

Peering Deployment

ISP/IXP 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 1st December 2021

Acknowledgements

- This material originated from the Cisco ISP/IXP Workshop Programme developed by Philip Smith & Barry Greene
 - I'd like to acknowledge the input from many network operators in the ongoing development of these slides
- 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

- Background & Requirements
- Equipment Requirements
- RPKI & IRR
- Peering Documentation
- Router Configuration Recommendations

Background

- The Value of Peering presentation covered why Peering was the #1 priority for a network operator in today's Internet
 - ▣ https://bgp4all.com/pfs/_media/workshops/02-value-of-peering.pdf
- Real-world – where to begin to set up peering?
 - What resources are needed?
 - What equipment is needed?
 - What are the routing protocol requirements?
- What does an end-site embarking on peering need to do?

Resource Requirements

- 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

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

Local Internet Exchange Point

- ❑ Defined as a public peering point serving the local Internet industry
- ❑ Local means where it becomes cheaper to interconnect with other ISPs at a common location than it is to pay transit to another ISP to reach the same consumer base
 - Local can mean different things in different regions!

Regional Internet Exchange Point

- These are also “local” Internet Exchange Points
- But also attract regional ISPs and ISPs from outside the locality
 - Regional ISPs peer with each other
 - And show up at several of these Regional IXPs
- Local ISPs peer with ISPs from outside the locality
 - They don’t compete in each other’s markets
 - Local ISPs don’t have to pay transit costs
 - ISPs from outside the locality don’t have to pay transit costs
 - Quite often ISPs of disparate sizes and influences will happily peer – to defray transit costs

Which IXP?

- How many routes are available?
 - What is traffic to & from these destinations, and by how much will it reduce cost of transit?
- 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

What should operators do?

- Many operators participate in their local IXP
 - 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) 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

Agenda

- Background & Requirements
- **Equipment Requirements**
- RPKI & IRR
- Peering Documentation
- Router Configuration Recommendations

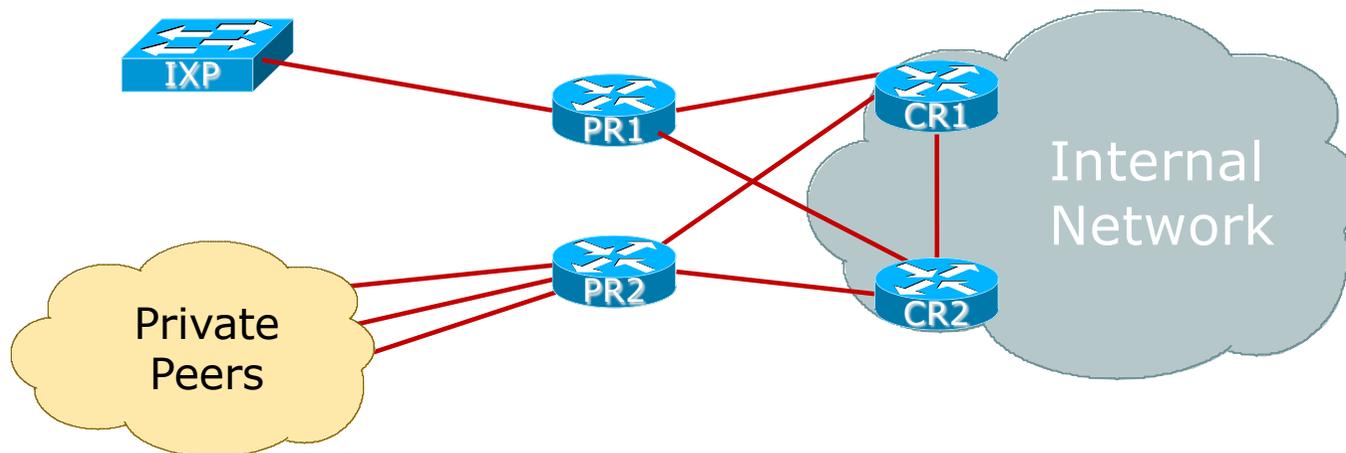
Equipment requirements

- 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)
- Peering router needs:
 - 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)

Equipment Requirements

□ Typical Scenario:

- Peering routers to peers/IXP, Core routers host internal network



Equipment Requirements

- Note separation of the two peering routers
 - IXP Peering router is quite often located at the IXP itself
- Peering Router function:
 - EBGP with peers
 - IBGP and OSPF/IS-IS with core devices
 - Traffic engineering/Policy implementation via BGP
 - Initial protection of the core network with packet filters

Equipment Requirements

- 1RU router is commonly chosen for IXP peering
 - Few interfaces needed
 - But high throughput needed
 - Examples: Juniper MX204, Cisco NCS540X

- **Note Well:**
 - Use a Router
 - Never a “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)

Agenda

- Background & Requirements
- Equipment Requirements
- RPKI & IRR
- Peering Documentation
- Router Configuration Recommendations

What now?

□ Status:

- Have obtained IPv4 and IPv6 address space from the RIR
- Have obtained an AS number from the RIR
- Have procured suitable peering routers
- Have finalised which IXP to participate at
 - Router procured and physically installed
- Have finalised which peers to interconnect with
 - Router procured and physically installed

RPKI: Signing ROAs

- When IPv4 and IPv6 address blocks are delegated, and the AS Number assigned, sign the ROAs
 - ROA stands for **Route Origin Authorisation**
 - A digital signature stating that a specific AS is authorised to originate a specific address block
 - Document this in your standard operational procedures
 - Don't forget to update the ROA if there are changes in address block size or origin AS
- How to sign ROAs?
 - Available via your RIR portal
 - Usually need to set up two factor authentication first

RPKI: Signing ROAs

- A typical ROA would look like this:

| | |
|-------------------|--------------|
| Prefix | 10.10.0.0/16 |
| Max-Length | /18 |
| Origin-AS | AS65534 |

- There can be more than one ROA per address block
 - Allows the operator to originate prefixes from more than one AS
 - Caters for changes in routing policy or prefix origin
 - (Allows your upstream to originate your address block from their AS until you are ready with your BGP)

Creating ROAs – Important Notes

- ❑ Always create ROAs for the aggregate and the individual subnets being routed in BGP
- ❑ Examples:
 - If creating a ROA for 10.10.0.0/16 **and** “max prefix” length is set to /16
 - ❑ There will only be a valid ROA for 10.10.0.0/16
 - ❑ If a subnet of 10.10.0.0/16 is originated, it will be state **Invalid**
 - If creating a ROA for 10.1.32.0/23 **and** “max prefix” length is set to /23
 - ❑ There will only be a valid ROA for 10.1.32.0/23
 - ❑ If 10.1.32.0/24 or 10.1.33.0/24 is originated, these will be state **Invalid**
 - If creating a ROA for 10.1.32.0/24 **and** “max prefix” length is set to /24
 - ❑ There will only be a valid ROA for 10.1.32.0/24
 - ❑ If 10.1.32.0/22 is originated, it will be state **NotFound**

Internet Routing Registry: Route Object

- A route object documents which AS number is originating the listed route
 - Superseded by a ROA
 - In fact, most RIRs now automatically create a route object in their IRR for each ROA that is signed
- Required by many major transit providers
 - They build their customer and peer filter based on the route-objects listed in the IRR
 - Referring to at least the 5 RIR routing registries and the RADB
 - Some operators run their own instance of the IRR as well
 - May require their customers to place a Route Object there (if not using the 5 RIR or RADB versions of the IRR)

Route Object: Examples

```
route:      100.64.0.0/24
descr:     ENTERPRISE-BLOCK
country:   ZZ
notify:    noc@yy.zz
mnt-by:    MAINT-ZZ-ENTERPRISE
origin:    AS64500
last-modified: 2018-09-18T09:37:40Z
source:    IRR
```

This declares that
AS64500 is the origin
of 100.64.0.0/24

```
route6:    2001:DB8:F:/48
descr:     ENTERPRISE-V6BLOCK
origin:    AS64500
notify:    noc@yy.zz
mnt-by:    MAINT-ZZ-ENTERPRISE
last-modified: 2010-07-21T03:46:02Z
source:    IRR
```

This declares that
AS64500 is the origin
of 2001:DB8:F::/48

AS Object: Purpose

- Documents peering policy with other Autonomous Systems
 - Lists network information
 - Lists contact information
 - Lists routes announced to neighbouring autonomous systems
 - Lists routes accepted from neighbouring autonomous systems
- Some operators pay close attention to what is contained in the AS Object
 - Some configure their border router BGP policy based on what is listed in the AS Object

AS Object: Example

```
aut-num:          AS64500
as-name:          ENTERPRISE-AS
descr:            Enterprise Network
country:          ZZ
import:           from AS64505  action pref=100;    accept ANY
export:           to AS64505    announce AS64500
import:           from AS64510  action pref=100;    accept ANY
export:           to AS64510    announce AS64500
<snip>
admin-c:          ENO1-ZZ
tech-c:           ENO1-ZZ
notify:           noc@yy.zz
mnt-by:           RIR-HM
mnt-lower:        MAINT-ZZ-ENTERPRISE
mnt-routes:       MAINT-ZZ-ENTERPRISE
last-modified:    2019-06-09T22:40:10Z
source:           IRR
```



Examples of inbound and
outbound policies – RPSL

Internet Routing Registry: Summary

- Route Object
 - Essential to have one
 - These days usually created when a ROA is signed
- AS Object
 - Not essential, but useful and informative
 - Shows operator's peering policy
 - And the ASNs connected to it

Agenda

- Background & Requirements
- Equipment Requirements
- RPKI & IRR
- Peering Documentation
- Router Configuration Recommendations

Types of Operator Peering Policies

- 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 a “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 the “Route Server” but will only peer bilaterally
- Restrictive Peering
 - Where an ISP decides who its peering partners are, and is generally not approachable to creating peering opportunities

Deciding on a Peering Policy

- Access/Hosting Provider
 - Generally these will peer with everyone and anyone, as it means they don't have to pay transit costs for traffic
 - Route Server peering at an IXP suits them
 - They don't have to maintain EBGP sessions with large numbers of peers
- Local/Regional Transit Provider & Content Provider/CDN
 - Generally these will have a selective peering policy, as they want to have direct contact with their peering partner
 - Business relationships, NOC contacts, Ops contacts, etc
 - They will not use the Route Server at IXPs
- Multi-National Transit Provider
 - Very unlikely to peer with anyone unless at their instigation
 - Any peering will be large volume & in multiple locations around the globe
 - Never at an IXP

The Peering Database

- The Peering Database documents operator peering policies
 - <https://www.peeringdb.com>
- All operators of ASNs 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 operators
 - Regional Peering Fora (European, Middle Eastern, Asian, Caribbean, Latin American)
 - Many countries now have their own Peering Fora

HKIX

| | |
|---|--|
| Organization | Hong Kong Internet eXchange Limited |
| Long Name | Hong Kong Internet Exchange |
| City | Hong Kong |
| Country | HK |
| Continental Region | Asia Pacific |
| Media Type | Ethernet |
| Protocols Supported | <input checked="" type="checkbox"/> Unicast IPv4 <input type="checkbox"/> Multicast <input checked="" type="checkbox"/> IPv6 |
| Notes  | |

Contact Information

| | |
|-----------------------|---|
| Company Website | https://www.hkix.net/ |
| Traffic Stats Website | https://www.hkix.net/hkix/stat/aggt/hkix-aggregate.html |
| Technical Email | noc@hkix.net |
| Technical Phone | +85239439900 |
| Policy Email | info@hkix.net |
| Policy Phone | +85239438800 |

LAN

| | |
|-------|-----------------------|
| MTU | 1500 |
| DOT1Q | <input type="radio"/> |
| IPv6 | 2001:7fa:0:1::/64 |
| IPv4 | 123.255.88.0/21 |

Local Facilities

| Facility ▼ | Country | City |
|---|-----------|-----------|
| CUHK | Hong Kong | Hong Kong |
| MEGA Two (iAdvantage Hong Kong) | Hong Kong | Hong Kong |
| MEGA-i (iAdvantage Hong Kong) | Hong Kong | Hong Kong |

Peers at this Exchange Point

| Peer Name ▼ ASN | IPv4 IPv6 | Speed Policy |
|--|---|------------------|
| ASGCNET HKIX Peering LAN 24167 | 123.255.91.53 2001:7fa:0:1::ca28:a135 | 10G Open |
| Asia Pacific Telecom HKIX Peering LAN 17709 | 123.255.91.86 2001:7fa:0:1::ca28:a156 | 10G Open |
| ASLINE HKIX Peering LAN 18013 | 123.255.92.13 2001:7fa:0:1::ca28:a20d | 10G Open |
| AT&T AP - AS2687 HKIX Peering LAN 2687 | 123.255.91.46 2001:7fa:0:1::ca28:a12e | 10G Selective |
| Automatic HKIX Peering LAN 2635 | 123.255.90.71 2001:7fa:0:1::ca28:a047 | 10G Open |
| Badoo Ltd HKIX Peering LAN 12678 | 123.255.90.220 None | 2G Open |
| Baidu HKIX Peering LAN 55967 | 123.255.90.131 2001:7fa:0:1::ca28:a083 | 10G Open |
| Baidu HKIX Peering LAN 55967 | 123.255.91.61 2001:7fa:0:1::ca28:a13d | 10G Open |
| Bayan Telecommunications Inc. HKIX Peering LAN 6648 | 123.255.91.45 2001:7fa:0:1::ca28:a12d | 3G Open |
| BGP Network Limited HKIX Peering LAN 64050 | 123.255.91.177 2001:7fa:0:1::ca28:a1b1 | 100G Open |
| BIGHUB-ISP HKIX Peering LAN 137989 | 123.255.90.207 2001:7fa:0:1::ca28:a0cf | 1G Open |
| BIGHUB-ISP HKIX Peering LAN | 123.255.91.98 | 10G |

Amazon.com Diamond Sponsor

| | |
|------------------------|--|
| Organization | Amazon.com |
| Also Known As | Amazon Web Services |
| Company Website | http://www.amazon.com |
| Primary ASN | 16509 |
| IRR as-set/route-set ? | AS-AMAZON |
| Route Server URL | |
| Looking Glass URL | |
| Network Type | Enterprise |
| IPv4 Prefixes ? | 5000 |
| IPv6 Prefixes ? | 2000 |
| Traffic Levels | Not Disclosed |
| Traffic Ratios | Balanced |
| Geographic Scope | Global |
| Protocols Supported | <input type="radio"/> Unicast IPv4 <input type="radio"/> Multicast <input checked="" type="radio"/> IPv6 <input type="radio"/> Never via route servers |
| Last Updated | 2019-12-29T14:56:38Z |
| Notes ? | <p>If you have a connectivity issue to Amazon then please visit:</p> <ul style="list-style-type: none"> IPv4: http://ec2-reachability.amazonaws.com/ IPv6: http://ipv6.ec2-reachability.amazonaws.com/ <p>And include detail on prefixes you think you have a problem with if you contact our Ops alias. This will reduce time with troubleshooting.</p> <p>The following Amazon US locations and associated IX's carry routes/traffic specific only to the services with infrastructure in that metro. For example, Jacksonville is CloudFront only, whereas Ashburn is CloudFront, EC2, S3, etc.)</p> <ul style="list-style-type: none"> Seattle Palo Alto San Jose Los Angeles Dallas |

Public Peering Exchange Points

| Exchange ▼ ASN | IPv4 IPv6 | Speed RS Peer |
|--|---|-------------------------------|
| AMS-IX 16509 | 80.249.210.100 2001:7f8:1::a501:6509:1 | 400G <input type="radio"/> |
| AMS-IX 16509 | 80.249.210.217 2001:7f8:1::a501:6509:2 | 400G <input type="radio"/> |
| AMS-IX Chicago 16509 | 206.108.115.36 2001:504:38:1:0:a501:6509:1 | 100G <input type="radio"/> |
| AMS-IX Hong Kong 16509 | 103.247.139.10 2001:df0:296::a501:6509:1 | 100G <input type="radio"/> |
| AMS-IX India 16509 | 223.31.200.29 2001:e48:44:100b:0:a501:6509:2 | 10G <input type="radio"/> |
| AMS-IX India 16509 | 223.31.200.30 2001:e48:44:100b:0:a501:6509:1 | 10G <input type="radio"/> |
| BBIX Osaka 16509 | 218.100.9.24 2001:de8:c:2:0:1:6509:1 | 40G <input type="radio"/> |
| BBIX Tokyo 16509 | 218.100.6.52 2001:de8:c::1:6509:1 | 200G <input type="radio"/> |
| BBIX Tokyo 16509 | 218.100.6.207 2001:de8:c::1:6509:2 | 200G <input type="radio"/> |
| BCIX BCIX Peering LAN 16509 | 193.178.185.95 2001:7f8:19:1::407d:1 | 200G <input type="radio"/> |
| BIX.BG Main 16509 | 193.169.198.87 2001:7f8:58::407d:0:1 | 100G <input type="radio"/> |
| RNIX | 194.53.172.122 | 100G |

Private Peering Facilities

| Facility ▼ ASN | Country City |
|--|-------------------|
| 151 Front Street West Toronto 16509 | Canada Toronto |
| 25 John Street / 250 Front Street West | Canada |

Telia Carrier

| | |
|------------------------|---|
| Organization | Telia Group |
| Also Known As | TeliaSonera, Telia, TSIC |
| Company Website | http://www.teliacarrier.com/ |
| Primary ASN | 1299 |
| IRR as-set/route-set ? | RIPE::AS-TELIANET RIPE::AS-TELIANET-V6 |
| Route Server URL | |
| Looking Glass URL | https://lg.telia.net/ |
| Network Type | NSP |
| IPv4 Prefixes ? | 426000 |
| IPv6 Prefixes ? | 40000 |
| Traffic Levels | 1 Tbps+ |
| Traffic Ratios | Balanced |
| Geographic Scope | Global |
| Protocols Supported | <input type="radio"/> Unicast IPv4 <input type="radio"/> Multicast <input checked="" type="radio"/> IPv6 <input type="radio"/> Never via route servers |
| Last Updated | 2020-02-05T11:43:25Z |
| Notes ? | <p>IPv4 + IPv6 Prefixes above would be actuals, not proposed max- prefix values.</p> <p>AS1299 is matching RPKI validation state and reject invalid prefixes from peers and customers. Our looking-glass marks validation state for all prefixes. Please review your registered ROAs to reduce number of invalid prefixes.</p> <p>All trouble ticket requests or support related emails should be sent to carrier-csc@teliacompany.com.</p> |

Peering Policy Information

| | |
|----------------|---|
| Peering Policy | https://www.teliacarrier.com/dam/jcr:d1e83942-3db1-4334-a5f8-431578633d26/Telia_Carrier_Global_Peering_Policy.pdf |
| General Policy | Restrictive |

Public Peering Exchange Points

| Exchange ▼ ASN | IPv4 IPv6 | Speed RS Peer |
|-------------------|--------------|------------------|
|-------------------|--------------|------------------|

No filter matches.
You may filter by **Exchange**, **ASN** or **Speed**.

Private Peering Facilities

| Facility ▼ ASN | Country City |
|--|--|
| 365 Data Centers Buffalo (BU1) 1299 | United States of America Buffalo |
| 365 Data Centers Detroit (DT1) 1299 | United States of America Southfield |
| 365 Data Centers Nashville (NA1) 1299 | United States of America Nashville |
| 365 Data Centers Tampa (TA1) 1299 | United States of America Tampa |
| 3U Rechenzentrum Berlin 1299 | Germany Berlin |
| Altus IT 1299 | Croatia Zagreb |
| Borovaya 57 1299 | Russia St. Petersburg |
| CE Colo Prague 1299 | Czechia Prague |
| CINECA - DC NaMeX 1299 | Italy Roma |
| COD BM-18 1299 | Russia St.Petersburg |
| Caldera21 1299 | Italy Milan |
| CarrierColo Berlin Luetzow (I/P/B/ site B) 1299 | Germany Berlin |
| Cologix MTL3 1299 | Canada Montreal |
| Cologix TOR1 1299 | Canada Toronto |

Screenshot

Peering Priorities

- Fully operational networks tend to:
 - Have “static” customers
 - Typical enterprise, small office, and home user
 - Have BGP customers
 - Customers who multihome between two or more providers
 - Peer at Internet Exchange Points
 - Bi-lateral peers
 - Peers via the IXP Route Server
 - Private Peer with various network operators
 - Buy transit from two or three upstream providers

Peering Priorities

- 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!
- What does this mean for setting policy?

Peering Priorities

- Setting local preferences on incoming BGP announced routes:

| Connection | Local Preference |
|----------------------|------------------|
| BGP Customers | 250 |
| Private Peers | 200 |
| IXP Bi-Lateral Peers | 175 |
| IXP RS Peers | 150 |
| (default) | 100 |
| Transit Providers | 50 |

Agenda

- Background & Requirements
- Equipment Requirements
- RPKI & IRR
- Peering Documentation
- Router Configuration Recommendations

Router Configuration Recommendations

- 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.

Peering Router: Configuration Recommendations

- 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

Peering Router: Configuration Recommendations

□ 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
!
```

Peering Router: Configuration Recommendations

□ General configuration:

■ Turn off/don't enable unneeded services including:

- DHCP server
- BOOTP server
- TFTP server
- HTTP & HTTPS servers
- Listeners for low TCP and UDP ports
- CDP/LLDP
- DHCP relay

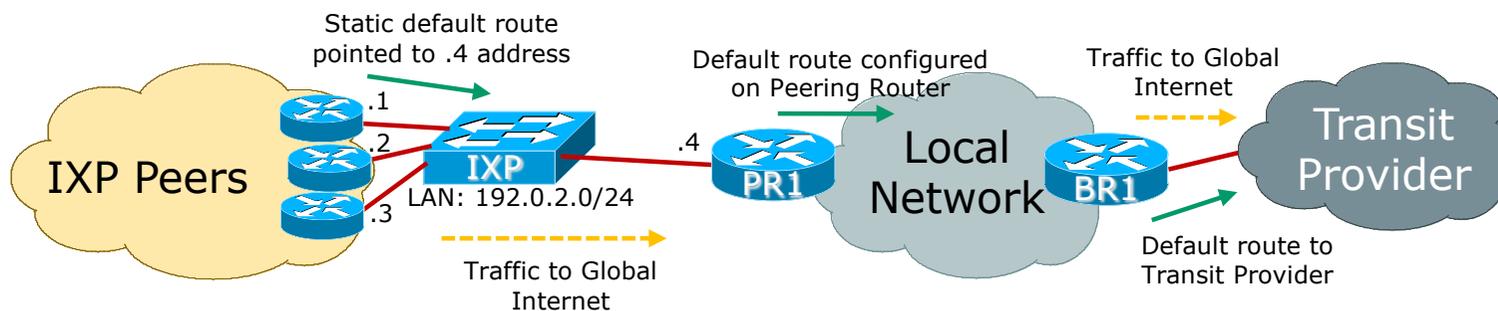
Cisco IOS Example:

```
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>
```

Peering Router: Configuration Recommendations

□ Routing configuration:

- Peering router only carries routes that peers should receive
 - No defaults (not in BGP, nor OSPF/IS-IS, nor 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)



Peering Router: Configuration Recommendations

- Routing configuration:
 - 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 unreachable
  no ipv6 unreachable
!
ip route 0.0.0.0 0.0.0.0 null0
ipv6 route ::/0 null0
```

Peering Router: Configuration Recommendations

□ Routing configuration:

- Never configure an IGP to look/listen for neighbours 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
  ...
!
```

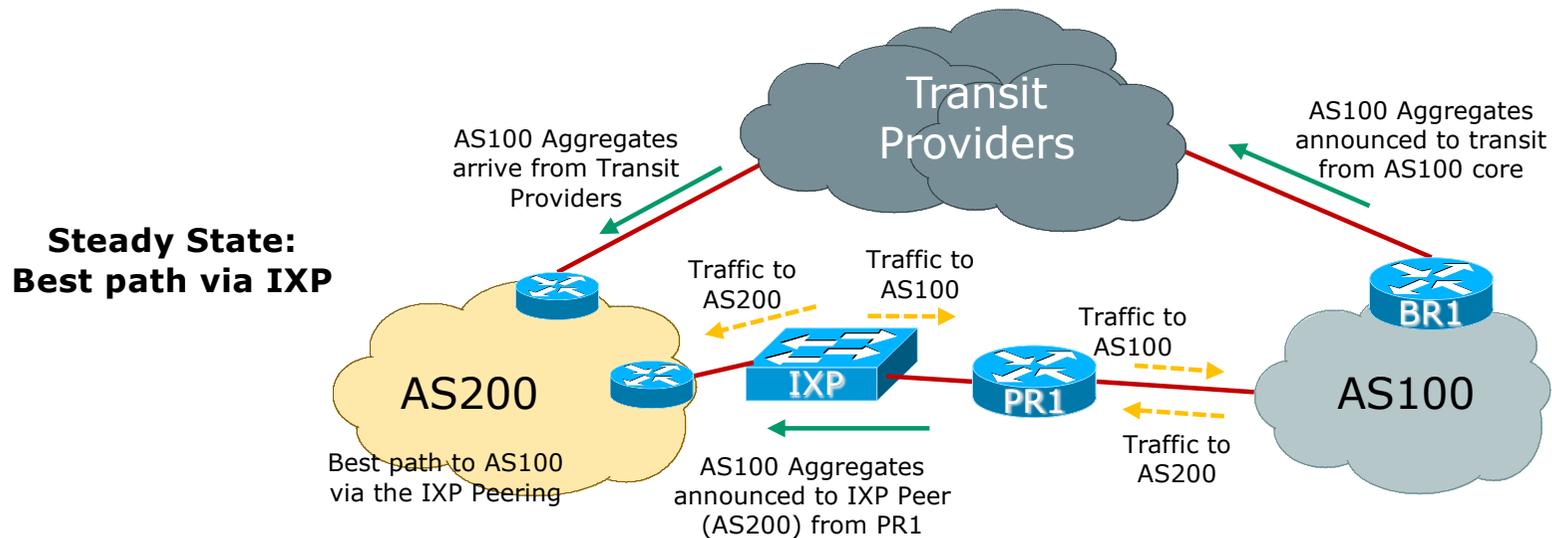
Cisco IOS IS-IS Example:

```
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
  ...
!
```

Peering Router: Configuration Recommendations

□ Routing Configuration:

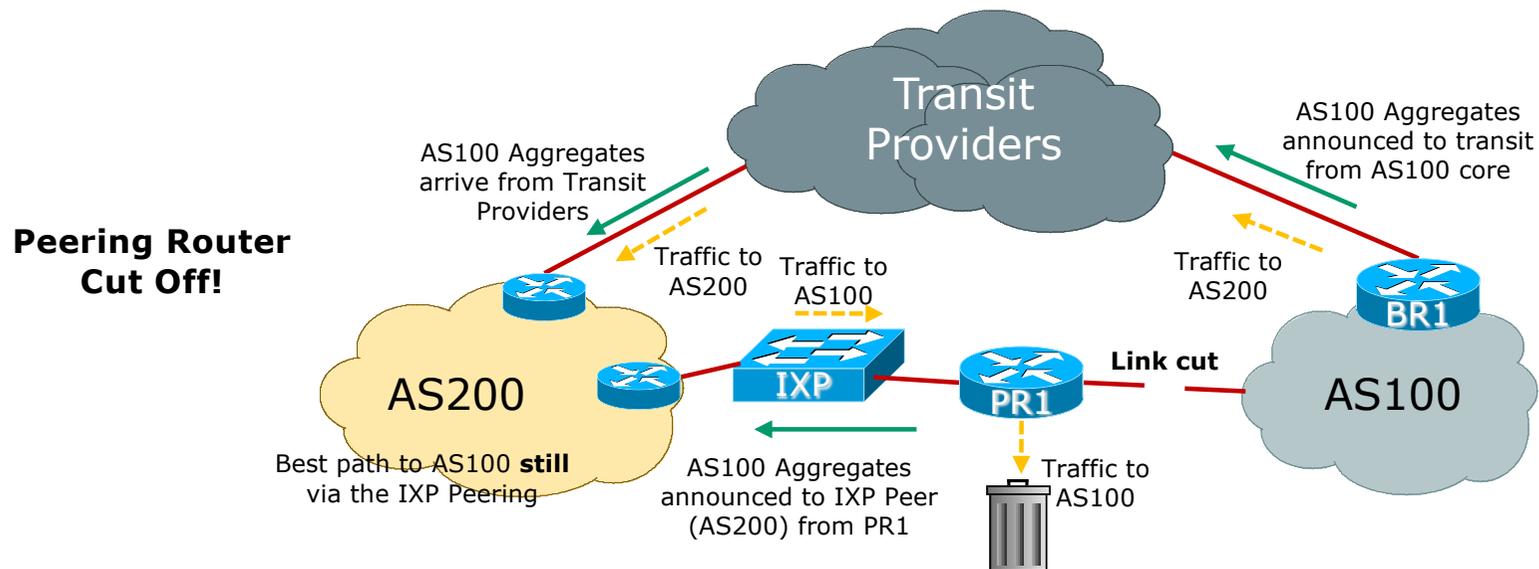
- Don't originate any prefixes into BGP on the IXP peering router
 - If this router is cut off from network core, it will still originate prefixes and likely still be best path, breaking your backup via your Transit Providers



Peering Router: Configuration Recommendations

□ Routing Configuration:

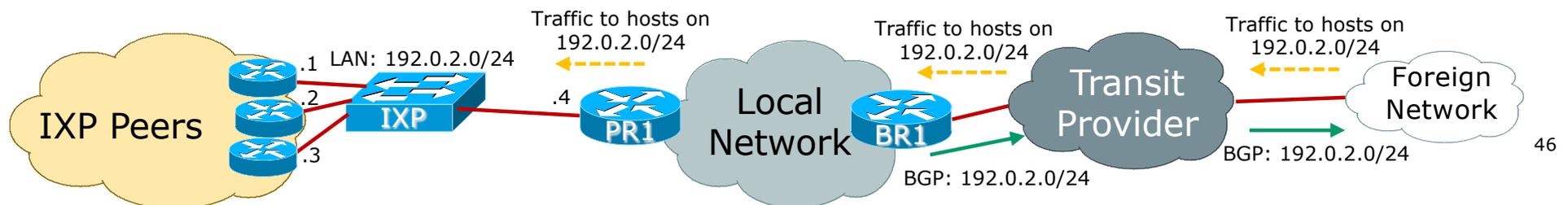
- Don't originate any prefixes into BGP on the IXP peering router
 - If this router is cut off from network core, it will still originate prefixes and likely still be best path, breaking your backup via your Transit Providers



Peering Router: Configuration Recommendations

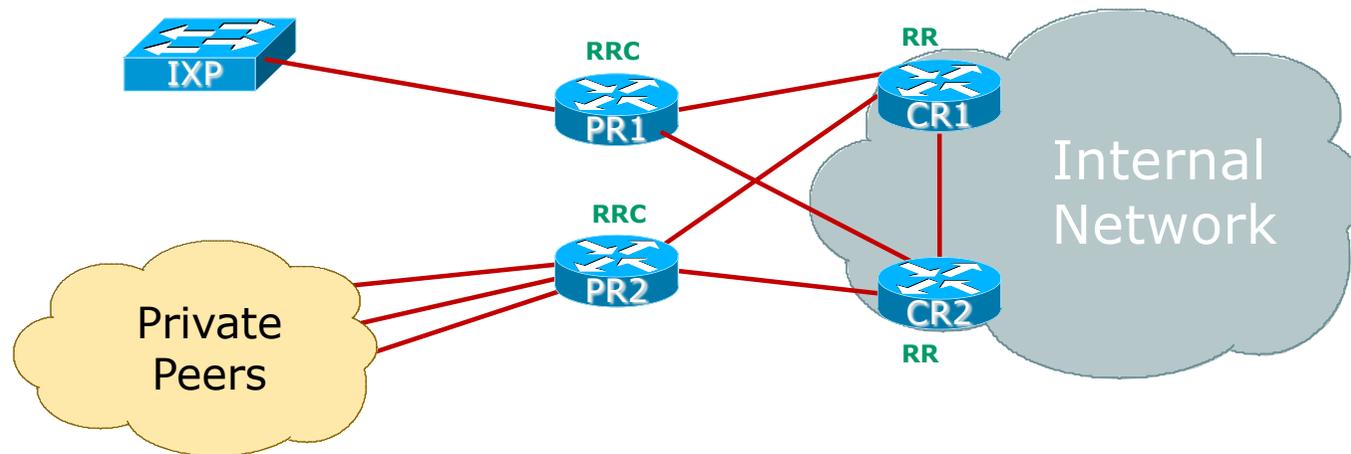
□ Routing Configuration:

- The IXP LAN subnet must never be carried in BGP
 - Carrying it in OSPF/IS-IS is okay so that traceroutes don't appear broken
 - Use the IBGP "next-hop-self" feature
 - If 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 ROA – but members need to do their part too!



Peering Router: BGP Configuration

- Use BGP Communities wherever possible!
- Make Peering Router a route reflector client (RRC)
 - Running core routers as Route Reflectors (RR) is standard practice
 - Only announce internal prefixes/aggregates to the Peering Router
 - Communities make this easy!
 - See https://bgp4all.com/pfs/_media/workshops/11-bgp-communities.pdf



Peering Router: BGP Configuration

- 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
 - Most vendors still do not adhere to this requirement!
 - 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)
 - (Cisco IOS: route-map and peer-group)

Peering Router: BGP Configuration

- Create suitable BGP policies:
 - 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 Internal BGP, Peering router needs to carry all customer routes, the aggregates, and subnets of the aggregates
 - **Note Well: Subnets of aggregates are not announced to external BGP peers**

Peering Router: EBGP Configuration

□ 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

Peering Router: IBGP Configuration

□ Cisco IOS IBGP 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
!
```

IOS does not send communities by default: send both standard and extended types

Core Router: IBGP Configuration

□ Cisco IOS IBGP Configuration Example

```
interface Loopback 0
  ip address 100.64.1.1 255.255.255.255
!
router bgp 64500
  neighbor 100.64.1.3 remote-as 64500
  neighbor 100.64.1.3 description IBGP with PR1 RR Client
  neighbor 100.64.1.3 send-community both
  neighbor 100.64.1.3 next-hop-self
  neighbor 100.64.1.3 update-source Loopback0
  neighbor 100.64.1.3 route-map partial-IBGP out
  neighbor 100.64.1.3 route-reflector-client
  ...
!
route-map partial-IBGP permit 5
  match community aggregate subnets bgp-customers
route-map partial-IBGP deny 10
!
```

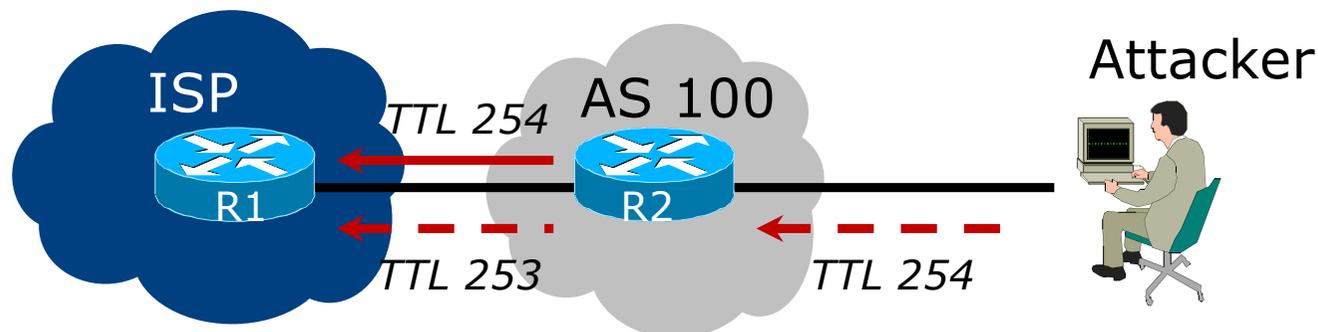
Pre-defined communities for AS64500
aggregates, subnets, and BGP customers

Peering Router: Other BGP Configuration

- Password on EBGP session
 - Often required by many operators
 - Often required by IXP Route Servers
- BGP TTL Hack (RFC5082)
 - Neighbour sets TTL to 255
 - Some operators require this
 - Needs to be done on both ends

```
neighbor 192.0.2.10 password s0m3th1ng5ecre7
```

```
neighbor 192.0.2.10 ttl-security hops 1
```



Peering Router: Other BGP Configuration

□ Strip out private & reserved ASNs

- Private range 64512-65534
- Private range 4000000000 upwards
- Documentation 64496 to 64511 and 65536-65551
- Cisco IOS has `neighbor 192.0.2.10 remove-private-AS`
 - Only works for original 16-bit private range
- None should appear on global Internet
- Note: some operators block all ASNs between 65536 and 131071, and from 458752 and above
 - RIRs are assigning from 131072 to 458751 only (for now)

Peering Router: Other BGP Configuration

□ Maximum Prefix Tracking:

- Set a limit on the number of prefixes expected from a peer
- Protects the network from accidental route leaks and misconfiguration by peers
- Used widely and considered a best operational practice
- Cisco IOS example:

```
neighbor 192.0.2.10 maximum-prefix <max> [restart N] [<threshold>] [warning-only]
```

- Where maximum-prefix is usually set to double what is expected from a peer
- Router will issue warnings at 75% of threshold
- Router will tear down peering once the number reaches the maximum
- And can optionally restart the BGP session at N minutes later (hoping that the configuration error has been fixed)

Peering Router: Route Origin Validation

- ❑ Check routes to ensure the origin AS is valid
- ❑ Aim is to defeat prefix hijacks & misoriginations
 - Covered in depth:
 - ❑ https://bgp4all.com/pfs/_media/workshops/02-rpki.pdf
 - In short:
 - ❑ Set up a validator (e.g. NLnetLabs Routinator 3000, NIC Mexico FORT)
 - ❑ Configure Peering (and all EBGP) Routers to talk with validator
 - ❑ Drop invalid routes (done by default in Cisco IOS)
 - ❑ Cisco IOS example:

```
router bgp 645000
  bgp rpki server tcp 10.0.0.3 port 3323 refresh 3600
```
- ❑ Note that some IXPs already do this on their Route Servers

Peering Security

□ 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



MANRS

Configuration Recommendations

- Most are considered industry best practices
 - 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)

Summary

- Background & Requirements
- Equipment Requirements
- RPKI & IRR
- Peering Documentation
- Router Configuration Recommendations

Peering Deployment



ISP/IXP Workshops