

Stories of IXP Development

Che-Hoo Cheng
Infrastructure & Development Director
APNIC

2019-01-09@SANOG33-Thimphu

Disclaimer

- This talk is more about my personal experience and observations from operating the IXP in HK before joining APNIC in Apr 2017
 - Plus a bit of my additional experience and observations from helping the development of a few IXPs in the region
- **The points to be presented may NOT represent the viewpoints of APNIC**
- Try not to name names if possible
- Try to be more interesting, and educational
- There is no “One Size Fits All”
 - Just to provide hints, not answers
- Cannot cover all scenarios here because of limited time

What is an Internet eXchange Point (IXP)?

- An IXP is a shared physical network infrastructure over which various Autonomous Systems can do easy peering with one another
 - One physical connection to IXP can be used for interconnections with multiple networks
 - More cost-effective and scalable
 - *ASes to be served by IXP include Internet Gateways, Internet Service Providers (ISPs), Research & Education (R&E) Networks, Cloud Service Providers, Content Providers and Content Delivery Network (CDN) Service Providers*

Benefits of IXP

- One main objective of an IXP is to keep local traffic local
 - Important to local Internet development
- Helps bypass 3rd-party network infrastructure for easy interconnection and direct traffic exchange among participating networks
 - Reduced cost – cheaper connectivity
 - Enhanced network performance – faster speed
 - Reduced latency – lower delay
- Helps encourage development of more local content and local applications
 - Helps local data centre business and other businesses
- Everybody is benefited
 - The gain for each may be different but all will gain
 - At the end, it is the most important that end users or consumers are benefited
- Often considered as Critical Internet Infrastructure locally, regionally or globally

IXPs are Layer-2 Networks

- Switched Ethernet
 - One physical connection for interconnections with multiple networks
 - Only routers are allowed to connect to the switching fabric directly usually
- IXP participants can do direct Bilateral Peering (BLPA) over the layer 2 infrastructure anytime
- With Route Server added to the layer 2 infrastructure, IXP participants can also do Multilateral Peering (MLPA) for easier interconnections among everybody
 - Traffic exchange is not going through the route server but direct
- Those called themselves “IXes” but serving layer-3 services are mostly transit providers

Value and Attractiveness of an IXP

- Proportional to the number of different networks (ASNs) connected and also the traffic volume
- Snowball effect after reaching critical mass
 - The initial period usually is the hardest
 - Most will take wait-and-see approach
 - Gradually will have good mix of networks of different types
 - E.g. Eyeballs vs Content

Success Factors of the IXP in HK

- Neutrality
 - Helped gain trust from the participants especially the early ones
 - But there is no 100% neutrality...
 - Competition from another university
 - After gaining critical mass, things are much easier
 - No need to do sales work at all
- Free Service Initially
 - In the first 10 years or so
 - Little hesitation for participants to connect
 - But cannot be free forever
- Started Early
 - Earlier than the incumbent telco starting its ISP business
 - They even asked for joining before they launched the service
 - History cannot be repeated that easily...

And also...

- Leveraging telecom deregulation in HK
- Leveraging existing networks
- Passion & persistence
 - And, there was incentive for doing it
- Adaptation to industry changes
 - E.g. opening up to unlicensed networks
- HK people have been enjoying fast local Internet connectivity since almost the beginning

Long-Term Misunderstandings

- Used to mention ">98% of traffic" a lot
- Government people and general public always think >98% of external traffic is going through the IXP in HK
 - How can that be possibly true?!
 - It is just wishful thinking of those people
- But the more accurate wordings should be:
 - The IXP in HK helps keep >98% of local traffic local

Other Misunderstandings

- The IXP supports Bilateral Peering since the beginning
 - Although it did emphasise Multilateral Peering in the early days
- The IXP is NOT the only IXP in HK
 - There are in fact multiple IXPs
 - The IXP is just the earliest and the biggest
 - The other IXPs together are not really small
 - Perhaps 70:30 in terms of traffic volume
 - But the IXP is the focus of people, most of the time

Multilateral Peering is evil?

- Mandatory MLPA established initially was meant to be for HK routes only
- Mandatory MLPA for HK routes did help attract some overseas ISPs to connect and then gradually made the IXP become Regional IXP
 - Personally think this was probably the most successful MLPA case
- Mandatory MLPA for HK routes was gradually “unmentioned” because of large content / CDN providers
 - Not big transit providers
 - Definitely not related to any other IXPs set up in HK
- Mandatory MLPA is not the norm all around the world now...
 - Large providers will find ways to get around it
- Personally do not like stripping away the ASN of the route server from AS Path as it helps identify the routes learned from MLPA more easily

Snowden Nightmare...

- Started from an article of his interview done in HK published at SCMP on 13 Jun 2013
 - Mentioned the name of the university while not mentioned the IXP at all...
 - But people still thought he was referring to the IXP
- A lot of reporters surrounded the main data centre hoping to find anomalies
- Grilled by media and politicians for months
- Enhanced physical security measures afterwards
 - Stopped all unessential data centre visits
- No findings of anything set up or done by the intelligence agency inside

Vulnerabilities of IXPs

- Proxy ARP
 - Why can't all router vendors have Proxy ARP disabled as default?
 - Cannot stop it totally because of possible human errors
 - Can only do regular monitoring by checking the ARP table
 - EVPN over VxLAN technology should be able to help but it is not a simple technology
- Unknown Unicast Flooding
 - May happen when there is asymmetric routing seen from an IXP
 - Can be mitigated by sending proactive ARP check to all active addresses every hour or so
 - EVPN over VxLAN technology should be able to help but it is not a simple technology
- Shared Buffer over Multiple Switch Ports
 - Can cause trouble to multiple connections when there is big congestion on one port
 - Unknown to innocent participants which do not have any congestion
 - Just be careful when choosing switch models
 - Also avoid switch models with small buffer

Vulnerabilities of Data Centres?

- Locations are known
 - Same for Landing Stations
 - Can easily be targets of physical attacks
- How can you better protect the fibre lead-ins and manholes which are outside of data centres?
- No such things as absolute security...
 - But let's still do our best

Visibility of Traffic?

- Support of layer-3 sFlow/NetFlow highly desirable for better visibility of traffic going through the IXP
 - It helps trouble-shooting and understanding of traffic profile/pattern a lot
 - Having visibility of just layer-2 data is of less use
- But participants and general public would be concerned about the perceived surveillance or monitoring
 - Should do the best not to give data away to 3rd-parties

Port Security Is Important

- The IXP in HK allows just one MAC address per port (physical or virtual)
 - Strictly one IPv4 address, one IPv6 address and one MAC address per port (physical or virtual)
 - Static MAC address for full control
 - “Violation Restrict” instead of “Violation Shutdown”
- Minimum protection to the layer-2 broadcast domain
- A few IXPs allow more MAC address per port but still a small number
- Should also do Ether-type filtering and broadcast/multicast traffic filtering

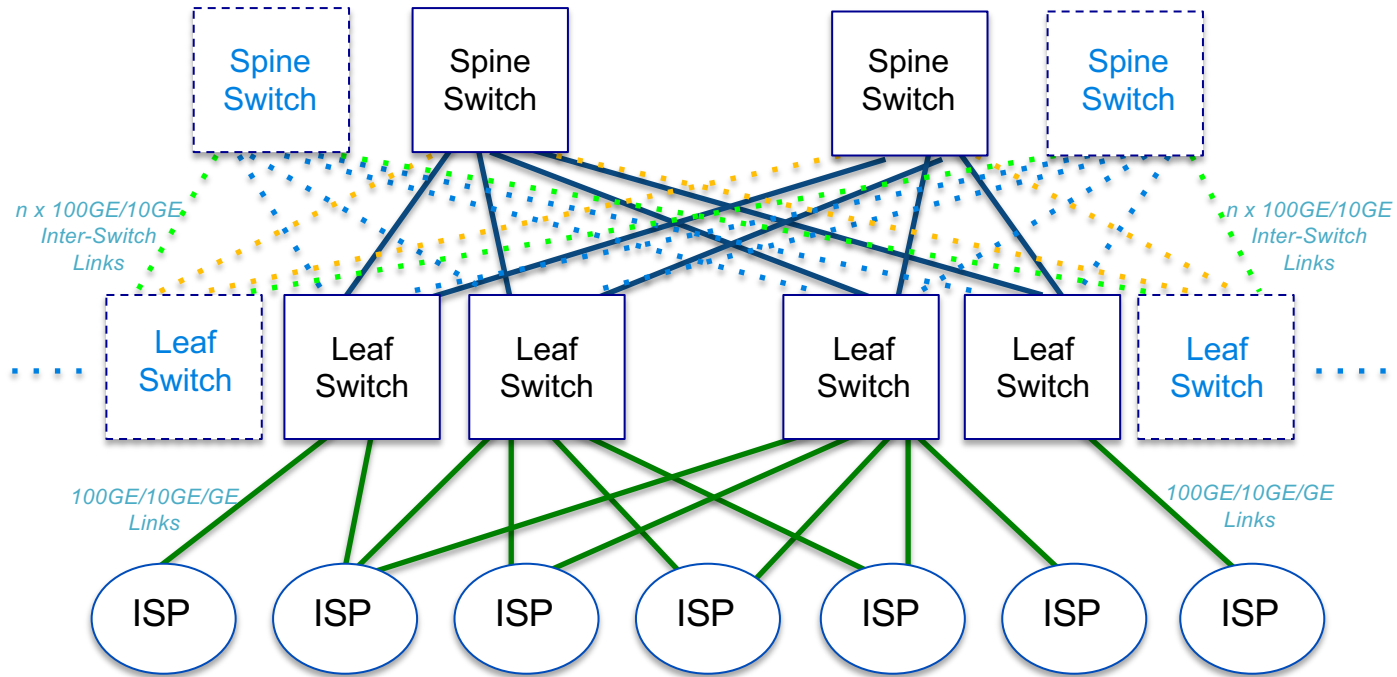
Remote Layer-2 Connections to IXP?

- More and more common nowadays
 - Some even from >1,000km away
- Using fibre-only connection is much easier, with much fewer issues
 - ZX/ZR/ZR4 are up to 70-80km
- Clear-Channel remote layer-2 circuits with full transparency are rare
 - Unless you are willing to pay more
 - Wasting a lot of effort to do trouble-shooting with carriers
- But IXPs cannot afford to not support them
 - As they want to have more business, sometimes through resellers
 - Unless their main business is data centre business

Scalability Issue

- IXPs were not supposed to have any packet loss in its infrastructure
 - And with very low latency too
- Become an issue when IXP grow beyond one switch
 - Due to not enough ports or expanding to multiple sites
- Inter-switch links are the risk
 - Over-subscription or not?
- Spine-and-leaf architecture helps a bit but not for all cases
 - Need to determine how much bandwidth from leaf to spine anyway
 - Still not ideal if there are adjacent leaf switches at one site
 - All traffic among 2 adjacent leaf switches has to go to the spine first?

Spine-and-Leaf Architecture



MRTG of Aggregate Traffic

- It is less sensitive of course
 - More an indication of the importance and the growth of an IXP
 - Should not neglect the huge difference between showing 5-min data and 1-min data
 - Should not neglect what traffic data is included – just the main broadcast domain or what?
- Usually incoming = outgoing
- If incoming > outgoing
 - Congestion at at least one port
 - May be DDoS attacks
- If outgoing > incoming
 - May have Unknown Unicast Flooding
- If sudden drop of large traffic volume
 - May have Proxy ARP problem
 - Usually happens when change of router / router software / router config

Other Observations from MRTG

- Situation in HK
 - Holiday Effect
 - Soccer Games Effect
 - Typhoon Effect
- Difference of Culture / Practices
 - E.g. HK vs Japan

IXPs and Data Centres

- They are natural partners
- Common situation in advanced metro cities
 - Multiple IXPs in one Data Centre
 - One IXP in multiple Data Centres
 - Should be the same layer-2 broadcast domain
 - Circuit cost is a burden
 - Healthy competition would be good
 - Customers have choices
 - Also for better resilience

IXP Models

- Developed economies vs developing economies
- Non-profit vs commercial
- Subsidized vs self-financed
- Government-led vs industry-led
- No one single model which can suit all situations
- Relative Neutrality is important

Commercial vs Non-Profit

- Commercial set-up is free to do anything
 - No need to care about neutrality too much
 - IXP is mostly a service to help other business
- Non-profit set-up tends to be more cautious
 - Neutrality is more important, at least to the target participants
 - Tend to be more independent
 - Tend to offer fewer services

IXP across Multiple Cities / Economies

- May not be good for maintaining neutrality
 - Considered as competing with participants which have presence in the same locations
- Commercial IXPs can take this business risk especially if this may help their other business
- But not so good for non-profit IXPs targeting all kinds of networks or providers
 - Those that see competition may not join and then it may affect the goal of “keeping local traffic local”

Interconnections of 2 or More IXPs

- What are the purposes of doing this?
- Not considered a good idea at Layer-2, especially if across cities or countries/economies
- Even at Layer-3, still need to be mindful of whether it affects the original purposes of each IXP involved

Advanced / Developed Economies

- IXPs are mostly business-oriented
 - Even for not-for-profit set-up
 - Less government involvement
- Multiple IXPs
 - Keen competition
- But if they cannot keep intra-economy traffic local, someone needs to step up
 - Government? Industry group? Customer pressure?

Developing Economies

- Some do not have any IXPs yet
- Local traffic does not stay local
 - A lose-lose situation for everybody
- IXPs can help Internet development a lot
 - Better to be non-for-profit set-up
 - May need to start with subsidized model
 - May not be a business at all
 - Help from government is mostly needed
 - Active participation of the biggest players is also very important

Examples of Pacific Islands

- Far from any other places
- External connectivity is very expensive
 - More submarine cables are being built for them
- Small markets because of small population
- Usually just a few ISPs but they may not be interconnected locally
- Local traffic across ISPs usually routed through US or Australia
- Local IXP is very much needed
- Observed immediate benefits on Day 1 of set-up of one Pacific Island IXP
 - Much improved latency and high volume of traffic
- Small land-locked economies have more or less similar issues

Politics Involved in Early IXP Development

- Usually larger ISPs like IXP less than smaller ISPs because smaller ISPs are mostly target customers of larger ISPs
- Larger ISPs refuse to connect to IXP making the value of IXP lower
- There are multiple possible mitigation options for that but in any case, larger ISPs need to collaborate
 - Separating access networks from Internet gateway or transit network
- If hurting the goal of “Keeping Local Traffic Local”, then it is lose-lose to everybody
- Government involvement may help or may hurt the case
 - It depends on the relationship between the industry and the government
 - Forcing large ISPs to do peering may not achieve the expected outcomes
- But having an IXP is NOT a magic wand to solve all the issues

Government Funding for IXPs?

- Is it good or bad?
- More needed during infancy stage of IXP development
- But for long-term, it is probably better to have bottom-up industry-led governance for IXP
 - Align with bottom-up multi-stakeholder approach
 - Need to have a long-term sustainable financial model

Possible Steps for IXP Development

- Can be gradual, step by step
- Layer-2 network is the bare minimal
 - Can use private IP addresses if small amount of participants
- Public IP addresses next
 - Legal entity issue
- Site resilience is **IMPORTANT** while equipment resilience is already included
 - Has to have site resilience sooner or later
- Route server(s) with ASN follows
 - RPKI consideration
- Other value added services
 - DNS: Root / TLDs / Recursive
 - Shared Content Caches?

Shared Content Caches Offered by IXP?

- A lot of misunderstandings about the use of caches
- A lot of local IXPs want to provide shared caches for their participants to increase their value
 - Cost recovery and cost sharing / accounting are major issues to them though
- Content / CDN providers are still sceptical about this model
 - They still mostly look at cache efficiency and traffic volume for justifications

IXP Participants

- Unfortunately, a lot of IXP participants do not make the best use of the IXP(s) they have connected
- IXP Participants without enough knowledge and skills may disrupt the operations of IXP from time to time
- IXP operators need to do a lot of education or push to their participants
- So, IXP engineers would be busy and dedicated resources would be needed
 - Volunteering type of operations mode cannot sustain for too long

Myth of Neutrality

- There is NO absolute neutrality
- Different organisation has different perspective of neutrality
 - A university?
 - A carrier-neutral data centre?
 - An IXP?
 - A government department?
 - A membership-based organisation?
- We can only be “very neutral” for a defined group of companies or organisations, but not for all...
- But maintaining higher relative neutrality is still better for IXPs

Which Models Can Sustain?

- Usual business model
 - IXP alone cannot make big money
 - Or IXP may just be a value added service
- Subsidized Model
 - Government funding may be more reliable?
- Model relying on sponsorship and/or volunteers
 - Most risky as sponsorship or support of volunteers is not guaranteed
- Membership-based Model
 - Open Membership vs Closed Membership
 - Proper governance is important
 - Most neutral but still need to have good financial model for long-term sustainability

Closing Remarks

- IXPs will continue to play a key role for easy interconnections among networks
 - Especially for developing economies
 - But IXP is NOT a magic wand to solve all the issues
- Need to find a suitable model for long-term sustainability
- Relative neutrality is still important
 - So still better to maintain it as much as possible
- After all, “Keeping Local Traffic Local” is the most important thing

