HKIX Platform Upgrade & Bilateral Peering

Che-Hoo CHENG 鄭志豪
Hong Kong Internet Exchange (HKIX)
The Chinese University of Hong Kong (CUHK)
02 MAR 2010

Introduction of HKIX *(1/2)*

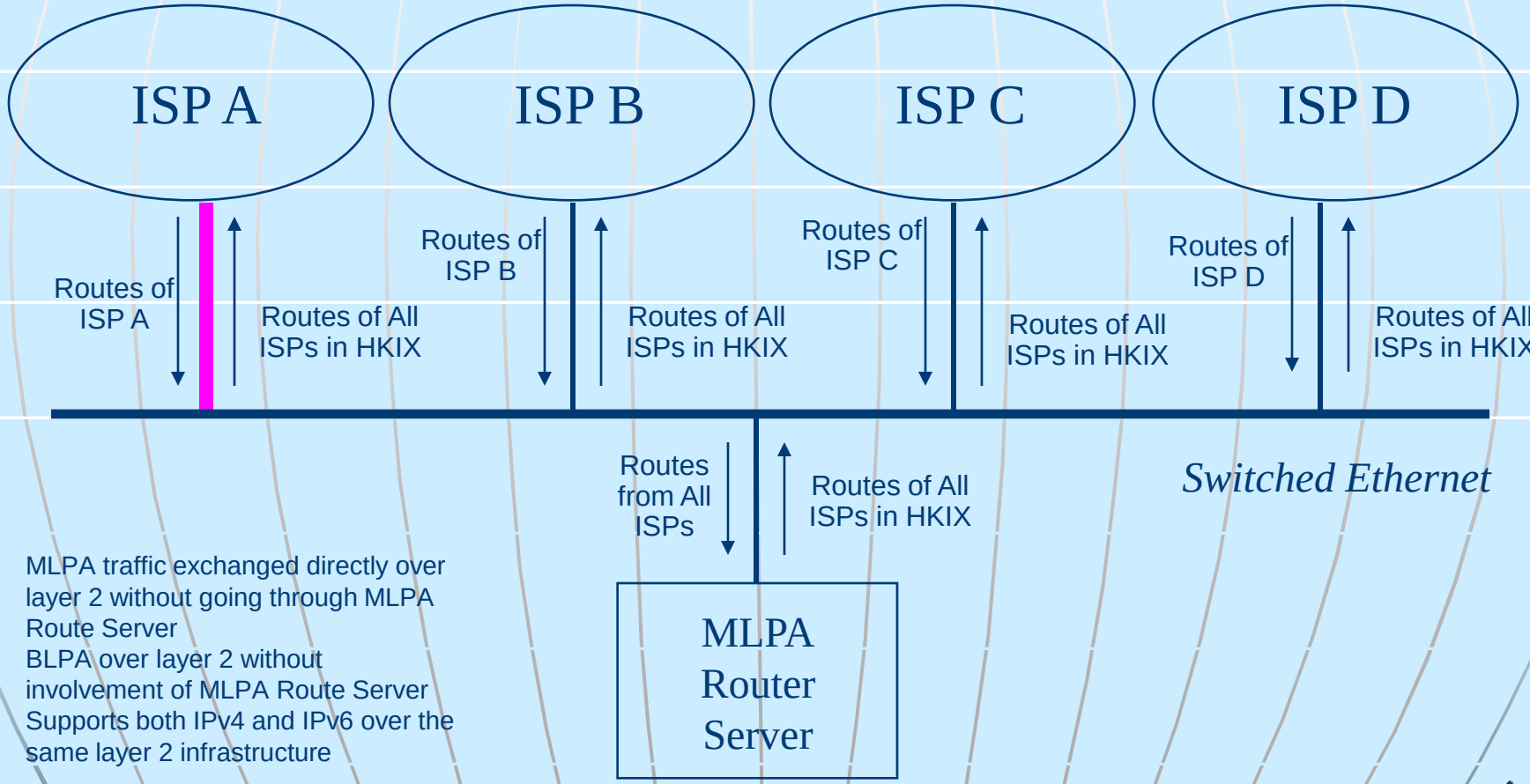
- HKIX is a Settlement-Free Layer-2 Internet Exchange Point (IXP), with mandatory Multi-Lateral Peering Agreement (MLPA) for Hong Kong routes
  - ISPs can interconnect with one another and exchange inter-ISP traffic at HKIX
  - HKIX is not a Transit Provider
- HKIX supports and encourages Bi-Lateral Peering Agreement (BLPA)
- HKIX was a project initiated and funded by ITSC of CUHK in Apr 1995 as a community service
  - Still owned, supported and operated by ITSC of CUHK

Introduction of HKIX *(2/2)*

- Two Main Sites for resilience:
  - HKIX1: CUHK Campus in Shatin
  - HKIX2: CITIC Tower in Central
- Our service is basically free of charge as we are **not-for-profit**
  - But there will be charge for 10GE port or many GE ports if traffic volume is not high enough to justify the resources
- Provide colo space for strategic partners such as root / TLD DNS servers & RIRs
- Considered as Critical Internet Infrastructure in Hong Kong
- We are confident to say that because of HKIX, more than 99% of intra-HK Internet traffic is kept within HK

- More information on www.hkix.net

HKIX Model — MLPA over Layer 2 (with BLPA support)



- MLPA traffic exchanged directly over layer 2 without going through MLPA Route Server
- BLPA over layer 2 without involvement of MLPA Route Server
- Supports both IPv4 and IPv6 over the same layer 2 infrastructure

Quick Updates *(1/3)*

- 1 x Cisco Nexus 7018 + 2 x Cisco Catalyst 6513 at HKIX1 and 1 x Cisco Catalyst 6513 at HKIX2
- Most connected to HKIX switches without co-located routers
  - Cross-border layer-2 Ethernet connections to HKIX possible
    - Ethernet over MPLS or Ethernet over SDH
- Officially allow overseas ISPs to connect
  - Local ISPs must have proper licenses
  - Those overseas ISPs may not have Hong Kong routes…
  - Major overseas R&E networks connected since 2008

Quick Updates (2/3)

- ~130 AS'es connected with IPv4 now
  - ~15 AS'es at both HKIX1 & HKIX2 for resilience
- ~24 10GE connections and >200 GE/FE connections
- >28,000 IPv4 routes carried by HKIX MLPA
  - More non-HK routes than HK routes
  - Serving intra-Asia traffic indeed
- Peak 5-min traffic >100Gbps now
- HKIX1 supports and encourages Link Aggregation (LACP)

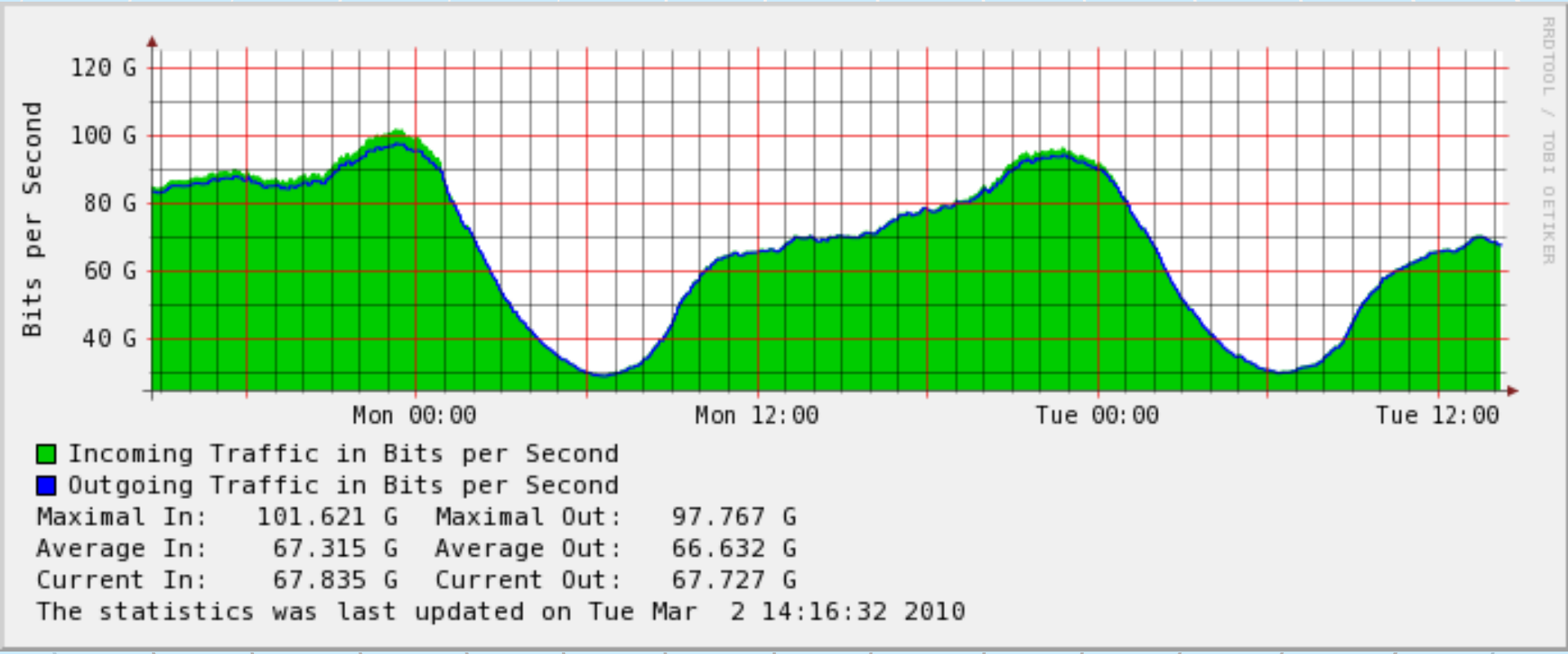
- A small POP in Mega-i with layer-2 GE links back to HKIX1 but it is for R&E network connections only

Quick Updates (3/3)

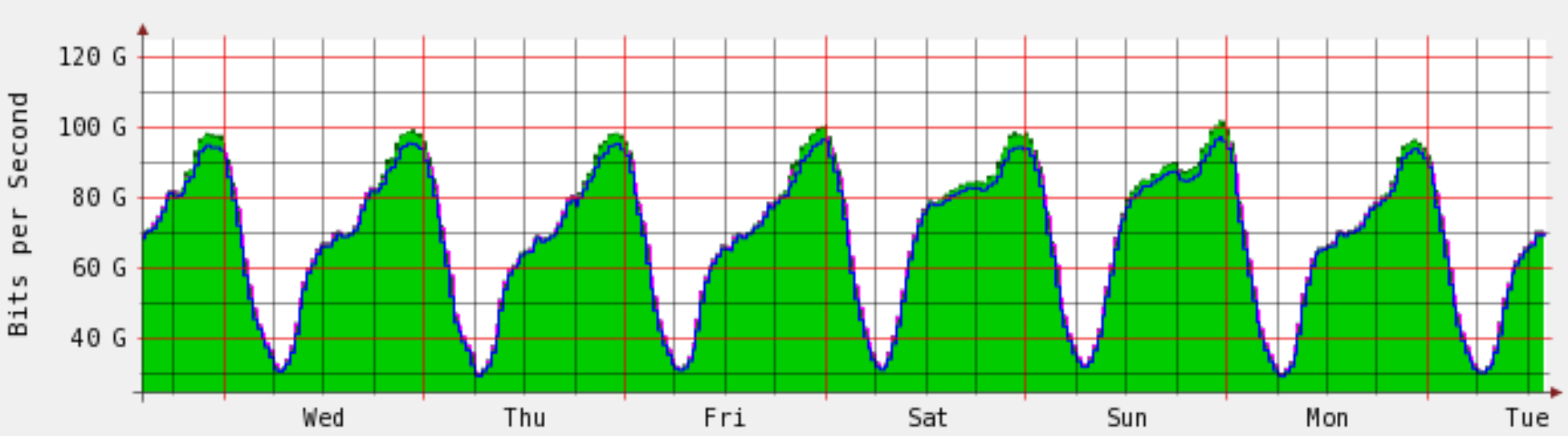
- Basic Set-up:
  - First 2 GE ports with no colo at HKIX1 and First 2 GE ports at HKIX2: Free of charge and no formal agreement
- Advanced Set-up:
  - 10GE port / >2 GE ports at either site / Colo at HKIX1: Formal agreement is needed and there will be colo charge and a small port charge unless aggregate traffic volume of all ports exceeds 50% (95th percentile)

- See http://www.hkix.net/hkix/connectguide.htm for details

Some Statistics *(1/3)*



Some Statistics (2/3)

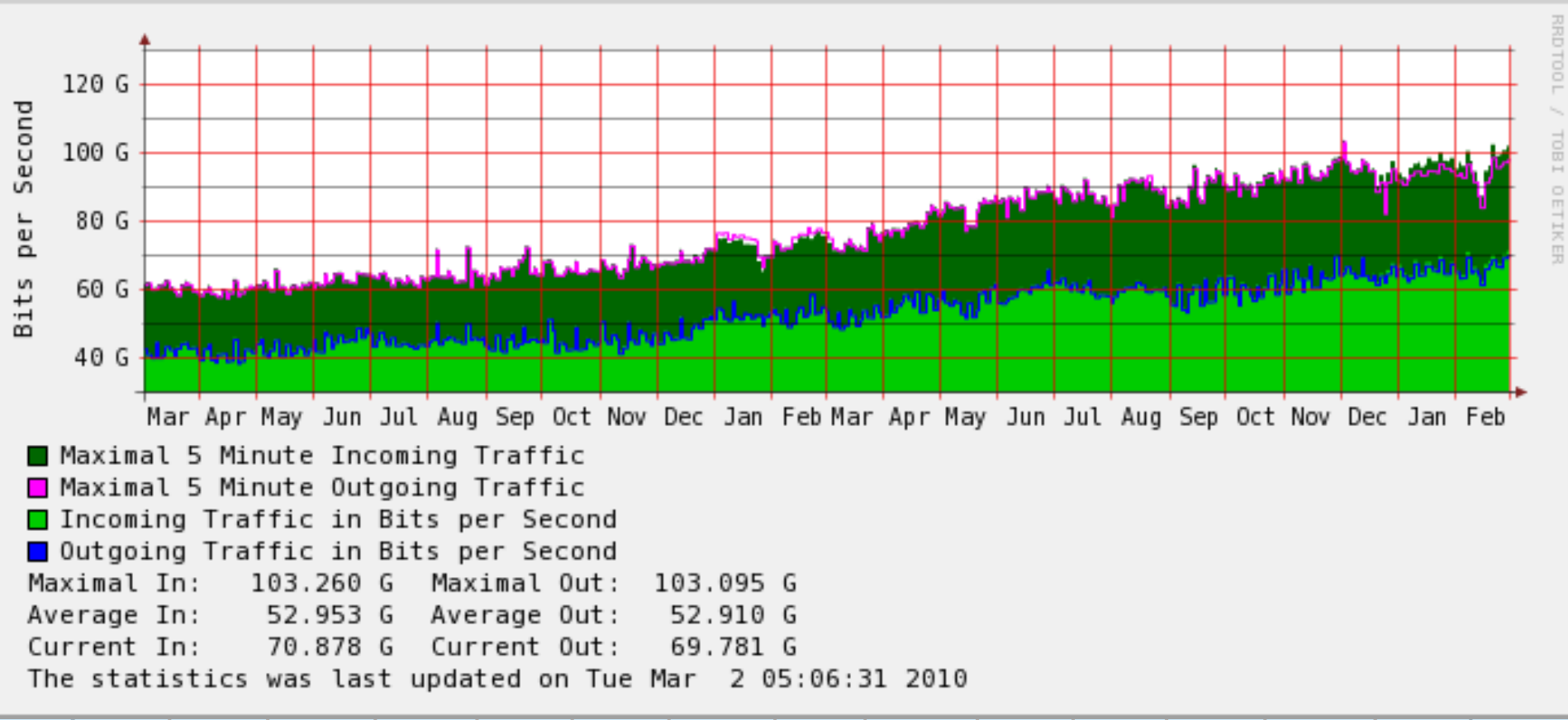


Maximal 5 Minute Incoming Traffic
Maximal 5 Minute Outgoing Traffic
Incoming Traffic in Bits per Second
Outgoing Traffic in Bits per Second

Maximal In:	101.621 G	Maximal Out:	97.767 G
Average In:	68.182 G	Average Out:	67.500 G
Current In:	69.288 G	Current Out:	69.255 G

The statistics was last updated on Tue Mar 2 14:11:31 2010

Some Statistics (3/3)



HKIX2 Currently

- Set up in 2004 as redundant site
- IX portion managed by CUHK

- Linked up to HKIX1 by 2 x 10GE links
- It is **Layer-3** connection so different broadcast domain from HKIX1
  - Same AS4635 MLPA
  - Participants cannot do BLPA across HKIX1 and HKIX2
  - But this is to be changed soon…

Implementation of New High-End Switch

- To sustain growth, HKIX needed a brand new high-end switch at the core (HKIX1)
  - To support >100 10GE ports
  - To support LACP with port security over GE & 10GE ports
  - To support sFlow or equivalent
- Cisco Nexus 7018 selected after extensive pre-tender POC tests and complicated tendering
- In production since 15 June 2009
- Migration of connections from 6513 to 7018 still in progress
  - Most 10GE connections have been migrated
- Have ordered another 7018 for resilience

Our New 7018

7018 Preparation (Before 15 Jun 2009)

- Non-standard equipment rack needed:
  - Delivery issue, installation issue and high price
- Chassis failure
- Port Security problem
  - Had to wait for NX-OS 4.2(1) with major fix on Port Security
- SFP+ contact problem: unplug->plug to solve
- DCNM software to manage 7018, with Windows server, needed to be upgraded at the same time as NX-OS for on-duty operators to disable port-security
  - SNMP to disable port-security not supported anymore
- ISSU seems working fine
- First IX customer so had good support from Cisco

Migration Issues (After 15 Jun 2009)

- 7018 in production since 15 Jun 2009
- Large participants' migration to new switch is a big issue
  - Layer 2 Netflow would help but we do not have it yet
- 6513 as central hub -> 7018 as central hub
- Inter-switch links 2x10GE -> 4x10GE
  - But we did not have enough 10GE ports on 6513's
  - 7018 does not support ER/ZR yet
- Xenpak changed to SFP+
  - Providing upgrade options to 10GE participants
  - Cabling patching done by fixed networks
- Concerns on migration by individual participants

IPv6 at HKIX

- CUHK/HKIX is committed to help Internet development in HK
- IPv6 supported by HKIX since Mar 2004
- Today, 42 AS'es have their IPv6 enabled at HKIX
  - >2,000 IPv6 routes served by MLPA
  - BLPA encouraged
- Dual Stack recommended
  - No need to have separate equipment and connection for IPv6 so easier to justify
  - But cannot know for sure how much IPv6 traffic in total
  - Should be lower than 1% of the total traffic
  - With the new switch, we should be able to have more detailed statistics later

New for IPv6 at HKIX

- HKIX can now support **IPv6-only** connections from commercial networks at MEGA-i
  - Max 1 x GE per participant
  - Must do BLPA with CUHK networks
  - This should help some participants try out IPv6 more easily
- More and more root / TLD servers on HKIX support IPv6

MLPA at HKIX

- Mandatory for Hong Kong routes only
- Our MLPA route servers do not have full routes
- We do monitor the BGP sessions closely
- ASN of Router Server: AS4635
  - AS4635 seen in AS Path
- IPv4 route filters implemented strictly
  - By Prefix or by Origin AS
  - But a few trustable participants have no filters except max number of prefixes and bogus routes filter
  - Accept /24 or shorter prefixes
- IPv6 route filter not implemented in order to allow easier interconnections
  - But have max number of prefixes and bogus routes filter
  - Accept /64 or shorter prefixes
- See http://www.hkix.net/hkix/route-server.htm for details

Bilateral Peering over HKIX

- **HKIX does support and encourage BLPA** as HKIX is basically a layer-2 IXP
- With BLPA, your can have better routes and connectivity
  - One AS hop less than MLPA
  - May get more routes from your BLPA peers than MLPA
- Do not blindly prefer routes learnt from HKIX's MLPA by using higher LocalPref
  - Doing more BLPA recommended
- Set up a record of your AS on www.peeringdb.com and tell everyone that you are on HKIX and willing to do BLPA
  - Also use it to find your potential BLPA peers
- Most content providers are willing to do bilateral peering
- Do set up bilateral peering with root / TLD DNS servers on HKIX to enjoy faster DNS queries

Participants from Other Asian Economies

- The number is increasing
- Those are among the top 5 ISPs in their corresponding economies and they are not really regional players so they do interconnections only in HK
- From Australia, Bhutan, India, Indonesia, Korea, Malaysia, Philippines, Qatar, Taiwan, Thailand and so on
- They seek for better interconnections and better connectivity
- They may be willing to do BLPA at HKIX so contact them for BLPA

- HKIX is indeed serving as an Asian IXP

Port Security

- Port Security implemented strictly
  - Also for LACP connections
- One MAC address / one IPv4 address / one IPv6 address per port (or LACP port channel)
- UFB (Unicast Flood Blocking) feature is important

- Some participants are unaware of this and do change of router / interface without notifying us

Link Aggregation (LACP)

- Having many connections to HKIX increases difficulties of traffic engineering
- May not be able to support many connections if you only have a few routers
  - Each router can only have one interface connecting to HKIX
- LACP is a solution to solve these issues when your traffic grows
- Now, 7018 at HKIX1 can support LACP
- However, please do check whether your circuit providers can provide clear channel Ethernet circuits to HKIX1 with enough transparency before you place orders
- Please also check whether your routers can support LACP

Other Operational Tips

- HKIX cannot help blackhole traffic because HKIX is basically a layer-2 infrastructure
- If there is scheduled maintenance, please notify hkix-noc@cuhk.edu.hk in advance so that we will not treat your BGP down message as failure
- Do monitor the growth of number of routes from our route server and adjust your max prefix settings accordingly
- Do monitor the utilization of your links closely and do upgrade before they are full
- When your link / BGP session is down, do also check with your circuit providers at the same time
- Do your own route / route6 / as-set objects on IRRDB and keep them up-to-date
  - APNIC RRDB is free if you are a member

To Be Done By June 2010

- HKIX1 broadcast domain / VLAN has been extended to HKIX2
  - To move all HKIX2 participants to HKIX1 VLAN which will involve change of IP addresses
- All IPv4 connections to migrate to 202.40.160/23 from 202.40.161/24 *(and 218.100.16/24)*:
  - Change of network mask only
- All IPv6 connections to migrate to 2001:7FA:0:1::/64 from 2001:7FA:0:1::CA28:A100/120 *(and 2001:7FA:0:1::DA64:1000/120)*:
  - Change of network mask only
- Support MLPA route server redundancy:
  - 202.40.161.1 (rs1.hkix.net) & 202.40.161.2 (rs2.hkix.net)
- Support 4-byte ASN

Our Goals

- To have one single HKIX broadcast domain to better support BLPA
- To have better resilience
- To sustain future growth
- To reduce confusion

HKIX1 RS Upgrade (20 Dec 2009)

- To fix IOS IPv6 bug (10 Nov 2009 incident) which caused route server reload
  - OS not upgraded for 5+ years
  - Still hybrid mode (CatOS + IOS)
  - CatOS upgrade (to support ZR) + IOS upgrade
  - Very much preparation beforehand
  - Supervisor module problem solved after unplug & plug
  - Still not native mode because hybrid mode upgrading to native mode with live connections too risky

HKIX2 Switch Upgrade (07 Feb 2010)

- Native IOS upgrade of 6513 at HKIX2
- For UFB (Unicast Flood Blocking) support
  - We had >500Mbps Unicast Flood traffic at HKIX2's new HKIX1 VLAN caused by asymmetric traffic and longer ARP table aging time than switch forwarding table aging time
- Also for 4-byte ASN support on redundant RS (rs2.hkix.net)
- We did prior test on spare 6513 first
- We had support on site and we did the upgrade remotely
- Layer 2 to HKIX1 beforehand with 2x10GE as VLAN trunk + LACP

To upgrade RS1 to IOS (Before 30 Jun 2010)

- To support 4-byte ASN
- Use spare 6513 to do RS1
- Move all UTP connections to 7018
- Move 10GE ER/ZR connections to spare 6513

Other Plans for 2010

- MLPA: Support daily automatic route filter updates from routing registry database (IRRDB)
- MLPA: Support more BGP community for easier traffic engineering
- Portal for Participants
  - Traffic statistics with data from Layer-2 Netflow
- Improve after-hour support

- **Suggestions are welcome**

Questions?