AFIX Technical Workshop: Session 4

Technical Aspects of Peering

Contents

Overview....1
Peering requirements....1
Peering: Step by Step....2
Peering arrangements and options....5
Option 1: Mandatory multi-lateral peering....5
Option 2: Bi-lateral peering....6
Option 3: Hybrid model....7
Exercises....8
Exercise 1....8
Exercise 2....11
Exercise 3: Multilateral / Route Server model....13

Overview

In the previous session we briefly covered dynamic routing and the protocols that are used to accomplish the task. In this section we focus more closely on how to set up peering in practice.

Peering requirements

- To peer with other organizations you need to have portable IP address space and an Autonomous System Number ("AS number" or "ASN"), both of which can be obtained from your local internet numbers registry. For Africa, this would be AfriNIC (http://www.afrinic.net/).

  It is important to understand what portable address space is and why it (and not just any address space) is needed for peering. Small ISPs and organizations usually receive address space from their upstream ISPs. This address space may not be used to peer with other ISPs at public peering points (IXPs) since their upstream ISPs are already advertising the address space to their peers and it is therefore not portable. (It is technically possible to make an exception to this rule, if you have the cooperation of your upstream ISP.)

  In order to peer, you need to obtain address space directly from your internet registry. Address space obtained from an internet registry is not used by anyone else on the internet and it is the responsibility of the recipient to advertise this address space to its upstream internet service providers and peers.

  An AS number is used to identify a set of routers under the same administrative control and sharing the same routing policy. Typically, all the routers within a single ISP are part of the same AS, and routers within a different ISP are part of a different AS. A network that connects through only one upstream provider does not need its own AS number, but is treated as part of the upstream provider's AS; this also applies to small ISPs that connect through only one upstream provider and do not have any peering links to neighbouring ISPs.

  All AS numbers from 64512 to 65536 are private and may be used by organizations on their own private networks, similar to private IP address space (such as 192.168.0.0/16, 10.0.0.0/8, and

172.16.0.0/12). When peering at a public internet exchange point, it is important to use a globally unique AS number (not a private AS number) for the same reasons that unique address space is required.

- Internet service providers have standardized on using BGP (Border Gateway Protocol) version 4 for peering. BGP is an exterior routing protocol, used between networks under different administrative control.

  The term “eBGP” or “exterior BGP” is often used to refer to the use of BGP between two different autonomous systems (such as between two different ISPs), while the term “iBGP” or “interior BGP” is often used to refer to the use of BGP inside a single autonomous system (such as between two routers inside the same ISP).

- Another important requirement is to check that your BGP router has enough memory to receive all the routes (also known as prefixes in the peering world) from all your peers. For instance if you are going to receive the global routing table, your router needs to store more than 165,000 routes (and their AS path information) which requires a considerable amount of memory. 256MB Memory is now regarded as an absolute minimum requirement for routers that receive the entire global routing table whereas 32MB memory should be enough to receive all the routes for the African continent.

Summary of peering requirements

- BGP4 capable router with enough memory (such as Cisco, Juniper, or Quagga)
- Your own unique autonomous system (AS) number
- Portable address space
- List of prefixes that will be advertised to peers (your address space and your customers' address space)
- For each peer, a list of prefixes that they will announce to you (their address space and their customers' address space)
- The AS number of each peer
- The IP addresses that will be used for the BGP connection between you and each peer (these would typically be the addresses used on the ethernet at the exchange point)

Armed with the above you are now ready to peer.

Peering: Step by Step

Let’s assume that you are an internet service provider called John Doe Communications and you would like to peer with an internet service provider called Expert Networks. Let's go through each step required to set up a peering link for John Doe Communications.

Step 1

Write down all the information listed above for each organization:

A:	Company Name:	John Doe Communications
	AS number:	AS100
	Address space:	12.1.1.0/24, 196.25.0.0/16
	Border router:	Cisco 2621
	BGP peer address:	192.0.2.5
B:	Company Name:	Expert Networks
	AS number:	AS200

Address space:	150.200.54.0/23
Border router:	Linux PC running Quagga
BGP peer address:	192.0.2.8

Step 2

Configure a loopback interface on the router. This is necessary in order to have a BGP peer with an interface that will always be up even if some of the physical interfaces on the router go down. You should always use loopback interfaces for iBGP, but hardly ever for eBGP.

```
interface Loopback0
ip address 12.1.1.10 255.255.255.255
```

Define filters to advertise and receive only the routes we want. This is very important. If this step is omitted any peer can flood your routing table with bogus entries. It can also cause your router to crash if too many prefixes are accepted by your router.

```
! “ip prefix-list AS100” allows routes for all networks that belong to  AS 100,
! including more-specific routes (provided the prefix length is no longer than /24)

ip prefix-list AS100 seq 5 permit 12.1.1.0/24
ip prefix-list AS100 seq 10 permit 196.25.0.0/16 le 24

! “ip prefix-list AS200” allows routes for all networks that belong to AS 200,
! including more-specific routes (provided the prefix length is no longer than /24)

ip prefix-list AS200 seq 5 permit 150.200.54.0/23 le 24
```

Configure basic parameters for BGP routing:

```
router bgp 100                ! “100” is our own AS number
```

By default BGP does not advertise a route until all routes within the AS have learned of the route through the IGP. Use the “no synchronization” command to turn off this unwanted historical behaviour.

```
no synchronization
```

By default BGP assumes that routing uses classful networks (class A, class B, class C), and attempts to convert some more-specific routes to classful routes. This is an unwanted historical behaviour. Use the “no auto-summary” command to turn it off.

```
no auto-summary
```

Log all changes such as BGP connections going down. These changes can be monitored by exporting the router logs to a syslog server (using other commands not shown here). Most ISPs have a central log server and have technicians monitoring all events.

```
bgp log-neighbour-changes
```

Arrange for your own networks to be imported into BGP. Do not use “redistribute” commands, because they make it too easy for unwanted routes to get into BGP. Use a “network” command for each prefix that you want in your BGP table. If the prefix is an aggregate which you have subnetted, then you also need a static pullup route to ensure that the aggregate route is always present.

```
! ensure that the 196.25.0.0/16 aggregate route is always present
ip route 196.25.0.0 255.255.0.0 null0 254          ! this is a static pullup route
```

```
! add your own networks to BGP
router bgp 100
network 12.1.1.0 mask 255.255.255.0
network 196.25.0.0 mask 255.255.0.0
```

For each BGP peer (also called a neighbour), we need multiple "neighbour" commands. Each such command specifies the neighbour's IP address. The first such command specifies the neighbour's AS number. We now set up a peering session with Expert Networks (AS 200, using IP address 1.2.3.4).

```
neighbour 1.2.3.4 remote-as 200
```

Add a description. If there are many neighbours defined, it is useful to find the appropriate neighbour when configuration changes have to be made by looking at these descriptions.

```
neighbour 1.2.3.4 description Expert Networks
```

This command instructs the router to set the gateways for all routes added to the routing table to itself. Always enable this when peering with other autonomous systems.

```
neighbour 1.2.3.4 next-hop-self
```

Instruct the router to store received updates. This allows us to update a BGP session without having to restart the session. (This uses extra memory. In IOS 12.0 or later, you can get a similar effect without using extra memory, with the BGP route refresh capability. Use "show ip bgp neighbor x.x.x.x" to check whether your peer supports this capability.)

```
neighbour 1.2.3.4 soft-reconfiguration inbound
```

Only advertise and accept routes allowed by our filters to prevent flooding of our routing table.

```
neighbour 1.2.3.4 prefix-list AS100 out
neighbour 1.2.3.4 prefix-list AS200 in
```

Step 3

Verify that everything is working as it should. The following commands can be used to diagnose problems with your BGP configuration:

```
! show a summary of peering sessions
show ip bgp summary
! show neighbour details
show ip bgp neighbours
! show routes received from neighbours
show ip bgp
! show routes received from neighbour 192.168.4.1 (before your "in" filter)
show ip bgp neighbours 192.168.4.1 received-routes
! show routes advertised to neighbour 192.168.4.3 (after your "out" filter)
show ip bgp neighbours 192.168.4.3 advertised-routes
! show all routes known to the kernel
show ip route
```

Peering arrangements and options

There are two major alternative ways of peering at an internet exchange point (IXP):

1. The IXP can implement mandatory multi-lateral peering using a route server.
2. Individual ISPs can conclude bi-lateral peering agreements, resulting in a full (eBGP) mesh if all participants choose to peer with all other participants, or in a partial mesh if some participants choose not to peer with each other.

It's also possible to use a combination of the two, creating a third, hybrid arrangement:

3. A combination of multi-lateral peering using a route server (usually new entrants) and bilateral peering (using full mesh).

Each of these options is considered in more detail below.

Option 1: Mandatory multi-lateral peering

Under this option, all the participants at the IXP peer with a central route server. This forces all the participants at the IXP to peer with each other and reduces the number of peering sessions that has to be maintained by each peer.

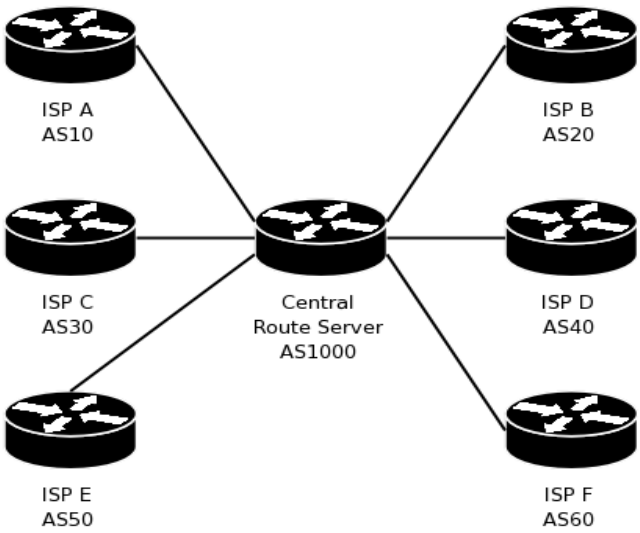
Route server vs route reflector: Getting the terminology right

A central route server is sometimes incorrectly referred to as a route reflector. Since this can cause a great deal of confusion, it is worth explaining the difference between them.

A **route reflector** is a concept within iBGP. Ordinarily, routes that a router learns through iBGP are not redistributed via iBGP to other routers in the same AS. This behaviour of iBGP results in a requirement that iBGP usually need to be configured in a full mesh, in which each router talks iBGP to every other router inside the same AS. In some cases (outside the scope of this workshop), a "route reflector" can be used to redistribute routes from one iBGP router to another iBGP router, and this can eliminate the need to configure a full mesh of iBGP neighbour relationships.

At an IXP, where you are peering with other autonomous systems, the protocol used is not iBGP but eBGP. Under this protocol, neighbours automatically distribute routes to all their eBGP neighbours (subject only to filters configured by the network administrator) – so a route reflector is not needed, and the concept of a route reflector does not exist in eBGP.

A **central route server,** on the other hand, is a router at an IXP, managed by the IXP itself, that all the participants at the IXP peer with – as in the diagram below:



The route server peering arrangement above has several advantages and disadvantages.

Advantages:

- All participants at the exchange point automatically peer with all other participants. This is an advantage because it encourages local exchange of traffic.
- The most complex configuration is centralised in the route server, where it can be managed by a small team of skilled people. This allows ISPs with less skilled staff to participate at the exchange point.
- It is easy for a new participant to connect to the exchange point. They need to configure only one eBGP session with the central route server.

Disadvantages:

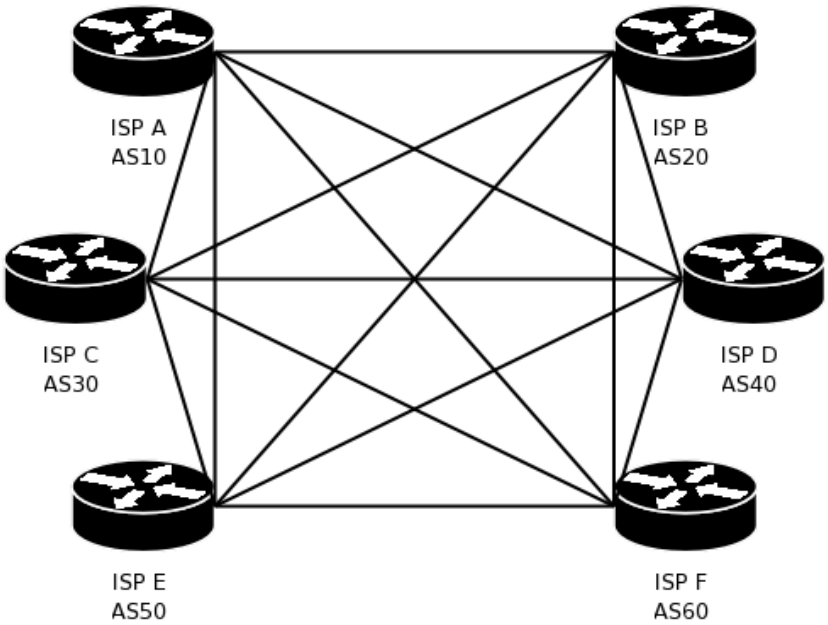
- All participants at the exchange point automatically peer with all other participants. This is a disadvantage because it prevents participants from making individual decisions about who to peer with.
- The central route server has a complex configuration, so the exchange point needs to appoint one or more skilled people to manage it.
- Each participant at the exchange point need to announce exactly the same routes to all other participants. This prevents more complex routing policies from being implemented.

If the number of participants at the IXP is very large, the central route server can't handle all the peering sessions. This situation can be avoided by creating a mesh of central route servers and letting each participant peer with one of the central route servers. The best option is to start with a single central route server as explained above and add more servers if necessary as the number of participants increases.

Option 2: Bi-lateral peering

An exchange point that allows bi-lateral peering will allow all the participants to negotiate their own peering arrangements with each other, and will not try to enforce peering between participants that do not want to peer with each other. The simple route server model mentioned above supports only multi-lateral peering agreements (in which each participant agrees to peer with every other participant),

If each participant at an exchange point chooses to peer with all other participants, then the result will be a mesh of eBGP peering sessions, as depicted below.



The bi-lateral peering arrangement above has several advantages and disadvantages.

Advantages:

- Each participant at the exchange point can choose whether or not to peer with each other participant. This is an advantage because it does not force any participant to do something that they do not wish to do.
- There is no central router that needs to be managed by the exchange point operator. All router management is performed by the ISPs that operate the routers.
- Each participant at the exchange point can choose to announce different routes to different peers (in addition to choosing whether or not to peer with each other participant). This allows complex routing policies, as agreed between the participants themselves without interference from the exchange point operator.

Disadvantages:

- Each participant at the exchange point can choose whether or not to peer with each other participant. This is a disadvantage because there may be participants who choose not to peer, and this can result in local traffic being routed over expensive international links.
- Each participant needs to manage a complex router configuration, which gets more complex as more peers are added.
- It is somewhat difficult for a new participant to connect to the exchange point. The new participant needs to negotiate multiple peering agreements with existing participants, and configure multiple BGP peering sessions.

Option 3: Hybrid model

It is possible to have both models operating simultaneously at an IXP, with some ISPs peering with the central route server and the others manually configuring their routers for bilateral peering with selected peers.

This is not the most desirable option, but in reality it can develop if, for example, both very large and very small ISPs are part of the same IXP. The large ISP might for business reasons prefer to treat the small ISP as a transit customer rather than a peer; bi-lateral agreements will give it the opportunity to control its relationships with other ISPs individually.

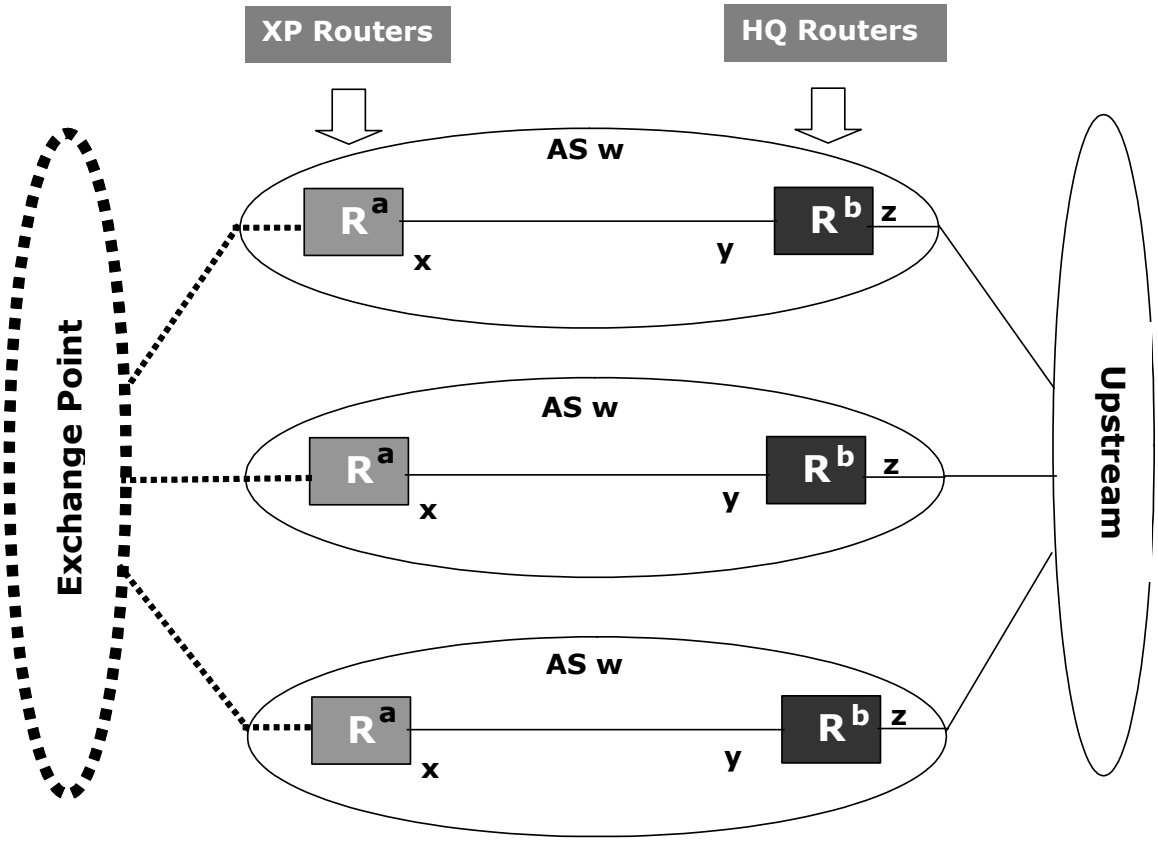
One simple option is to start with a single central route server and multi-lateral peering, then let individual business demands determine what happens next.

Exercises

Exercise 1

In this exercise, we will simulate the situation where there are multiple ISPs, each ISP is connected to an upstream provider, and all the ISPs are in the process of connecting to an exchange point. We will configure iBGP routing between the existing router ("HQ") at the ISP's headquarters and the new router ("XP") at the exchange point.

In exercise 1, we will configure iBGP between the routers inside each ISP, but we will not configure eBGP between one ISP and another ISP (that will happen in exercise 2). For exercise 1, **the XP router is not yet connected to the exchange point ethernet** (that will also happen in exercise 2).



Begin with three ISPs configured according to the diagram above. Each has an AS Number ("w"), a router ("HQ") at their headquarters, and a router ("XP") at the exchange point, with a serial line connecting the XP router to the HQ router. The HQ router is connected to an upstream provider.

1. Ensure that you do not have any cables still plugged in or router configurations left over from previous exercises.
2. Each ISP needs an AS number ("w"). This will be assigned by the instructors.
3. Each ISP needs an IP address range. These will be assigned by the instructors (or just use the same addresses that you had from the exercises in a previous session). Each ISP will manage the details of what happens inside their own address range.
4. Instructors will manage a router that acts as the upstream provider.
5. Connect the cables to link the XP router to HQ router, and to link the HQ router to the upstream provider. Do not connect the XP router to the exchange point ethernet (that will happen in exercise 2).

6. Assign a /30 subnet for the link between the HQ router and the XP router, and assign IP addresses “x” and “y” to the relevant interfaces on the two routers.
7. Assign a single IP address (a /32 prefix) for the loopback interface “a” in the XP router, and another IP address for the loopback interface “b” in the HQ router.
8. The upstream provider (instructors) will tell you what IP addresses to use for the connection between your HQ router (“z”) and the upstream provider.
9. Add standard configuration lines that are used on almost every Cisco router

```
hostname hq.example.net
ip subnet-zero
no ip source-route
no ip domain-lookup
ip classless
no cdp run
```

10. Configure IP addresses on interfaces “x”, “y” and “z”.

```
interface Serial0                    ! specify interface name
 description serial from HQ to XP
 encapsulation ppp
 ip address y.y.y.y 255.255.255.252      ! specify IP address and netmask
interface Ethenet0/0
 description ethernet to upstream provider
 ip address z.z.z.z m.m.m.m              ! specify IP address and netmask
 no ip redirects
 no ip proxy-arp
```

11. Configure a static default route from the HQ router to the upstream provider.

```
ip route 0.0.0.0 0.0.0.0 u.u.u.u   ! “u.u.u.u” represents upstream provider
```

12. Configure loopback interfaces at “**a**” and “**b**”

```
interface loopback0
 ip address a.a.a.a 255.255.255.255     ! “a.a.a.a” is the IP address
```

13. Configure some kind of interior routing protocol (using either OSPF or static routes) so that your HQ router can ping loopback interface “a” on your XP router, and so that your XP router can ping loopback interface “b” on your HQ router.

    Use techniques learned in previous exercises

14. Configure iBGP between the loopback interfaces “a” and “b”.

```
router bgp wwww          ! “wwww” represents your AS number
 no synchronization
 no auto-summary
 bgp log-neighbor-changes
 ! Remote IP address is the other router's loopback interface “b”.
 ! Remote AS number is same as local AS number (“w”)
 neighbor b.b.b.b remote-as wwww
 neighbor b.b.b.b update-source loopback0
```

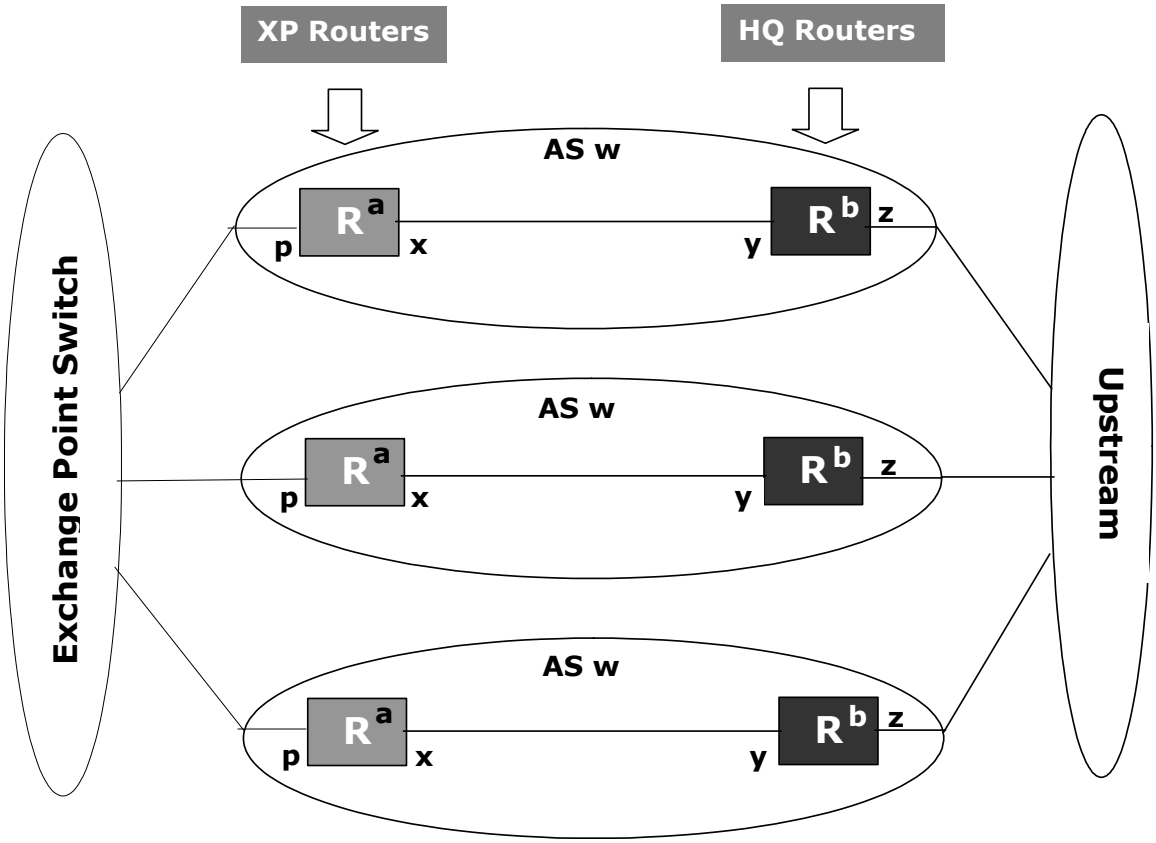
15. Add BGP “network” statements and static pullup routes on the HQ router. Do not also do this on the XP router.

```
! static pullup route for network “n.n.n.n” netmask “m.m.m.m”
ip route n.n.n.n m.m.m.m null0 254
! add network “n.n.n.n” netmask “m.m.m.m” to BGP
router bgp wwww
  network n.n.n.n mask m.m.m.m
```

16. Verify that the “XP” router learns all your routes from the “HQ” router via iBGP. Use “show ip bgp summary”, “show ip bgp”, “show ip bgp neighbor”, “show ip route”.

Exercise 2

At the end of exercise 1, we have several ISPs that each have a router at the exchange point, but the exchange point is not working yet. In exercise 2, we proceed to make the exchange point work using a bilateral peering model.



1. Start with everything set up as it was at the end of exercise 1.
2. The exchange point operator (instructors) will allocate an IP address block for use at the exchange point itself, and will provide an ethernet switch at the exchange point.
3. Ask the exchange point operator (instructors) which port on the switch you may connect to, and what IP address you should use (for interface “p” on your XP router). The exchange point operator will want to know your AS number (“w”) and organisation name.
4. Make an information sheet showing:
    - Your organisation name
    - Your AS number (“w”)
    - The IP address you use at the exchange point (“p”)
    - The IP address ranges (prefixes) that you will announce to peers at the exchange point. This includes all your customers networks.
5. Obtain copies of all your peers' information sheets.
6. Connect an ethernet cable from your XP router to your assigned port on the exchange point switch, and configure the IP address on interface “p” of your XP router.

7. Define filters to:
    - Allow your routes, deny others.

        ip prefix-list my-routes seq 5 permit n.n.n.n/16 le 24

    - Allow peer's routes, deny others. There will be several such filters; one for each peer at the exchange point. Each filter may have several lines; one for each prefix that the peer may announce to you.

        ip prefix-list peer-AS200 seq 5 permit n.n.n.n/20 le 24

        ip prefix-list peer-AS200 seq 10 permit x.x.x.x/24

    - Deny RFC1918, 127.0.0.0, and other bogus nets; allow others

        ip prefix-list no-bogons seq 5 deny 127.0.0.0/8 le 32

        ip prefix-list no-bogons seq 10 deny 0.0.0.0/8 le 32

        ip prefix-list no-bogons seq 15 deny 10.0.0.0/8 le 32

        ip prefix-list no-bogons seq 20 deny 172.16.0.0/12 le 32

        ip prefix-list no-bogons seq 25 deny 192.168.0.0/16 le 32

        ! ... add more lines here for other bogus nets you might know about

        ip prefix-list no-bogons seq 1000 permit 0.0.0.0/0 le 24

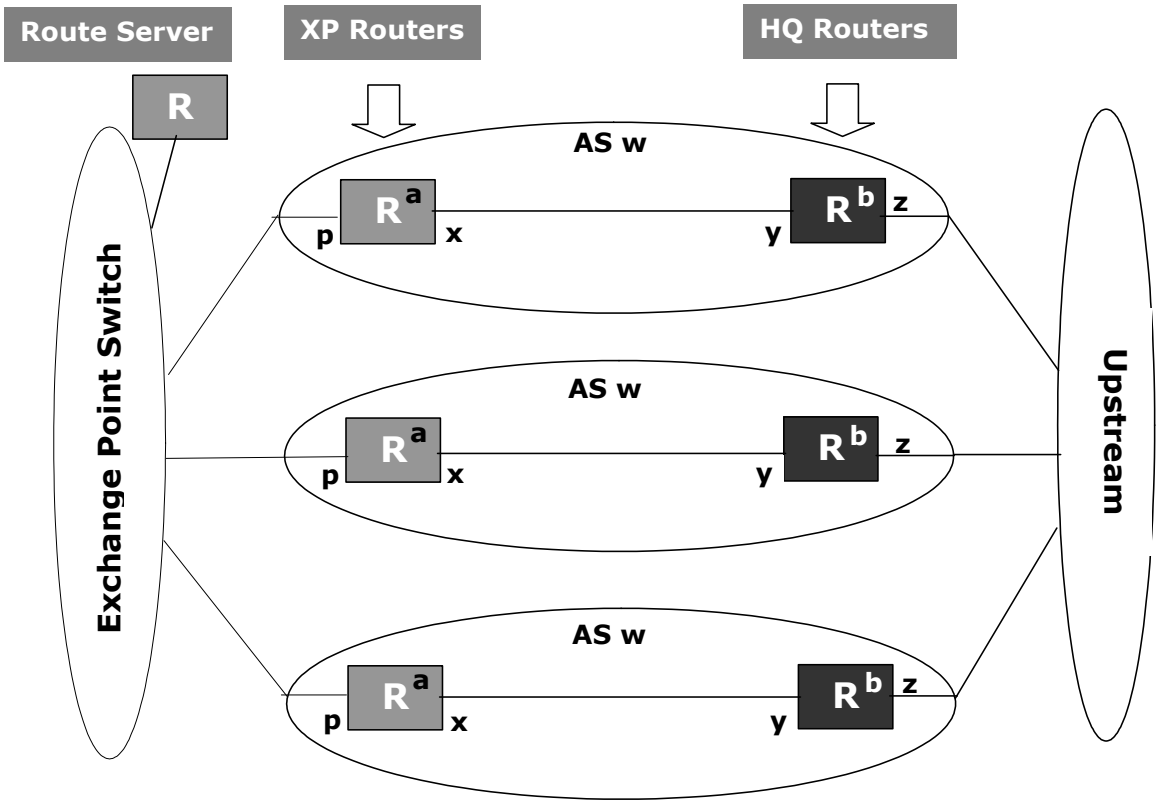
8. For each peer at the exchange point, add bgp “neighbor” commands to talk to them. In each case, use an outgoung filter that allows your routes and denies others, and use an incoming filter that allows that particular peer's routes and denies others.

```
router bgp 100                          ! Your AS number
 neighbor p.p.p.p remote-as 200         ! Their IP address and AS number
 neighbor p.p.p.p description Expert Networks, phone 123-4567
 neighbor p.p.p.p next-hop-self
 neighbor p.p.p.p soft-reconfiguration inbound
 neighbor p.p.p.p prefix-list my-routes out
 neighbor p.p.p.p prefix-list peer-AS200 in
 neighbor p.p.p.p maximum-prefix 100
```

9. Verify using “show ip bgp summary”, “show ip bgp”, “show ip bgp neighbor”, “show ip route”, “ping”, and “traceroute”.

Exercise 3: Multilateral / Route Server model

At the end of exercise 1, we have several ISPs that each have a router at the exchange point, but the exchange point is not working yet. In exercise 3, we proceed to make the exchange point work using a route server model. You can do exercise 3 instead of exercise 2, or as well as exercise 2, depending on time available.



1. Start with everything set up as it was at the end of exercise 1. If you skipped exercise 2, then your routers should already be set up exactly correctly. If you didn't skip exercise 2, then you will have to start by undoing all the bgp “neighbor” commands that you added in exercise 2. (Don't worry about undoing the rest of the commands you added in exercise 2.)

```
router bgp 100              ! your AS number
 no neighbor p.p.p.p        ! undo everything related to this neighbour
 no neighbor q.q.q.q        ! also remove this other neighbour
```

2. Instructors will operate the route server, which is connected to the exchange point ethernet. They will tell you the AS number and IP address of the route server.

3. Proceed exactly as in exercise 2, except, when you get to adding “bgp neighbor” commands, do not add a BGP session to peer with any other ISP. Instead, add a BGP session (“neighbor” commands) to peer with the route server. Use an outgoing filter that allows your routes and denies other routes. Use an incoming filter that allows all routes except for bogus routes. (Note that the filters were stricter in exercise 2, but are easier to set up in this exercise.)

```
router bgp 100                      ! your AS number
 neighbor s.s.s.s remote-as ssss    ! route server's IP address and AS number
 neighbor s.s.s.s description Route Server at IXP
```

```
neighbor s.s.s.s prefix-list no-bogons in
neighbor s.s.s.s prefix-list my-routes out
neighbor s.s.s.s next-hop-self
neighbor s.s.s.s soft-reconfiguration inbound
neighbor s.s.s.s maximum-prefix 2000
```

4. Verify using “show ip bgp summary”, “show ip bgp”, “show ip bgp neighbor”, “show ip route”, “ping”, and “traceroute”.