Best Practices for ISPs

Che-Hoo Cheng
CUHK/HKIX
2014.08.01

What Providers Care About

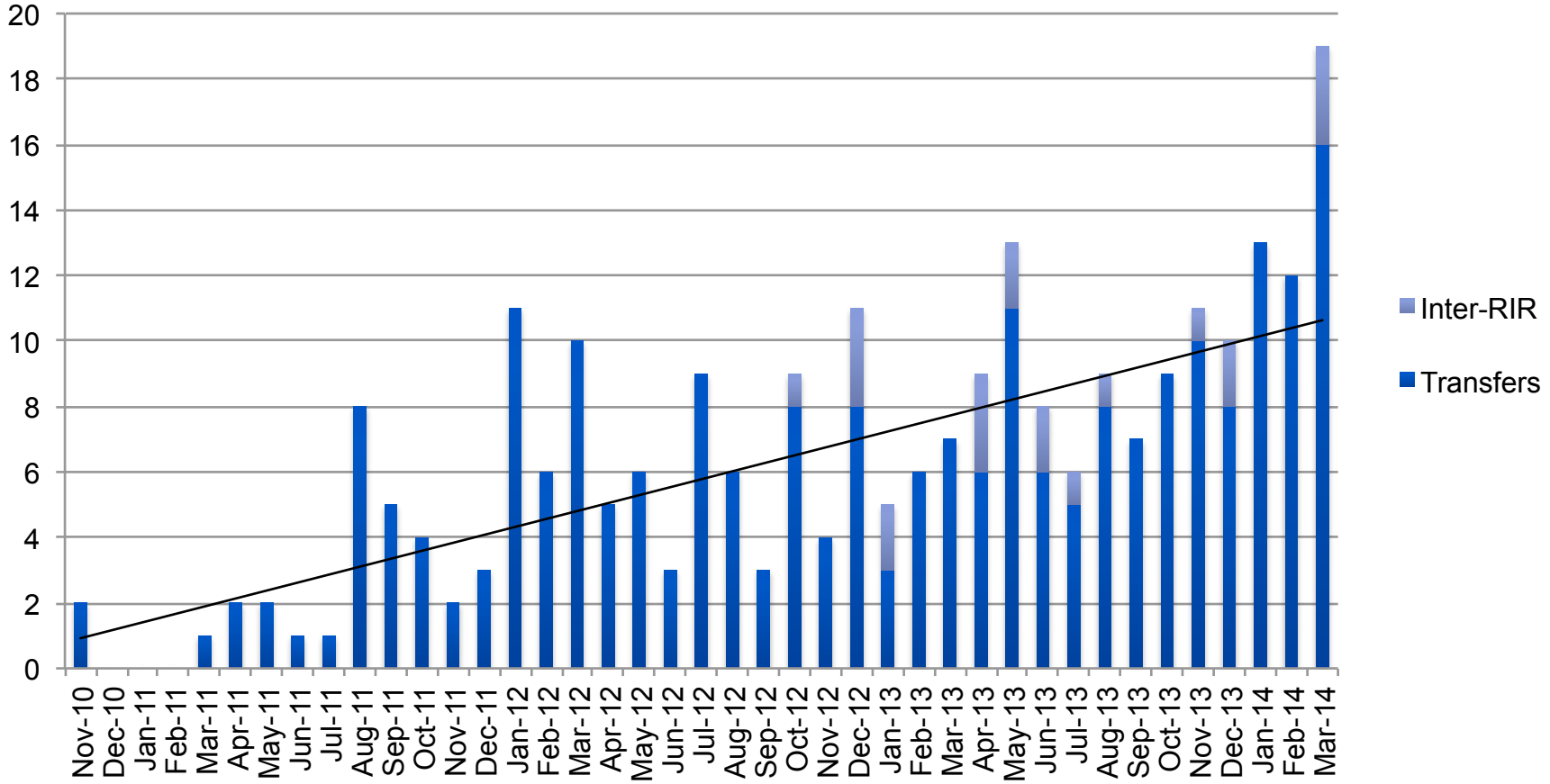
- Cost / Performance / Resilience / Interconnections / Efficiency / Scalability / **Security / Stability / Reliability**

- The market is highly competitive
- All providers are searching for their own niche positions which make them look better than their competitors in order to survive

IPv4 Addresses

- Running out globally
- New policy for address distribution from IANA returned pool (APNIC prop-105)
  - One more /22 for each APNIC member
  - In addition to one /22 from the last /8 pool
- Still have to demonstrate the needs
- New APNIC members can enjoy this also

IPv4 Address Transfers by APNIC

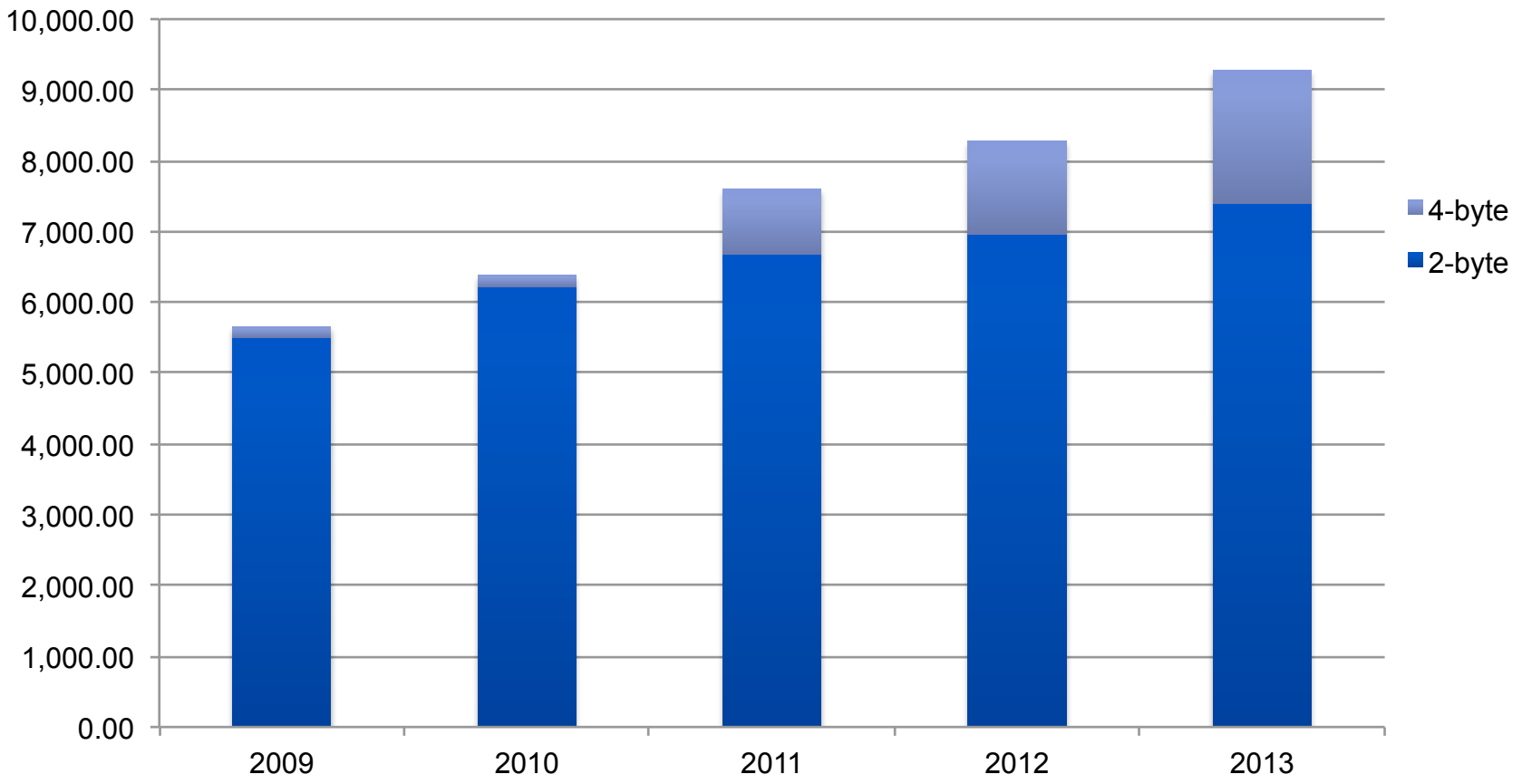


Source: APNIC

Trend of IPv4 Addresses

- Growing Trend is more and more end-users (enterprises) are getting their own portable IPv4 addresses
  - Up to 2 x /22 directly from APNIC plus buying from market
  - Still need to demonstrate needs
  - Easy referral at MyAPNIC for ISPs to refer customers to join APNIC as members

ASN Delegations by APNIC

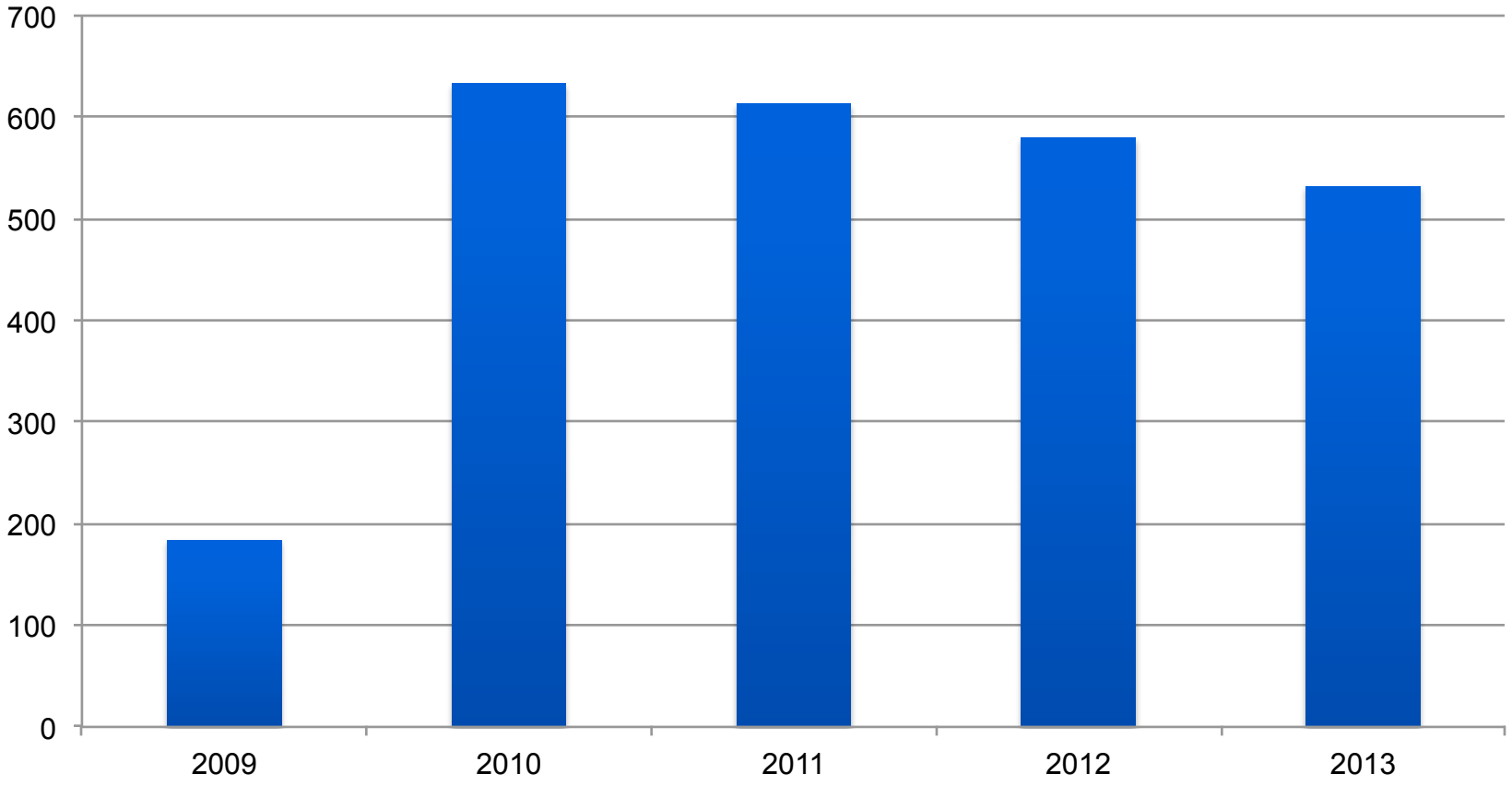


Source: APNIC

ASN

- 4-byte ASN (Autonomous System Number) is slowly becoming the norm
- APNIC is assigning more 4-byte ASN than 2-byte ASN now
- Support from equipment vendors is mature
- More and more large enterprises are running BGP with their own ASN and IP addresses
  - For multi-homing
- ASN is now transferrable (APNIC prop-107)
  - Still need to justify the use based on ASN policy

IPv6 Delegations by APNIC



Source: APNIC

IPv6 Deployment

- Deployment is growing but slowly although IPv4 addresses are running out globally
- Total IPv6 traffic vs total Internet traffic is still within single digit percentage-wise
- Access providers are most reluctant to deploy
  - Possibility of IPv4 transfers and wide-spread use of NATs (of various kinds) are holding them back
- Accelerated deployment will only be observed when the cost of keeping running IPv4 is higher than the cost of deploying IPv6

Broadband Penetration

- Japan, Hong Kong and South Korea have the highest broadband penetration within Asia
- These 3 economies also have the highest average connection speed within Asia *(source: Akamai)*
- FTTH plays a big role here
- Hong Kong ISP market is largely market driven
  - Broadband ISPs do not have incentive to do FTTH for low-density buildings
  - 4G/LTE is an option for those low-density buildings
- Mobile: 4G/LTE usage is growing very fast because of proliferation of smartphones and mobile hotspots

100G in Operations

- For supporting FTTH and continuous growth of HD video content, backbone links need to be even faster
- Multiple 10G's may not be enough at certain locations
- Having higher and higher demands for 100G as backbone links
- Prices of 100G optics are dropping slowly but gradually and this helps the adoption of 100G

Network Expansion to Overseas

- To improve overall connectivity and performance for customers
  - It is a global trend to go overseas for better interconnections
- Set-up equipment (POPs) at major Internet hub locations and do interconnections
  - Tokyo, Hong Kong and Singapore are the main hubs in Asia
  - But other economies are trying to join the club
- Alternate model for access providers is to connect to IXPs remotely by using Ethernet over SDH/MPLS
  - Some special providers provide such remote IXP connection services specifically
- But for cloud/content services providers, they have to set up servers everywhere in order to get closer to the end users
  - Same for anycast DNS service providers
  - This helps the data center business all around the world

Data Centers

- With the high growth of cloud service providers, CDN service providers and big content providers, data centers around the world are running out of space
  - Anycast DNS service providers and TLD registries/registrars also need space globally but their need is relatively small
- More data centers are being built
- Facilitate easy private interconnections within data centers
- For data centers with multiple locations, they tend to provide carrier services for their customers across different locations

IXPs

- IXPs continue to play a key role for interconnections among ISPs and other internet players
- **IXPs must have enough spare capacity so that they are not vulnerable to DDoS themselves**
- Larger IXPs are mostly serving global market
  - Support of 100G is becoming essential
  - Use of 100G starts from inter-switch links
- Some IXPs to expand overseas with independent layer-2 infrastructure
  - Some even provide layer-3 transit services (full or partial transit)
- **IXPs and data centers are natural partners**

DNS

- More and more TLDs (some are IDN-TLDs) being approved by ICANN
  - TLD registries and registrars need good global infrastructure
  - They tend to use anycast more and more
- Anycast is important to improve resilience of authoritative DNS infrastructure
  - Not just for root/TLDs but also for individual DNs
  - Not just globally but also locally
  - We need more anycast DNS service providers which have good infrastructure world-wide and locally
- DNS infrastructure is something very special and very critical
  - Becoming the main targets of many recent attacks!!!
  - Traditional network admins and system admins do not put much energy on DNS infrastructure
  - Need real DNS professionals to run it
- DNSSEC adoption rate is still low

DDoS Attacks

- With enhanced Internet infrastructure from backbones to edges, there are more and more large-scale DDoS attacks on Internet of different types
  - DNS Amplification
  - NTP Amplification
  - TCP SYN Flood
  - Random DNS queries on targeted domain names
    - Relevant DNS servers are suffered
- HK suffered a lot recently
- ISPs and Network Operators MUST follow the best practices!!!

Best Practices for ISPs

- More like guidelines for engineers
- On technical and operational parts
- Not something very static
- Technologies and the industry are changing very fast
- Engineers should update themselves continuously
- Start from: http://www.ietf.org/rfc/bcp/bcp-index.txt

- End Goal – To help make Internet more secure, stable and reliable

BCP38/RFC2827 – Network Ingress Filtering

- Defeating Denial of Service Attacks which employ IP Source Address Spoofing
  - Ingress traffic filtering at the periphery of Internet connected networks will reduce the effectiveness of source address spoofing denial of service attacks
  - Sources of attacks can be traced more easily
  - Reflection type of attacks can be mitigated largely
- Should be done at both the ISPs and edge networks

BCP84/RFC3704 – Ingress Filtering for Multihomed Networks

- Aimed at ISPs and edge networks
  - Should be stricter when closer to edge
- Ingress Access List
  - checks the source address of every message received
  - Quite manual
- Strict Reverse Path Forwarding (Strict RPF)
  - Source address is looked up in the Forwarding Information Base (FIB)
  - Very strict so may not suit asymmetric routing / multi-homed scenarios
- Feasible Path Reverse Path Forwarding / Loose Reverse Path Forwarding / Loose Reverse Path Forwarding Ignoring Default Routes
- Ingress Filtering at Multiple Levels
  - Use Loose RPF When Appropriate
- Both the ISPs and edge networks should verify that their own addresses are not being used in source addresses in the packets coming from outside their network

Packet Filtering Principles

- Filter as close to the edge as possible
- Filter as precisely as possible
- Filter both source and destination where possible

BCP171/RFC6441 – Time to Remove Filters for Previously Unallocated IPv4 /8s

- There are no more unallocated IPv4 /8s at IANA
- Some network administrators might want to continue filtering unallocated IPv4 addresses managed by the RIRs
  - This requires significantly more granular ingress filters and the highly dynamic nature of the RIRs' address pools means that filters need to be updated on a daily basis to avoid blocking legitimate incoming traffic
- Network operators may deploy filters that block traffic destined for Martian prefixes [BCP153/RFC6890]
- http://www.team-cymru.org/Monitoring/BGP/

Best Practices for Ingress/Egress Prefix Filtering

- For Ingress Prefix Filtering:
  - Don't accept RFC1918/RFC6890 prefixes
  - Don't accept your own prefixes
  - Don't accept default (unless you need it)
  - Don't accept prefixes longer than /24
  - Set ingress max prefix limit for peers/IXPs
- For Egress Prefix Filtering:
  - Announce only your own and your customers' prefixes to your upstream providers and peers/IXPs
    - Don't announce default, prefixes belonging to upstream providers/peers/IXPs that you directly connect to and all other prefixes!!!
  - You may announce default and/or full routes to your downstream customers (if they need them) but not the others
- http://www.team-cymru.org/Monitoring/BGP/

BCP16/RFC2182 – Selection and Operation of Secondary DNS Servers

- Authoritative Servers
  - Usually one primary/master server and multiple secondary/slave servers
  - Which one is the real master may not be known externally
  - *Using stealth/hidden master is recommended*
- Servers should be placed at both topologically and geographically dispersed locations on the Internet
- *Using anycast for DNS is becoming the norm*
  - *Not just for root or TLDs but also for individual domain names which have high demands*
  - *Not just globally but also locally*
- *Do not forget about reserved zones*

BCP126/RFC4786 – Operation of Anycast Services

- Multiple nodes sharing the same IP address
  - Coarse distribution of load across nodes
  - Mitigate non-distributed DoS attacks by localizing damage
  - Constraint of DDoS attacks
  - Improve query response time
  - Good for serving DNS queries
- Routing to determine which node to use
- Local scope anycast vs global scope anycast

BCP169/6382 – Unique Origin Autonomous System Numbers per Node for Globally Anycasted Services

- In order to be able to better detect changes to routing information associated with critical anycasted resources, globally anycasted services with partitioned origin ASNs SHOULD utilize a unique origin ASN per node where possible
- Discrete origin ASNs per node provide a discriminator in the routing system that would enable detection of leaked or hijacked instances more quickly and would enable operators that so choose to proactively develop routing policies that express preferences or avoidance for a given node or set of nodes associated with an anycasted service

RFC140/RFC5358 – Preventing Use of Recursive Nameservers in Reflector Attacks

- Due to small query-large response potential of the DNS system, it is easy to yield great amplification of the source traffic as reflected traffic towards the victims
- Amplification factor (response packet size / query packet size) could be up to 100
- Nameserver operators to provide recursive name lookup service to only the intended clients:
  - Disable Open Recursive Servers!!!
  - IP address based authorization
  - Incoming interface based selection
- **Turn recursion off complete on Authoritative Servers!!!**
  - **Keep recursive and authoritative services separate as much as practical**

BCP91/RFC3901 – DNS IPv6 Transport Operational Guidelines

- To avoid name space fragmentation
- Every recursive name server SHOULD be either IPv4-only or dual stack
  - This rules out IPv6-only recursive servers. However, one might design configurations where a chain of IPv6-only name server forward queries to a set of dual stack recursive name server actually performing those recursive queries.
- Every DNS zone SHOULD be served by at least one IPv4-reachable authoritative name server
  - This rules out DNS zones served only by IPv6-only authoritative name servers

BCP162/RFC6302 – Logging Recommendations for Internet-Facing Servers

- NAT is being used widely to preserve IPv4 addresses
  - Multiple nodes sharing one IPv4 address
- Still need to support abuse mitigation or public safety requests under such scenarios
- It is RECOMMENDED that Internet-facing servers logging incoming IP addresses from inbound IP traffic also log:
  - The source port number
  - A timestamp, RECOMMENDED in UTC, accurate to the second, from a traceable time source (e.g., NTP [RFC5905])
  - The transport protocol (usually TCP or UDP) and destination port number, when the server application is defined to use multiple transports or multiple ports

BCP46/RFC3013 – Recommended ISP Security Services and Procedures

- A good summary as of Year 2000
- Computer Security Incident Response Team (CSIRT)
  - *Abuse / Incident Response Team (IRT) contacts on APNIC database*
- Appropriate Use Policy (AUP)
  - Should be clear in stating what sanctions will be enforced in the event of inappropriate behavior
- Registry Data Maintenance
  - Internet Routing Registry (IRR) and APNIC databases
- Ingress/Egress Filtering on Source Address
- Route Filtering
- Disable Directed Broadcast as default [BCP34/RFC2644]
- No Open Mail Relay [BCP30/RFC2505]
- SMTP Service Extension for Authentication [RFC2554]

Other Good Practices for Networks

- Disable Proxy ARP
- No transit over IXPs
- BGP with MD5
- Help blackhole DDoS traffic closer at the sources
- Use of PeeringDB for interconnections
- RPKI (Resource Public Key Infrastructure)
  - Not production yet
  - Should keep an eye on the development (via APNIC)
  - Should participate in the testing

Additional Measures

- Harden your own network
  - Physical Security
  - Close all unnecessary services
  - Secure network management (syslog, snmp, tftp)
  - Secure remote access (ssh, vpn)
  - Strong authentication (2FA)
- Network Monitoring
  - Netflow
  - Passive DNS
- Proactively scan/detect vulnerable servers/equipment/devices on your networks and for your customers
  - Disable open recursive DNS servers and open NTP servers
- Response-Rate-Limiting on authoritative DNS servers
- **Always have the same security measures for IPv6 as IPv4!!!**
- **Start deploying DNSSEC!!!**

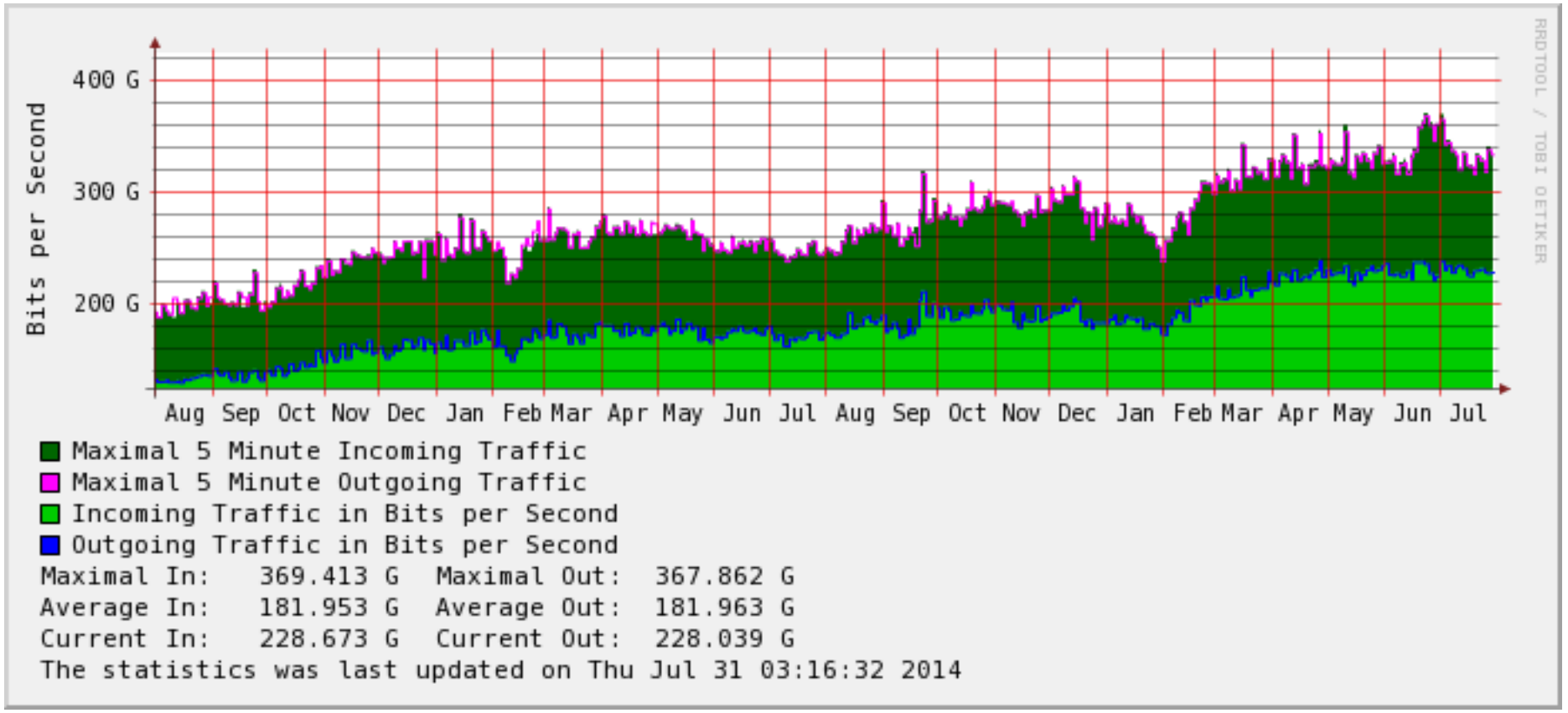
What is HKIX?

- HKIX is a public Internet Exchange Point (IXP) in Hong Kong
- HKIX is the main IXP in HK where various networks can interconnect with one another and exchange traffic
  - Not for connecting to the whole Internet
- HKIX was a project initiated by ITSC (Information Technology Services Centre) of CUHK (The Chinese University of Hong Kong) and supported by CUHK in Apr 1995 as a community service
  - Still fully supported and operated by CUHK
- HKIX serves both commercial networks and R&E networks
- The original goal is to keep intra-HongKong traffic within Hong Kong

HKIX Today

- Supports both MLPA (Multilateral Peering) and BLPA (Bilateral Peering) over layer 2
- Supports IPv4/IPv6 dual-stack
- Accessible by all local loop providers
- Neutral among ISPs / telcos / local loop providers / data centers / content providers / cloud services providers
- More and more non-HK participants
- >240 ASNs connected
- >370 connections in total
  - ~130 10GE connections
- ~370Gbps (5-min) total traffic at peak
- Annual Traffic Growth = 30% to 40%

Yearly Traffic Statistics



Help Keep Intra-Asia Traffic within Asia

- We have almost all the Hong Kong networks
- So, we can attract participants from Mainland China, Taiwan, Korea, Japan, Singapore, Malaysia, Thailand, Indonesia, Philippines, Vietnam, India, Bhutan, Qatar and other Asian countries
- We now have more non-HK routes than HK routes
  - On our MLPA route servers
  - Even more non-HK routes over BLPA
- We do help keep intra-Asia traffic within Asia
- In terms of network latency, Hong Kong is a good central location in Asia
  - ~50ms to Tokyo
  - ~30ms to Singapore
- HKIX is good for intra-Asia traffic

CUHK's Vision

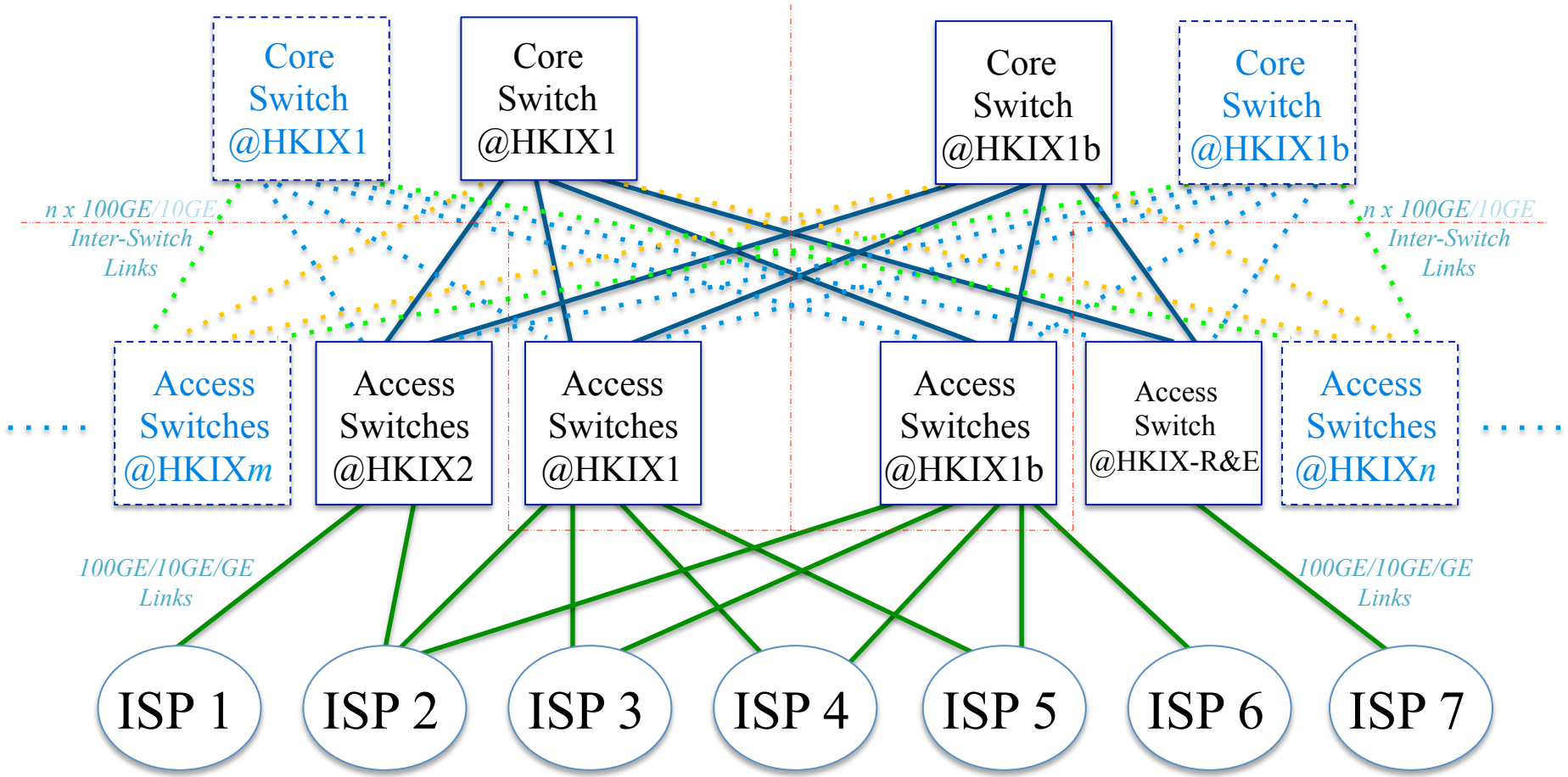
- CUHK has a strategic uniqueness in running HKIX in a long-term
- While CUHK does not have a service provider role, we are still obligated to continue managing it as a public service
- HKIX is very much like road infrastructure and airport in Hong Kong
- Support from HKSAR Government is needed to make it prosper, and to maintain it as an Asian internet hub

- **HKSAR Government has provided one-off funding for capital expenses of network equipment upgrade in 2013-14**

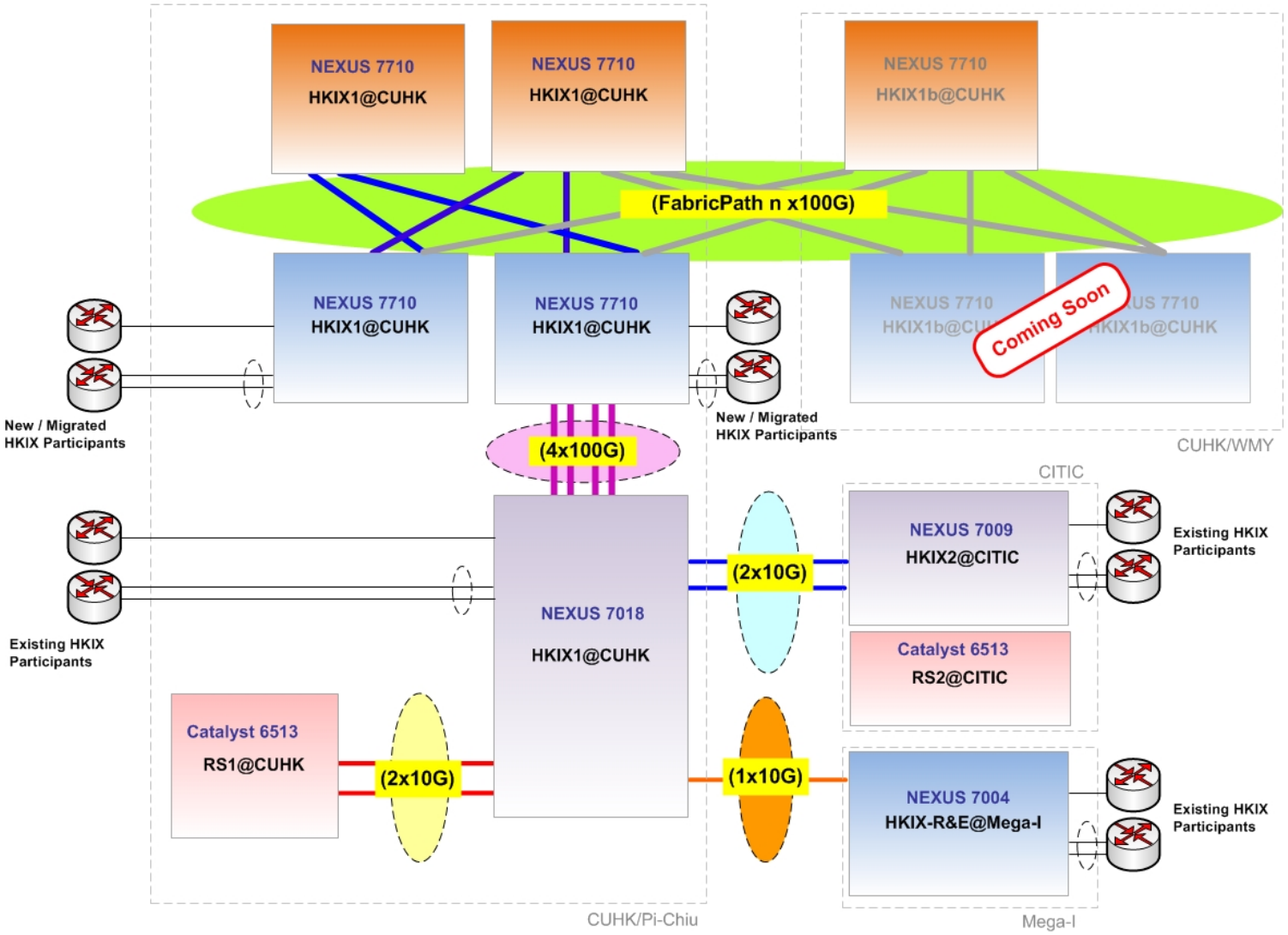
HKIX in 2013-14

- Have started simple port charge model since Jan 2013
- Maintain as not-for-profit operations
  - Target for fully self-sustained operations for long-term sustainability
- Deploying new highly-scalable two-tier dual-core architecture within CUHK by 4Q2014 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 using FabricPath**

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



HKIX Network Diagram (July 2014)



NOGs

- Network Operators' Groups (NOGs) are being established everywhere in Asia
  - To exchange knowledge and information
    - Best practices, new trends and so on
  - To enhance overall quality of Internet infrastructure
    - Performance, security, stability and so on
  - To help do trouble-shooting and solve problems together when needed
- Regional NOGs
  - NANOG, APRICOT/APOPS, SANOG, MENOG
- Local NOGs
  - JANOG, AusNOG, NZNOG, MYNOG, SGNOG, IDNOG, BDNOG are all active
- **HKNOG has been formed**
  - **Did half-day trial events twice (HKNOG 0.1 & 0.2)**
  - **Will have full-day HKNOG 1.0 meeting on Sep 1**
    - **Check www.hknog.net**
  - **To provide a repository of useful information for engineers in HK, such as up-to-date Best Practices for ISPs and other useful tips and trouble-shooting tools**

Useful Resources

- http://www.ietf.org/rfc/bcp/bcp-index.txt
- https://www.nanog.org/resources
- http://bgp.he.net
- https://www.ripe.net/data-tools/stats/ris
- http://www.routeviews.org
- http://www.team-cymru.org/Services/Bogons/
- https://atlas.ripe.net/dnsmon/

Let's work together to make Internet more secure, stable and reliable.

Thank you!