CLOUD**X**CHANGE

Network Peering Issue

Bahirul Rizki Ariyanto

ari@cloudxchange.id

Agenda

- Common Misconfigurations in IXP Connections

- Routing Optimization

- Peering Traffic Analysis

- Next-Generation IP Peer Scalability on IXP

Common Misconfigurations in IXP Connections

Common Misconfigurations ⚠️

Case:

Not received any prefixes from IXP route servers ?

```
BGP table version 192
RIB entries 186, using 33 KiB of memory
Peers 1, using 723 KiB of memory

Neighbor        V         AS   MsgRcvd   MsgSent   TblVer   InQ OutQ  Up/Down  State/PfxRcd
103.225.171.45  4     133339        70         5        0     0    0  00:00:43             0
```

Common Misconfigurations ⚠️

Cause 1 (enforce-first-as):

CXC-IX is Layer 2 IXP, first as-path (route servers) is not visible on prefixes received, IXP members need to configure "`no enforce-first-as`" on some vendors.

Needed on some vendors so router can receive prefixes without AS133339 being first in path

```
router bgp 100
    no bgp enforce-first-as
    neighbor x.x.x.a remote-as 133339
    neighbor x.x.x.a route-map cxc-rs1-in in
    neighbor x.x.x.a route-map cxc-rs1-out out
```

Common Misconfigurations ⚠️

Cause 2 (bgp filter):

route-map import on member filtered with match first-as of IXP. There's no need this filtered since CXC-IX is Layer 2 IXP.
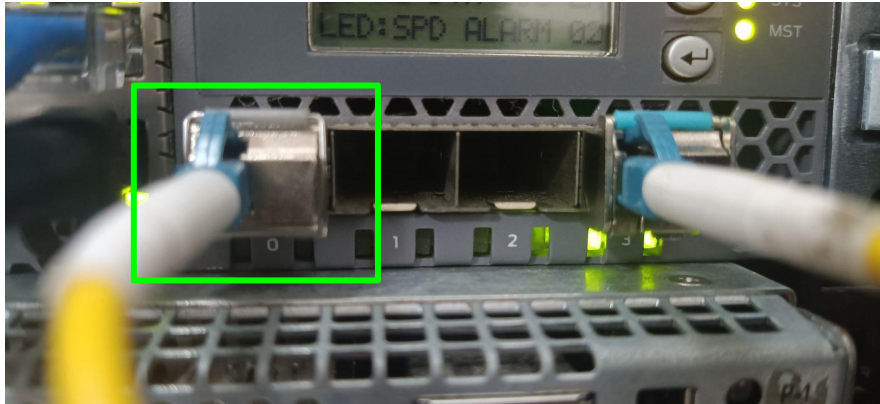
Filtered prefix match with first-as to 133339, **no need** this on Layer 2 ixp.

```
bgp as-path access-list cxc seq 5 permit ^133339_
```

Common Misconfigurations ⚠️

Case:

Port to IXP blocked or down ?

Common Misconfigurations ⚠️

Cause 1 (Blocked by bpdu-guard):

CXC-IX implement bpdu-guard to protect against unwanted stp change from member ports.
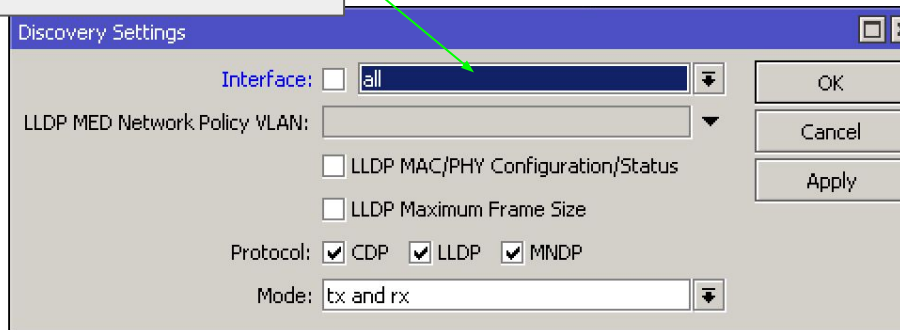
Filter bpdu on port facing to IXP.

```
spanning-tree bpdufilter enable
```

Common Misconfigurations ⚠️

Cause 2 (Blocked by unwanted broadcast traffic):

CXC-IX filter broadcast traffic on member ports. Please disable protocol like CDP, LLDP, MNDP, etc.

Filter MNDP interface **not use** "all".

Please **disable** discovery protocol like cdp..

Discovery Settings

Interface: ☐	all ▼
LLDP MED Network Policy VLAN:	▼
	☐ LLDP MAC/PHY Configuration/Status
	☐ LLDP Maximum Frame Size
Protocol:	☑ CDP ☑ LLDP ☑ MNDP
Mode:	tx and rx ▼

OK
Cancel
Apply

```
interface Ethernet0/0

 description To CXC-IX

 ............

 no cdp enable

 spanning-tree bpdufilter enable
```
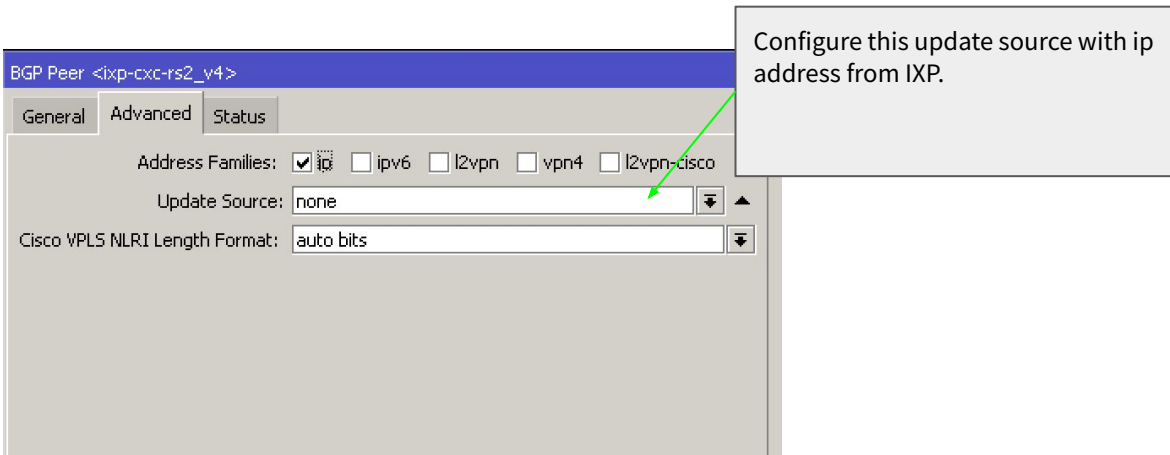
Common Misconfigurations ⚠️

Case:

BGP session is not established on Mikrotik Router OS ?

■■ ■ ixp-cxc-rs2_v4	default	103.225.171.45	133339 active

Common Misconfigurations ⚠️

Cause (update-source ip not defined):

Some case in Mikrotik Router OS need to define update-source on BGP peer.

Configure this update source with ip address from IXP.

Routing Optimization

Routing Optimization ⚙️

Case 1:

Cannot access specific web/prefix?

This site can't be reached

domain.com took too long to respond.

Try:
- Checking the connection
- Checking the proxy and the firewall

ERR_CONNECTION_TIMED_OUT

Details   Reload

Routing Optimization ⚙️

Case 2:

High latency to local ISP/Content/Network?
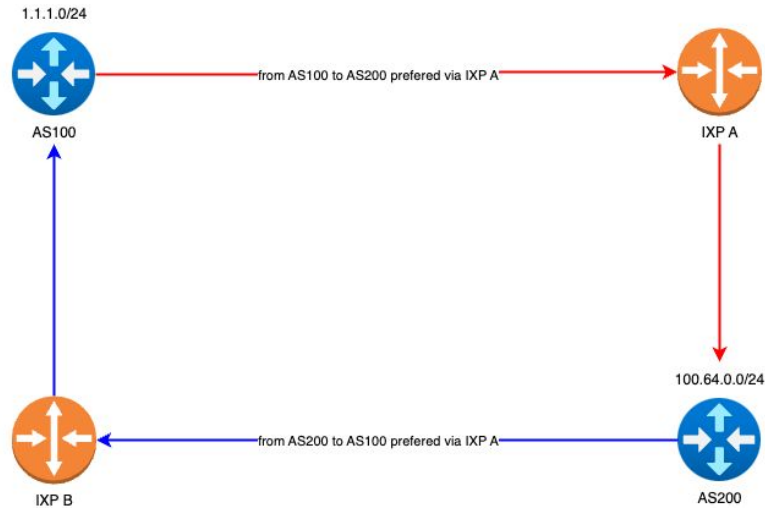
```
PING 10.10.10.1 (10.10.10.1): 56 data bytes
64 bytes from 10.10.10.1: icmp_seq=0 ttl=56 time=40.425 ms
64 bytes from 10.10.10.1: icmp_seq=1 ttl=56 time=40.683 ms
64 bytes from 10.10.10.1: icmp_seq=2 ttl=56 time=40.058 ms
64 bytes from 10.10.10.1: icmp_seq=3 ttl=56 time=40.198 ms
64 bytes from 10.10.10.1: icmp_seq=4 ttl=56 time=40.547 ms

--- 1.1.1.1 ping statistics ---
```

Routing Optimization ⚙️

Cause (asymmetric routing) :
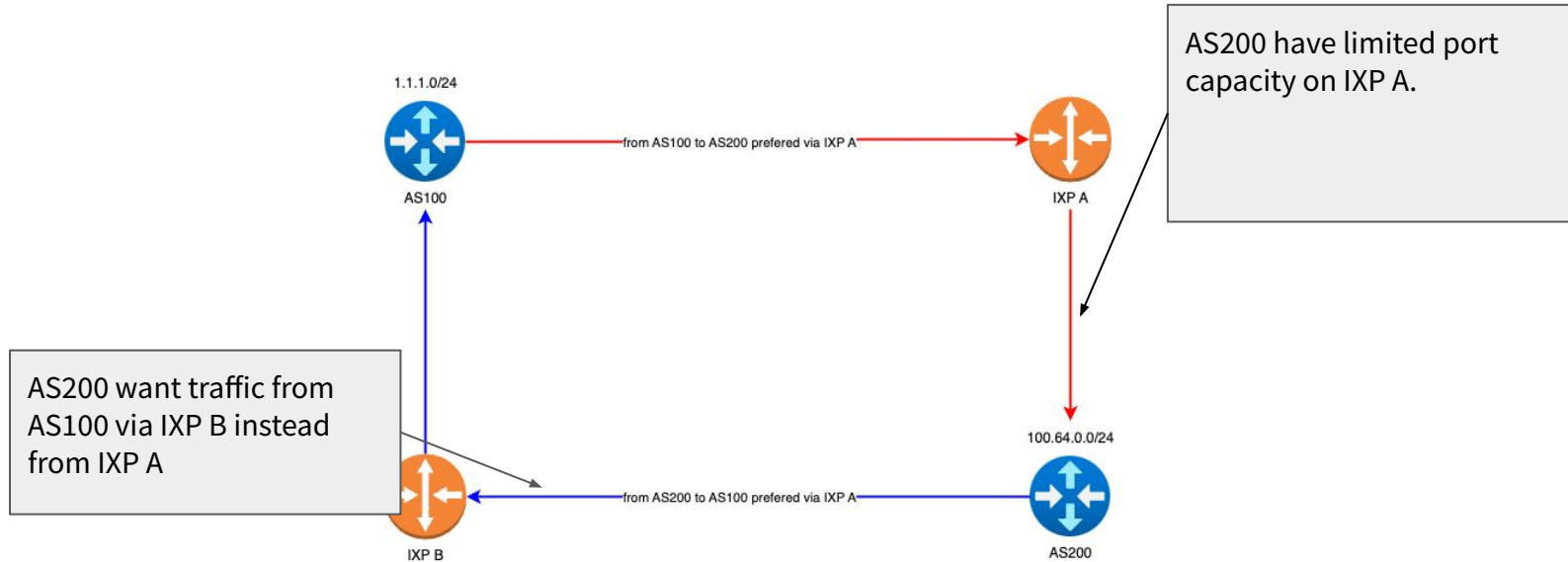
Routing Optimization ⚙️

Solution (control advertise on IXP using BGP communities):

CXC-IX support BGP community advertisement control since 2019, with prepend, discard/drop to specific $AS and blackhole[new].

👉 https://cloudxchange.id/rs

Routing Optimization ⚙️

Solution (control advertise on IXP using BGP communities):

AS200 have limited port capacity on IXP A.

1.1.1.0/24

from AS100 to AS200 prefered via IXP A

IXP A

AS100

AS200 want traffic from AS100 via IXP B instead from IXP A

100.64.0.0/24

from AS200 to AS100 prefered via IXP A

IXP B

AS200

Routing Optimization ⚙️

Solution (control advertise on IXP using BGP communities):

2. AS100 will not receive prefix AS200 from IXP A.

3. AS100 receive prefix AS200 from IXP B.

1. AS200 advertise their prefix tagged with discard community (to specific AS100): 0:100

1.1.1.0/24

from AS100 to AS2 ~~X~~ prefered via IXP A

IXP A

100.64.0.0/24

from AS200 to AS100 prefered via IXP A

AS200

IXP B

AS100
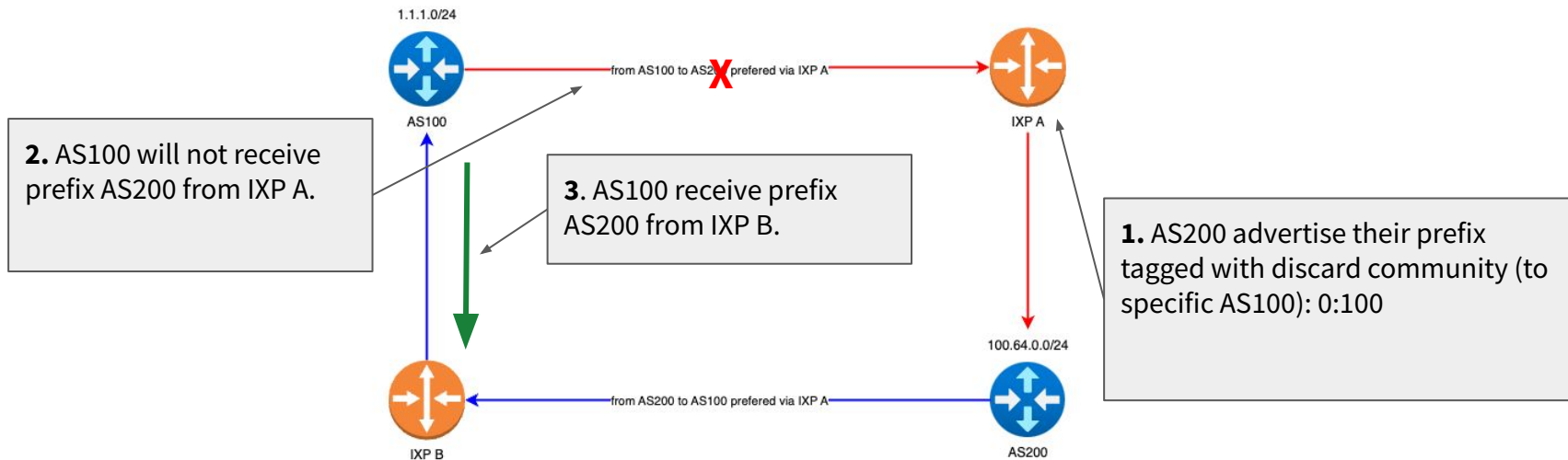
Routing Optimization ⚙️

Solution (control advertise on IXP using BGP communities):

With BGP Communities control IXP members have ability to control how their routes are announced to other multilateral participants, tool like looking glass also helpful for check and validation prefix advertisement to IXP.
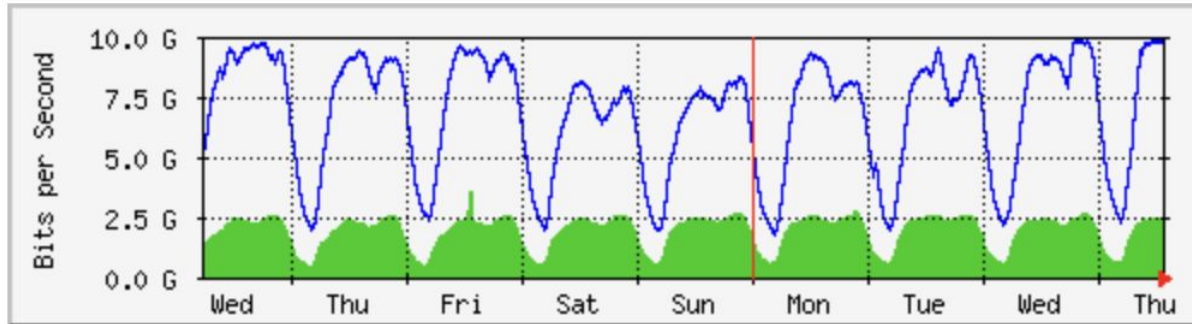
IXP	Members (based on PeeringDb)	BGP Community Control	Looking Glass
JKT-IX	578	Yes	YES
IIX	559	Yes	YES
OpenIXP	518	No	YES
CXC-IX (cloudXchange)	245	Yes	YES
BIX	157	Yes	YES

Peering Traffic Analysis

Peering Traffic Analysis 📈

Case:

Port traffic is full, but don't know where traffic goes 🙆 ?

Peering Traffic Analysis 📈

Solution (sflow analyzer):

IXP members can use sflow analyzer for further peering traffic analysis.

- **Optimize network performance.**
- **Enhance traffic management.**
- **Ensure efficient use of network resources.**

There are available tools from paid to open-source.

Peering Traffic Analysis 📈

Example sflow configuration:

Cisco example configuration

```
feature sflow
sflow sampling-rate 4096
sflow max-sampled-size 128
sflow counter-poll-interval 20
sflow max-datagram-size  1400
sflow counter-poll-interval 30
sflow collector-ip <collector-IP> vrf default
sflow collector-port 6343
sflow agent-ip <router-ip>
sflow data-source interface ethernet 1/5
```
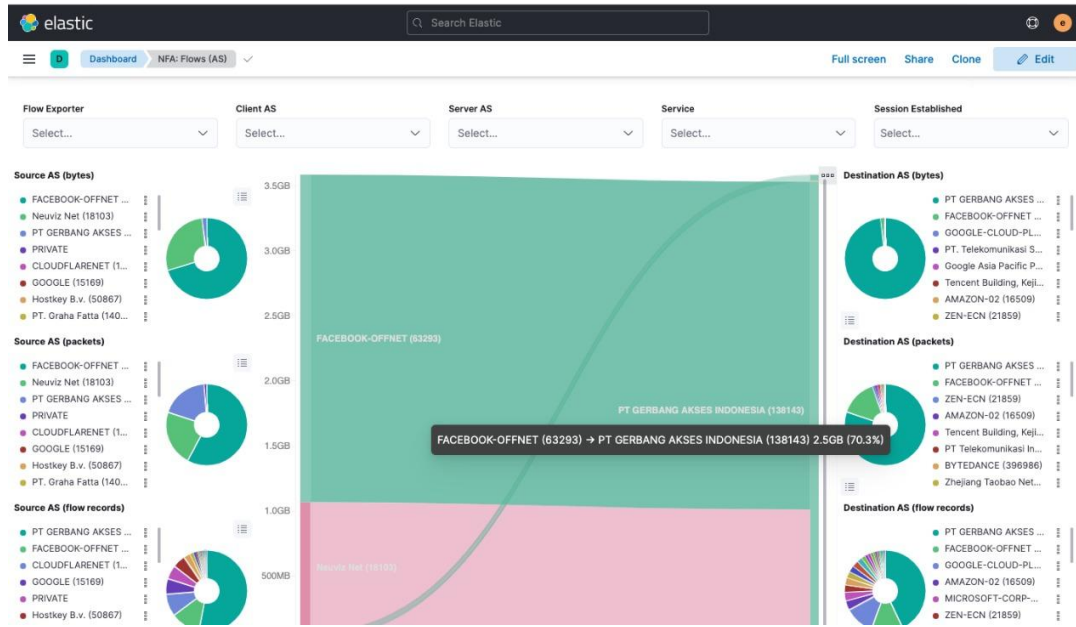
Junos example configuration

```
[edit protocols]
set sflow collector 10.204.32.46 udp-port 5600
set sflow interfaces ge-0/0/0
set sflow polling-interval 20
set sflow sample-rate egress 1000
```
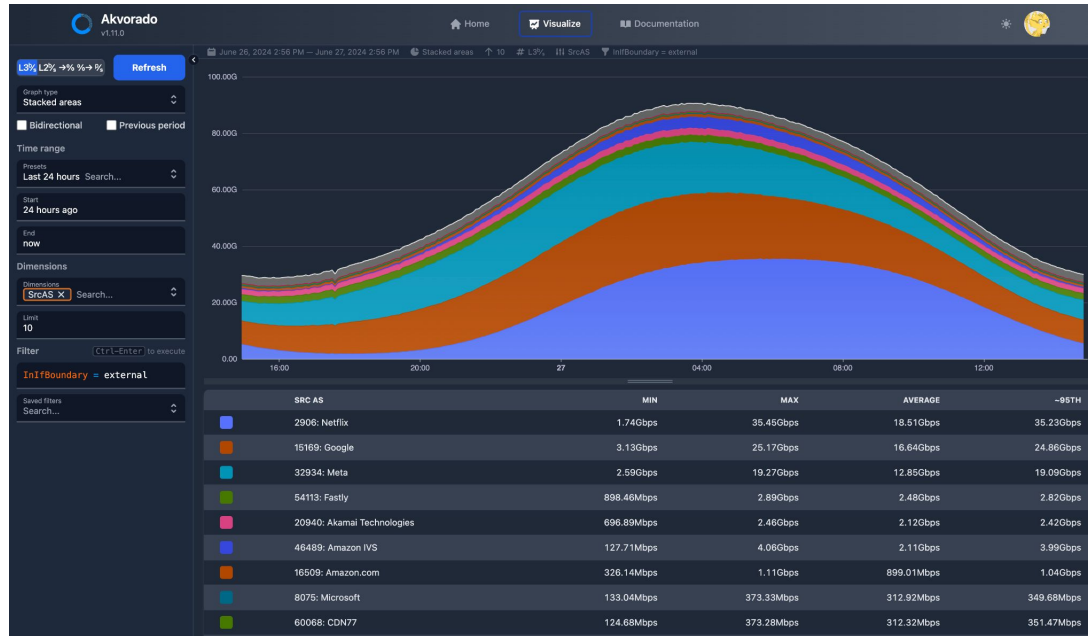
Peering Traffic Analysis 📈

Tools:

ElastiFlow (https://github.com/robcowart/elastiflow)

Peering Traffic Analysis 📈

CLOUD **X** CHANGE

Tools:

Akvorado (https://github.com/akvorado/akvorado)

NexGen IP Peer on IXP

NexGen IP Peer on IXP

Some IXP have limited ipv4 address space and this lead to unscalable peering LAN, we see this happen few times when IXP renumbered or expand peering LAN address.

- CXC-IX (CGK) expand the peering LAN from /24 to /23

- JKT-IX expand the peering LAN from /23 to /22

- IIX (Jakarta) renumbered to new peering LAN.

NexGen IP Peer on IXP

The problem of expanding or renumbered peering LAN

- There will be new ASN over the time, can we expand the existing peering LAN ? is this scalable ?
- Selective peering member need reconfigure their session unless they have automation
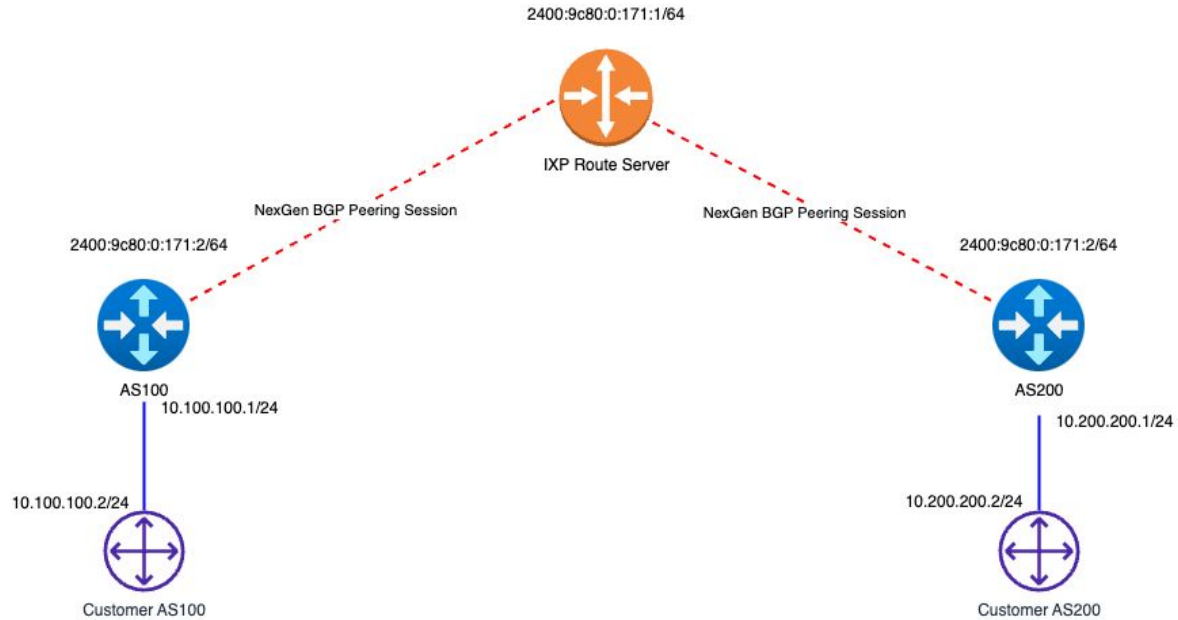
NexGen IP Peer on IXP

RFC 8950 for next cluster development

Advertising IPv4 Network Layer Reachability Information (NLRI) with an IPv6 Next Hop

https://datatracker.ietf.org/doc/html/rfc8950

NexGen IP Peer on IXP

Example Topology

NexGen IP Peer on IXP

Peers (Members) Configuration

```
router bgp 100
 .....
 neighbor 2400:9c80:0:171::1 remote-as 133339
 neighbor 2400:9c80:0:171::1 description rs1
 neighbor 2400:9c80:0:171::1 capability extended-nexthop
 .....
```

Enable extended-nexthop capability to use RFC 8590

(accept/send ipv4 prefix with ipv6 nexthop)

```
...
route-map out-cxc permit 10
 match ip address prefix-list my_prefix
 set ipv6 next-hop global 2400:9c80:0:171::2
 set ipv6 next-hop prefer-global
exit
...
```

Set route-map ipv4 prefix list with ipv6 nexthop (prefer-global)

NexGen IP Peer on IXP

Received Route

```
frr# show ip bgp
BGP table version is 14, local router ID is 103.225.171.2, vrf id 0
Default local pref 100, local AS 100
Status codes:  s suppressed, d damped, h history, * valid, > best, = multipath,
               i internal, r RIB-failure, S Stale, R Removed
Nexthop codes: @NNN nexthop's vrf id, < announce-nh-self
Origin codes:  i - IGP, e - EGP, ? - incomplete
RPKI validation codes: V valid, I invalid, N Not found

   Network          Next Hop            Metric LocPrf Weight Path
*> 10.100.100.0/24  0.0.0.0                  0         32768 i
*> 10.200.200.0/24  2400:9c80:0:171::3
                                             0             0 200 i

Displayed  2 routes and 2 total paths
frr# sh ip bgp summary

IPv4 Unicast Summary (VRF default):
BGP router identifier 103.225.171.2, local AS number 100 vrf-id 0
BGP table version 14
RIB entries 3, using 552 bytes of memory
Peers 1, using 716 KiB of memory

Neighbor          V        AS  MsgRcvd  MsgSent  TblVer  InQ OutQ  Up/Down State/PfxRcd  PfxSnt Desc
2400:9c80:0:171::1 4   133339       86       84       0    0    0 00:37:10           1       1 rs1

Total number of neighbors 1
frr#
```

AS100 receive ipv4 route with ipv6 nexthop address

NexGen IP Peer on IXP

Reachability Test

```
[[admin@MikroTik] > tool traceroute 10.200.200.2 src-address=10.100.100.2
 # ADDRESS               LOSS SENT    LAST      AVG     BEST   WORST
 1 10.100.100.1           0%    7    3.1ms     3.4      3.1     4.1
 2 10.200.200.1           0%    7     6ms      6.1      5.8     6.7
 3 10.200.200.2           0%    7     9ms      10       9      11.8
-- [Q quit|D dump|C-z pause]
```

Test from AS100 to AS200

BGP IP Peer on IXP only use IPv6 address 😱

```
[[admin@MikroTik] > tool traceroute 10.100.100.2 src-address=10.200.200.2
 # ADDRESS               LOSS SENT    LAST      AVG     BEST   WORST
 1 10.200.200.1           0%    6    2.8ms     3.2      2.8     3.6
 2 10.100.100.1           0%    6    5.8ms     5.8      5.6     6.2
 3 10.100.100.2           0%    6    8.7ms      9       8.6     9.8
-- [Q quit|D dump|C-z pause]
```

Test from AS200 to AS100

Extended Nexthop Capability

RFC 8950	Mikrotik (ROS 6.49.6)	Mikrotik v7	FRR 8.2.2	Bird 2.0.12
Extended Nexthop Capability	⛔	❓	✅	✅

Q&A and Short Discussion

If you have any questions or comments, please raise your hand ✋

CLOUD**X**CHANGE

Thank you!