Indian Network Operators Group

23 MARCH 2023, IIT DELHI

ISP AND IXP

An Ecosystem

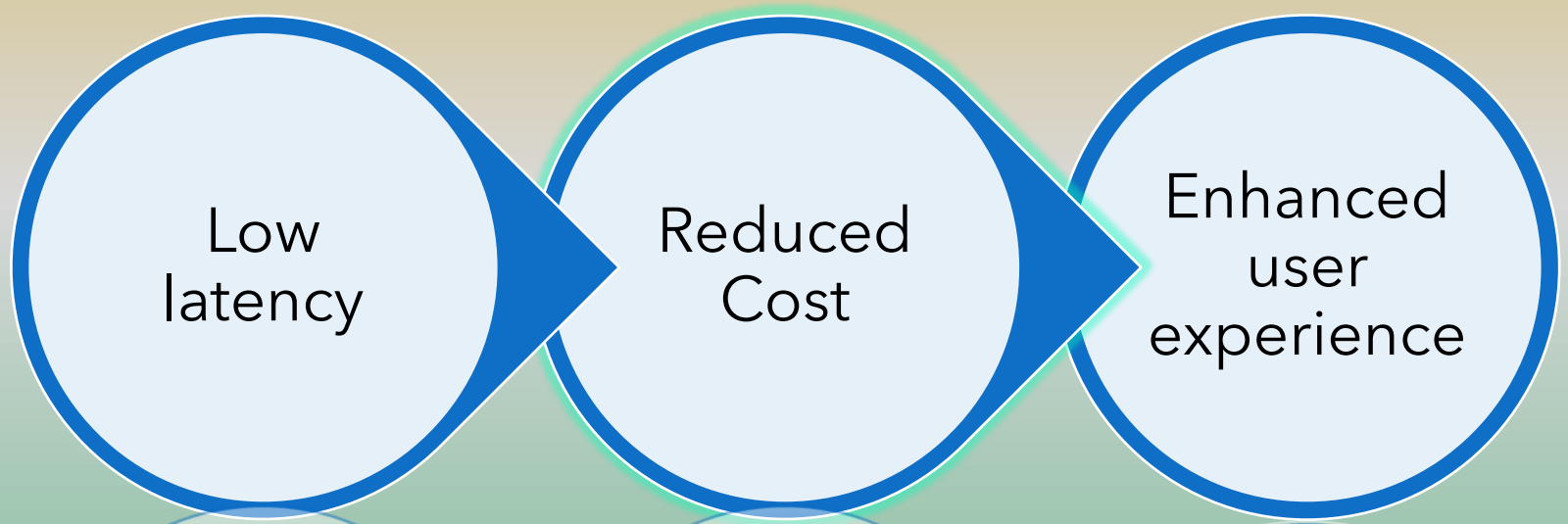
Manoranjan Tiwary

Agenda for Workshop

- How IXP works and how to get maximum benefit from an IXP
- Find a suitable IXP for your ASN
- A simple setup to configure BGP with IXP
- Best Practice to configure IXP port
- Examine BGP community (RFC 1997), large community (RFC 8092), and extended communities (RFC 4360) and how they are implemented in IXPs with the aid of looking glasses for IXPs.BGP filter for an IXP link
- We will use Mikrotik as a sample ISP IX router. You can use Juniper/Cisco/Huawei/Nokia any for your real environment
- Analyze some major content providers' cases and practices that may have an impact on your traffic.
- Receiving assistance from an IXP:- What you should expect and what you must accomplish on your own

manoranjan@mtiwary.in

Benefits of connecting to an IXP



manoranjan@mtiwary.in

What is an IXP

- An Internet Exchange Point (IXP) is a physical location where multiple networks link and exchange traffic.
- IXPs facilitate the direct exchange of traffic between networks, helping to reduce latency and increase network efficiency.
- IXPs are network infrastructure that supports the development of an ecosystem by permitting networks to connect with one another. It is a Layer 2 fabric (However L3 IXPs are also present but it is not common and obsolete)
- It is a switching fabric that makes it possible for traffic to be exchanged directly across various networks that are linked to the IXP, allowing them to do so without the need for intermediary providers or long-haul networks.

manoranjan@mtiwary.in

Components of an IXP infrastructure

- DWDM/CWDM
- Switches
- Servers for Router
- Servers to host Software to manage Infra –like IXP manager ETC

manoranjan@mtiwary.in

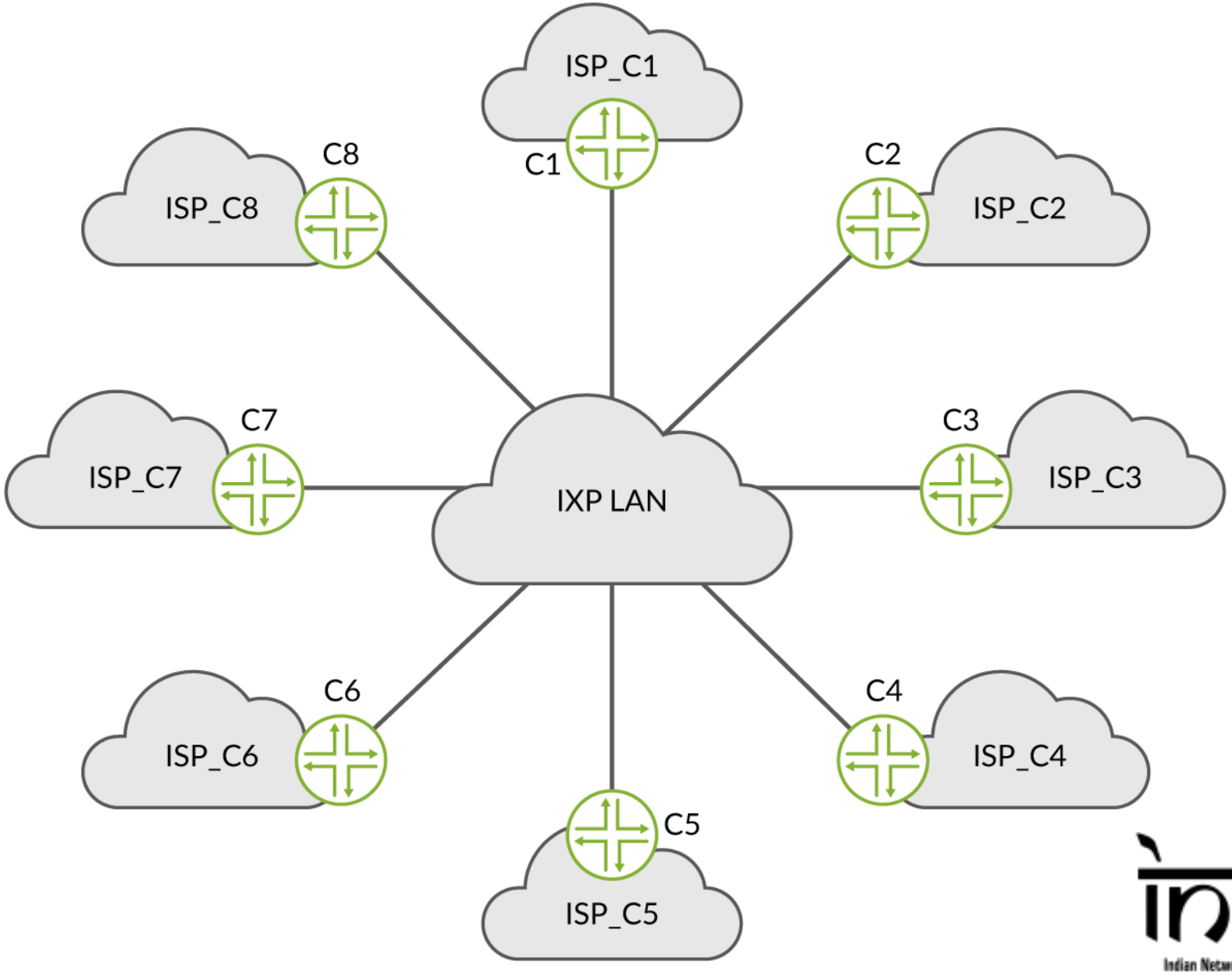
How crucial IXP is

- IXPs allow ISPs to exchange traffic with other networks in a more direct and efficient manner, lowering costs and boosting performance.
- The ISP and IXP ecosystem contributes to ensuring that all users have access to the internet, regardless of their location or the networks they use.
- An IXP reduces requirement of multiple circuits and multiple BGP sessions for an ISP
- An IXP helps to prevent tromboning by providing a direct interconnection point between ISPs and network operators, allowing them to exchange traffic directly and avoid unnecessary hops through third-party networks

manoranjan@mtiwary.in

Indian Network Operators Group

IXP



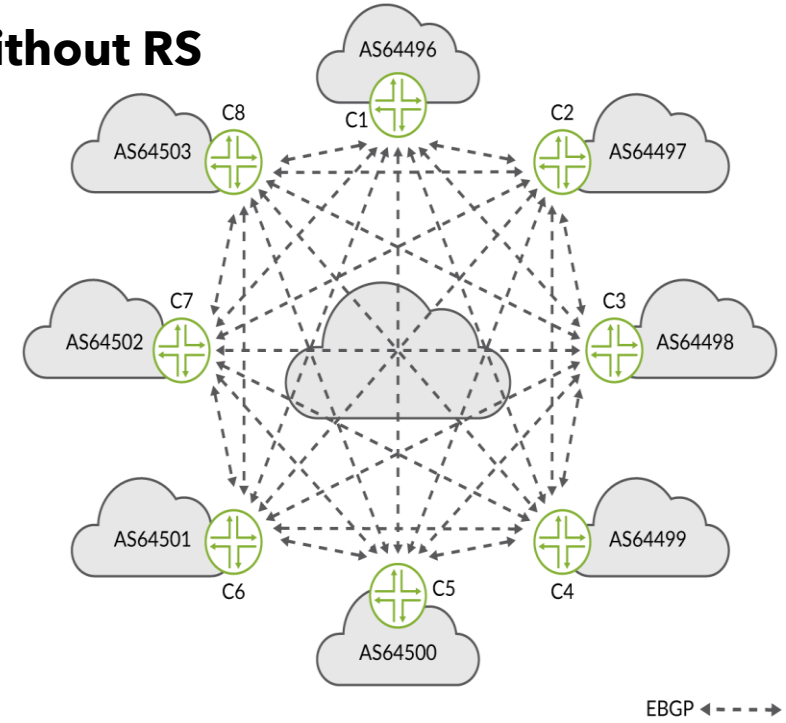
Indian Network Operators Group

BGP with Route Servers

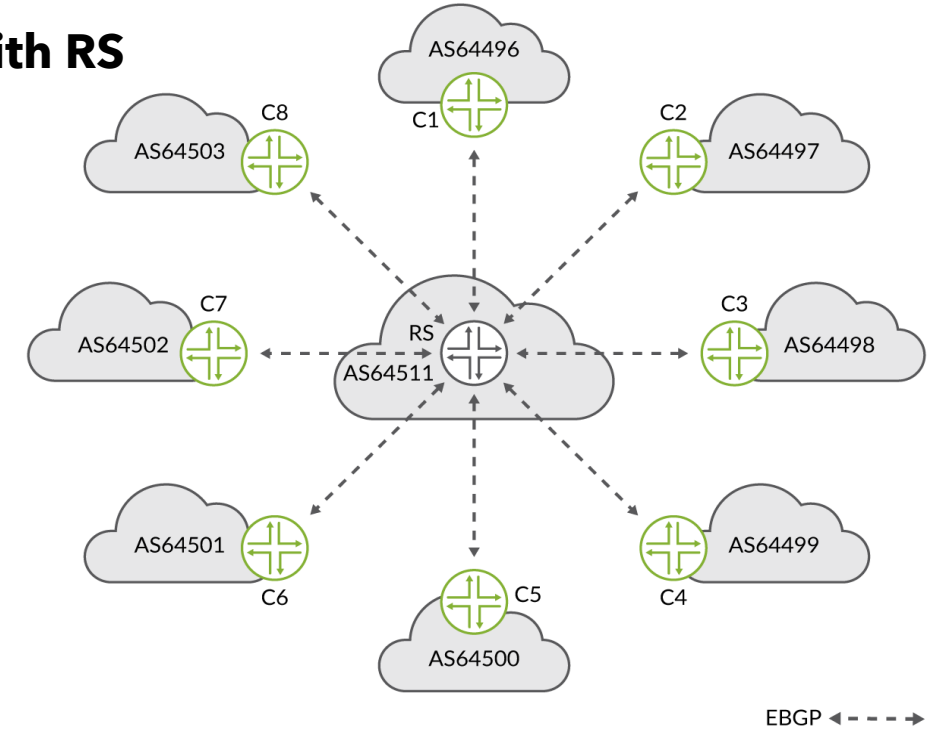
- As the route server does not insert its own ASN into the AS path of relayed prefix announcements, you must specify "no bgp enforce-first-as (IOS, IOS-XE) / bgp enforce-first-as disable (IOS-XR) and for Huawei "undo check-first-as"
- At IXPs, only announce your prefixes and those of your customers
- You should never use the default route towards an IXP peering router
- IXPs gives you flexibility to control your routes advertisement to peers
- BGP communities in IXPs is a common practice that provide greater flexibility and control for connected peers in managing their routing policies
- All major IXPs in India provide their connected peers with the feature of community-based routing policies. IXPs have documented their policies and processes for using communities, which can be found on their websites or by getting in touch with their support teams
- BGP Community-based policies are effective and simple to adopt

manoranjan@mtiwary.in

BGP without RS



BGP with RS



An IXP's Route Server simplifies the method of exchanging routing information between multiple networks

It reduces administrative overhead, and improves scalability and fault tolerance.

By using a Route Server, the IXP is able to scale more easily as the number of connected networks grows

Each network only needs to establish a single peering session with the Route Server.

This illustration is courtesy of Juniper Documentation.

https://www.juniper.net/documentation/en_US/day-one-books/topics/topic-map/internet-exchange-point-overview.html

manoranjan@mtiwary.in

Indian Network Operators Group

Sample BGP Configuration with Route Server

```
##Cisco

router bgp your-asn
 bgp router-id your-router-id
 bgp log-neighbor-changes
 bgp listen range 0.0.0.0/0 peer-group INOG-RS
 neighbor INOG-RS peer-group
 neighbor INOG-RS remote-as XXXXX
 neighbor INOG-RS description INOG-Test Route Server
 neighbor INOG-RS version 4
 neighbor INOG-RS transport connection-mode active
 neighbor INOG-RS route-reflector-client
 neighbor INOG-RS next-hop-self
 neighbor INOG-RS soft-reconfiguration inbound
 neighbor INOG-RS route-map TO-RS out
!
!
ip prefix-list TO-RS seq 10 permit 192.168.110.0/24
ip prefix-list TO-RS seq 20 permit 192.168.111.0/24
ip prefix-list TO-RS seq 30 permit 192.168.112.0/24
!
route-map TO-RS permit 10
 match ip address prefix-list TO-RS
!
```

manoranjan@mtiwary.in

- You should not advertise BOGON routes to IXP or any other ILL link
- To avoid route leaks, configure your filters properly and set the maximum number of prefixes you expect from any peer (IXPs are already setting maximum number of prefixes for each peer)
- Use appropriate export and import filters

Check http://www.iana.org/assignments/iana-ipv4-special-registry/iana-ipv4-special-registry.xhtml for IPs that should not be announced on IX/ILL

Address Block	Name	RFC
0.0.0.0/8	"This network"	[RFC791], Section 3.2
0.0.0.0/32	"This host on this network"	[RFC1122], Section 3.2.1.3
10.0.0.0/8	Private-Use	[RFC1918]
100.64.0.0/10	Shared Address Space	[RFC6598]
127.0.0.0/8	Loopback	[RFC1122], Section 3.2.1.3
169.254.0.0/16	Link Local	[RFC3927]
172.16.0.0/12	Private-Use	[RFC1918]
192.0.0.0/24 [2]	IETF Protocol Assignments	[RFC6890], Section 2.1
192.0.0.0/29	IPv4 Service Continuity Prefix	[RFC7335]
192.0.0.8/32	IPv4 dummy address	[RFC7600]
192.0.0.9/32	Port Control Protocol Anycast	[RFC7723]
192.0.0.10/32	Traversal Using Relays around NAT Anycast	[RFC8155]
192.0.0.170/32, 192.0.0.171/32	NAT64/DNS64 Discovery	[RFC8880][RFC7050], Section 2.2
192.0.2.0/24	Documentation (TEST-NET-1)	[RFC5737]
192.31.196.0/24	AS112-v4	[RFC7535]
192.52.193.0/24	AMT	[RFC7450]
192.88.99.0/24	Deprecated (6to4 Relay Anycast)	[RFC7526]
192.168.0.0/16	Private-Use	[RFC1918]
192.175.48.0/24	Direct Delegation AS112 Service	[RFC7534]
198.18.0.0/15	Benchmarking	[RFC2544]
198.51.100.0/24	Documentation (TEST-NET-2)	[RFC5737]
203.0.113.0/24	Documentation (TEST-NET-3)	[RFC5737]
240.0.0.0/4	Reserved	[RFC1112], Section 4
255.255.255.255/32	Limited Broadcast	[RFC8190] [RFC919], Section 7

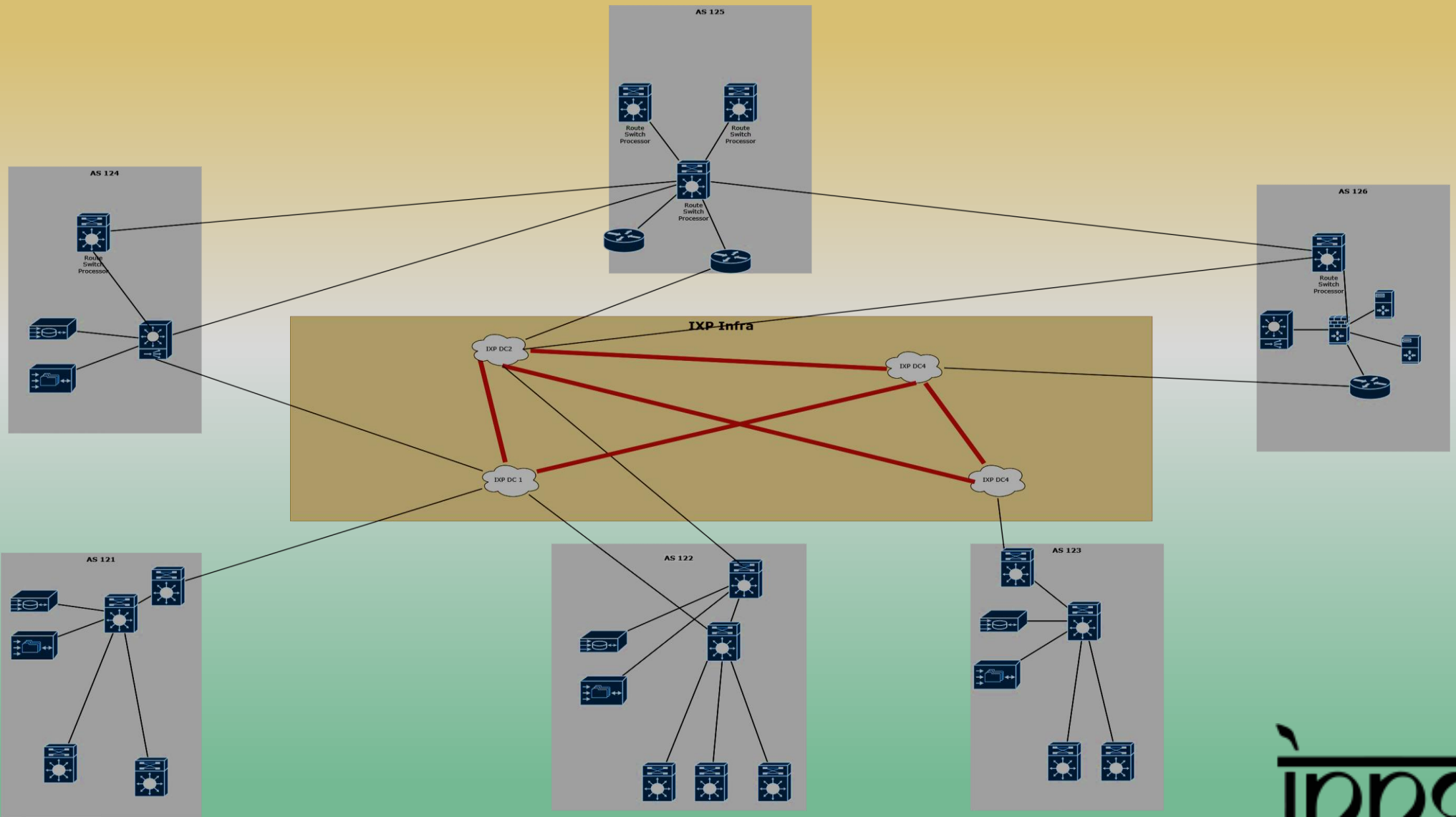
manoranjan@mtiwary.in

- Sample BGP to filter BOGONs

```
##Cisco

ip prefix-list bogons deny 0.0.0.0/8
ip prefix-list bogons deny 10.0.0.0/8
ip prefix-list bogons deny 100.64.0.0/10
ip prefix-list bogons deny 127.0.0.0/8
ip prefix-list bogons deny 169.254.0.0/16
ip prefix-list bogons deny 172.16.0.0/12
ip prefix-list bogons deny 192.0.0.0/24
ip prefix-list bogons deny 192.0.2.0/24
ip prefix-list bogons deny 192.88.99.0/24
ip prefix-list bogons deny 192.168.0.0/16
ip prefix-list bogons deny 198.18.0.0/15
ip prefix-list bogons deny 198.51.100.0/24
ip prefix-list bogons deny 203.0.113.0/24
ip prefix-list bogons deny 224.0.0.0/4
ip prefix-list bogons deny 240.0.0.0/4
ip prefix-list bogons permit 0.0.0.0/0 le 32
ip prefix-list bogons permit 0.0.0.0/0 ge 24 ip prefix-list bogons deny any
!
ip prefix-list ixp-prefixes deny 192.168.0.0/16
ip prefix-list ixp-prefixes deny 172.16.0.0/12
ip prefix-list ixp-prefixes permit any
!
route-map block-bogons permit 10
 match ip address prefix-list bogons
!
route-map block-ixp-prefixes permit 10
 match ip address prefix-list ixp-prefixes
!
router bgp <your ASN>
 neighbor <neighbor IP> remote-as <neighbor ASN>
 neighbor <neighbor IP> route-map block-bogons in
 neighbor <neighbor IP> route-map block-ixp-prefixes in
 neighbor <neighbor IP> route-map block-ixp-prefixes out
```

Decide where to connect



manoranjan@mtiwary.in

Indian Network Operators Group

Choose Interconnection Point
PeeringDB is here to assist.

- Find an IXP /DC where you can reach with least effort and best latency and uptime
- Check available peers and make a decision based on your traffic requirement
- Make BGP with RS and other members whom you think it is beneficial
- Check Routing table of your desired content providers via IXP Looking glass
- Check Latency from your NLD provider for the interconnection points from your end

Now you have data to decide what is good for you

manoranjan@mtiwary.in

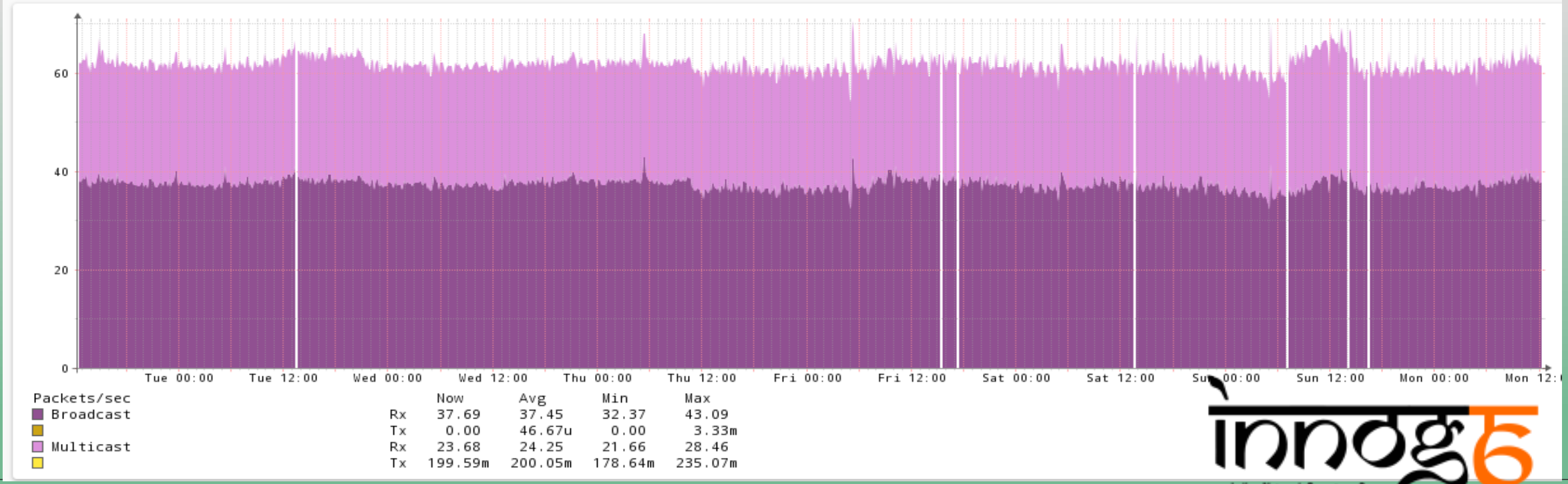
Get Maximum benefit from an IXP

- Check Proper port configuration for Broadcast storm since you are in a big broadcast domain and any misconfiguration at any peers end could disturb your traffic.
- Check how your routes advertised
- Check with Looking glass if any issue is observed by RS for your advertised prefixes
- Now IXPs are using RPKI and ROA based filters so it is good to verify it
- Tools/Portals you can use:- LG of your IXP , https://rpki-validator.ripe.net/ , https://rpki.cloudflare.com/
- Use Proper community for filtering your routes and learned routes

manoranjan@mtiwary.in

Best practices for IXP Ports / Why it's crucial

- You are on a large broadcast domain
- Your IXP connection will offer you multiple MAC addresses.
- Watch your link since a peer's misconfiguration could cause you to receive unwanted ARP, Broadcast, or Multicast traffic.
- A sample for IXP Broadcast traffic is below with /24 Subnet and 200 Peers



Indian Network Operators Group

What an IXP is supposed to do

- A transit between DCs with least possible latency
- Follow best practices for an IXP https://www.manrs.org/2020/12/ixp-peering-platform-an-environment-to-take-care-of/

manoranjan@mtiwary.in

- Configure Broadcast storm rate for your edge device where IXP is connected

Juniper

```
set ethernet-switching-options storm-control-profiles broadcast-storm-control interface all
```

Cisco

```
switch(config-if)# storm-control broadcast level 10
```

Mikrotik

```
/interface ethernet switch set ether1 broadcast-limit=1000
```

Huawei

```
[Switch-GigabitEthernet0/0/1] broadcast-suppression pps 1000
```

Commands may vary depending on the device model; consult docs.

Disable Link Layer Discovery Protocols (LLDPs) like NDP, CDP, LLDP, and MDP

- Do not enable proxy ARP on IXP interfaces since this may affect traffic for the entire IXP ecosystem.
- Disable Link Layer Discovery Protocols (LLDPs) like NDP, CDP, LLDP, and MDP on your IXP ports
- MAC address changes are infrequent in IXP environments, increasing ARP cache timeout will reduce ARP traffic.
- Disable STP on IX port
- To reduce the chance of route hijacking and to ensure proper routing, check your route objects, AS Set, ROA, and RPKI entry.

manoranjan@mtiwary.in

ROA and RPKI
Why is it critical for BGP?

- BGP works on trust and any mistake may cost a lot. It is good now IXPs are enabling ROA/RPKI filters

BORDER GATEWAY PROTOCOL INSECURITY —

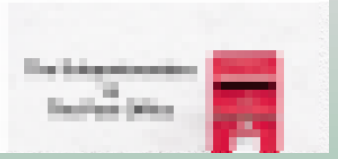
How 3 hours of inaction from Amazon cost cryptocurrency holders $235,000

For 2nd time in 4 years, Amazon loses control of its IP space in BGP hijacking.

DAN GOODIN - 9/23/2022, 11:34 PM

Google BGP route leak was accidental, not hijacking

Despite early speculation, experts concluded the BGP route leak that sent Google traffic through China and Russia was due to an accidental misconfiguration and not malicious activity.



By **Michael Heller**, Senior Reporter — Published: **16 Nov 2018**

Outage Analysis

ThousandEyes caught the entirety of the incident caused when JSC RTComm.RU AS 8342 improperly announced the 104.244.42.0/24 IP prefix owned by Twitter. Since JSC RTComm.RU AS 8342 is not the prefix owner, this is a prefix mis-origination, often referred to as a route hijacking—though it is important to note that the term "hijacking" here is a formal designation that does not imply malicious intent, but simply the origination of an unowned prefix without authorization. In Figure 2, we see JSC RTComm announcing Twitter's prefix to its BGP peer, MTS PJSC AS 8359, which in turn propagated the route to its peers.

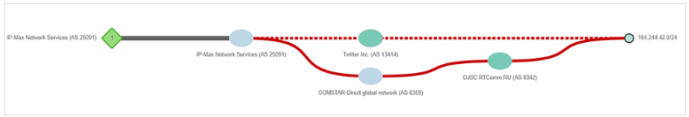


Figure 2. Start of BGP advertisement of 104.244.42.0/24 by RTComm (AS 8342)

Traffic that was destined for Twitter was rerouted for some users and instantly began to fail. Figures 3 and 4 show the sudden and immediate impact to network traffic and Twitter site access.

https://www.ripe.net/manage-ips-and-asns/resource-management/rpki/resource-certification-roa-management#---route-origin-authorisations--roas-

manoranjan@mtiwary.in

Indian Network Operators Group

How you can monitor your IXP link

- For Proper Monitoring of link use NMS and Sflow/Netflow/Jflow analyser
- Establish appropriate alerting tools to gather information about any unexpected behavior.

Some useful tools are PRTG , Observium, Librenms, Zabbix, Solarwinds solutions, Dynamite-nsm, Loggly ,Papertrail Etc.

- Check portals provided by your IXP for your port statistics your SLA graphs
- Keep an eye for any potential planned maintenance about any major content provider or IXP infrastructure.
- It is good practice to create NQA/SLA alerts for your IXP link to check any issue with link

manoranjan@mtiwary.in

BGP communities are made available by IXPs so you can manage prefix advertisement using RS.

- The BGP community is a tool used to tag or mark BGP routes with extra data, enabling more controllability over route selection and advertisement.
- A BGP community is a 32-bit value that can be assigned to a BGP route by a router or an autonomous system.
- Typically, BGP communities are used for traffic engineering, route filtering, and policy-based routing.

Extended BGP Community:

- Extended BGP communities are consists of a type field, a subtype field, and a value field. They are 64-bit values.
- The type and subtype fields define the meaning of the community value, whereas the value field includes the actual community value.
- For more sophisticated routing policies and traffic engineering, such as MPLS VPNs and multicast routing, extended BGP communi ties are employed.

Large BGP Community:

- Large BGP community is a further extension of the BGP community attribute that allows for even larger community values.
- Large BGP communities are 96-bit values that consist of three 32-bit fields.
- The first two fields are similar to the type and subtype fields in the extended BGP community, while the third field is a value field that can be used to carry additional information.
- Large BGP communities are designed to provide more flexibility in tagging BGP routes, particularly for service providers and large-scale networks.

manoranjan@mtiwary.in

Implementations

This page tracks **Large BGP Communities** implementations. This information is used to compile the Specification's RFC 7942 section. Updates should be emailed here.

BGP Speakers

Check http://largebgpcommunities.net/implementations/ for updated information

Vendor	Software	Status	Details
Arista	EOS	✓ Done!	EOS 4.21.3F
Cisco	NX-OS	✓ Done!	NX-OS 10.3(1)F
Cisco	IOS XE	✓ Done!	IOS XE 17.4
Cisco	IOS XR	✓ Done!	IOS XR 6.3.2
cz.nic	BIRD	✓ Done!	BIRD 1.6.3 (commit)
Extreme	NetIron	✓ Done!	NetIron 06.3.00
Extreme	SLX-OS	✓ Done!	SLX-OS 18r.2.00_v1
ExaBGP	ExaBGP	✓ Done!	PR482
FreeRangeRouting	frr	✓ Done!	Issue 46 (commit)
Juniper	Junos OS	✓ Done!	Junos OS 17.3R1
MikroTik	RouterOS	✓ Done!	ROSv7
Nokia	SR OS	✓ Done!	SR OS 16.0.R1
nop.hu	freeRouter	✓ Done!	
OpenBSD	OpenBGPD	✓ Done!	OpenBSD 6.1 (commit)
OSRG	GoBGP	✓ Done!	PR1094
rtbrick	Fullstack	✓ Done!	FullStack 17.1
Quagga	Quagga	✓ Done!	Quagga 1.2.0 (875)
Ubiquiti	EdgeOS	Planned	Feature Requested (maybe 2.0?)
VyOS	VyOS	✓ Done!	Vyos 1.2.0

manoranjan@mtiwary.in

Explore Looking Glasses of Indian IXPs

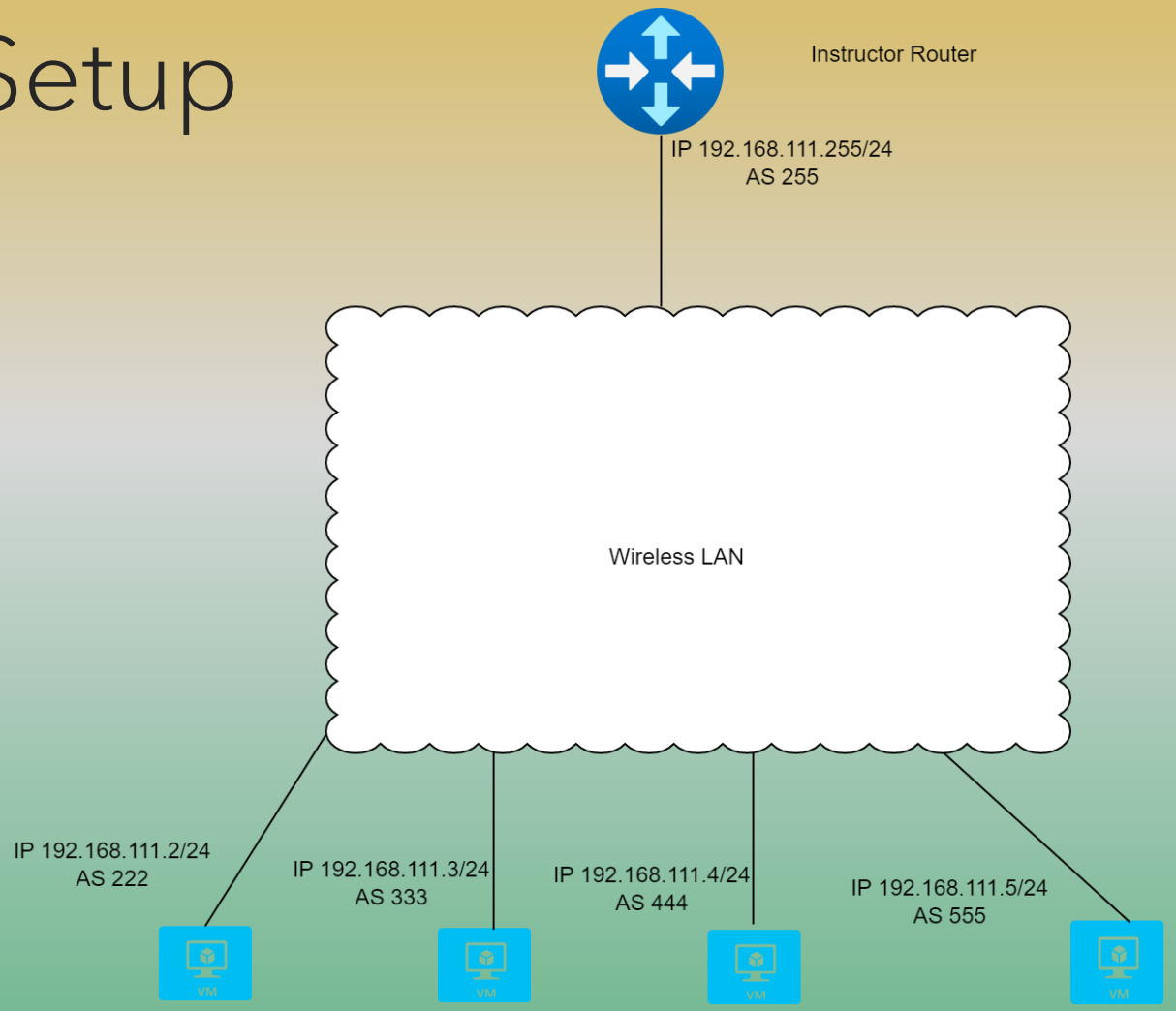
- https://lg.ams-ix.net/routeservers/mum-rs1-v4
- https://lg.extreme-ix.org/lg
- https://alice-lg.theixp.net/
- https://lg.nixi.in/

manoranjan@mtiwary.in

manoranjan@mtiwary.in

Indian Network Operators Group

LAB Setup



manoranjan@mtiwary.in

Indian Network Operators Group

- Configure VirtualBox
- Download Mikrotik and use instructions to get a live Mikrotik CHR router https://wiki.mikrotik.com/wiki/Manual:CHR_VirtualBox_installation
- For IP address and AS use your seat number for Ex:- for seat 5 it is 192.168.111.5/24 and AS 555

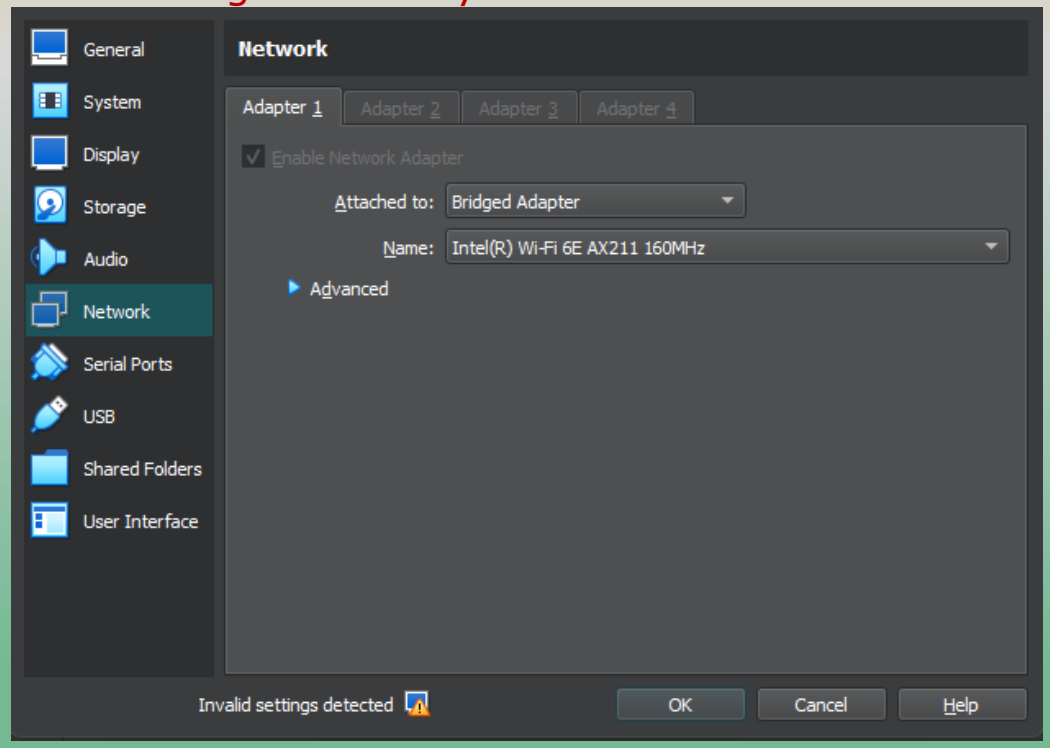
So use 192.168.111.X/24 where X is your seat number

AS for Seat 1 to 9 :- 111,222……999 i.e XXX

For Seat 10-99 1110,1111,1112……1199 i.e. 11XX Last 2 digits will be your seat number

Setup Virtual BOX adapter in Bridge mode so that you can connect to other members

After the server is online, you can connect via Winbox or, if you prefer, CLI.



Indian Network Operators Group

manoranjan@mtiwary.in

- Every Member will advertise Prefixes as:-

  Seat number 1:- 10.0.1.0/24
  Seat Number 2:- 10.0.2.0/24
  .

  .

  .

  Seat number X:- 10.0.X.0/24

- Configure BGP with Trainer Router it will work like a RS so you will receive all members prefixes from this router.
- Note since it is working as RS it will not advertise its own AS to any Peer. Check and verify this
- RS will advertise you prefixes with following BGP community added to it 254:<Peer AS>
  So you will receive Desk 2 Members Prefixes as 254:111 and desk 99 as 254:1199

manoranjan@mtiwary.in

- **Sample Configuration**

- Configuration
  # Configure BGP peer and local router information
- /routing bgp peer
- add remote-address=192.168.111.254 remote-as=254 name=peer1
- /routing bgp instance
- set default as=XXX router-id=192.168.111.X

- # Advertise network 10.0.1.0/24 via BGP
- /routing bgp network
- add network=10.0.1.0/24 synchronize=no

- # Filter prefixes received with BGP community 254:YYY
- /routing filter
- add action=discard chain=bgp-in prefix=0.0.0.0/0 bgp-communities=!254:YYY

manoranjan@mtiwary.in

LET'S MAKE SOME CONFIGURATION CHANGES BASED ON QUESTIONS

Try to Answer

What are some of the advantages of peering with an IXP rather than relying solely on transit providers for internet connectivity?

manoranjan@mtiwary.in

Indian Network Operators Group

Special Cases for some content providers

- **Google**:-
  If you are getting traffic from Google PNI and migrating to an IXP for any reason . You need to contact Google for shifting traffic from PNI to IX. It will not shift automatically.
- If you have multiple links for Google set BGP Community as specified by google to set priorities

https://support.google.com/interconnect/answer/9664829?hl=en

- **Akamai** :-
  Always keep in mind that for Akamai, latency is the primary factor taken into account. So, always look for the best IXP point from which to obtain the lowest latency

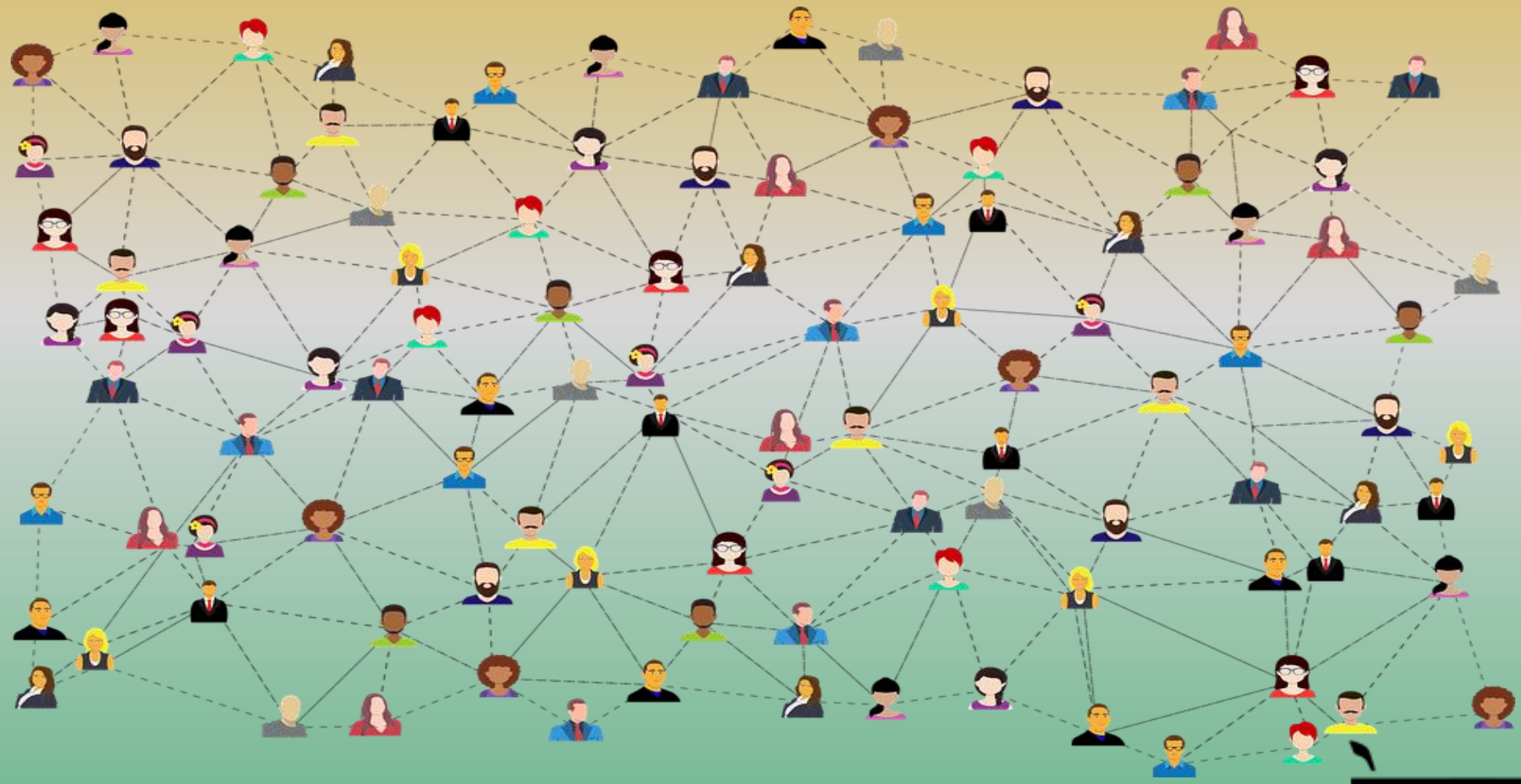
manoranjan@mtiwary.in

- Find out which major content providers are not using RS for BGP.
- Amazon and Microsoft, for example, are not currently exchanging routes via RS. If you want to receive traffic from Microsoft and Amazon, you must configure bilateral sessions.
- Microsoft has specific requirements to qualify for bilateral BGP, which can be found on their website

https://learn.microsoft.com/en-us/azure/internet-peering/policy

manoranjan@mtiwary.in

Time for Discussion



manoranjan@mtiwary.in

inog6
Indian Network Operators Group

Thank you for participating in this IXP workshop. My hope is that you leave with a greater understanding of the crucial role that IXPs play in the evolution of the Internet, as well as the benefits they provide to end users, network operators, and content providers.

manoranjan@mtiwary.in

Indian Network Operators Group

manoranjan@mtiwary.in

Indian Network Operators Group