BGP Workshop

How to configure BGP for peering

Humberto Galiza
Network Development Engineer ..::.. humbertogaliza@gmail.com

AONOG/AOPF 2021 - Tutorial session 1 - 26th Nov 2021

Disclaimer

Legal & Acknowledgments

- This presentation is intended for educational purposes only and do not replace independent professional judgment.
- Opinions expressed are solely my **own** and **do not** express the views, opinions, products or technologies of **my current or past employers**.
- The information being shared in the presentation haven't hammered any NDAs.
- All the references mentioned are public information and can be found on the Internet or as form of Academic Papers and/or IETF Request For Comments (RFCs).
- Some of the slides used in this presentation belong to or were adapted from Philip Smith's BGP4ALL/NSRC material [1] as well as APNIC BGP training [2], and both are under Creative Commons 4.0 license.
  - A huge thanks to the original creators for their contributions to the Internet community!

Agenda

What to expect?

- Background & Motivation: who wants to peer with whom?
- Peering Requirements
- Configuration Recommendations & Best practices

Brasil Peering Forum

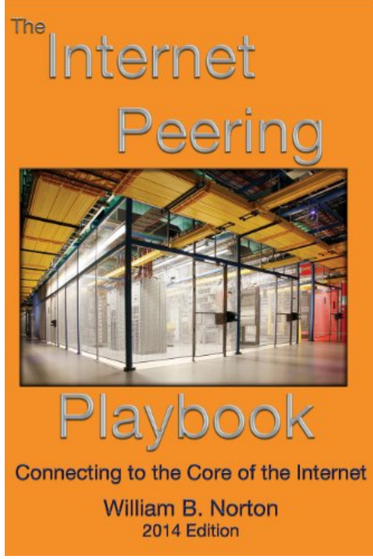
O Brasil Peering Forum é um NOG (Network Operators Group) onde vários profissionais trabalham com o objetivo de fazer uma Internet Brasileira melhor.

https://wiki.brasilpeeringforum.org

Maior e **melhor** repositório de artigos técnicos das áreas de redes & Internet. Visite, contribua, participe!

Motivation

Who wants to peer with whom?



https://amz.run/504z

Today, the vast majority of content and resources consumed by end-users is available by peering:

- The multi-national content providers (Twitter, Google, Facebook, etc)
- The multi-national “cloud” providers (AWS, GCP, Azure, etc)
- Private cross connects (e.g.: PNI)
- Internet Exchange Points (IXPs)

Network Operators Goal

Minimise the cost of operating the business

Transit

- ISP has to pay for circuit (international or domestic) and data (usually per Mbps)
- Repeat for each transit provider
- Significant cost of being a service provider

Peering

- ISP shares circuit cost with peer (private) or runs circuit to public peering point (one off cost)
- No need to pay for data
- Reduces transit data volume, therefore reducing cost

Peering 101 [1]

Foundations: Border Gateway Protocol

BGP (RFC 4271) is quite literally the protocol that makes the Internet work

- The Internet is essentially a set of interconnected **Autonomous System**'s **(AS)**.
- A unique **ASN** (AS number) is allocated to each AS for use in BGP routing.

A **peer** is another AS with which the local network has agreed to exchange locally sourced routes and traffic

Private peer

- Private link between two providers for the purpose of interconnecting

Public peer

- Internet Exchange Point, where providers meet and freely decide who they will interconnect with

- **Recommendation: peer as much as possible!**

https://twitter.com/iwiizkiid/status/1445737084178690048

Peering 101 [2]

By interconnection type

Private Network Interconnection - PNI

- Where two network operators agree to interconnect their networks, and exchange their respective routes, for the purpose of ensuring their customers can reach each other directly over the peering link

Settlement Free Interconnection - SFI

- No traffic charges
- The most common form of peering

Paid Peering

- Where two operators agree to exchange traffic charges for a peering relationship

By agreement type

Multilateral Peering Exchange (MLPE)

- Takes place at Internet Exchange Points, where operators all peer with each other via a Route Server

Bi-lateral Peering

- Very similar to Private Peering, but usually takes place at a public peering point (IXP)

Mandatory Multilateral Peering

- Where operators are forced to peer with each other as condition of IXP membership
- Strongly discouraged: Has no record of success

Peering 101 [3]

By policy

Open Peering

- Where an ISP publicly states that they will peer with all parties who approach them for peering
- Commonly found at IXPs where ISP participates via the Route Server

Selective Peering

- Where an ISP's peering policy depends on the nature of the operator who requests peering with them
- At IXPs, operator will not peer with RS but will only peer bilaterally

Restrictive Peering

- Where an ISP decides who its peering partners are, and is generally not approachable to considering peering opportunities

At the end, it is all about relationship ;-)

The Peering Database documents ISPs/OTTs peering policies

- https://www.peeringdb.com

All AS operators should register in the PeeringDB

- All operators who are considering peering or are peering must be in the PeeringDB to enhance their peering opportunities

Participation in peering fora is encouraged too

- Global Peering Forum (GPF) – (for North American peering)
- The African Peering & Interconnection Forum - www.afpif.org
- Many countries now have their own Peering Fora (e.g., ngPIF)

peeringdb.com/ix/421

PeeringDB

Search here for a network, IX, or facility.

Advanced Search

Register or Login

Angola IXP

Peers	Connections	Open Peers	Total Speed	% with IPv6
12	16	6	79G	62

Organization	Angola IXP
Also Known As	IXP - Angola
Long Name	Angola Internet Exchange point
City	Luanda
Country	AO
Continental Region	Africa
Media Type	Ethernet
Service Level	Not Disclosed
Terms	Not Disclosed
Last Updated	2021-05-13T09:34:03Z
Notes	

Contact Information

Company Website	http://www.angola-ixp.ao/
Traffic Stats Website	https://monitor.angola-ixp.ao/cacti/graph.php?local_graph_id=208&rra_id=0
Technical Email	rogerio@velonet.net
Technical Phone	
Policy Email	
Policy Phone	
Sales Email	
Sales Phone	

LAN

MTU	1500
IX-F Member Export URL Visibility	Private

Peers at this Exchange Point

Filter

Peer Name IPv4	ASN IPv6	Speed
Angola IXP Route Servers 196.223.1.8	37383 2001:43f8:390::8	10G
Angola IXP Route Servers 196.223.1.9	37383 2001:43f8:390::9	10G
Facebook Inc AS63293 196.223.1.51	63293 2001:43f8:390::33	20G
Gransy s.r.o. - anycastdns.cz 196.49.6.15	60592	1G
IP WORLD 196.223.1.22	328111	
ISC F-ROOT LAD1 196.223.1.49	33558 2001:43f8:390::31	1G
ISC F-ROOT LAD1 196.223.1.50	33558 2001:43f8:390::32	1G
MSTELCOM, SARL 196.223.1.11	17400	1G
Packet Clearing House 196.223.1.31	3856	1G
Packet Clearing House AS42 196.223.1.30	42	1G
Paratus Africa	33763	10G

peeringdb.com/ix/1007

PeeringDB

Search here for a network, IX, or facility.

Advanced Search

Register or Login

angonix

Peers	Connections	Open Peers	Total Speed	% with IPv6
21	24	7	223G	79

Organization	angonix
Also Known As	
Long Name	Angola Internet Exchange Point
City	Luanda
Country	AO
Continental Region	Africa
Media Type	Ethernet
Service Level	Not Disclosed
Terms	Not Disclosed
Last Updated	2021-04-08T09:52:36Z
Notes	

Contact Information

Company Website	http://angonix.net/
Traffic Stats Website	http://angonix.net/about-angonix/statistics/
Technical Email	support@angonix.net
Technical Phone	+244923166814
Policy Email	support@angonix.net
Policy Phone	
Sales Email	
Sales Phone	

LAN

MTU	0
IX-F Member Export URL Visibility	Private

Prefixes

Peers at this Exchange Point

Filter

Peer Name IPv4	ASN IPv6	Speed	Policy
Angola Cables 196.11.234.1	37468 2001:43f8:9d0::925c:0:1	40G	Selective
angonix Route Servers 196.11.234.252	327788 2001:43f8:9d0:0:5:6c:0:1	40G	Open
angonix Route Servers 196.11.234.253	327788 2001:43f8:9d0:0:5:6c:0:2	40G	Open
Banco BAI 196.11.234.28	36936	1G	Selective
CMC Networks 196.11.234.14	25818	1G	Selective
Congo Telecom 196.11.234.26	37451 2001:43f8:9d0::924b:a:1	1G	Selective
Facebook Inc AS63293 196.11.234.45	63293 2001:43f8:9d0::f73d:0:1	20G	Selective
G8 196.11.234.22	28329 2001:43f8:9d0::6ea9:0:1	1G	Selective
i3D.net 196.11.234.19	49544 2001:43f8:9d0::c188:0:1	1G	Selective
IP WORLD 196.11.234.13	328111		Open
LINK BARATO.COM TELECOMUNICACOES 196.11.234.24	268331 2001:43f8:9d0::4182:b:1	1G	Selective
Microsoft 196.11.234.40	8075 2001:43f8:9d0::1f8b:0:40	10G	Selective

Real-world: where to begin set up peering?

- What resources are needed?
- What equipment is needed?
- What are the routing protocol requirements?

Resource Requirements

Assumptions

- Operators who are embarking with peering for the first time presumably:
  - Already have their own IP address space
  - Already have their own ASN
  - Already use BGP to talk with their upstream service providers
- If the operator only has a static connection to a single upstream provider, there is more work to be done to prepare the network for peering
  - Consult these two presentations for more information
    - **https://bgp4all.com/pfs/_media/workshops/06-transitioning-to-bgp.pdf**
    - **https://bgp4all.com/pfs/_media/workshops/10-multihoming-deployment.pdf**

Strategy [1]

Private or Public Peering?

Private peering

- Scaling issue, with costs, number of providers, and infrastructure provisioning

Public peering

- Makes sense the more potential peers there are (more is usually greater than “two”)

Which public peering point?

- Local Internet Exchange Point: great for local traffic and local peers
- Regional Internet Exchange Point: great for meeting peers outside the locality, might be cheaper than paying transit to reach the same consumer base

Which IXP?

How many routes are available?

- What is traffic to & from these destinations, and by how much will it reduce cost of transit
  - NetFlow/sFlow can help here

What is the cost of co-lo space?

- If prohibitive or space not available, pointless choosing this IXP

What is the cost of running a circuit to the location?

- If prohibitive or competitive with transit costs, pointless choosing this IXP

What is the cost of remote hands/assistance?

- If no remote hands, doing maintenance is challenging and potentially costly with a serious outage

Strategy [2]
What should operators do?

Many operators participate in their local IXP - I'd say, it is a MUST!

- Keeps local traffic local
- Reduces latency & transit costs for local traffic
- Gives best experience to the end-user for content

Many operators also purchase connectivity (bandwidth/capacity) to Regional IXPs

- Bandwidth as IPLC (international private leased circuit)
  - **NOT** buying transit to the Regional IXP
- And establish peering across the IX fabric
- And establish PNI with major content operators for Cache fill

Equipment requirements [1]

A dedicated peering router is required

Peering can be done from existing core or border (connecting to upstream) routers, but there are risks involved with that

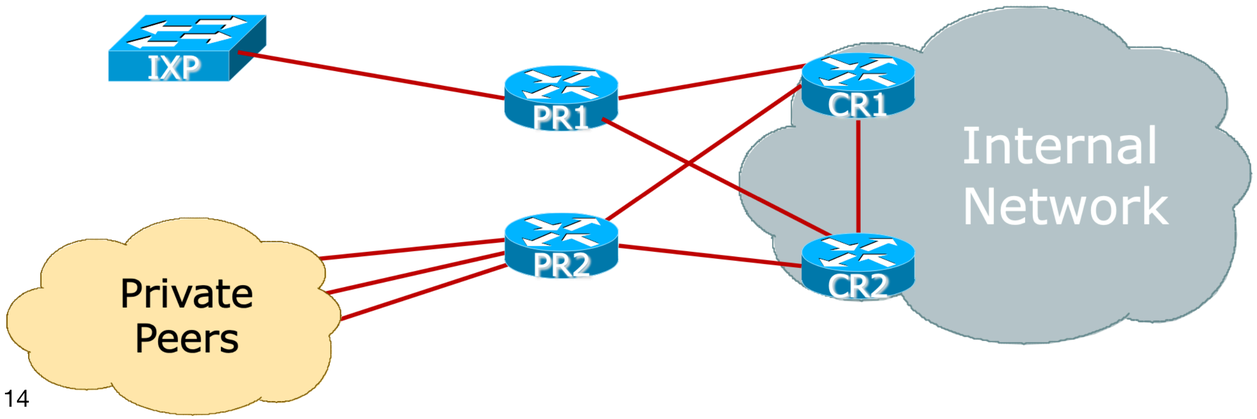
Consider separating routers used for private peering from those used to connect to Internet Exchange Points (IXP)

Requirements

- To be able to support BGP
- To be able to handle the expected traffic volume
- Sufficient external interfaces to connect to peers (or the IXP)
- Two or more internal interfaces
  - Common today for border & peering routers to have at least four ethernet ports (one used external facing, the other three internal facing)

Role Functions

- eBGP with peers
- iBGP and IGP (OSPF/IS-IS) with core devices
- Traffic engineering/Policy implementation via BGP
- Initial protection of the core network with packet filters



Equipment requirements [2]

1RU router is commonly chosen for IXP peering

- Few interfaces needed
- But high throughput needed
- Examples: Juniper MX204, Cisco NCS 540X, etc
- **Note Well:**
  - Use a Router
  - Never and L3 switch
    - Very hard (if not impossible) to disable all the L2 features of an ethernet switch to make it work as an IXP peering router
    - FIB limits could be challenging (for bigger IXPs)

Peering priorities [1]
What does this mean for setting routing policy?

Transit providers are last resort

- They **cost money**!

Internet Exchange Point peers are a priority

- No cost traffic interconnect via a third party L2 infrastructure
- Bi-lateral peers are higher priority than those via the Route Server

Private peers are higher priority than IXP peers

- Direct interconnect does not involve a third party
- Can be deemed “more reliable” and “higher capacity” than the IXP, therefore more dependable

BGP and static customers are of highest priority of all

- They **earn money**!

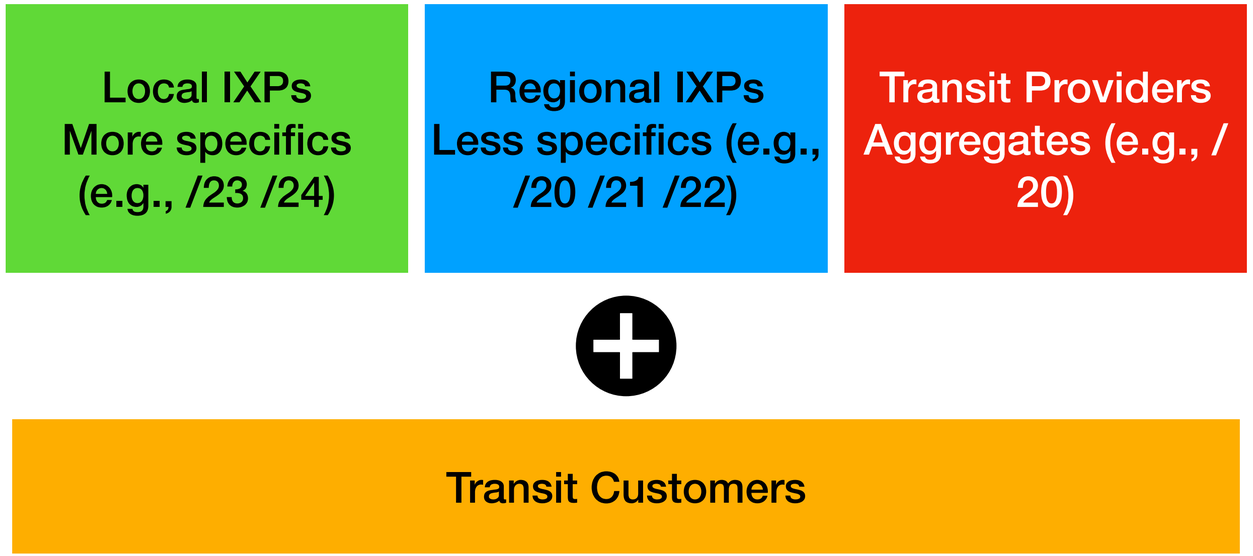
Peering priorities [2]

What does this mean for setting routing policy?

Setting preferences for outbound traffic (BGP received routes)

Connection	Local Preference
Customers	250
Private Peers	200
IXP Bi-Lateral Peers	175
IXP RS Peers	150
Default	100
Transit/Upstreams	50

Setting preferences for inbound traffic (BGP announced routes)



Configuration Recommendations [1]

- Internet Exchange Points usually have “rules” for new members connecting to their IXP fabric
  - Consult the Euro-IX Best Current Operational Practice pages:
    - **https://www.euro-ix.net/en/forixps/set-ixp/ixp-bcops/**
    - Especially the technical recommendations
- Private Peers will usually have requirements for interconnection as well
  - Some form of “contract” document or agreement, which will include technical recommendations, contact details etc.

Configuration Recommendations [2]

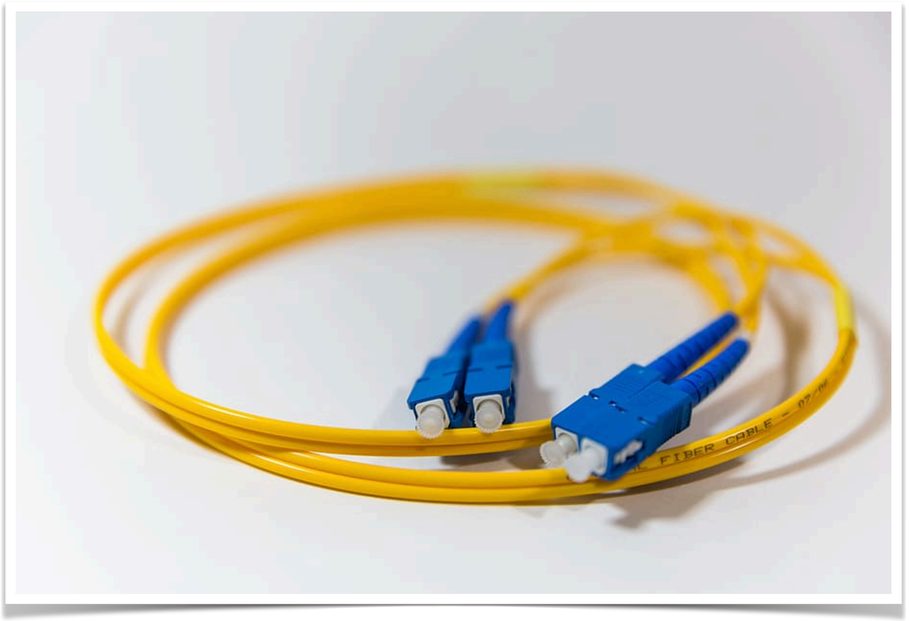
Physical interface connecting to an IXP

Cat5E (or Cat6) cable if:

- Physically close to the IXP (same room, adjacent rack)
- 100Mbps or 1Gbps link
- Switch supports it

Single mode fibre patch:

- To IXP switch if in same facility
- To transmission equipment if IXP is remote
- Use SFP if 1Gbps, SFP+ if 10Gbps, etc



Fibre optics are almost always preferred and are relatively inexpensive

Usually the IXP will supply the SFP needed for their switch

Configuration Recommendations [3]

Physical Interface configuration notes

- Use the LAN subnet address (IPv4/IPv6) provided by the IXP
- Disable:
  - Proxy ARP
  - Forwarding of Directed Broadcasts
  - Sending of ICMP Redirect messages
  - All discovery protocols (eg CDP, LLDP)
  - IPv6 Neighbour Discovery:
    - Router Advertisements
    - IPv6 Routing Prefix Advertisement

Cisco IOS Example:

```
interface Gig 0/0/1
description IXP LAN
ip address 192.0.2.10 255.255.255.0
no ip redirects
no ip proxy-arp
no ip directed-broadcast
no cdp enable
ipv6 address 2001:DB8:1:1::a/64
no ipv6 redirects
ipv6 nd prefix default no-advertise
ipv6 nd ra suppress all
!
```

Configuration Recommendations [4]

Disable unnecessary services

- Turn off/don’t enable unneeded services including:
  - DHCP server
  - BOOTP server
  - TFTP server
  - HTTP & HTTPS servers
  - Listeners for low TCP/UDP ports
  - CDP/LLDP
  - DHCP relay

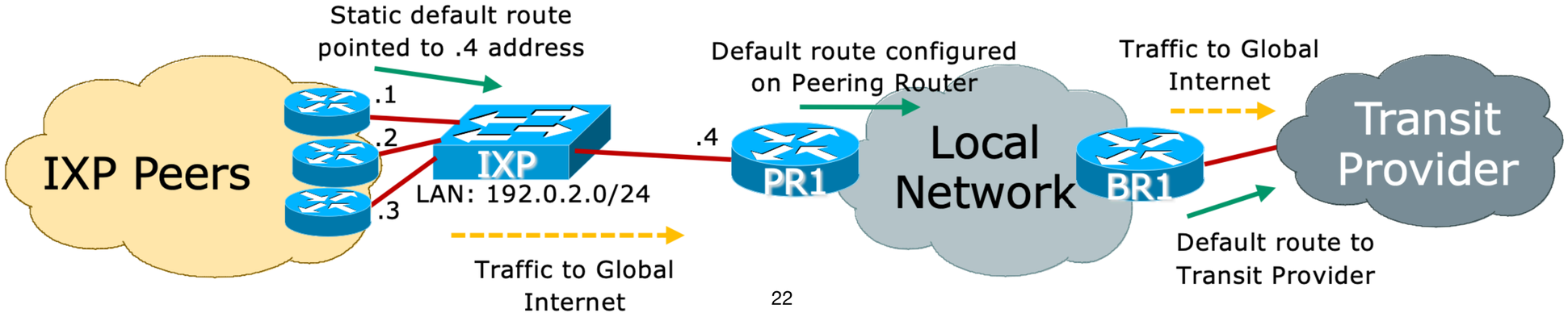
Cisco IOS Example:

```
no service dhcp
no ip bootp server
no tftp-server <Argument>
no ip http server
no ip http secure-server
no service tcp-small-servers
no service udp-small-servers
no cdp run
interface Gigabit 0/0/1
  no ip helper-address <DHCP server>
```

Routing best practices [1]
Routing Configuration

Peering router only carries routes that peers should receive

- No defaults (in BGP, or OSPF/IS-IS, or static pointing to core)
- No full BGP table
  - This is so that peers can't accidentally/deliberately (?) transit your network by pointing a default route at your router
  - (Packet filters could be used, but that's both a denial of service vector and potentially a severe burden on CPU based routers)



Routing best practices [2]

Routing Configuration [cont]

Point default route to the null (discard) interface

- Disable ICMP unreachable messages being sent
- Incoming packets with no specific entries in the forwarding table will be silently discarded
  - Much more efficient than packet filtering

Cisco IOS Example:

```
interface Null0
  no ip unreachables
  no ipv6 unreachables
!
ip route 0.0.0.0 0.0.0.0 null0
ipv6 route ::/0 null0
```

Routing best practices [3]

Routing Configuration [cont]

Never configure an IGP on the peering interfaces

- Especially for IXPs!
- Avoids accidental leakage of internal routes
- Avoids potentially malicious traffic on the peering LAN
- Check with your vendor implementation how to do this

Cisco IOS OSPF Example:

```
interface Gigabit 0/0/1
 description IXP LAN
 ip address 192.0.2.10 255.255.255.0
 ipv6 address 2001:DB8:1:1::a/64
 ...
!
router ospf 100
 passive-interface Gigabit 0/0/1
 ...
!
```

Cisco IOS IS-IS Example:

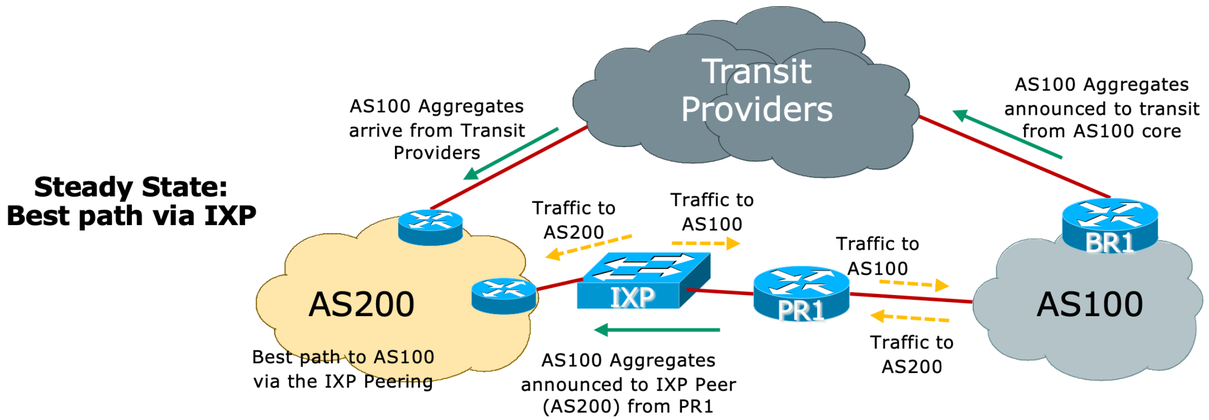
```
interface Gigabit 0/0/1
 description IXP LAN
 ip address 192.0.2.10 255.255.255.0
 ipv6 address 2001:DB8:1:1::a/64
 ...
!
router isis ISP
 passive-interface Gigabit 0/0/1
 ...
!
```

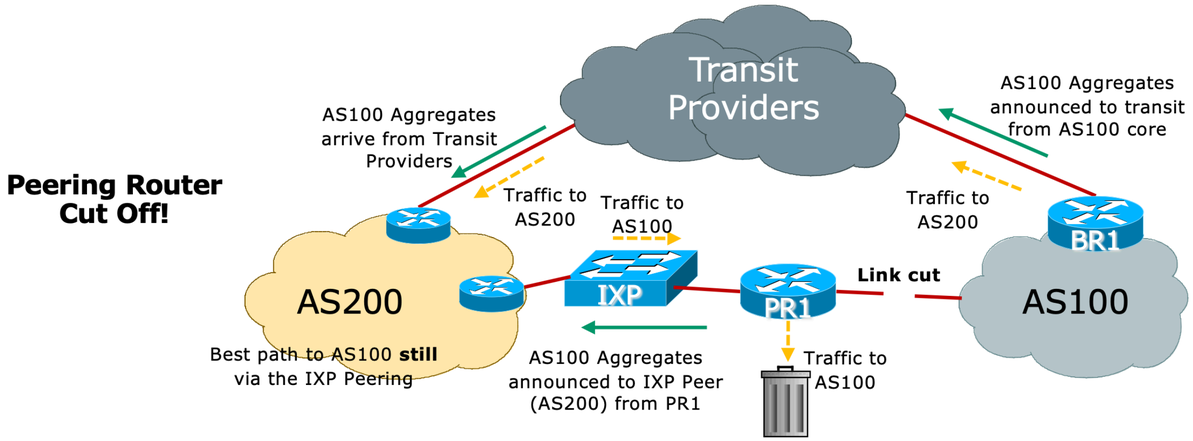
Routing best practices [4]

Routing Configuration [cont]

Don’t originate any prefixes into BGP on the IXP peering router

- If this router is cut off from network core, it will still originate prefixes and likely still be best path, breaking your backup via your Transit Providers



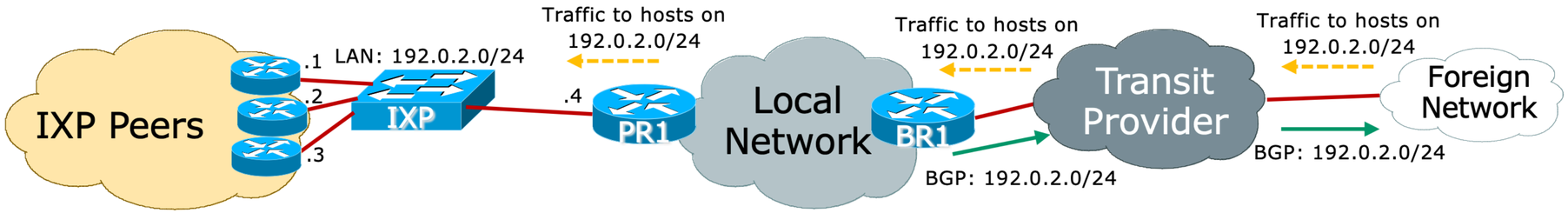


Routing best practices [5]

Routing Configuration [cont]

The IXP LAN subnet must never be carried in BGP

- Use the iBGP “next-hop-self” feature
- Carrying it in IGP is okay so that traceroutes don’t appear broken
- If the IXP LAN carried in iBGP, chances are it might leak to your eBGP and out to the Global Internet, **which means**:
  - Other networks can now transit your network to get access to all IXP peers!!
  - Because IXP LAN is publicly known - and it takes little trial and error to work out which peer is on which IXP address
- Some IXPs are now signing their IXP LAN with the AS0 (zero) ROA - but members need to do their part too!

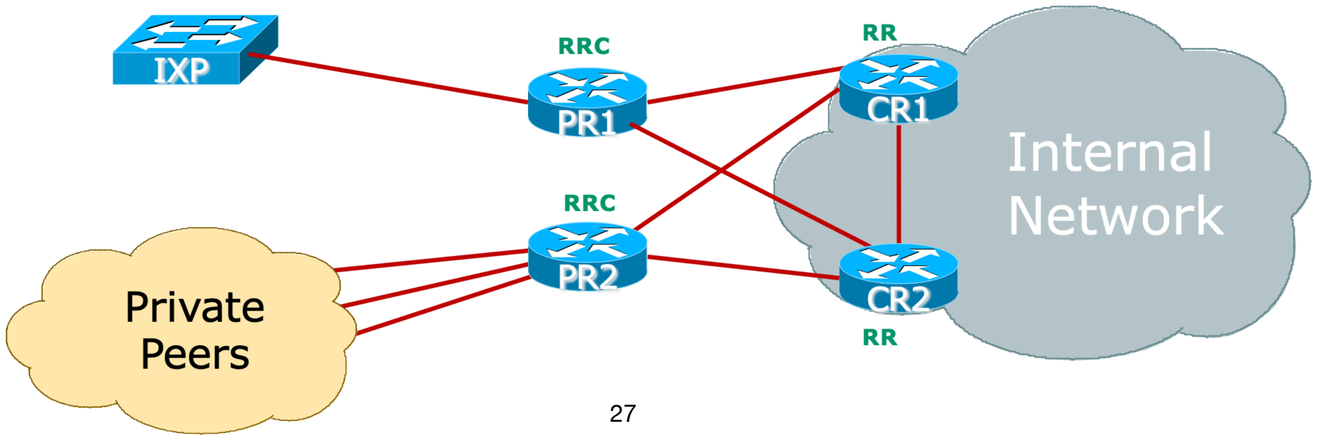


BGP configuration tips [1]

BGP Communities

Use BGP Communities wherever possible!

- Make Peering Router a Route-Reflector-Client (RRC)
  - Running core routers as Route Reflectors (RR) (or having dedicated devices for this function) is standard practice
  - Only announce internal prefixes/aggregates to the Peering Router
    - Community make this task easy!
    - See: https://bgp4all.com/pfs/_media/workshops/11-bgp-communities.pdf



BGP configuration tips [2]

eBGP configuration example

- Cisco IOS eBGP configuration Example

```
interface Loopback 0
 ip address 100.64.1.3 255.255.255.255
!
router bgp 64500
 neighbor 100.64.1.1 remote-as 64500
 neighbor 100.64.1.1 description IBGP with Core1 RR
 neighbor 100.64.1.1 send-community both
 neighbor 100.64.1.1 next-hop-self
 neighbor 100.64.1.1 update-source Loopback0
 neighbor 100.64.1.2 remote-as 64500
 neighbor 100.64.1.2 description IBGP with Core2 RR
 neighbor 100.64.1.2 send-community both
 neighbor 100.64.1.2 next-hop-self
 neighbor 100.64.1.2 update-source Loopback0
!
```

BGP configuration tips [3]

Create suitable BGP policies

Always filter all inbound and outbound BGP announcements!

- RFC8212 reminds what default policy should be in the absence of filters
  - Default policy: accept nothing, send nothing
  - Unfortunately most vendors still do not adhere to this requirements, see: https://github.com/bgp/RFC8212

Outbound is going to be same for every peer at IXP

- Create a policy statement to be shared amongst all peers
- Basically matching the communities that get out to peers
  - Aggregates, any BGP customers, etc
- Cisco IOS: route-map and peer-group

BGP configuration tips [4]

Create suitable BGP policies [cont]

Inbound policy is going to have two parts:

- A per-peer prefix filter
- A uniform policy for all peers:
  - Setting Local Preference High
  - Assign a specific “IXP" community
- Remember the Local Preference values in the Peering Priorities discussed earlier

- For Internet BGP, Peering router needs to carry all customer routes, the aggregates, and subnets of the aggregates

BGP configuration tips [5]

eBGP configuration example

- Cisco IOS eBGP configuration Example

```
router bgp 64500
 neighbor 192.0.2.10 remote-as 64505
 neighbor 192.0.2.10 description Bi-lateral Peering with Peer-10
 neighbor 192.0.2.10 prefix-list PEER-10 in
 neighbor 192.0.2.10 route-map IXP-peers-in in
 neighbor 192.0.2.10 route-map IXP-peers-out out
!
ip prefix-list PEER-10 permit <prefixes from Peer-10>
!
route-map IXP-peers-in permit 5
 set local-preference 175
 set community 64500:1200
route-map IXP-peers-in deny 10
!
route-map IXP-peers-out permit 5
 match community aggregates bgp-customers
route-map IXP-peers-out deny 10
```

The community for prefixes learned from IXP peers (for example)

Pre-defined communities for AS100 aggregates and BGP customers

BGP configuration tips [6]

Other BGP Configuration

Password on eBGP session

- Often required by many operators
- Often required by IXP Route Servers

Strip out private & reserved ASNs

- Private range 64512-65534
- Private range 4000000000 upwards
- Documentation 64496 to 64511 and 65536-65551
- Cisco IOS has `neighbor 100.64.1.1 remove-private-as`
- None should appear on global Internet
- Note: some operators block all ASNs from 458752 and above
- RIRs are assigning from 131072 to 458751 only (for now)

Final thoughts

Implement the MANRS recommendations

- 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 on access network
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

BGP configuration advice are all part of BGP best operational practice recommendations

- Many operators are more strict than even what is covered here!
- MANRS compliance is vitally important for the wellbeing of the Internet
- When peering, remember:
  - Don’t misuse the interconnects with your peers
  - Don’t leave your network open to misuse by your peers
  - Don’t abuse the interconnect infrastructure (IXP)

Thanks for your patience & happy peer!

References & Credits

1. https://bgp4all.com/pfs/workshops/start
2. https://wiki.apnictraining.net/bgp-20210630-online