GRAND DUCHY OF LUXEMBOURG
Ministry of Foreign Affairs

Directorate for Development Cooperation

European Union Africa
Infrastructure Trust Fund

Internet Exchange Point Design

IXP Technical Design, Technical Resources and Value Added Services

Internet Society

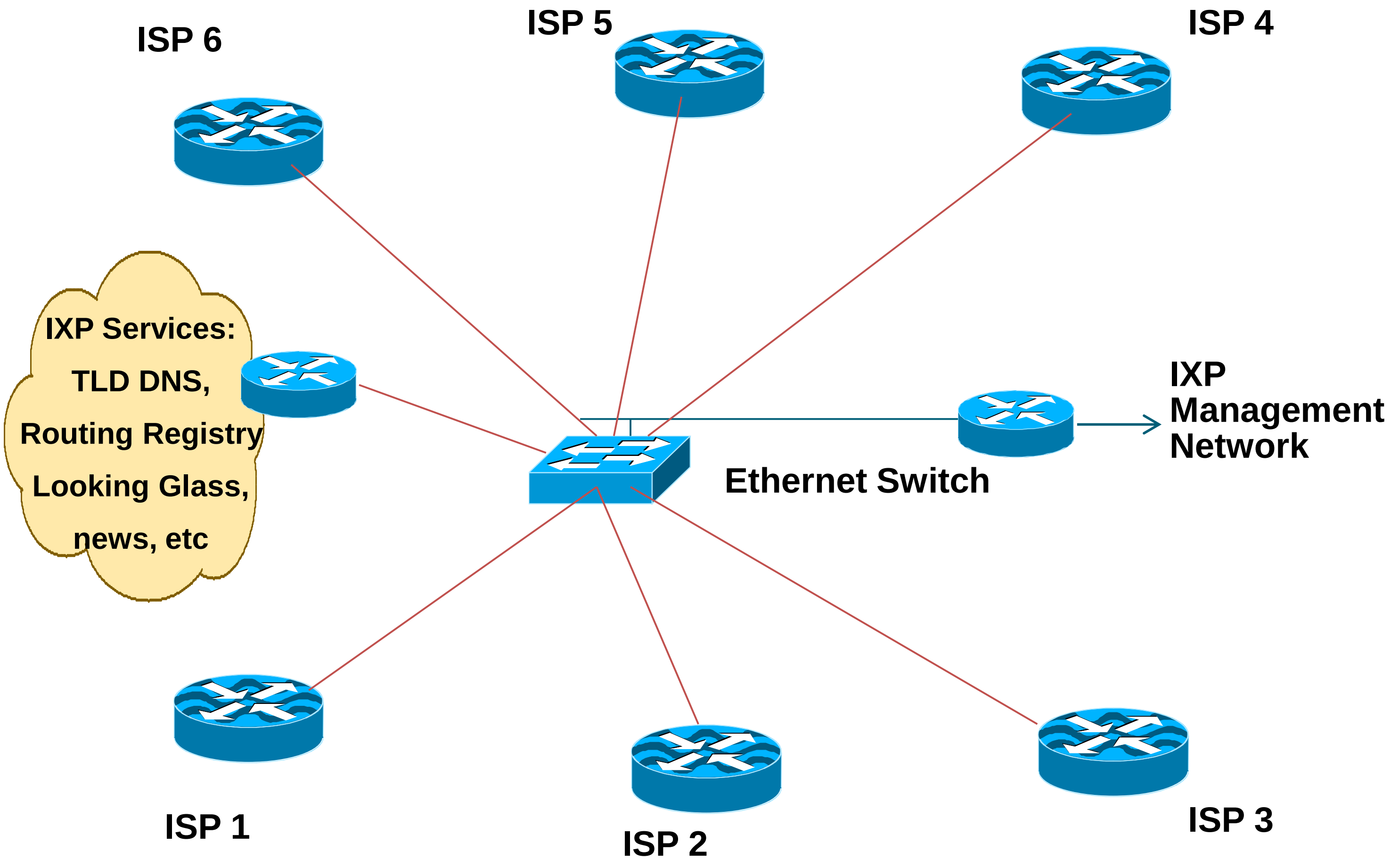
IXP Design

- Layer 2 Exchange Point
- Design Considerations
- Route Collectors & Servers
- What can go wrong?

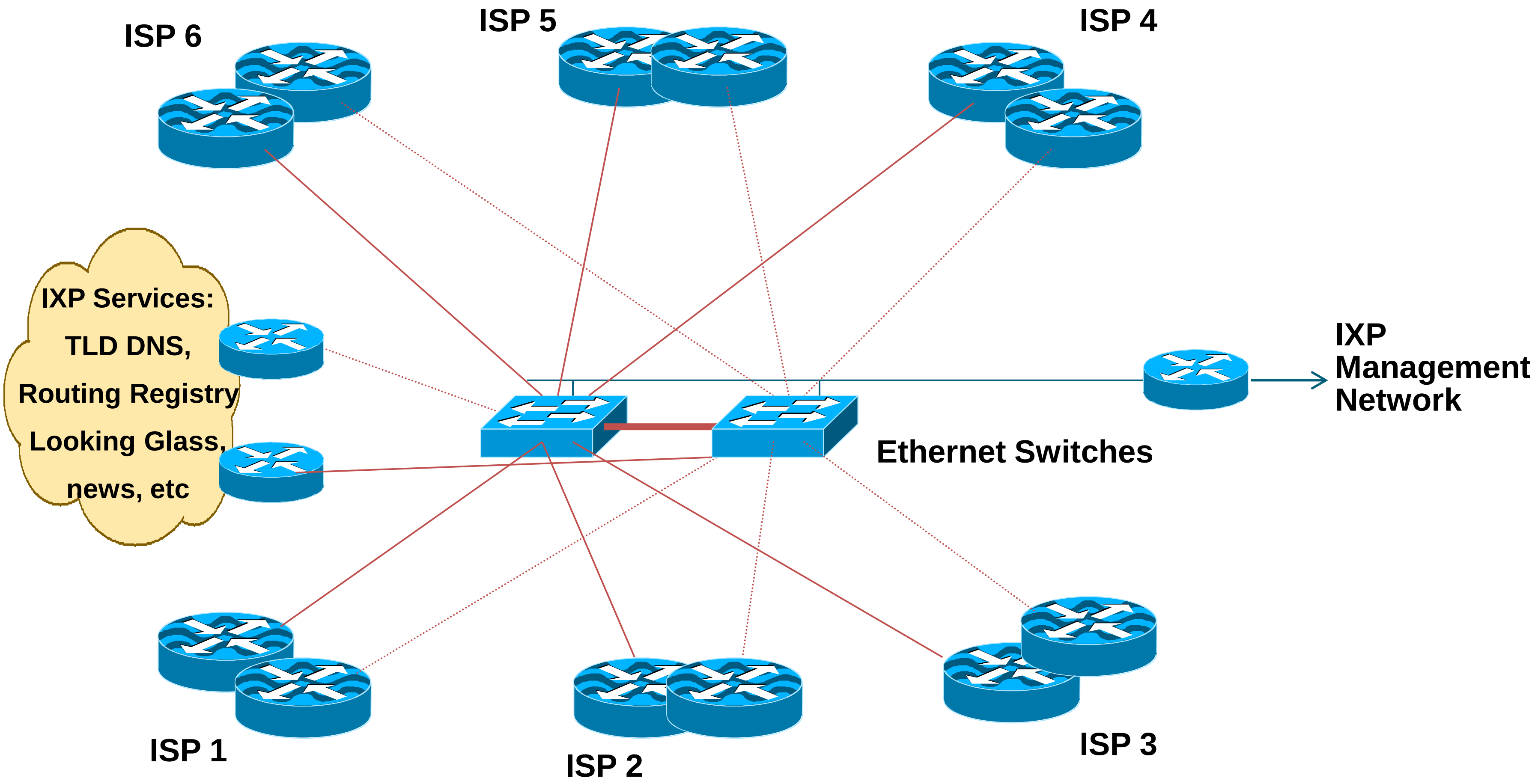
Layer 2 Exchange

The traditional IXP

Layer 2 Exchange



Layer 2 Exchange



Layer 2 Exchange

- Two switches for redundancy
- ISPs use dual routers for redundancy or loadsharing
- Offer services for the “common good”
  - Internet portals and search engines
  - DNS TLD, News, NTP servers
  - Routing Registry and Looking Glass

Layer 2 Exchange

- Requires neutral IXP management
  - usually funded equally by IXP participants
  - 24x7 cover, support, value add services
- Secure and neutral location
- Configuration
  - private address space if non-transit and no value add services
  - ISPs require AS, basic IXP does not

Layer 2 Exchange

- Network Security Considerations
  - LAN switch needs to be securely configured
  - Management routers require TACACS+ authentication, vty security
  - IXP services must be behind router(s) with strong filters

Layer 2 versus Layer 3

- Layer 3
  - IXP team requires good BGP knowledge
  - Rely on 3rd party for BGP configuration
  - Less freedom on who peers with whom
  - Usually competes with IXP membership
  - Tends to be distributed over wide area

Layer 2 versus Layer 3

- Layer 2
  - IXP team does not need routing knowledge
  - Easy to get started
  - More complicated to distribute over wide area
  - ISPs free to set up peering agreements with each other as they wish

Layer 2 versus Layer 3 Summary

- Layer 2 is a **REAL** internet exchange point
- Layer 3 is marketing concept used by Transit ISPs
  - **Is NOT a real IXP**

IXP Design Considerations

Exchange Point Design

- The IXP Core is an Ethernet switch
- Has superseded all other types of network devices for an IXP
    - From the cheapest and smallest 12 or 24 port 10/100 switch
    - To the largest 192 port 10GigEthernet switch

Exchange Point Design

- Each ISP participating in the IXP brings a router to the IXP location
- Router needs:
  - One Ethernet port to connect to IXP switch
  - One WAN port to connect to the WAN media leading back to the ISP backbone
  - To be able to run BGP

Exchange Point Design

- IXP switch located in one equipment rack dedicated to IXP
  - Also includes other IXP operational equipment
- Routers from participant ISPs located in neighbouring/adjacent rack(s)
- Copper (UTP) connections made for 10Mbps, 100Mbps or 1Gbps connections
- Fibre used for 10Gbps and 40Gbps

Peering

- Each participant needs to run BGP
  - They need their own AS number
  - **Public** ASN, **NOT** private ASN
- Each participant configures external BGP directly with the other participants in the IXP
  - Peering with all participants

    or
  - Peering with a subset of participants

Peering (more)

- Mandatory Multi-Lateral Peering (MMLP)
  - Each participant is required to peer with every other participant as part of their IXP membership
  - **Has no history of success** — the practice is strongly discouraged
- Multi-Lateral Peering (MLP)
  - Each participant peers with every other participant (usually aided by a Route Server)
- Bi-Lateral Peering
  - Participants set up peering with each other according to their own requirements and business relationships
  - This is the most common situation at IXPs today

Routing

- ISP border routers at the IXP generally should NOT be configured with a default route or carry the full Internet routing table
  - Carrying default or full table means that this router and the ISP network is open to abuse by non-peering IXP members
  - Correct configuration is only to carry routes offered to IXP peers on the IXP peering router
- Note: Some ISPs offer transit across IX fabrics
  - They do so at their own risk – see above

Routing (more)

- ISP border routers at the IXP should not be configured to carry the IXP LAN network within the IGP or iBGP
  - Use next-hop-self BGP concept
- Don’t generate ISP prefix aggregates on IXP peering router
  - If connection from backbone to IXP router goes down, normal BGP failover will then be successful

Address Space

- Some IXPs use private addresses for the IX LAN
  - Public address space means IXP network could be leaked to Internet which may be undesirable
  - Because most ISPs filter RFC1918 address space, this avoids the problem
- Some IXPs use public addresses for the IX LAN
  - Address space available from the RIRs
  - IXP terms of participation often forbid the IX LAN to be carried in the ISP member backbone

AfriNIC Policy on IXP Address Space

- The End-User Assignments policy caters for IXPs Public Address space under Critical infrastructure
- It requires the IXP to be a layer 2 IXP with 3 ISPs connected and have clear and open policy for joining
- The minimum allocation for critical infrastructure is /24 of IPv4 and /48 for IPv6

Hardware

- Try not to mix port speeds
  - if 10Mbps and 100Mbps connections available, terminate on different switches (L2 IXP)
- Don’t mix transports
  - if terminating ATM PVCs and G/F/Ethernet, terminate on different devices
- Insist that IXP participants bring their own router
  - moves buffering problem off the IXP
  - security is responsibility of the ISP, not the IXP

Services Offered

- Services offered should not compete with member ISPs (basic IXP)
  - e.g. web hosting at an IXP is a bad idea unless all members agree to it
- IXP operations should make performance and throughput statistics available to members
  - Use tools such as MRTG to produce IX throughput graphs for member (or public) information

Services to Offer

- ccTLD DNS
  - the country IXP could host the country’s top level DNS
  - e.g. “SE.” TLD is hosted at Netnod IXes in Sweden
  - Offer back up of other country ccTLD DNS
- Root server
  - Anycast instances of I.root-servers.net, F.root-servers.net etc are present at many IXes
- Usenet News
  - Usenet News is high volume
  - could save bandwidth to all IXP members

Services to Offer

- Route Collector
  - Route collector shows the reachability information available at the exchange
  - Technical detail covered later on
- Looking Glass
  - One way of making the Route Collector routes available for global view (e.g. www.traceroute.org)
  - Public or members only access

Services to Offer

- Content Redistribution/Caching
  - For example, Akamised update distribution service
- Network Time Protocol
  - Locate a stratum 1 time source (GPS receiver, atomic clock, etc) at IXP
- Routing Registry
  - Used to register the routing policy of the IXP membership (more later)

Introduction to Route Collectors

What routes are available at the IXP?

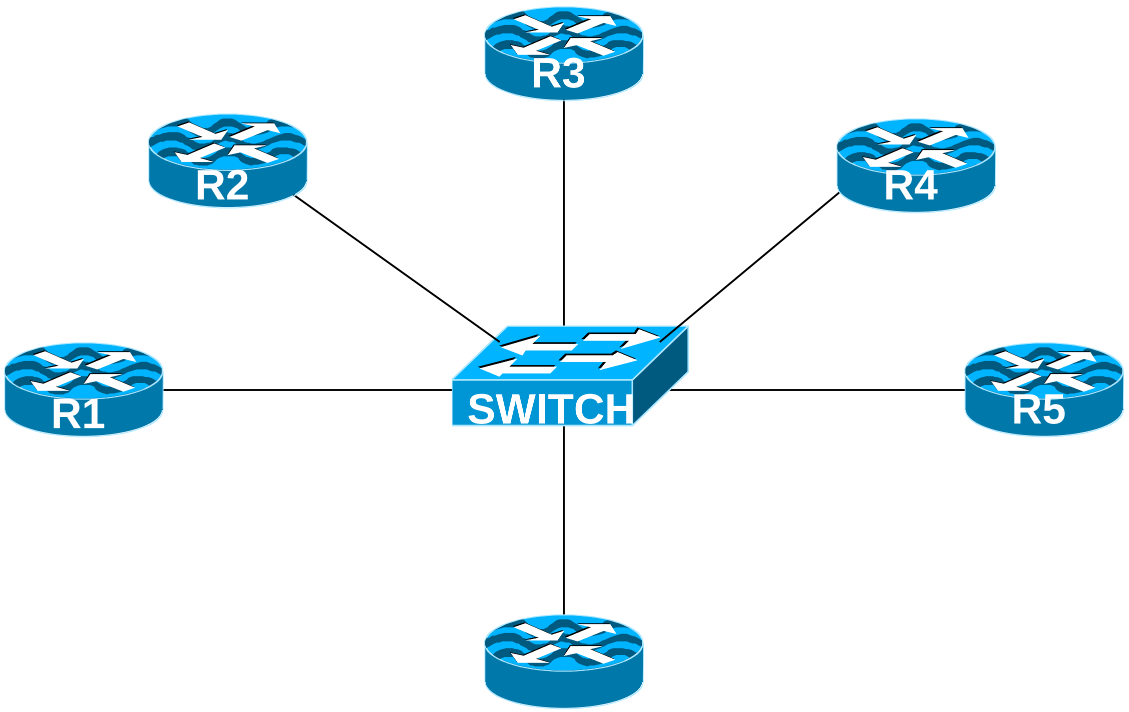
What is a Route Collector?

- Usually a router or Unix system running BGP
- Gathers routing information from service provider routers at an IXP
  - Peers with each ISP using BGP
- Does not forward packets
- Does not announce any prefixes to ISPs

Purpose of a Route Collector

- To provide a public view of the Routing Information available at the IXP
  - Useful for existing members to check functionality of BGP filters
  - Useful for prospective members to check value of joining the IXP
  - Useful for the Internet Operations community for troubleshooting purposes
    - E.g. **www.traceroute.org**

Route Collector at an IXP



Route Collector

Route Collector Requirements

- Router or Unix system running BGP
  - Minimal memory requirements – only holds IXP routes
  - Minimal packet forwarding requirements – doesn’t forward any packets
- Peers eBGP with every IXP member
  - Accepts everything; Gives nothing
  - Uses a private ASN
  - Connects to IXP Transit LAN
- “Back end” connection
  - Second Ethernet globally routed
  - Connection to IXP Website for public access

Route Collector Implementation

- Most IXPs now implement some form of Route Collector
- Benefits already mentioned
- Great public relations tool
- Unsophisticated requirements
  - Just runs BGP

Introduction to Route Servers

How to scale very large IXPs

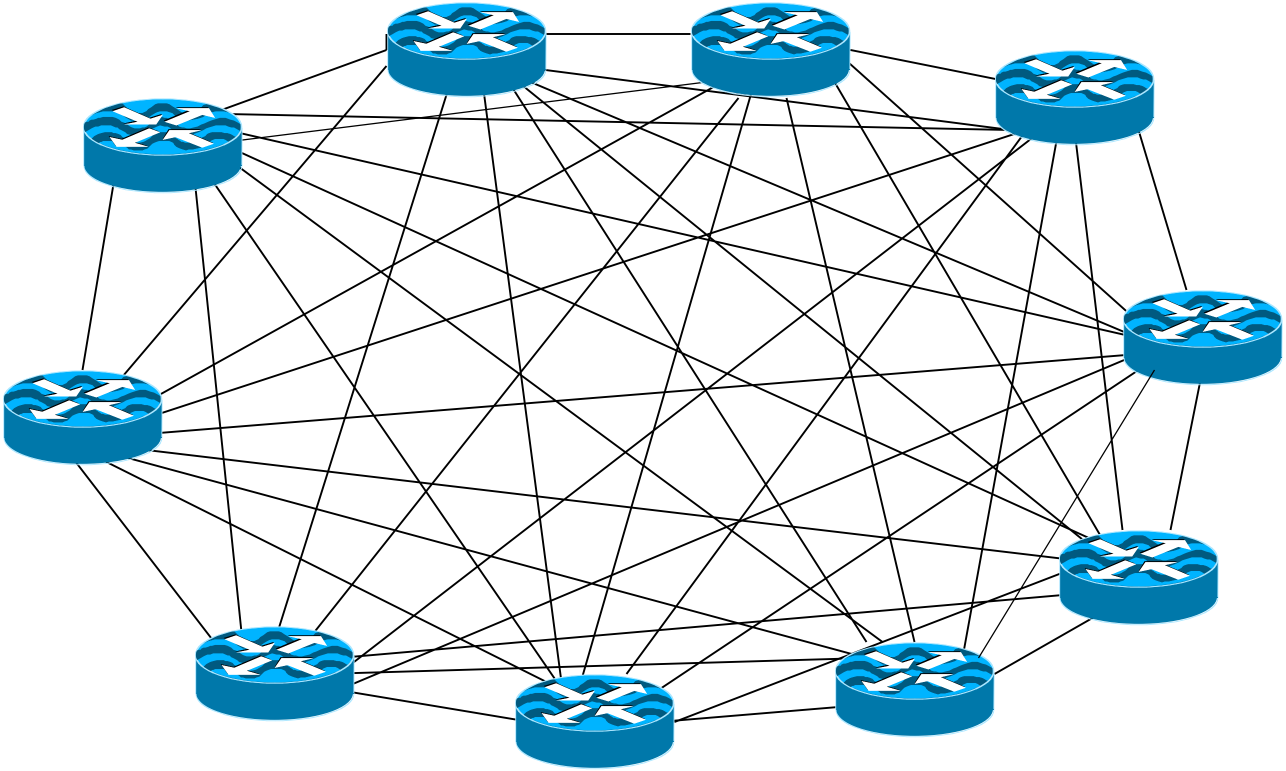
What is a Route Server?

- Has all the features of a Route Collector
- But also:
  - Announces routes to participating IXP members according to their routing policy definitions
- Implemented using the same specification as for a Route Collector

Features of a Route Server

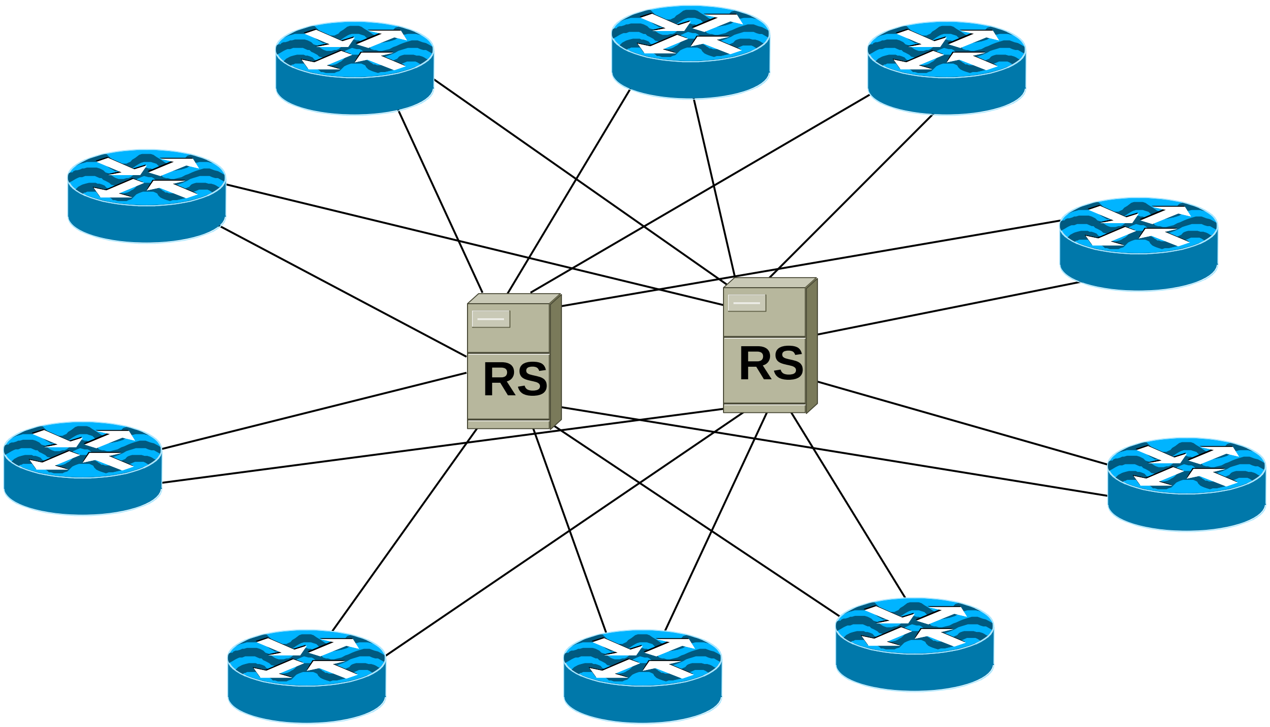
- Helps scale routing for large IXPs
- Simplifies Routing Processes on ISP Routers
- Optional participation
    - Provided as service, is NOT mandatory
- Does result in insertion of RS Autonomous System Number in the Routing Path
- Optionally uses Policy registered in IRR

Diagram of N-squared Peering Mesh



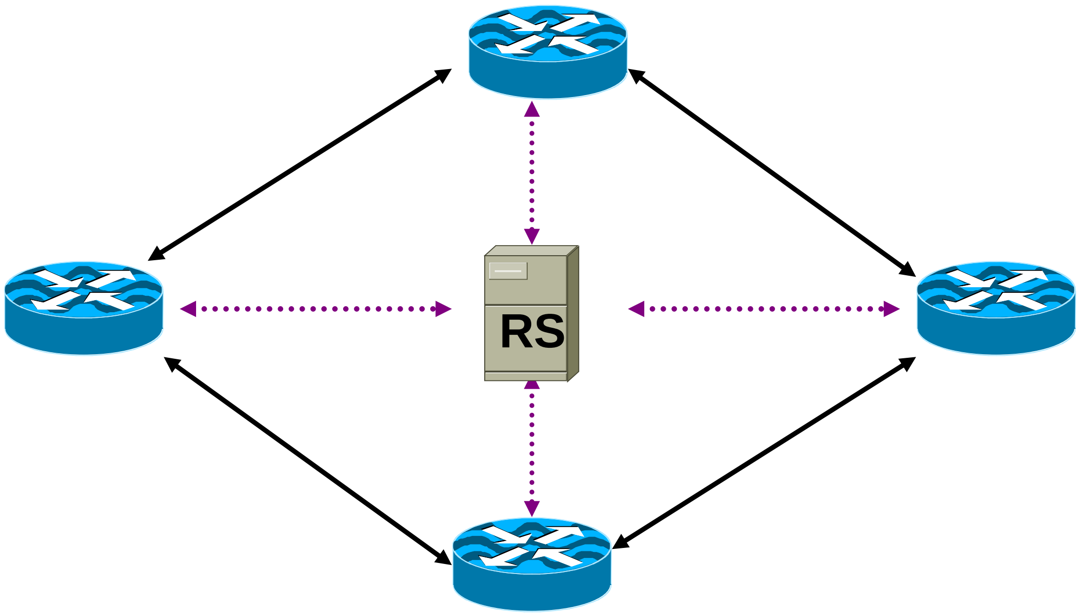
- For large IXPs (dozens for participants) maintaining a larger peering mesh becomes cumbersome and often too hard

Peering Mesh with Route Servers



- ISP routers peer with the Route Servers
  - Only need to have two eBGP sessions rather than N

RS based Exchange Point Routing Flow



↔ TRAFFIC FLOW

◀⋯▶ ROUTING INFORMATION FLOW

Advantages of Using a Route Server

- Helps scale Routing for very large IXPs
- Separation of Routing and Forwarding
- Simplify Routing Configuration Management on ISPs routers

Disadvantages of using a Route Server

- ISPs can lose direct policy control
  - If RS is only peer, ISPs have no control over who their prefixes are distributed to
- Completely dependent on 3rd party
  - Configuration, troubleshooting, etc…
- Insertion of RS ASN into routing path
  - Traffic engineering/multihoming needs more care
- These are major disadvantages
  - Usually out-weigh the advantages

Typical usage of a Route Server

- Route Servers may be provided as an **OPTIONAL** service
  - Most common at large IXPs (>50 participants)
  - Examples: LINX, TorIX, AMS-IX, etc
- ISPs peer:
  - Directly with significant peers
  - With Route Server for the rest

Things to think about...

- Would using a route server benefit you?
  - Helpful when BGP knowledge is limited (but is NOT an excuse not to learn BGP)
  - Avoids having to maintain a large number of eBGP peers
  - But can you afford to lose policy control? (An ISP not in control of their routing policy is what?)

What can go wrong…

The different ways IXP operators harm their IXP…

What can go wrong? Concept

- Some Service Providers attempt to cash in on the reputation of IXPs
- Market Internet transit services as “Internet Exchange Point”
  - “We are exchanging packets with other ISPs, so we are an Internet Exchange Point!”
  - So-called Layer-3 Exchanges — really Internet Transit Providers
  - Router used rather than a Switch
  - Most famous example: SingTelIX

What can go wrong? Competition

- Too many exchange points in one locale
  - Competing exchanges defeats the purpose
- Becomes expensive for ISPs to connect to all of them

- An IXP:
  - is **NOT** a competition
  - is **NOT** a profit making business

What can go wrong? Rules and Restrictions

- IXPs try to compete with their membership
  - Offering services that ISPs would/do offer their customers
- IXPs run as a closed privileged club e.g.:
  - Restrictive membership criteria (closed shop)
- IXPs providing access to end users rather than just Service Providers
- IXPs interfering with ISP business decisions e.g. Mandatory Multi-Lateral Peering

What can go wrong?
Technical Design Errors

- Interconnected IXPs
  - IXP in one location believes it should connect directly to the IXP in another location
  - Who pays for the interconnect?
  - How is traffic metered?
  - Competes with the ISPs who already provide transit between the two locations (who then refuse to join IX, harming the viability of the IX)
  - Metro interconnections work ok (e.g. LINX)

What can go wrong?
Technical Design Errors

- ISPs bridge the IXP LAN back to their offices
  - “We are poor, we can’t afford a router”
  - Financial benefits of connecting to an IXP far outweigh the cost of a router
  - In reality it allows the ISP to connect any devices to the IXP LAN — with disastrous consequences for the security, integrity and reliability of the IXP

What can go wrong? Routing Design Errors

- Route Server implemented from Day One
  - – ISPs have no incentive to learn BGP
  - – Therefore have no incentive to understand peering relationships, peering policies, &c
  - – Entirely dependent on operator of RS for troubleshooting, configuration, reliability
    - RS can’t be run by committee!
- Route Server is to help scale peering at LARGE IXPs

What can go wrong? Routing Design Errors

- iBGP Route Reflector used to distribute prefixes between IXP participants
- Claimed Advantage (1):
  - Participants don’t need to know about or run BGP
- Actually a Disadvantage
  - IXP Operator has to know BGP
  - ISP not knowing BGP is big commercial disadvantage
  - ISPs who would like to have a growing successful business need to be able to multi-home, peer with other ISPs, etc — these activities require BGP

What can go wrong?
Routing Design Errors (cont)

- Route Reflector Claimed Advantage (2):
  - Allows an IXP to be started very quickly
- Fact:
  - IXP is only an Ethernet switch — setting up an iBGP mesh with participants is no quicker than setting up an eBGP mesh

What can go wrong?
Routing Design Errors (cont)

- Route Reflector Claimed Advantage (3):
  - IXP operator has full control over IXP activities
- Actually a Disadvantage
  - ISP participants surrender control of:
    - Their border router; it is located in IXP's AS
    - Their routing and peering policy
  - IXP operator is single point of failure
    - If they aren't available 24x7, then neither is the IXP
    - BGP configuration errors by IXP operator have real impacts on ISP operations

What can go wrong? Routing Design Errors (cont)

- Route Reflector Disadvantage (4):
  - – Migration from Route Reflector to “correct” routing configuration is highly non-trivial
  - – ISP router is in IXP’s ASN
    - Need to move ISP router from IXP’s ASN to the ISP’s ASN
    - Need to reconfigure BGP on ISP router, add to ISP’s IGP and iBGP mesh, and set up eBGP with IXP participants and/or the IXP Route Server

More Information

Exchange Point Policies & Politics

- AUPs
  - Acceptable Use Policy
  - Minimal rules for connection
- Fees?
  - Some IXPs charge no fee
  - Other IXPs charge cost recovery
  - A few IXPs are commercial
- Nobody is obliged to peer
  - Agreements left to ISPs, not mandated by IXP

Exchange Point etiquette

- Don’t point default route at another IXP participant
- Be aware of third-party next-hop
- Only announce your aggregate routes
  - Read RIPE-399 first

    www.ripe.net/docs/ripe-399.html
- Filter! Filter! Filter!
  - And do reverse path check

Exchange Point Examples

- LINX in London, UK
- TorIX in Toronto, Canada
- AMS-IX in Amsterdam, Netherlands
- SIX in Seattle, Washington, US
- PA-IX in Palo Alto, California, US
- JPNAP in Tokyo, Japan
- DE-CIX in Frankfurt, Germany
- HK-IX in Hong Kong

…

- All use Ethernet Switches

Features of IXPs (1)

- Redundancy & Reliability
  - Multiple switches, UPS
- Support
  - NOC to provide 24x7 support for problems at the exchange
- DNS, Route Collector, Content & NTP servers
  - ccTLD & root servers
  - Content redistribution systems such as Akamai
  - Route Collector – Routing Table view

Features of IXPs (2)

- Location
  - – neutral co-location facilities
- Address space
  - – Peering LAN
- AS Number
  - – If using Route Collector/Server
- Route servers (optional, for larger IXPs)
- Statistics
  - – Traffic data – for membership

More info about IXPs

- http://www.pch.net/documents
  - Another excellent resource of IXP locations, papers, IXP statistics, etc
- http://www.telegeography.com/ee/ix/index.php
  - A collection of IXPs and interconnect points for ISPs

Summary

- L2 IXP – most commonly deployed
  - The core is an ethernet switch
  - ATM and other old technologies are obsolete
- L3 IXP – nowadays is a marketing concept used by wholesale ISPs
  - Does not offer the same flexibility as L2
  - Not recommended unless there are overriding regulatory or political reasons to do so
  - **Avoid!**

Acknowledgement and Attribution

This presentation contains content and information originally developed and maintained by the following organisation(s)/individual(s) and provided for the African Union AXIS Project

Cisco ISP/IXP Workshops

Philip Smith: - pfsinoz@gmail.com

APNIC **www.apnic.net**

GRAND DUCHY OF LUXEMBOURG
Ministry of Foreign Affairs

Directorate for Development Cooperation

European Union Africa
Infrastructure Trust Fund

Internet Exchange Point Design

End

Internet Society™