



# HKIX Updates at HKNNOG 1.1

*Che-Hoo CHENG  
CUHK/HKIX  
17 Apr 2015*



# 20<sup>th</sup> Anniversary of HKIX



- HKIX started with thin coaxial cables in Apr 1995
  - *Gradually changed to UTP cables / fibers with switch(es)*
    - *low-end -> high-end*
    - *One switch -> multiple switches*
- Participants had to put co-located routers at HKIX sites in order to connect
  - *Until Metro Ethernet became popular*
- It was a free service
  - *Now a fully chargeable service*



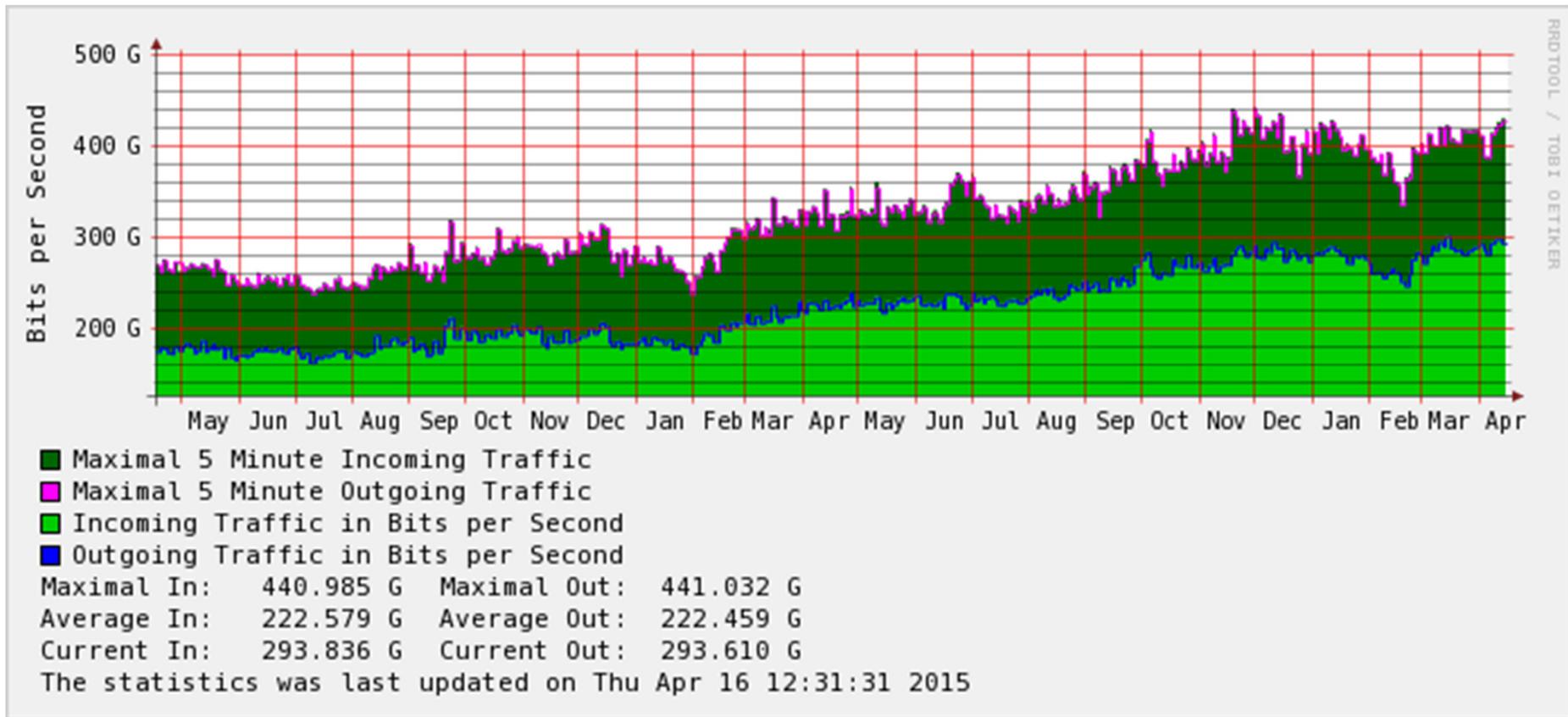
# HKIX Today



- Supports both MLPA (Multilateral Peering) and BLPA (Bilateral Peering) over layer 2
- Supports IPv4/IPv6 dual-stack
- Neutral among ISPs / telcos / local loop providers / data centers / content providers / cloud services providers
- More and more non-HK participants
- >240 ASNs connected
- >430 connections in total
  - 2 x 100GE + >180 x 10GE connections
- ~440Gbps (5-min) total traffic at peak
- Annual Traffic Growth = 30% to 40%



# Yearly Traffic Statistics





# Charging Model



- An evolution from free-of-charge model adopted at the very beginning, to penalty-based charging model based on traffic volume for curbing abuse, to now simple port charge model for fairness and sustainability
- **Have started simple port charge model since 01 Jan 2013**
  - See <http://www.hkix.net/hkix/Charge/ChargeTable.htm>
- Still not for profit
  - HKIX Ltd (100% owned by CUHK) to sign agreement with participants
  - Target for fully self-sustained operations for long-term sustainability



# HKIX Port Charge Table (effective 01 Jul 2014)



Standard Charges		NRC		MRC	
Port Size	Interface	HKD	USD	HKD	USD
E/FE/GE	UTP or SMF	Waived		936	120
10GE	LR	17,940	2,300	7,800	1,000
	ER	39,000	5,000		
	ZR	62,400	8,000		
100GE	LR4	117,000	15,000	46,800	6,000

\* New E/FE(10M/100M) connections are no longer supported.

\*\* Existing E/FE(10M/100M) connections have to be decommissioned or upgraded to GE or above by 30 Jun 2015

Save-IP Discount (applied <b>ONLY</b> if IP address is <b>NOT</b> needed for the port)		Reduction of MRC for each port	
Port Size	Conditions	HKD	USD
10GE	With LACP enabled; <b>NOT</b> applied to the 1st LACP port which needs IP	-780	-100
100GE	With LACP enabled; <b>NOT</b> applied to the 1st LACP port which needs IP	-4,680	-600

\* No such discount for E/FE/GE connections and NRC

Volume Discount (applied under same ASN and under same contract <b>ONLY</b> )		Reduction of MRC for each port	
Port Size	Conditions	HKD	USD
10GE	Applied to the <b>5th</b> 10GE port and onwards	-780	-100
100GE	Applied to the <b>3rd</b> 100GE port and onwards	-4,680	-600

\* No such discount for E/FE/GE connections and NRC

**REMARKS:**

**NRC**=Non-Recurring Charge (Non-Refundable)

**MRC**=Monthly Recurring Charge



# The Recent Upgrade

- A new highly-scalable two-tier dual-core architecture within CUHK by taking advantage of the new data center inside CUHK campus
  - HKIX1 site + HKIX1b site as Core Sites
    - Fiber distance between 2 Core Sites: <2km
  - Provide site/chassis/card resilience
  - Support 100GE connections
  - Scalable to support >6.4Tbps total traffic using 100GE backbone links primarily and FabricPath
- **Ready to support HKIX2/3/4/5/6/etc as Satellite Sites having Access Switches only which connect to Core Switches at both Core Sites**

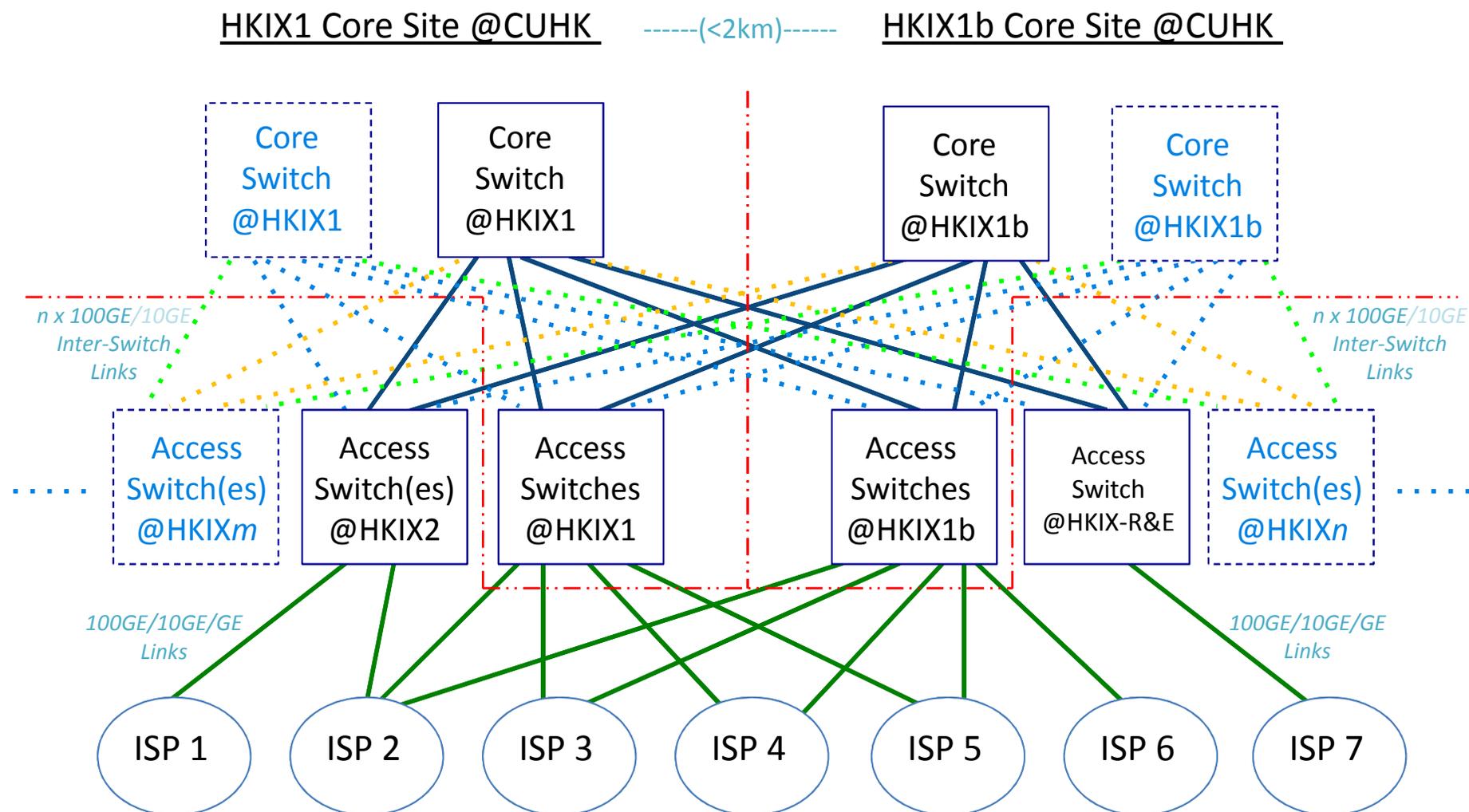


# The Design

- Dual-Core Two-Tier Design for high scalability
  - Have to sustain the growth in the next 5 years (to support >6.4Tbps traffic level)
  - Core Switches at 2 Core Sites (HKIX1 & HKIX1b) only
    - No interconnections among core switches
  - Access Switches to serve connections from participants at HKIX1 & HKIX1b
    - Also at Satellite Sites HKIX2/3/4/5/6/etc
    - Little over-subscription between each access switch and the core switches
  - FabricPath (TRILL-like) used among the switches for resilience and load balancing
- Card/Chassis/Site Resilience
  - LACP not supported across chassis though (card resilience only)
- 100GE optics support
  - LR4 for <=10km and ER4-lite for <=25km (2Q2015)
  - Support by local loop providers is key
- Port Security still maintained (over LACP too)
  - Only allows one MAC address / one IPv4 address / one IPv6 address per port (physical or virtual)
- Have better control of Unknown-Unicast-Flooding traffic and other storm control



# HKIX Dual-Core Two-Tier Architecture For 2014 and Beyond





# The Migration

- New switches in production at HKIX1 site starting Mar 2014
  - Parallel run with old switches, interconnected with the new switches using multiple 100GE links
  - All new connections have been on new access switches since their production
    - While existing connections are being moved to the new access switches one by one
- By early Aug 2014, all 10GE connections had been migrated
- HKIX1b site put into production in Nov 2014 to provide true site resilience
- Remaining GE connections **MUST** be moved to new switches
  - Deadline is **30 Jun 2015**
  - No E/FE support starting then
- RS1 & HKIX-R&E have been moved to the new architecture
- HKIX2 will be moved to the new architecture
- The remaining old switch will be decommissioned after **30 Jun 2015**



# The Upgrade in General

- The upgrade was considered successful after very much hard work
- Support 100GE connections starting Oct 2014
- Provide site / chassis / card resilience options for HKIX participants
- Support easy expansion with high scalability
- Support fast deployment for satellite sites



# FabricPath

- We adopt leaf-and-spine topology for high scalability
  - Avoid connecting participant ports on core switches
- The Spanning Tree Protocol (STP) domains do not cross into the FabricPath network
- Layer 2 gateway switches, which are on the edge between the CE and the FabricPath network, must be the root for all STP domains that are connected to a FabricPath network
- Load balancing is working fine
  - Even with odd number of links
- Transparent to participants (i.e. no BGP down) when adding/removing inter-switch links



# One Very Critical Point for an IXP



- An IXP must NOT be vulnerable to DDoS attack itself
  - Congestion at one port must NOT cause trouble to any other ports
- Network QoS Policy - Congestion Control Mechanisms
  - Default is “Burst optimized” which is not good for IXP because of sharing of buffer by multiple ports
  - “Mesh optimized” is more suitable for IXP



# Proxy ARP Threat

- Proxy ARP **MUST be disabled** all the times on the interface connecting HKIX switches
- Participants **MUST** provide the “show ip interface” output for verification before HKIX turning up the switch ports
- Will use Dynamic ARP Inspection (DAI) to maintain static ARP list
  - But not used yet as it is manual
  - Need to input a few commands for this instead of just one command



# DDoS Mitigation Trial

- Accept /32's with BGP community 4635:666 from HKIX route servers for MLPA
- Do black-holing by null route the next-hop of HKIX address 202.40.160.66 to null interface
- When under DDoS attack:
  - Signal other participants by advertising /32's under your managed address blocks with 4635:666 to do black-holing at their networks
  - Relevant sites will be sacrificed but other sites and networks will be saved
- >50 participants did the test
  - All of them had positive results
- Will declare production soon
- Will seek collaboration with similar initiatives for maximum outcome



# IPv4 Address Renumbering and Route Servers Upgrade



**Migration Date: 12-15 Jun 2015 (Fri-Mon)**

## IPv4 Address Renumbering

- Network mask will be changed to /21 from /23, for accommodating future growth
- ALL participants must change to **NEW** IPv4 addresses, away from **202.40.160/23**
- **Parallel run of old and new IPv4 addresses only during the 4-day migration period, having learnt from experience of other IXPs**
- MLPA: New route servers will support new IPv4 addresses while existing route servers will support old addresses, but IPv6 will be handled separately
- BLPA: Individual participants have to coordinate with their peering partners directly
- *No change to IPv6 addresses*

## Route Servers Upgrade

- The two existing route servers running on 6513s will be decommissioned
- New route servers will be installed at HKIX1 and HKIX1b (the two HKIX core sites)
- More route server features will be supported later



# IPv4 Address Renumbering and Route Servers Upgrade

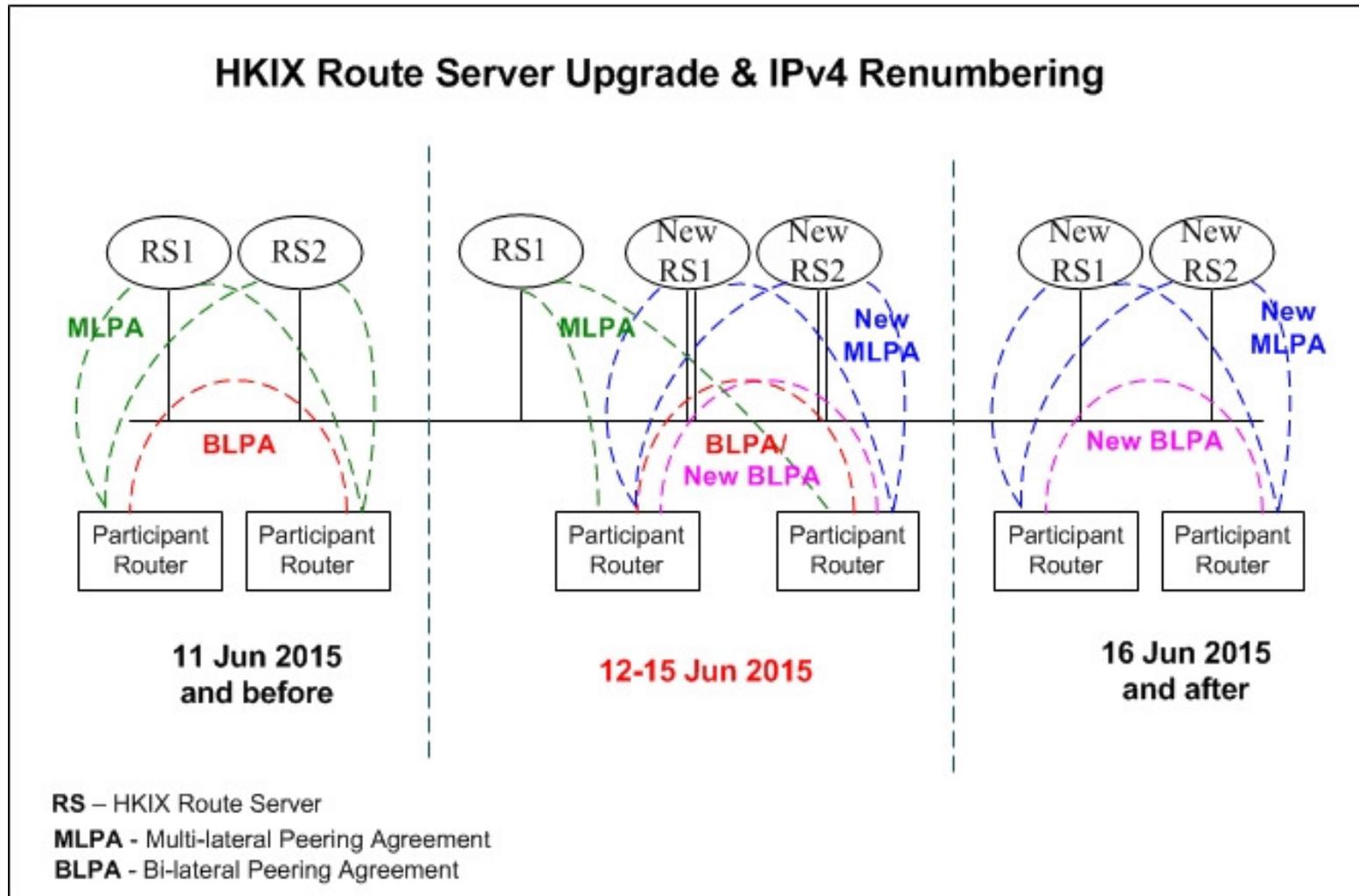


## Schedule

DATE	TASKS
11 MAR 2015	HKIX announces IPv4 renumbering and route server upgrade
11 APR 2015	2nd reminders to ALL HKIX participants
04 MAY 2015	HKIX provides the information of NEW IP address block (/21) and publish the mapping of current IPs (/23) and new IPs (/21) on HKIX web site
11 MAY 2015	Final reminders to ALL HKIX participants
12 JUN 2015	HKIX setup the new route servers for MLPA peering
<b>12-15 JUN 2015</b>	<b>HKIX participants change to new IPs and setup MLPA peering with new route servers. Re-establish the BLPA peering (if any) with individual participant(s) with the corresponding IPs in new /21</b>
16 JUN 2015	HKIX decommission the old route servers and IP address block 202.40.160/23 ceased



# IPv4 Address Renumbering and Route Servers Upgrade





# Setting up Multiple HKIX Satellite Sites



- To be named HKIX2/3/4/5/6/etc
- Allow participants to **connect to HKIX more easily at lower cost** from those satellite sites
- Open to all commercial data centres which fulfil minimum requirements so as to maintain neutrality which is the key success factor of HKIX
  - ISO27001 requirement
  - Minimum size requirements
  - Requirements on circuits connecting back to the two HKIX core sites
  - Non-exclusive
- Intend to create win-win situation with satellite site partners
- **Contact us if you are interested in this**
- *NOTE: HKIX1 and HKIX1b (the two HKIX core sites) will continue to serve participants directly*



# Planned Work in 2015-2016



- Control Proxy ARP
- Advanced Route Server functions
- Better support for DDoS Mitigation
- More L2 ACL on HKIX peering LAN
- Portal for HKIX participants
  - Port info and traffic statistics
  - Self-service port security update
  - Network maintenance schedule
- Improved after-hour support
- ISO27001



**Thank you!**