

BGP Case Studies

ISP Workshops



These materials are licensed under the Creative Commons Attribution-NonCommercial 4.0 International license (<http://creativecommons.org/licenses/by-nc/4.0/>)

Last updated 21st May 2018

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

Agenda

- Peering Priorities
- Transit Provider Peering at an IXP
- 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

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
Paid Transit	50

Agenda

- Peering Priorities
- Transit Provider Peering at an IXP
- 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

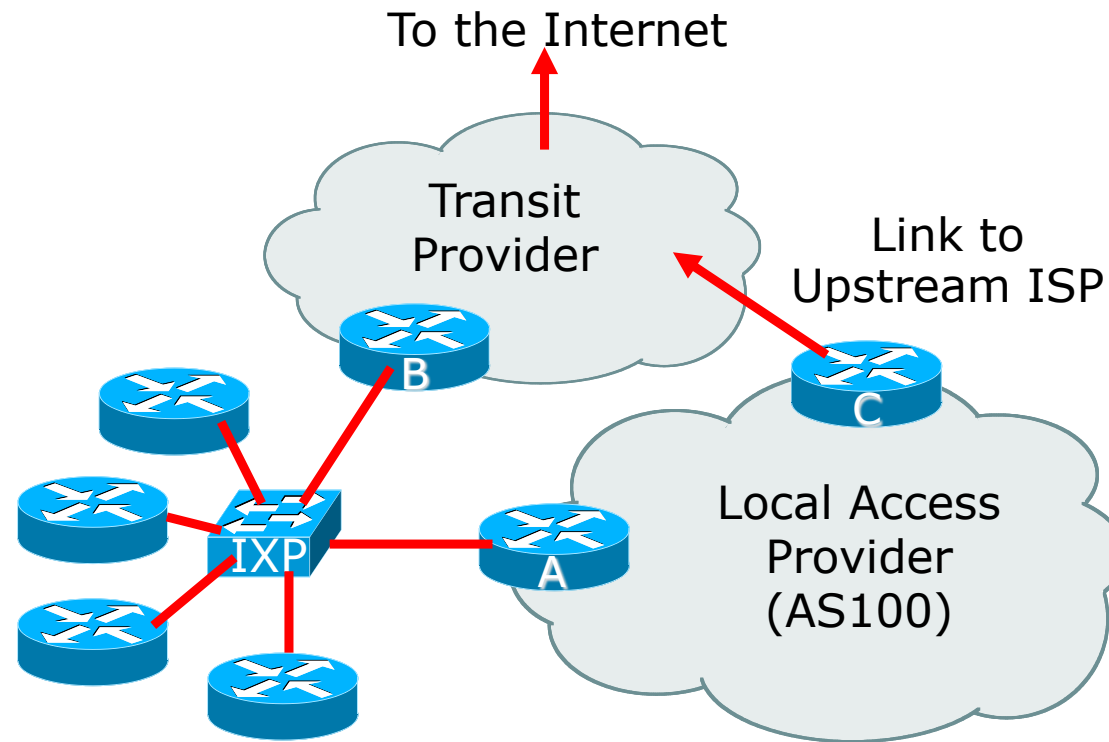
IXP Peering: When a Transit Provider is Also a Peer



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

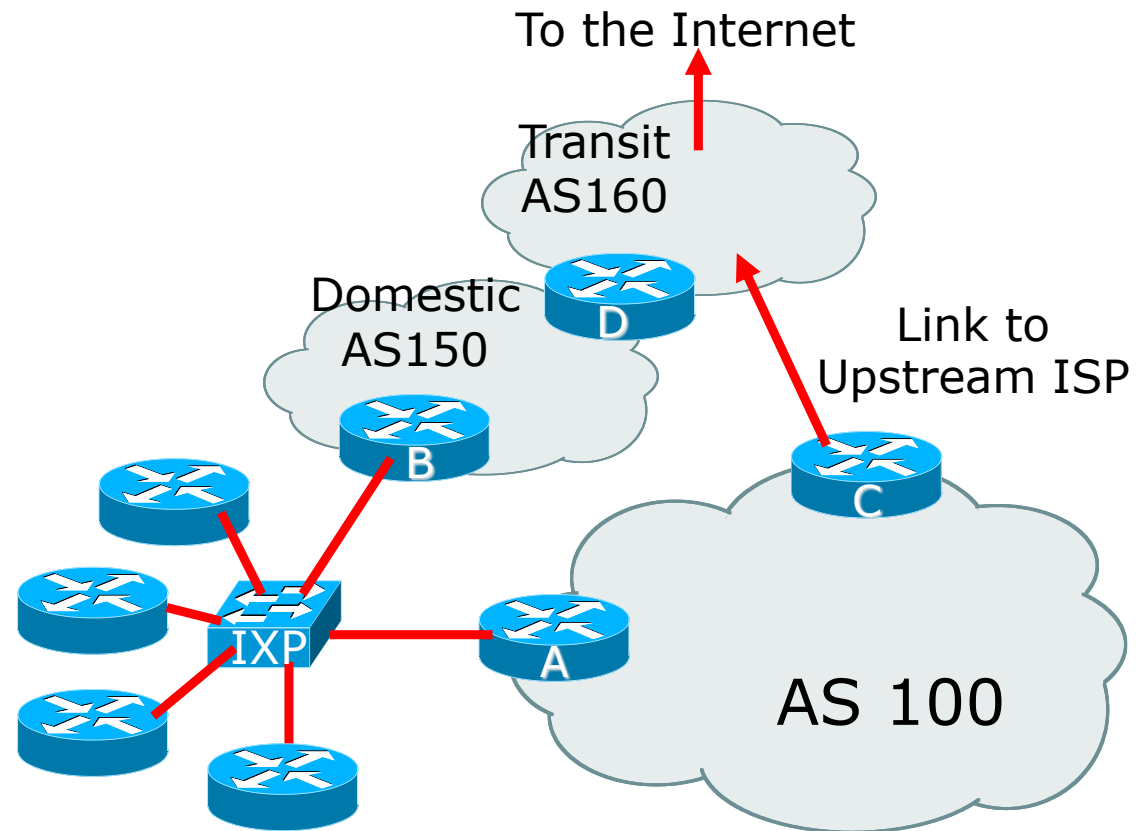
IXP Peering, when Transit Provider is also a Peer



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



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
- 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

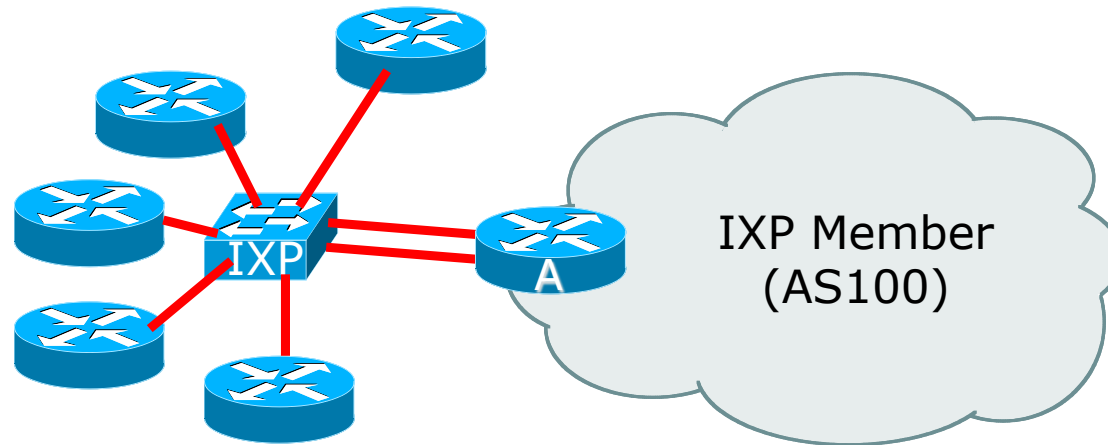
Traffic Engineering over two interfaces connected to one IXP



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



- Diagram shows case of two ethernet links from one router
 - Could also have two routers connecting to the IX

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 not possible if the member opts to provision two routers 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
- 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

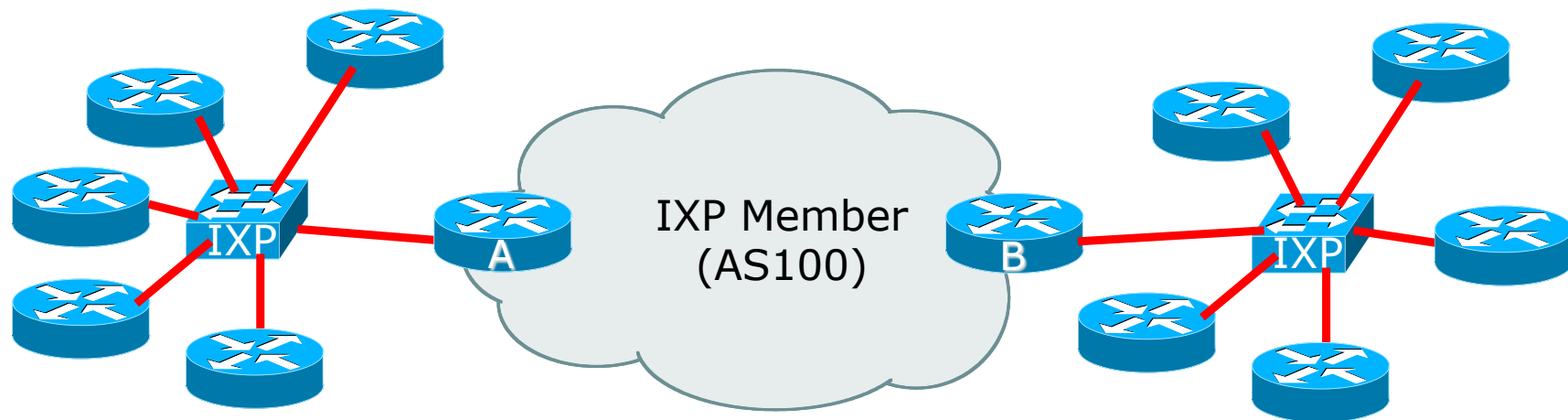
Traffic Engineering when connected to two IXPs



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

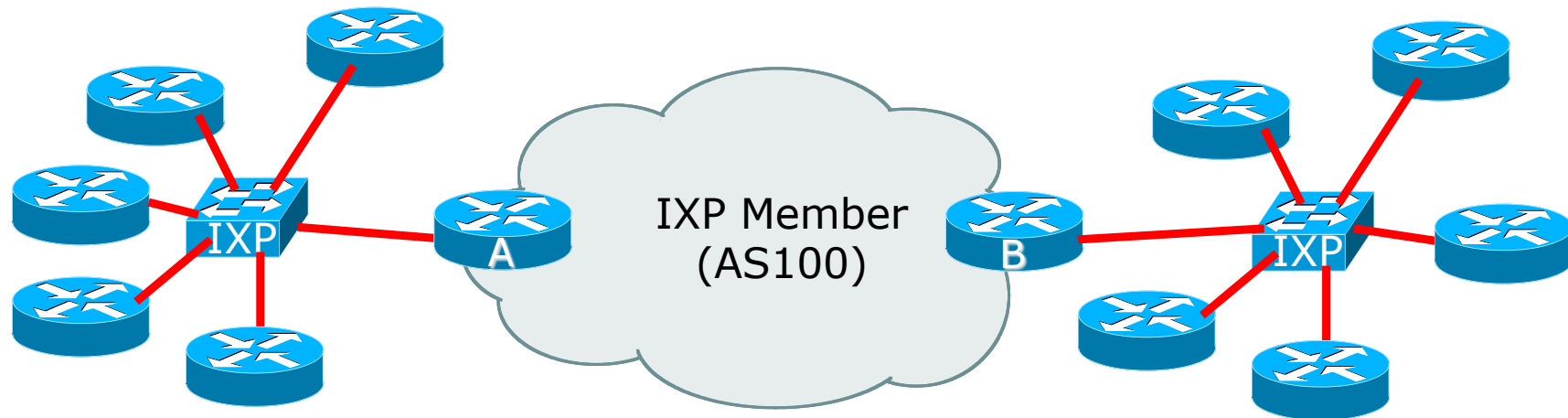
Peering at two local IXPs

- 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

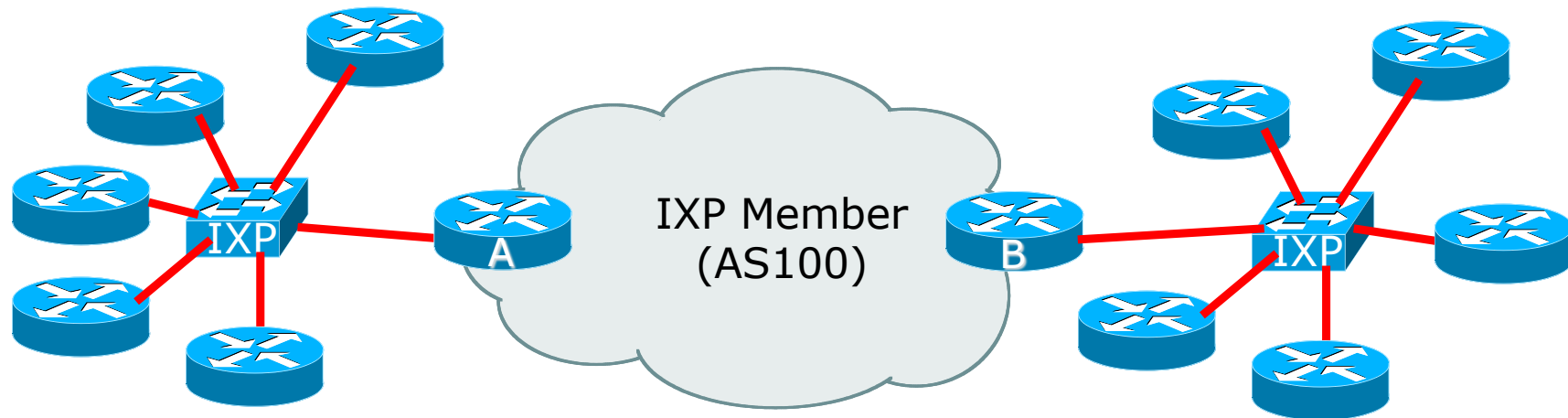
Peering at one local IXP and one regional IXP

- 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 after all 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
- 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

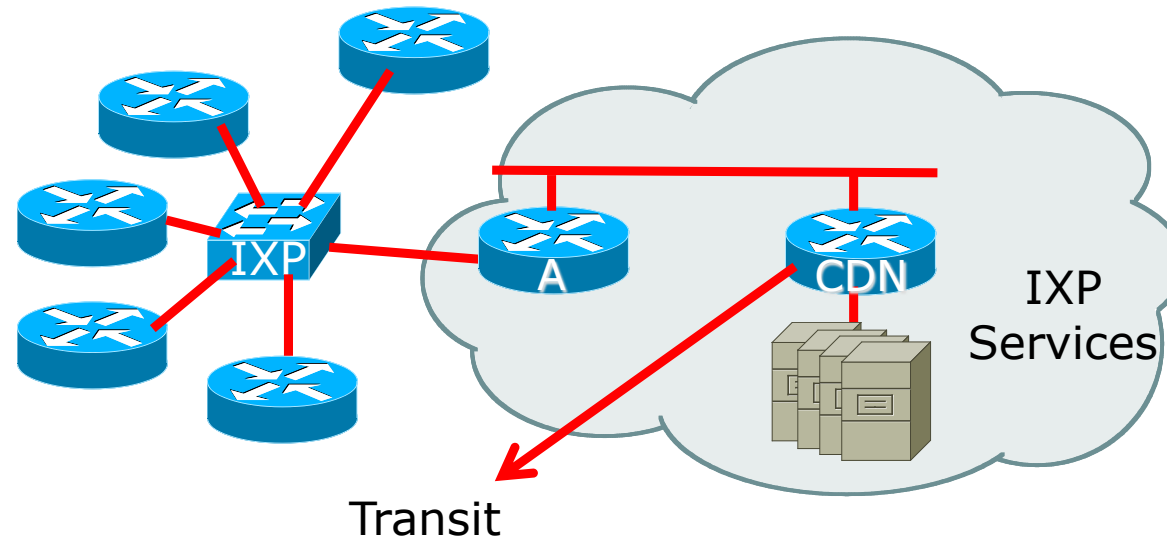
Traffic Engineering and CDNs



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

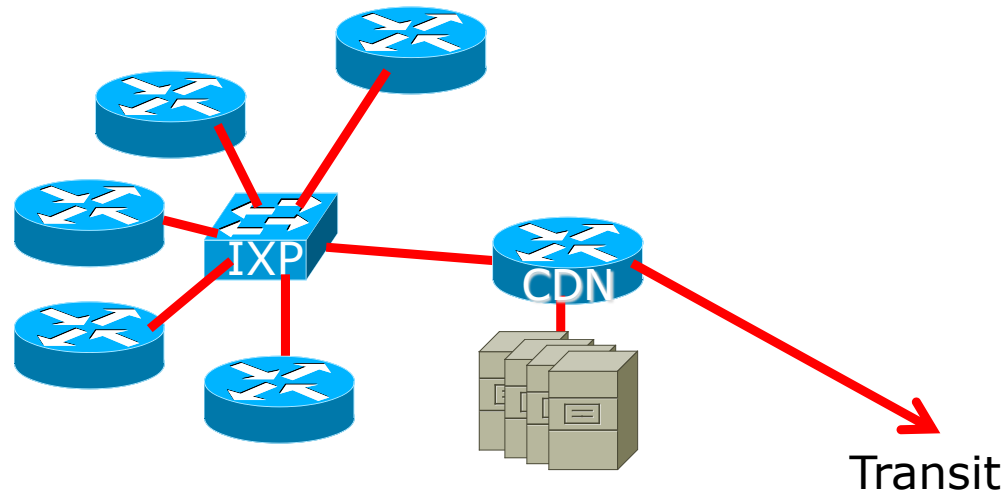
□ BGP configuration:

- 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

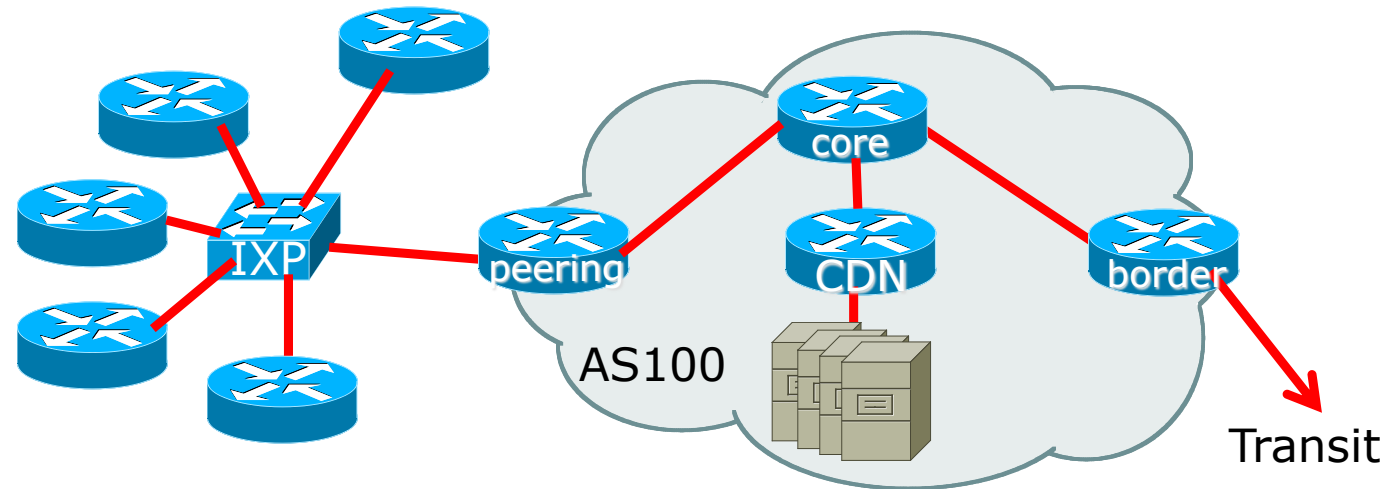


- Diagram shows content provider peering directly at the IXP
 - Transit connection for Cache Fill

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

BGP Case Studies



ISP Workshops