Get your hands dirty with BGP

Iljitsch van Beijnum

RIPE-70
Amsterdam, May 2015

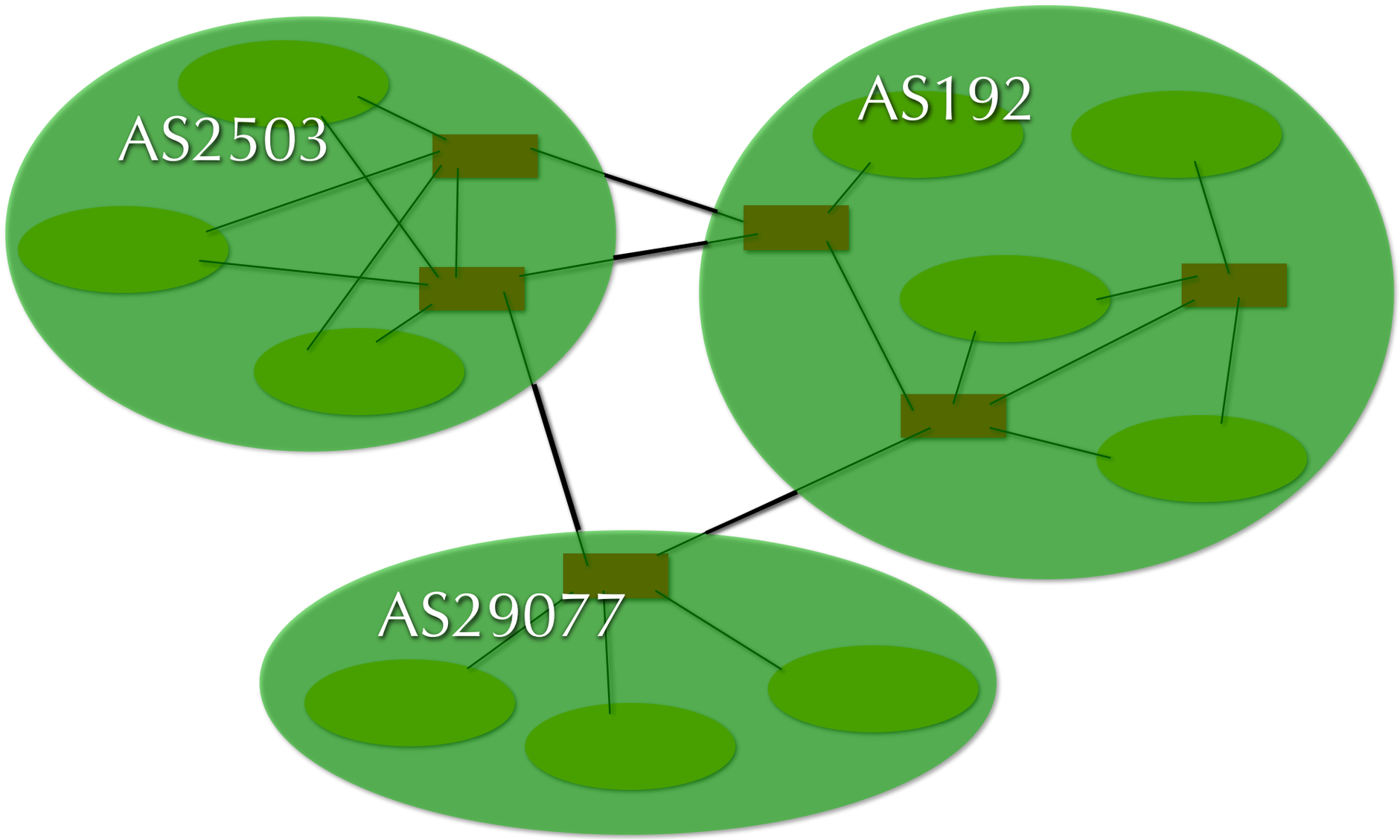
BGP

- "Border Gateway Protocol"
- Gateway = router
- Border gateway / border router: router between our own internal network and external networks
- This makes BGP an "Exterior Gateway Protocol" (external routing protocol)
- As opposed to "Interior Gateway Protocols" such as RIP, OSPF, EIGRP and IS-IS

BGP vital stats

- BGP works over TCP port 179
- Configure BGP "neighbors" manually!
- BGP works with large address blocks and "autonomous systems" (ASes)
- Hardware: Cisco, Juniper, Brocade (Foundry), Riverstone, Extreme, etc.
- Software: GNU Zebra, Quagga, OpenBGPd, ZebOS, GateD, BIRD, etc.

Autonomous Systems



How BGP works

- Routers try to connect on TCP port 179
- After connection, send capabilities
  - if capabilities don't match, disconnect, error
- Send copies of best routes (ones the router itself uses) for all destinations to neighboring router
  - if allowed by policy!
- Then, only send keepalives; updates if there is a change

How BGP works (3)

AS 123, 10.0.0.1

	Network	Next Hop	Path
>	65.0.3.0/24	84.0.4.1	18 286 4323 i
>	72.8.0.0/16	93.6.0.4	1239 i

AS 456, 172.16.0.1

	Network	Next Hop	Path
>	65.0.3.0/24	17.3.58.1	4323 i
>	72.8.0.0/16	93.6.0.4	1239 i

AS 123, 10.0.0.1

	Network	Next Hop	Path
	65.0.3.0/24	84.0.4.1	18 286 4323 i
>		172.16.0.1	456 4323 i
>	72.8.0.0/16	93.6.0.4	1239 i
		172.16.0.1	456 1239 i

AS 456, 172.16.0.1

	Network	Next Hop	Path
	65.0.3.0/24	10.0.0.1	123 18 286 4323 i
>		17.3.58.1	4323 i
	72.8.0.0/16	10.0.0.1	123 1239 i
>		93.6.0.4	1239 i

withdraw

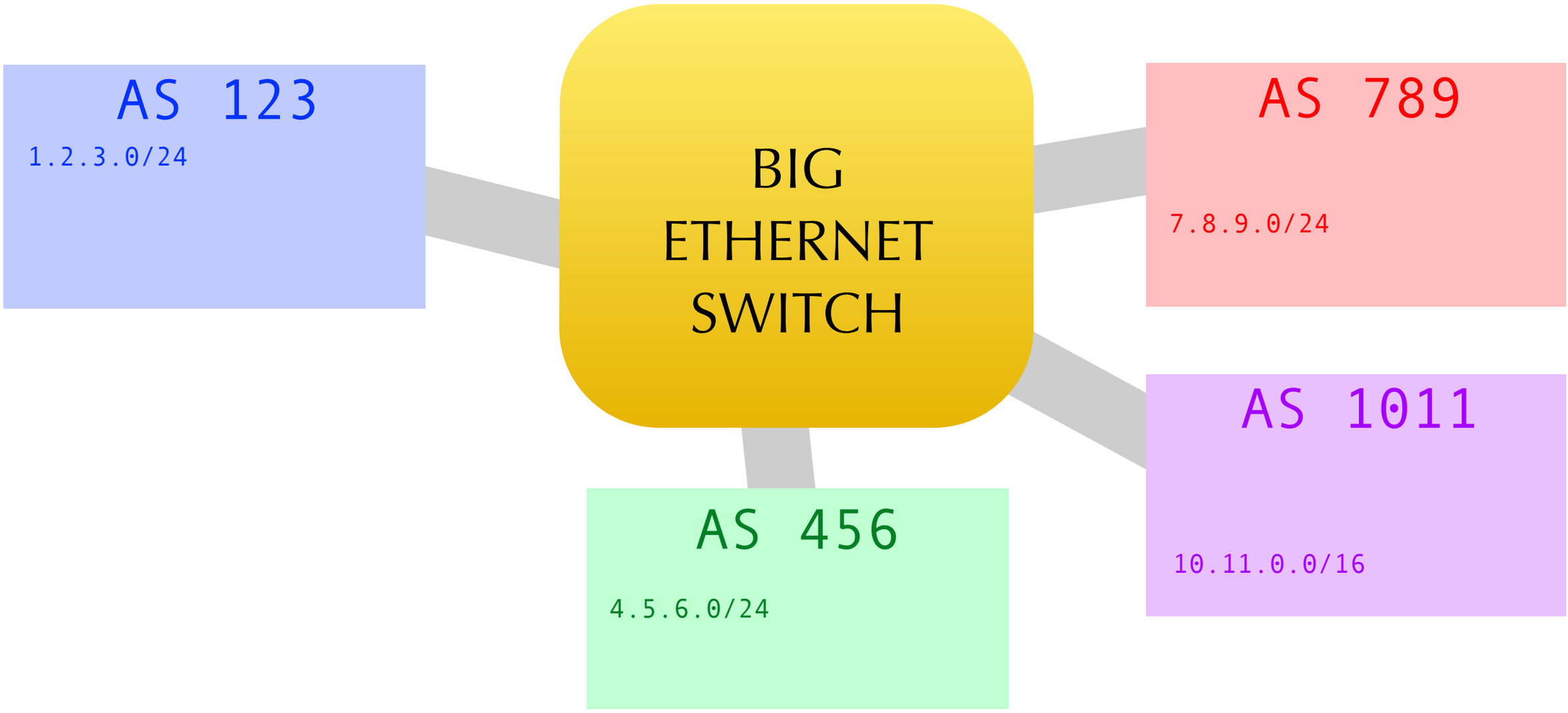
AS 123, 10.0.0.1

	Network	Next Hop	Path
	65.0.3.0/24	84.0.4.1	18 286 4323 i
>		172.16.0.1	456 4323 i
>	72.8.0.0/16	93.6.0.4	1239 i
		172.16.0.1	456 1239 i

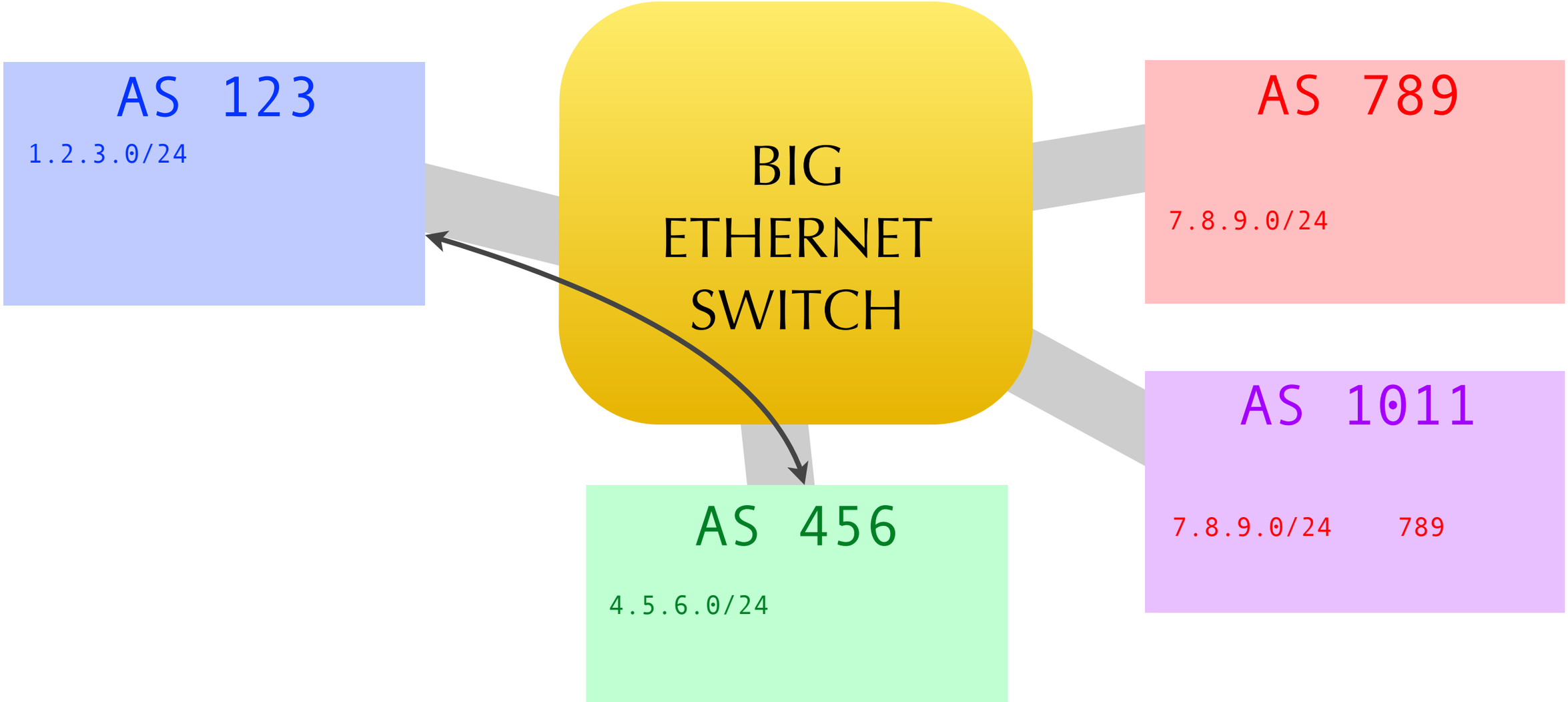
AS 456, 172.16.0.1

	Network	Next Hop	Path
>	65.0.3.0/24	17.3.58.1	4323 i
	72.8.0.0/16	10.0.0.1	123 1239 i
>		93.6.0.4	1239 i

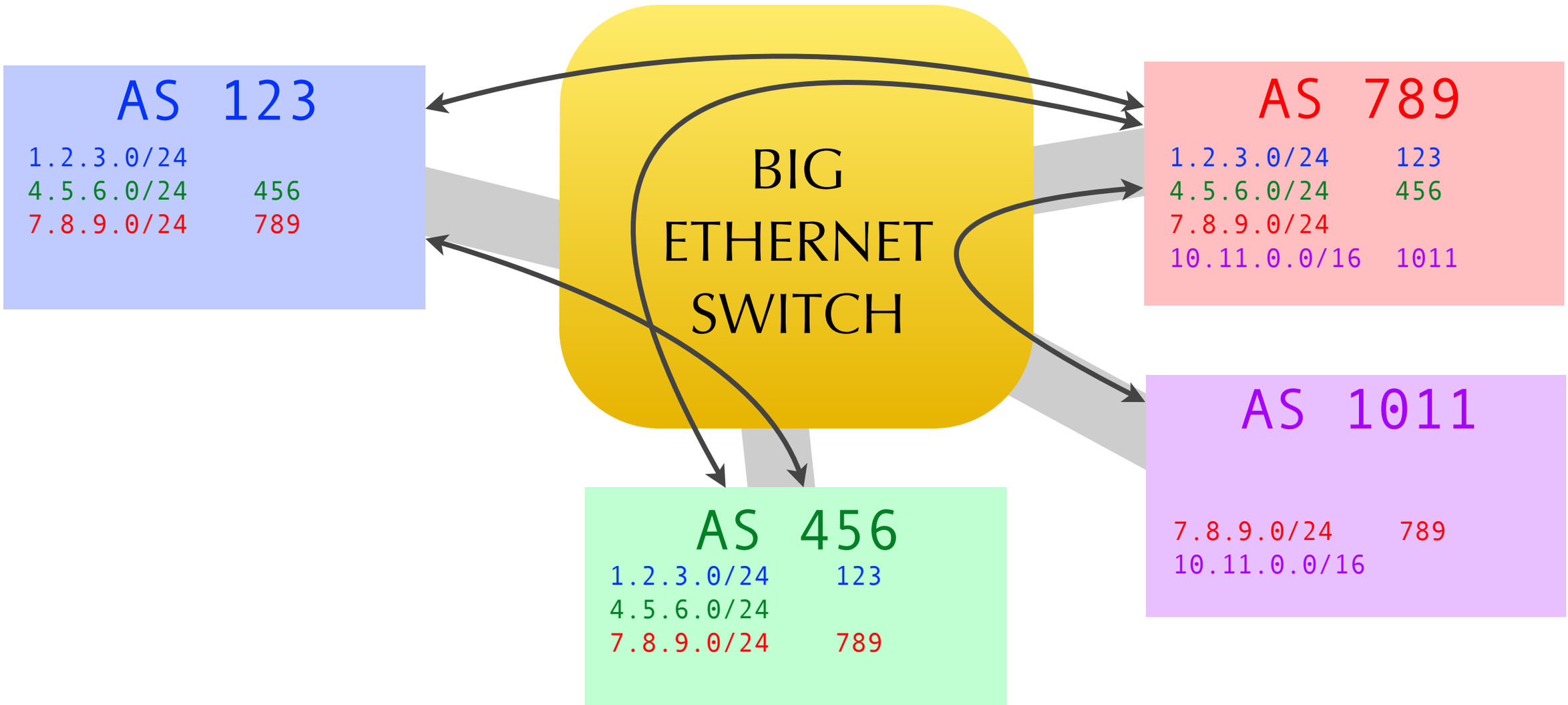
Internet exchange



IX peering


AS 123
1.2.3.0/24
BIG
ETHERNET
SWITCH
AS 789
7.8.9.0/24
AS 1011
7.8.9.0/24 789
AS 456
4.5.6.0/24

IX peering (2)


AS 123
1.2.3.0/24
4.5.6.0/24 456
7.8.9.0/24 789
BIG ETHERNET SWITCH
AS 789
1.2.3.0/24 123
4.5.6.0/24 456
7.8.9.0/24
10.11.0.0/16 1011
AS 1011
7.8.9.0/24 789
10.11.0.0/16
AS 456
1.2.3.0/24 123
4.5.6.0/24
7.8.9.0/24 789

Path attributes

- Information attached to a prefix in BGP:
  - Next hop: where the packets go
  - AS path: all intermediate AS numbers
  - Local preference: indicates... local preference
  - Multi Exit Discriminator (MED) or "metric": indicates neighboring AS's preference
  - Community: two numbers (702:120), no fixed meaning

Configuring BGP

```
!
router bgp 123
 network 192.0.2.0/24
 neighbor 64.51.2.33 remote-as 65065
 neighbor 64.51.2.33 description Transit ISP
 neighbor 223.223.223.90 remote-as 456
 neighbor 223.223.223.90 description IX peer
!
```

Configuring BGP

```
!
router bgp 123
 neighbor 3ffe:9500:3c:74::10 remote-as 65065
 no neighbor 3ffe:9500:3c:74::10 activate
!
address-family ipv6
 neighbor 3ffe:9500:3c:74::10 activate
 network 2001:db8::/32
 exit-address-family
!
```

BGP route selection

1. Prefer the path with the largest WEIGHT
2. Prefer the path with the largest LOCAL_PREF
3. Prefer the path that was locally originated via a network or aggregate BGP subcommand, or through redistribution from an IGP
4. Prefer the path with the shortest AS_PATH
5. Prefer the path with the lowest origin type
6. Prefer the path with the lowest multi-exit discriminator (MED)
7. Prefer external (eBGP) over internal (iBGP) paths
8. Prefer the path with the lowest IGP metric to the BGP next hop
9. (...)
10. When both paths are external, prefer the path that was received first (the oldest one)
11. Prefer the route coming from the BGP router with the lowest router ID
12. If the originator or router ID is the same for multiple paths, prefer the path with the minimum cluster ID length
13. Prefer the path coming from the lowest neighbor address

http://www.cisco.com/warp/public/459/25.shtml

Most important

1. Prefer the path with the largest WEIGHT
2. Prefer the path with the largest LOCAL_PREF
3. Prefer the path that was locally originated via a network or aggregate BGP subcommand, or through redistribution from an IGP
4. Prefer the path with the shortest AS_PATH
5. Prefer the path with the lowest origin type
6. Prefer the path with the lowest multi-exit discriminator (MED)
7. Prefer external (eBGP) over internal (iBGP) paths
8. Prefer the path with the lowest IGP metric to the BGP next hop
9. (...)
10. When both paths are external, prefer the path that was received first (the oldest one)
11. Prefer the route coming from the BGP router with the lowest router ID
12. If the originator or router ID is the same for multiple paths, prefer the path with the minimum cluster ID length
13. Prefer the path coming from the lowest neighbor address

http://www.cisco.com/warp/public/459/25.shtml

Most important

- Prefer the path with the largest LOCAL_PREF
- Prefer the path with the shortest AS_PATH
- Prefer the path with the lowest multi-exit discriminator (MED)
- Communities!

Influence BGP

- "Route map": if-then construction, if route satisfies match condition, then execute set action
- Possible conditions:
  - destination falls within a range of IP addresses
  - regular expression on AS path
  - community
  - or simply everything from a certain neighbor

Route map actions

- Set local preference
- Make AS path longer by prepending it
- Set or adjust MED
- Add or remove communities
- Or filter out the route altogether

Increase local pref

```
!
router bgp 123
 neighbor 223.223.223.90 remote-as 456
 neighbor 223.223.223.90 description IX peer
 neighbor 223.223.223.90 route-map loc-pref in
!
route-map loc-pref permit 10
 set local-preference 110
!
```

Prepend path

```
ip as-path access-list 25 permit _1103_
!
route-map selective-prep permit 10
 match as-path 25
 set as-path prepend 123 123
!
route-map selective-prep permit 20
 set as-path prepend 123
!
```

Prefix lists

- Filter route destination (IP address ranges)

```
!
router bgp 123
 neighbor 223.223.223.90 remote-as 65456
 neighbor 223.223.223.90 prefix-list export out
 neighbor 223.223.223.90 prefix-list import in
!
ip prefix-list export permit 192.0.2.0/24
!
ip prefix-list import deny 192.0.2.0/24 le 32
ip prefix-list import deny 223.223.222.0/23 le 32
ip prefix-list import permit 0.0.0.0/0 le 24
!
```

Prefix lists (2)

```
!
router bgp 123
 neighbor 2223:2223::6:5456 remote-as 65456
!
address-family ipv6
 neighbor 2223:2223::6:5456 activate
 neighbor 2223:2223::6:5456 prefix-list export out
 neighbor 2223:2223::6:5456 prefix-list import in
!
ipv6 prefix-list export permit 2001:db8::/32
!
ipv6 prefix-list import deny 2001:db8::/32 le 128
ipv6 prefix-list import deny 2223:2223::/64 le 128
ipv6 prefix-list import permit ::/0 le 48
!
```

Filtering

- Don't provide transit by mistake
- AS path filter list: filter on AS
- Prefix list: filter on IP addresses

- Unfortunately, full scale incoming route filtering largely infeasible on internet exchange
- So everyone must have good outgoing filters!
- Incoming: at least reject own block and IX prefix
- Also, as a safety net: maximum prefixes

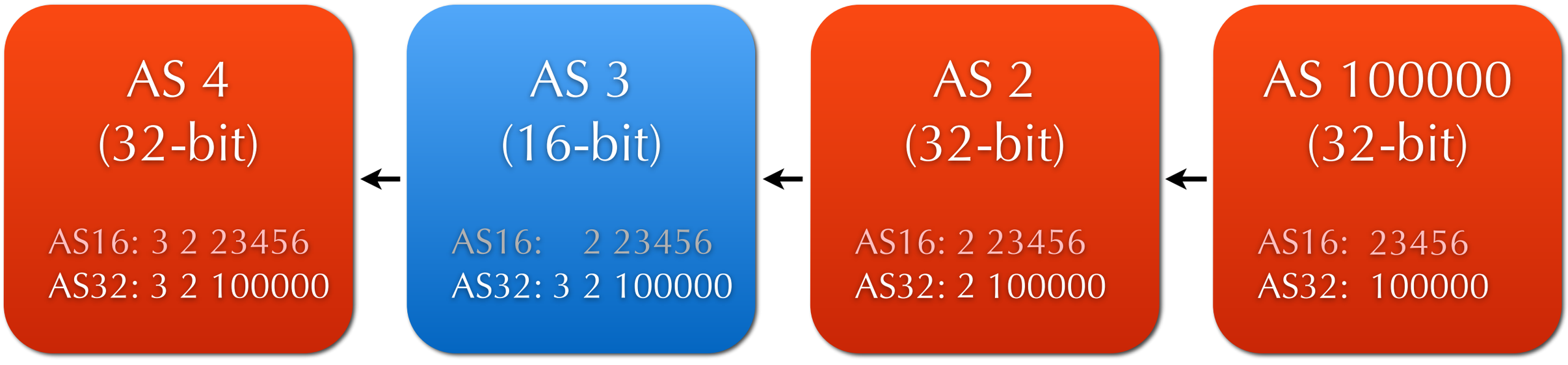
BGP security

- Protect BGP TCP sessions:
  - today: with RFC 2385 MD5 password option
  - in the future: IPsec?
- Protect data inside BGP:
  - today: huge filters (based on routing registry?)
  - soon/now: address space certificates (rPKI)
  - past (future?): Secure BGP or Secure Origin BGP

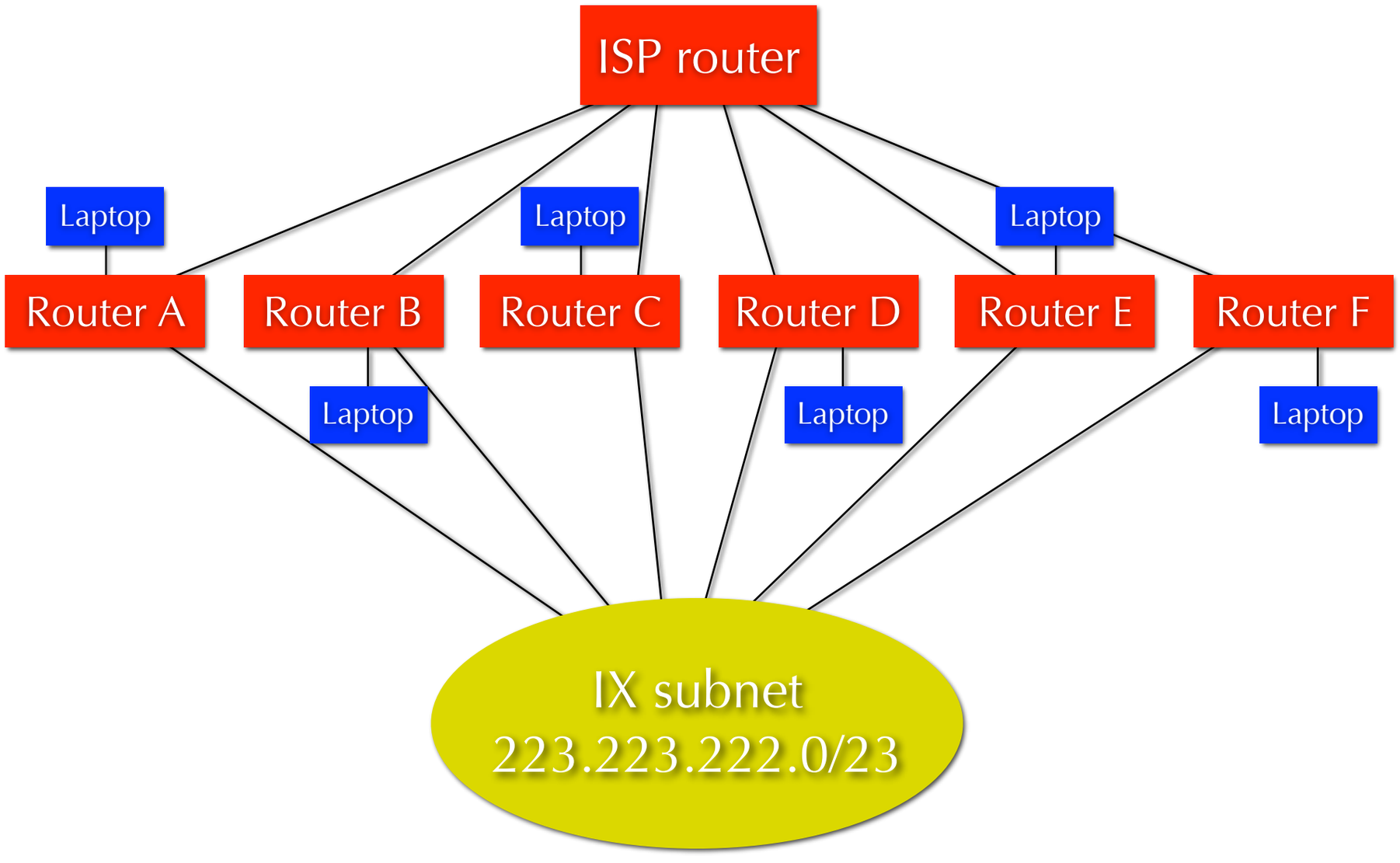
32-bit AS numbers

- AS numbers originally 16 bits: 64500 usable
  - in use now: 59146 (with ± 3050/year new)
- So make 32 bits
  - hide 32-bit AS path in new attribute
  - 16-bit only routers see AS 23456
- 32-bit AS numbers:
  - in use now: 10694 (with ± 3000/year new)

32-bit AS numbers (2)



Our test network



IP addresses

- You get a letter/number, like G = 71
- Configs for your VM(s): www.bgpexpert.com/course.php
  - AS number: 9997<u>1</u>
  - prefix: 99.<u>71</u>.0.0/20 / 2001:960:7bf:<u>71</u>00::/56
  - admin address: 99.<u>71</u>.0.1/24
  - addresses for transit: 10.0.<u>71</u>.2/30 / 2001:960:7bf:<u>71</u>::2
  - IX: 223.223.223.<u>65</u>/23 / 2001:960:7bf:223::9:99<u>71</u>/64

Exercises

- Feel free to experiment on your own!
- Or do these exercises:
  - set up a transit BGP session to 10.0.XX.1 (no filters yet!)
  - find two other people to peer with, set up BGP sessions towards them
  - increase the local preference on your peering prefixes
  - set up filtering for your peering sessions
  - set up MD5 passwords on your BGP sessions

Common commands

- Show status neighbors: `show ip bgp sum` / `show bgp sum`
- Routes to/from neighbor: `sh ip bgp nei 1.2.3.4` / `2001:abc::1` `advertised-routes` / `paths` / `routes` / `received-routes` / `received`
- Show BGP table: `show ip bgp` / `show bgp` / `show bgp ipv6 unicast`
- Show prefix details: `show ip bgp 10.0.0.0`
- Reset session: `clear ip bgp 12345` / `1.2.3.4`
  - or * for all sessions, also with "in" or "out"

Thanks for listening!

http://www.bgpexpert.com/

iljitsch@bgpexpert.com