs
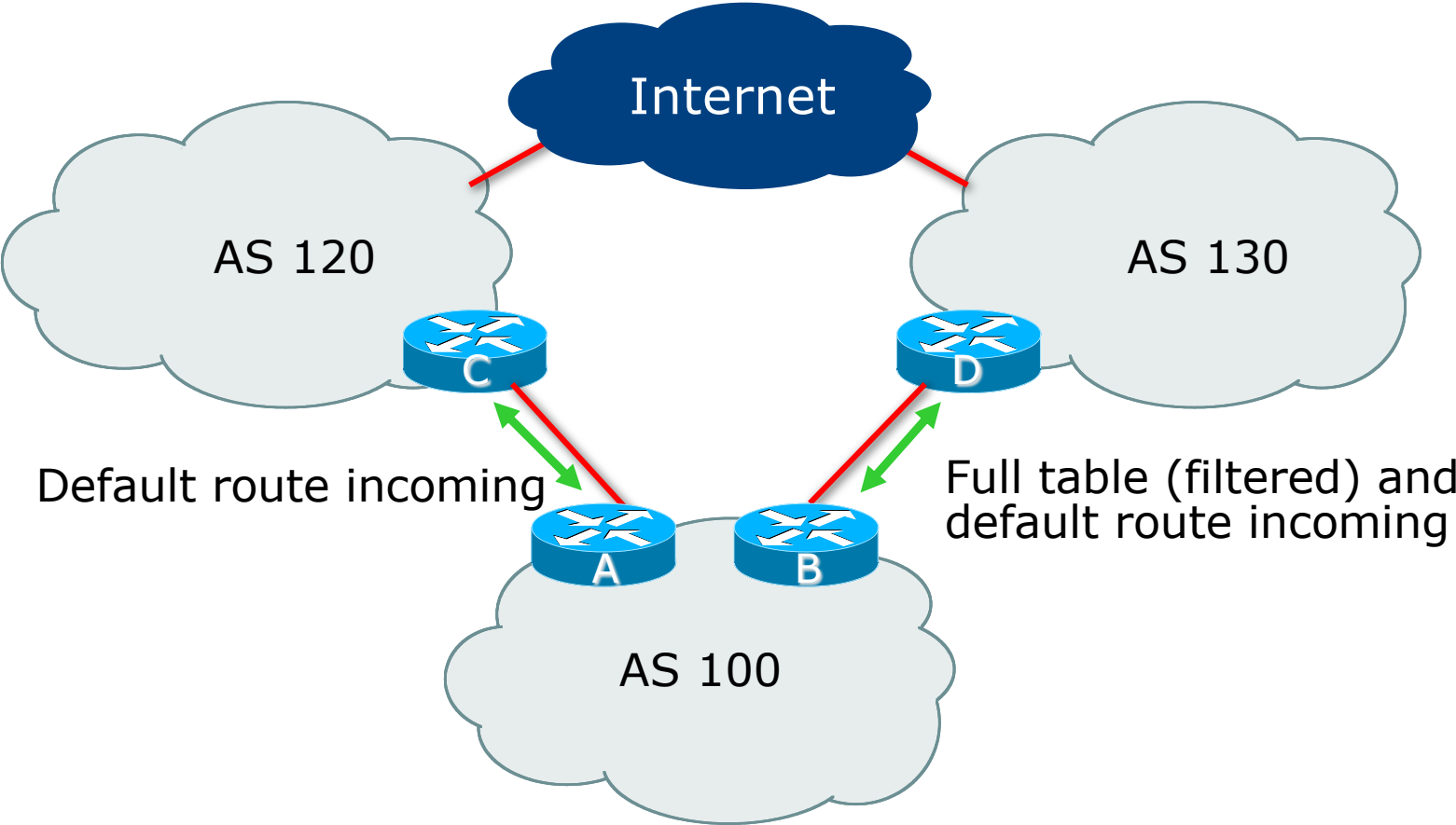- Multihoming with Partial Routes: propagating the default using the IGP

Multihoming with Partial Routes: default route propagation using IS-IS

Default routes in IS-IS

- When using BGP for multihoming, and using only partial routes from upstream providers, it is desirable to propagate a default route across the internal network using IS-IS
  - There isn't a need to carry default in BGP unless providing transit to other operators
    - And in this case you'd opt for full routes in the default free zone
  - A default route in IBGP means only one of the multihoming exit paths is used, whereas operators may want to load balance exit traffic based on "nearest exit" derived from the IGP

Connectivity

Internet

AS 120

AS 130

C

D

Default route incoming

Full table (filtered) and
default route incoming

A

B

AS 100

Partial Routes

- AS120 sends just a default route
- AS130 sends the full BGP table and a default route
- As described in the "Multihoming: Outbound Traffic Engineering" presentation
  - Sufficient routes from AS130 are accepted so as to ensure
    - Close to symmetric inbound/outbound traffic flow
    - Close to equal outbound link loading
  - For the remaining destinations, default route is via AS120, propagated by IBGP
    - And if AS120 link fails totally, then AS130 link takes the entire load

Handling the default route

- With IBGP, the default route is the BGP best path
  - Failure of one path will inevitably result in the default route disappearing from the IBGP while updates propagate from the other Border Router
- Operators may also wish to balance the default route by nearest exit
- Propagating the default route in the IGP:
  - Works around the "disappearing default" failure mode
  - Allows default by nearest exit, if desired

Connectivity
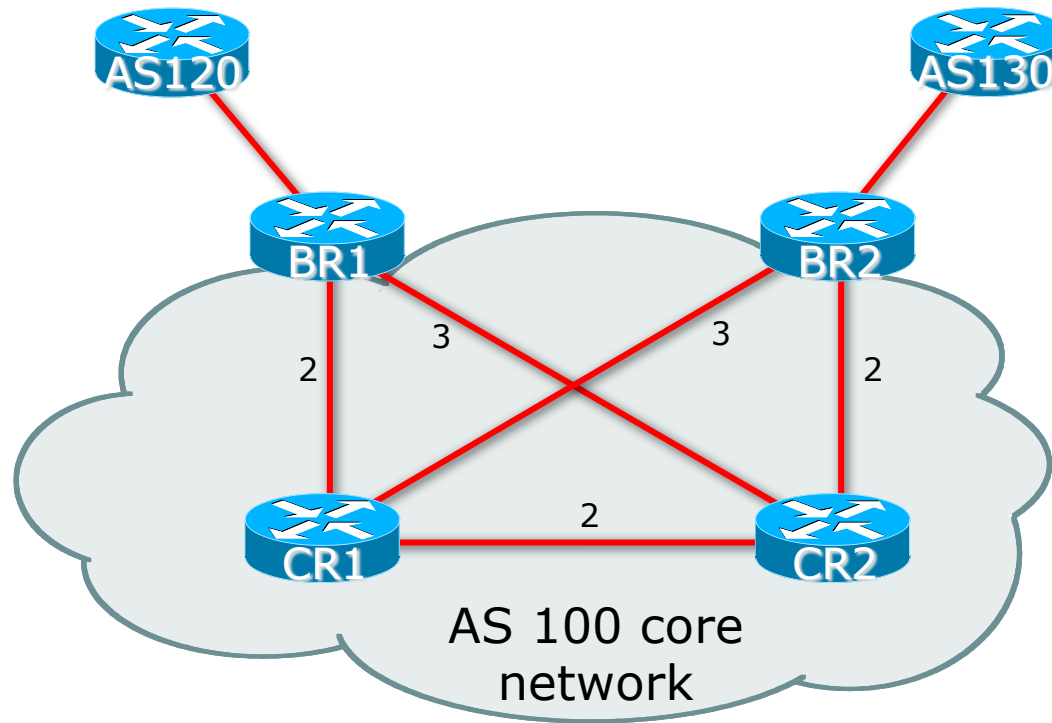
IGP metrics:
→ 2 for direct
→ 3 for crossover

BR1 / AS120 configuration

- AS120 sends the default route
  - Default is tagged with the `no-advertise` community
    - This means it will not propagated to IBGP peers
- BR1 router configuration includes:
  - IGP originating a default route to neighbours if and only if the default is heard by EBGP from the configured EBGP neighbour
  - Setting BGP so that default EBGP protocol distance is lower than that of any IGP (Cisco's default)
    - Failure to do this would mean that the default route originated by the other Border Router into IS-IS would take precedence over the external default learned by EBGP

BR1 / AS120 configuration

```
interface GigabitEthernet1/0
 description Link to AS120 Upstream
 ip address 100.66.10.1 255.255.255.252
!
router isis as100
 <snip>
 default-information originate route-map DEFAULT-ORIG
!
router bgp 100
 address-family ipv4
  distance bgp 20 200 200
  neighbor RR peer-group
  neighbor RR remote-as 100
  neighbor RR description iBGP with RR
  neighbor RR update-source Loopback0
  neighbor RR send-community
  neighbor RR next-hop-self
```

BR1 / AS120 configuration

```
router bgp 100
 address-family ipv4
  neighbor 100.64.0.3 peer-group RR
  neighbor 100.64.0.3 description CR1 Route Reflector
  neighbor 100.64.0.4 peer-group RR
  neighbor 100.64.0.4 description CR2 Route Reflector
  neighbor 100.66.10.2 remote-as 120
  neighbor 100.66.10.2 description AS120 upstream
  neighbor 100.66.10.2 prefix-list AGGREGATE out
  neighbor 100.66.10.2 route-map AS120-IN in
  neighbor 100.66.10.2 activate
!
ip prefix-list DEFAULT permit 0.0.0.0/0
ip prefix-list AGGREGATE permit 100.64.0.0/19
!
```

BR1 / AS120 configuration

```
ip access-list standard BGP-NH-AS120
 remark AS120 BGP speaker
 permit 100.66.10.2
!
route-map DEFAULT-ORIG permit 10
 match ip address prefix-list DEFAULT
 match ip next-hop BGP-NH-AS120
 set metric 10
!
route-map AS120-IN permit 10
 description Tag default with no-advertise
 match ip address prefix-list DEFAULT
 set community no-advertise
!
route-map AS120-IN deny 20
 description Drop all other prefixes
!
```

BR2 / AS130 configuration

- AS130 sends the full BGP table and default route
  - Default is tagged with the `no-advertise` community
    - This means it will not propagated to IBGP peers
  - Remaining prefixes are filtered to accept ones useful for traffic engineering
- BR2 router configuration includes:
  - IGP originating a default route to neighbours if and only if the default is heard by EBGP from the configured EBGP neighbour
  - Setting BGP so that default EBGP protocol distance is lower than that of any IGP (Cisco's default)

BR2 / AS130 configuration

```
interface GigabitEthernet1/0
 description Link to AS130 Upstream
 ip address 100.67.3.2 255.255.255.252
!
router isis as100
 <snip>
 default-information originate route-map DEFAULT-ORIG
!
router bgp 100
 address-family ipv4
  distance bgp 20 200 200
  neighbor RR peer-group
  neighbor RR remote-as 100
  neighbor RR description IBGP with RR
  neighbor RR update-source Loopback0
  neighbor RR send-community
  neighbor RR next-hop-self
```

BR2 / AS130 configuration

```
router bgp 100
 address-family ipv4
  network 100.64.0.0 mask 255.255.224.0
  neighbor 100.64.0.3 peer-group RR
  neighbor 100.64.0.4 peer-group RR
  neighbor 100.67.3.1 remote-as 130
  neighbor 100.67.3.1 prefix-list AGGREGATE out
  neighbor 100.67.3.1 route-map AS130-IN in
  neighbor 100.67.3.1 activate
!
ip prefix-list DEFAULT permit 0.0.0.0/0
ip prefix-list AGGREGATE permit 100.64.0.0/19
!
ip route 100.64.0.0 255.255.224.0 null0 250
!
ip as-path access-list 1 permit <snip>
```

81

BR2 / AS130 configuration

```
ip access-list standard BGP-NH-AS130
 remark AS130 BGP speaker
 permit 100.66.10.2
!
route-map DEFAULT-ORIG permit 10
 match ip address prefix-list DEFAULT
 match ip next-hop BGP-NH-AS130
 set metric 10
!
route-map AS130-IN permit 10
 description Tag default with no-advertise
 match ip address prefix-list DEFAULT
 set community no-advertise
!
route-map AS130-IN permit 20
 match ip as-path 1
!
```

Outcomes: BR1 & BR2

- The BGP table on the BR1 router would look something like this:

```
BR1#sh ip bgp
     Network          Next Hop          Metric LocPrf Weight Path
 *>  0.0.0.0          100.66.10.2                           0 120 i
 * i 100.64.0.0/19    100.64.0.3             0     100      0 i
 *>i                  100.64.0.4             0     100      0 i
...
```

- The BGP table on the BR2 router would look something like this:

```
BR2#sh ip bgp
     Network          Next Hop          Metric LocPrf Weight Path
 *>  0.0.0.0          100.67.3.1                            0 130 i
 * i 100.64.0.0/19    100.64.0.3             0     100      0 i
 *>i                  100.64.0.4             0     100      0 i
...
```

Outcomes: BR1 & BR2

- The Routing table on the BR1 router would be similar to:

```
BR1#sh ip route
Gateway of last resort is 100.66.10.2 to network 0.0.0.0

B*      0.0.0.0/0 [20/0] via 100.66.10.2, 06:45:35
B       100.64.0.0/19 [200/0] via 100.64.0.4, 06:07:39
...
```

- The Routing table on the BR2 router would be similar to:

```
BR2#sh ip route
Gateway of last resort is 100.67.3.1 to network 0.0.0.0

B*      0.0.0.0/0 [20/0] via 100.67.3.1, 06:48:07
B       100.64.0.0/19 [200/0] via 100.64.0.4, 22:57:03
...
```

Outcomes

□ Notice that the best path for the default route on each border router is via the directly attached upstream

□ And on the core routers, for example CR1's Routing table:

```
CR1#sh ip route
Gateway of last resort is 100.64.1.2 to network 0.0.0.0

i*L2  0.0.0.0/0 [115/12] via 100.64.1.2, 06:50:46, GigabitEthernet2/0
...
```

■ Default learned by IS-IS from BR1 (link metric 2 plus configured metric of 10)

■ Outcome will be similar on CR2, with default learned from BR2

Outcomes: Core Router IS-IS RIB

☐ CR1's IS-IS RIB:

```
CR1#sh isis rib
...
0.0.0.0/0
  [115/L2/12] via 100.64.1.2(GigabitEthernet2/0), from 100.64.0.1, tag 0, LSP[3/125]
              (installed)
  [115/L2/13] via 100.64.1.6(GigabitEthernet3/0), from 100.64.0.2, tag 0, LSP[4/119]
```

■ Default routes are learned by IS-IS from both Border Routers
  ☐ The route with the lowest metric (12) is installed in the Global RIB

☐ CR2's IS-IS RIB will look similar

■ Lowest metric path will be to BR2 instead

Traffic flow

Traffic follows nearest
exit default route

Traffic follows nearest
exit default route
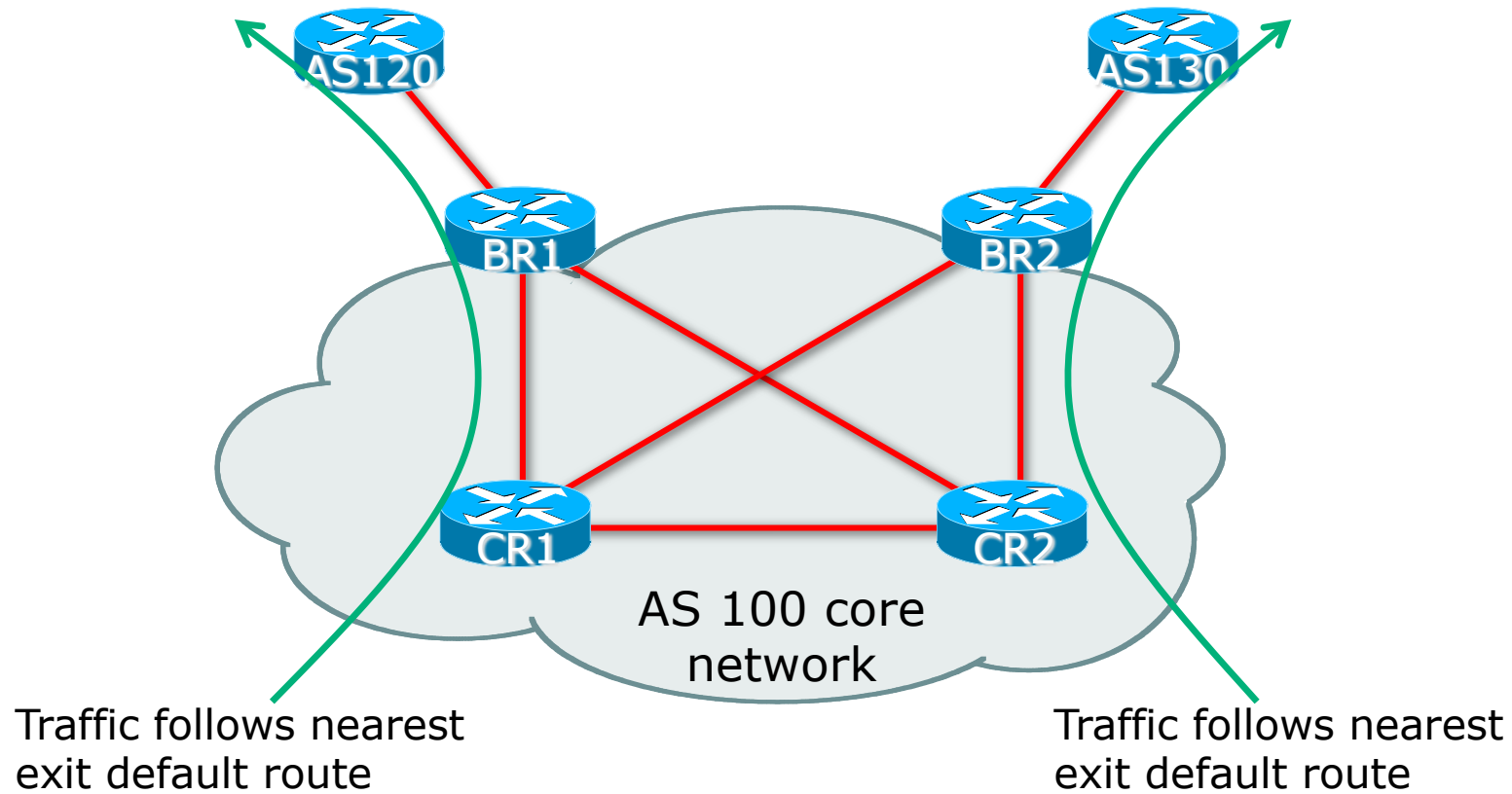
Option: Preferring one Border exit over another

- Some organisations prefer to have one exit as primary path, with the other simply acting as the backup
    - May be receiving some BGP routes via their 2nd link
- Solution: Halve the value of the IS-IS metric in the **DEFAULT-ORIG** route-map on the router that is the preferred exit
    - Example: Making BR2 the preferred default exit:

```
route-map DEFAULT-ORIG permit 10
 match ip address prefix-list DEFAULT
 match ip next-hop BGP-NH-AS130
 set metric 5
 !
```

Deployment Scenarios

- Configuration described here are suitable for and have been deployed at many different organisations:
  - Small ISPs carrying defaults and partial BGP table from their upstream providers
  - Enterprises multihoming between two ISPs
  - Research & Education institutions multihoming between their R&E provider and their commodity Internet provider

Summary

- Propagation of the default via IGP allows for:
  - Swift failover if either external BGP session drops
  - Sending traffic to nearest exit if this is desired

- Much more effective than propagating default by IBGP
  - Note the Cisco IOS specific requirements:
    - EBGP route distance needs to be lower than any IGP on border router
    - IS-IS needs a route-map to conditionally propagate the default route

BGP Case Studies

ISP Workshops