Building Internet eXchange Point (IXP) Network with MikroTik

Lay Minh (Makito)
CCIE # 47682, MikroTik Certified Trainer, MikroTik Consultant
Presented On: February 16, 2014
Publication Date: March 12, 2017

About Me

- **Lay Minh (Makito)**
  - CCIE # 47682
  - MikroTik Certified Trainer & Consultant
  - Experiences:
    - 12 years in ISP industry since 2005
    - Billing solutions for service providers
    - ISP core network design and operation
  - Certifications:
    - **Juniper** JNCIA-Junos, JNCIS-SP, JNCDA
    - **VMware** VCA6-NV
    - MikroTik MTCNA, MikroTik MTCRE, MikroTik MTCWE, MikroTik MTCTCE, MikroTik MTCUME, MikroTik MTCINE
  - Areas of interest: BGP, MPLS, IPv6

Agenda

- About Internet eXchange
- What is IXP?
- IXP Implementation
- Lab

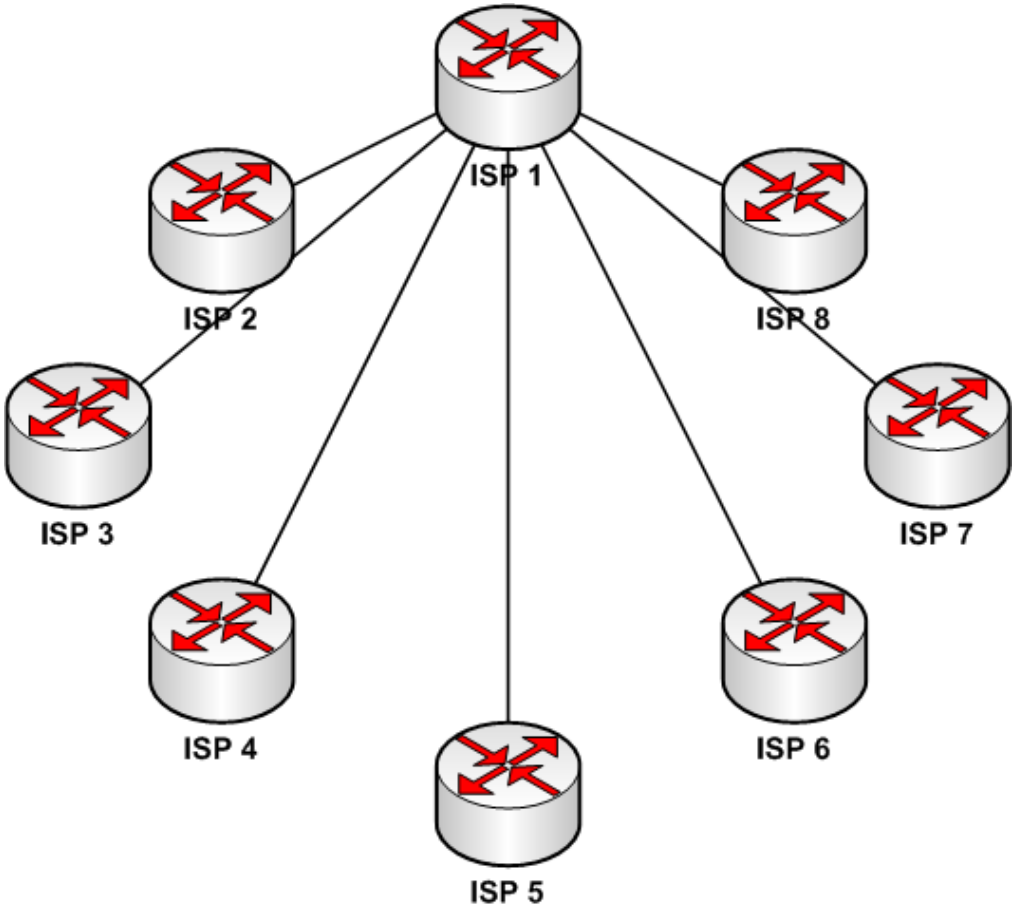
About Internet eXchange

- Each Internet Service Provider (ISP) has various upstream providers, ISPs pay to their upstream providers for IP transit.

- Without internet exchange, ISPs can only send everything to their upstream providers, so traffic between local ISPs would travel through international links, which causes following problems:
  - High Cost: IP transit cost on international links
  - High Latency: Packets traveled to overseas and come back
  - Low Stability: Unexpected issues on middle hops

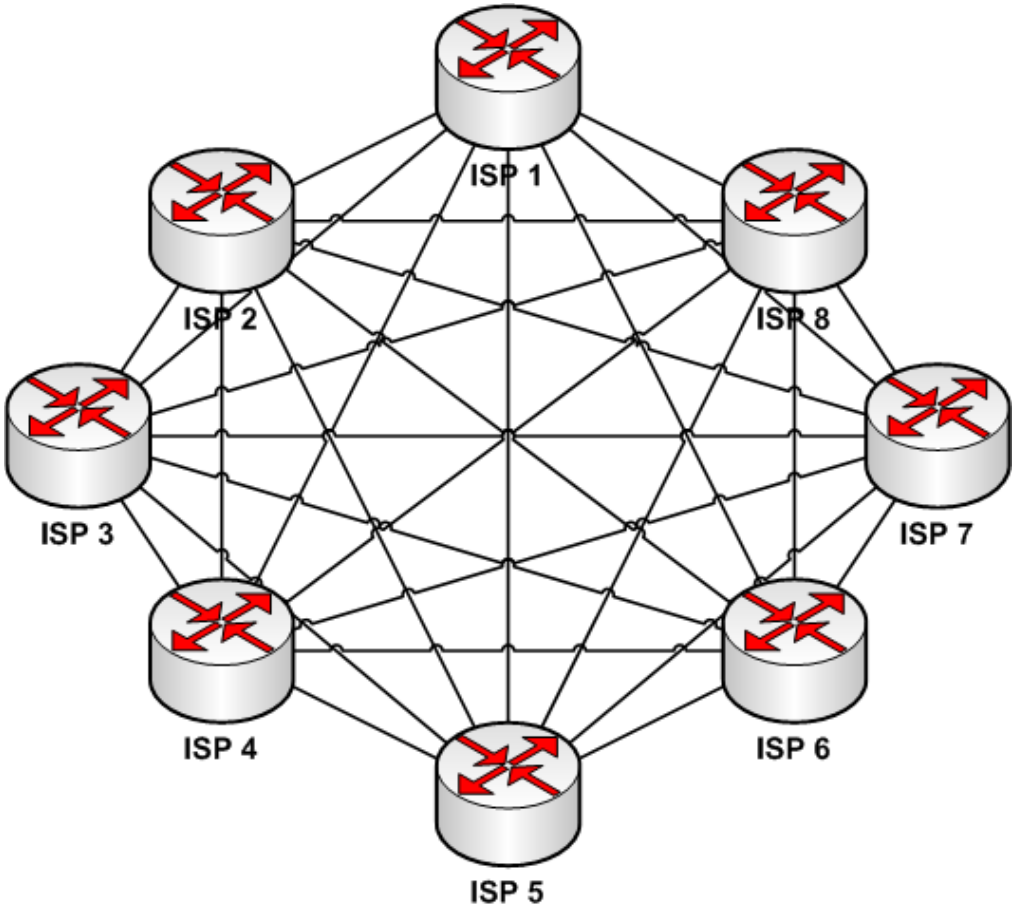
How to Exchange Traffic?

- Some ISPs run private circuits to other ISPs, like this:



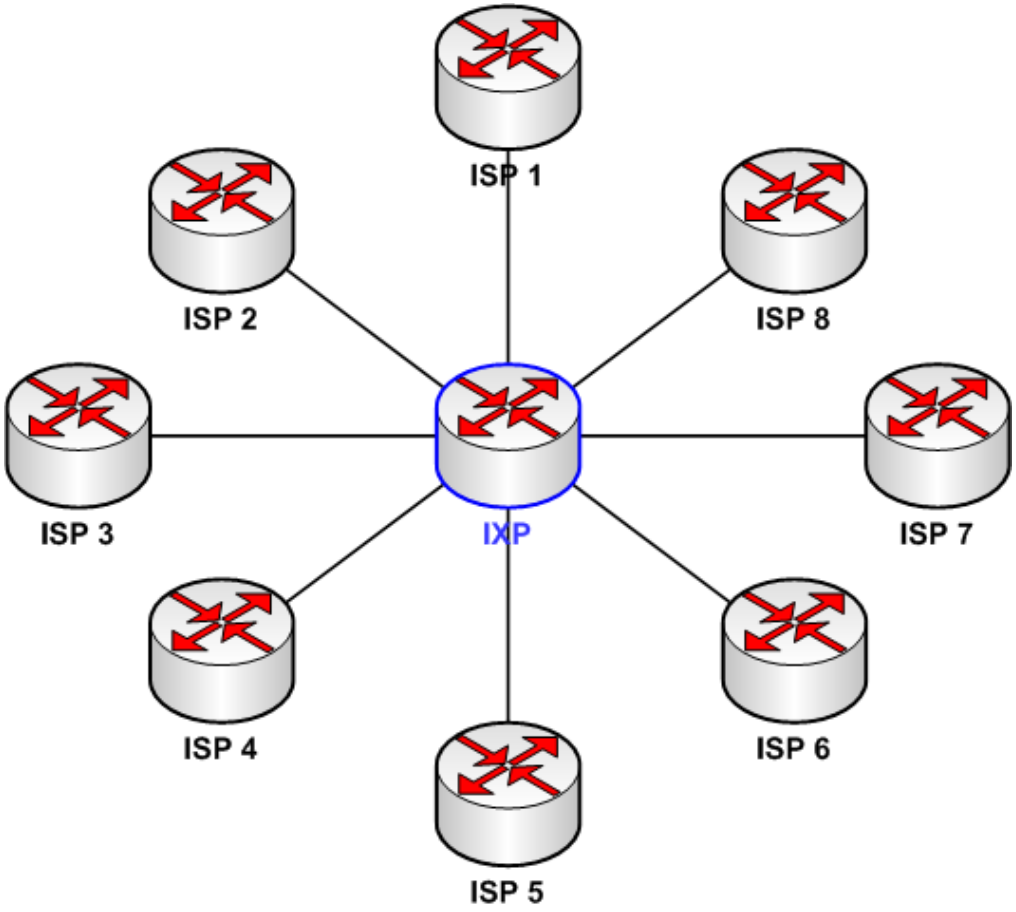
How to Exchange Traffic? (Cont.)

- **But it won't be a good idea if everyone did this…**



How to Exchange Traffic? (Cont.)

- **We need an IXP for this scenario:**



What is IXP?

- IXP stands for Internet eXchange Point, is a physical infrastructure through which ISPs exchange internet traffic between their networks.
- IXP is like a hub, which connects various ISPs together with minimum cabling costs.
- IXPs are usually run by non-profit organizations or universities, and located at a neutral location, which most ISPs can easily participate.

IXP Implementation

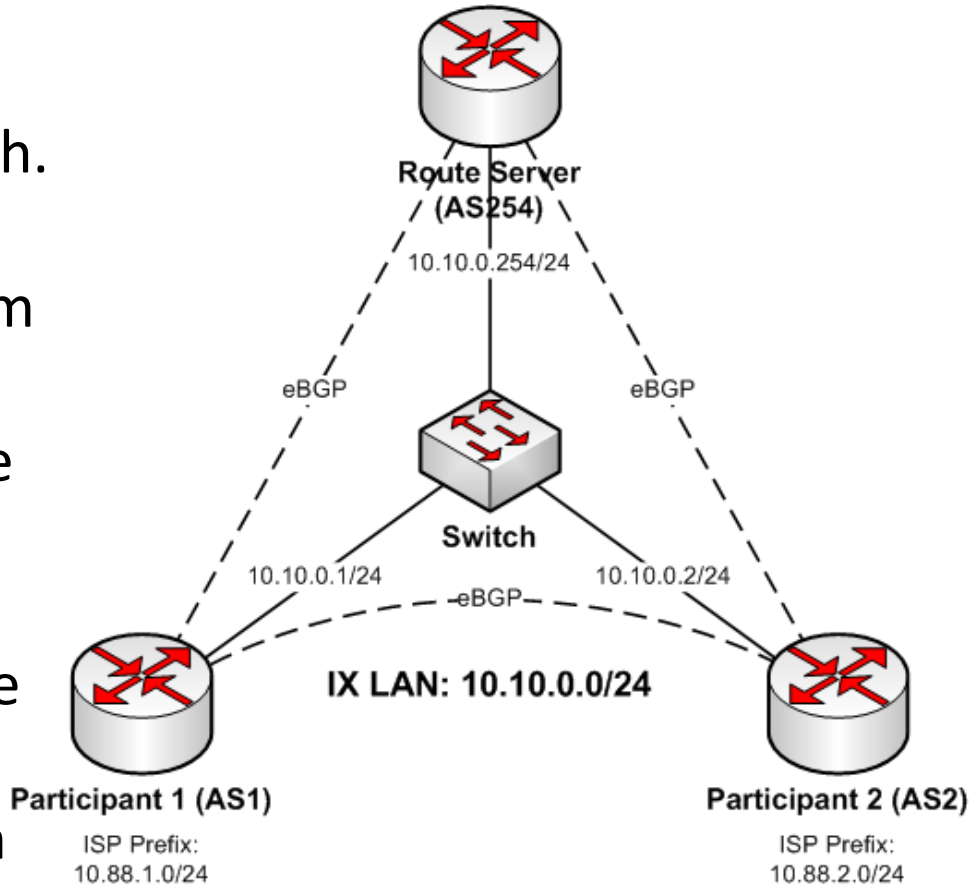
- Peering policies are various between IXPs.
- Mostly IXP participants are allowed to advertise own ISP prefixes and downstream ISP's prefixes only, advertising internet routing table or routes received from peers in the same IX are prohibited.
- Route server is present in some IXPs to add more flexibilities and availabilities.
- Peering with route server can be "Selective" or "Required", it depends on the IXP's peering policies.

IXP Implementation (Cont.)

- Benefits of route server:
  - Route server receives routes from everyone, and advertises all best routes to everyone without changing the next hop address, so traffic between participants are always switched
  - Peering with route server is a good option for new participants to get all routes before they have any peering agreement with others
  - If participants have their direct peer already, then route server still can be used for backup purpose, in case their direct peer down, they still can have failover possibility

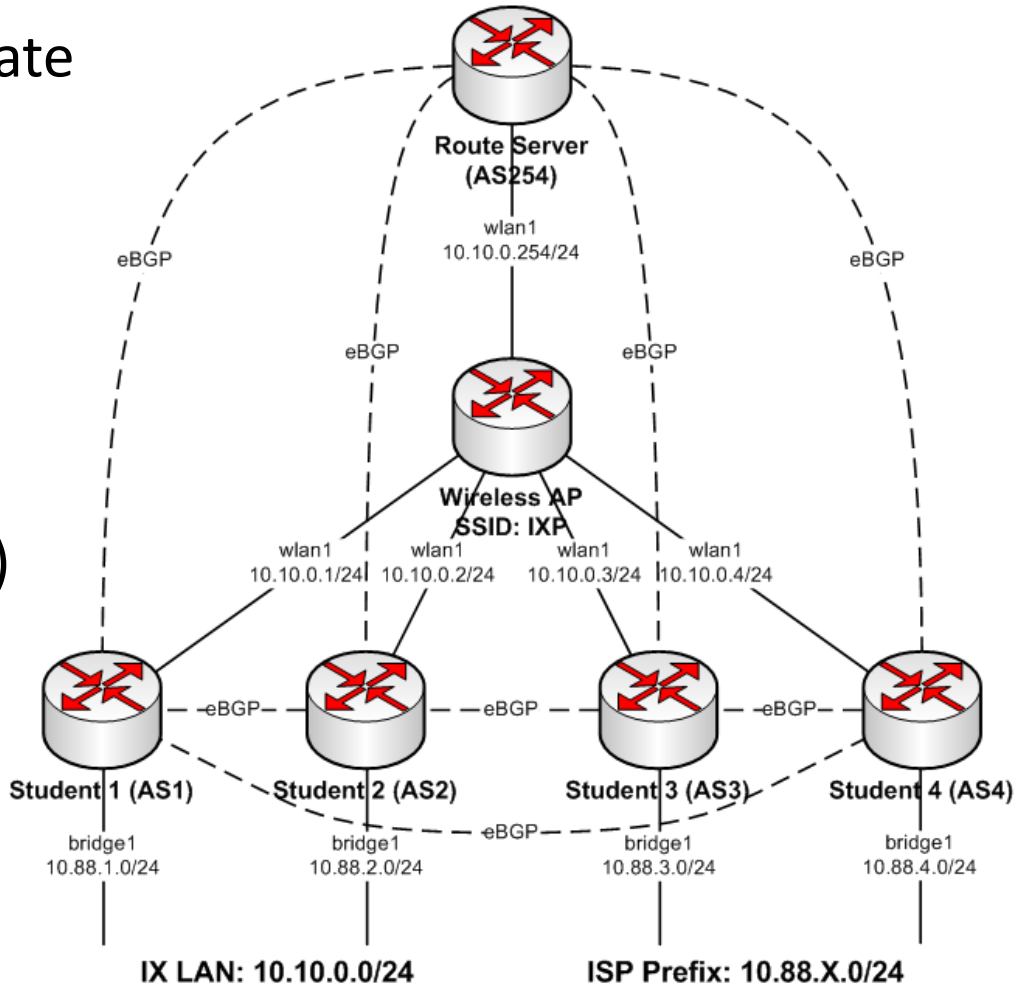
IXP Implementation (Cont.)

- All participants are connected to the IXP switch.
- Each participant was assigned an IP address from the IX LAN.
- All of them are in the same LAN.
- Participants run eBGP between them to exchange routes.
- Participants run eBGP with route server to get all routes of the entire IX network.



Lab Topology

- We use class AP to simulate the IX switch.
  - SSID: IXP
- Route server (My router)
  - ASN: 254
  - IP: 10.10.0.254/24
- Participants (Your router)
  - ASN: X
  - wlan1 IP: 10.10.0.X/24
  - ISP Prefix: 10.88.X.0/24
  - bridge1 IP: 10.88.X.1/24
  - X = Your seat number



Lab Instruction

1. Reset your router configuration.
2. Configure your router according to the lab topology.
3. Setup eBGP peer with your classmate who sits next to you:
   - Student 1 peers with student 2
   - Student 2 peers with student 3…etc.
   - The last student peers with student 1
4. Setup eBGP peer with route server.

Lab Instruction (Cont.)

5. Filter your outbound advertisement, only your ISP prefix should be advertised to peers.
6. Verify your BGP peers, everyone should have 3 eBGP peers:
   - Classmate who you go to peer with
   - Classmate who comes to peer with you
   - Route server
7. Verify your BGP advertisement and routing table.
8. Test connectivity to everyone's ISP prefix.

BGP

- BGP Stands for Border Gateway Protocol.
- Runs on TCP protocol port 179.
- Path vector protocol.
  - A path vector protocol defines a route as a pairing between a destination and the attributes of the path to that destination
- BGP is standard protocol for ISPs and IXPs to exchange internet routes.

BGP (Cont.)

- Autonomous System (AS) is the cornerstone of BGP.
  - Collection of networks with same routing policy
  - Usually under single ownership, trust and administrative control
  - Identified by a unique 32-bit integer (ASN)
- iBGP: peering between routers within the same AS.
- eBGP: peering between routers from different AS.
- By default, eBGP will advertise all BGP best routes to peers, so we need to use route filter to control the advertisement.

Route Server
BGP Configuration

- Create BGP instance.

```
/routing bgp instance add name=AS254 as=254 router-id=10.10.0.254
```

- Setup eBGP peer with everyone.

```
/routing bgp peer
add name=EBGP-AS1 instance=AS254 remote-address=10.10.0.1 remote-as=1
add name=EBGP-AS2 instance=AS254 remote-address=10.10.0.2 remote-as=2
add name=EBGP-AS3 instance=AS254 remote-address=10.10.0.3 remote-as=3
add name=EBGP-AS4 instance=AS254 remote-address=10.10.0.4 remote-as=4
add name=EBGP-AS5 instance=AS254 remote-address=10.10.0.5 remote-as=5
add name=EBGP-AS6 instance=AS254 remote-address=10.10.0.6 remote-as=6
add name=EBGP-AS7 instance=AS254 remote-address=10.10.0.7 remote-as=7
add name=EBGP-AS8 instance=AS254 remote-address=10.10.0.8 remote-as=8
…
```

Student Router BGP Configuration

- Create BGP instance.

```
/routing bgp instance add name=ASX as=X router-id=10.10.0.X
```

- Advertise your ISP prefix.

```
/routing bgp network add network=10.88.X.0/24
```

- Configure route filter “EBGP-OUT” to advertise only your prefix.

```
/routing filter add chain=EBGP-OUT prefix=10.88.X.0/24 action=accept
/routing filter add chain=EBGP-OUT action=discard
```

- **X** = Your seat number

Student Router BGP Configuration (Cont.)

- Setup eBGP peer with your classmate, filter your BGP advertisement with route filter “EBGP-OUT”.
  - You will need to repeat this step two times, since there are two direct peers for each student:
    - Classmate who you go to peer with
    - Classmate who comes to peer with you

```
/routing bgp peer add name=EBGP-ASY instance=ASX remote-
address=10.10.0.Y remote-as=Y out-filter=EBGP-OUT
```

- Setup eBGP peer with route server , filter your BGP advertisement with route filter “EBGP-OUT”.

```
/routing bgp peer add name=EBGP-RS instance=ASX remote-address=10.10.0.254
remote-as=254 out-filter=EBGP-OUT
```

- **X** = Your seat number
- **Y** = Your classmate seat number

Route Server Verification

- Verify BGP peers.
  - One eBGP peer for each student

```
[admin@RS] > /routing bgp peer print
Flags: X - disabled, E - established
 #    INSTANCE        REMOTE-ADDRESS                                  REMOTE-AS
 0 E  AS254           10.10.0.1                                       1
 1 E  AS254           10.10.0.2                                       2
 2 E  AS254           10.10.0.3                                       3
 3 E  AS254           10.10.0.4                                       4
```

- Verify BGP advertisements.
  - Redistribute best routes to everyone

PEER	PREFIX	NEXTHOP	AS-PATH	ORIGIN	LOCAL-PREF
EBGP-AS1	10.88.3.0/24	10.10.0.3	3	igp	
EBGP-AS1	10.88.2.0/24	10.10.0.2	2	igp	
EBGP-AS1	10.88.4.0/24	10.10.0.4	4	igp	
EBGP-AS2	10.88.3.0/24	10.10.0.3	3	igp	
EBGP-AS2	10.88.1.0/24	10.10.0.1	1	igp	
EBGP-AS2	10.88.4.0/24	10.10.0.4	4	igp	
EBGP-AS3	10.88.1.0/24	10.10.0.1	1	igp	
EBGP-AS3	10.88.2.0/24	10.10.0.2	2	igp	
EBGP-AS3	10.88.4.0/24	10.10.0.4	4	igp	
EBGP-AS4	10.88.3.0/24	10.10.0.3	3	igp	
EBGP-AS4	10.88.1.0/24	10.10.0.1	1	igp	
EBGP-AS4	10.88.2.0/24	10.10.0.2	2	igp	

Route Server Verification (Cont.)

- Verify routing table.
  - Has routes from everyone

```
[admin@RS] > /ip route print
Flags: X - disabled, A - active, D - dynamic,
C - connect, S - static, r - rip, b - bgp, o - ospf, m - mme,
B - blackhole, U - unreachable, P - prohibit
 #      DST-ADDRESS        PREF-SRC        GATEWAY            DISTANCE
 0 ADC  10.10.0.0/24       10.10.0.254     wlan1                     0
 1 ADb  10.88.1.0/24                       10.10.0.1                20
 2 ADb  10.88.2.0/24                       10.10.0.2                20
 3 ADb  10.88.3.0/24                       10.10.0.3                20
 4 ADb  10.88.4.0/24                       10.10.0.4                20
```

Student Router Verification

- We use Student 2 as example in this slide.
- Verify BGP peers.
  - Everyone should have 3 eBGP peers:
    - Classmate who you go to peer with
    - Classmate who comes to peer with you
    - Route server

```
[admin@AS2] > /routing bgp peer print
Flags: X - disabled, E - established
 #    INSTANCE        REMOTE-ADDRESS                                  REMOTE-AS
 0 E  AS2             10.10.0.1                                       1
 1 E  AS2             10.10.0.3                                       3
 2 E  AS2             10.10.0.254                                     254
```

- Verify BGP advertisements.
  - Should be only 1 prefix advertised per peer, which is your ISP prefix (10.88.X.0/24)

```
[admin@AS2] > /routing bgp advertisements print
PEER      PREFIX              NEXTHOP          AS-PATH    ORIGIN     LOCAL-PREF
EBGP-AS1  10.88.2.0/24        10.10.0.2                   igp
EBGP-AS3  10.88.2.0/24        10.10.0.2                   igp
EBGP-RS   10.88.2.0/24        10.10.0.2                   igp
```

Student Router Verification (Cont.)

- Verify routing table.
  - You should see everyone's ISP prefixes in your routing table
  - There will be two prefixes that have two BGP routes each, because:
    - You received one from direct peer with your classmate
    - You received another one from route server
    - Routes from route server always have lower priority than direct peer, because of longer AS path, so they stay inactive as long as there is the same prefix from direct peer

```
[admin@AS2] > /ip route print
Flags: X - disabled, A - active, D - dynamic,
C - connect, S - static, r - rip, b - bgp, o - ospf, m - mme,
B - blackhole, U - unreachable, P - prohibit
 #      DST-ADDRESS        PREF-SRC        GATEWAY            DISTANCE
 0 ADC  10.10.0.0/24       10.10.0.2       wlan1                     0
 1 ADb  10.88.1.0/24                       10.10.0.1                20
 2  Db  10.88.1.0/24                       10.10.0.1                20
 3 ADC  10.88.2.0/24       10.88.2.1       bridge1                   0
 4 ADb  10.88.3.0/24                       10.10.0.3                20
 5  Db  10.88.3.0/24                       10.10.0.3                20
 6 ADb  10.88.4.0/24                       10.10.0.4                20
```

Student Router Verification (Cont.)

- See details of the prefix.
  - Route from route server has AS254 in the AS path, which made its AS path longer than the route from direct peer

```
[admin@AS2] > /ip route print detail where dst-address=10.88.3.0/24
Flags: X - disabled, A - active, D - dynamic,
C - connect, S - static, r - rip, b - bgp, o - ospf, m - mme,
B - blackhole, U - unreachable, P - prohibit
 4 ADb  dst-address=10.88.3.0/24 gateway=10.10.0.3
        gateway-status=10.10.0.3 reachable via  wlan1 distance=20 scope=40
        target-scope=10 bgp-as-path="3" bgp-origin=igp received-from=EBGP-AS3

 5  Db  dst-address=10.88.3.0/24 gateway=10.10.0.3
        gateway-status=10.10.0.3 reachable via  wlan1 distance=20 scope=40
        target-scope=10 bgp-as-path="254,3" bgp-origin=igp
        received-from=EBGP-RS
```

Student Router Verification (Cont.)

- Test connectivity to everyone's ISP prefix.

```
[admin@AS2] > /ping 10.88.1.1 src-address=10.88.2.1 count=1
HOST                                     SIZE TTL TIME  STATUS
10.88.1.1                                  56  64 1ms
    sent=1 received=1 packet-loss=0% min-rtt=1ms avg-rtt=1ms max-rtt=1ms
HOST                                     SIZE TTL TIME  STATUS

[admin@AS2] >
[admin@AS2] > /ping 10.88.3.1 src-address=10.88.2.1 count=1
HOST                                     SIZE TTL TIME  STATUS
10.88.3.1                                  56  64 2ms
    sent=1 received=1 packet-loss=0% min-rtt=2ms avg-rtt=2ms max-rtt=2ms
HOST                                     SIZE TTL TIME  STATUS

[admin@AS2] >
[admin@AS2] > /ping 10.88.4.1 src-address=10.88.2.1 count=1
HOST                                     SIZE TTL TIME  STATUS
10.88.4.1                                  56  64 2ms
    sent=1 received=1 packet-loss=0% min-rtt=2ms avg-rtt=2ms max-rtt=2ms
HOST                                     SIZE TTL TIME  STATUS
```

Questions & Answers

If you have any questions, please feel free to ask!

The End

Thanks for your attention!

Contact Me

makito@informationbeam.net

Skype: akn_makito