ISP SERVICES AND PEERING

TCOM610-BGP
Tawfiq Khan
George Mason University

ISP Services

- Dedicated Internet Access
  - Speeds from 56 kbps to OC12 and OC48
  - Predictable access, always available
  - Most expensive access option
- Frame Relay and ATM Access
  - Greater range of access speeds due to statistical multiplexing and therefore pricing options
  - Either statically provisioned (mostly for primary access) or dynamically allocated (often for backup)
- VPN Services
  - Virtual private network connecting office and branches in different locations

Dial-up Service

- Connectivity on demand
  - Traditional POTS lines
  - Limited speed, up to 56 kb/s
- ISDN
  - BRI at 128 kb/s and PRI at 1.5 Mb/s
  - Expensive and no longer offered for residential customers (still available for small business)
- Wholesale Dial
  - Offered on behalf of other dial networks like AOL
  - Provide modem port, aggregate and deliver traffic to designated facility

Broadband: DSL and CM

- DSL and variants
  - Uses existing phone copper wiring
  - Typically 384 kb/s up, 1.5 Mb/s down
  - Higher speeds available now
  - Distance from central office to site is critical
- Cable modems
  - Uses cable TV coax cable, often with optical transport to neighborhood (hybrid fiber/coax – HFC)
  - Typically 500 kb/s up, 2 Mb/s down
  - DOCSIS standards for modem technology
  - Combined with other services like VoIP

Dedicated Hosting Services

- Hosting sites
  - Large, reliable server farms
  - High speed internal networks
  - High speed external access
  - Secure physical locations
- Sites located at convenient locations on the provider network

Other Services

- Email and news
- Web service and database (ASP)
- Virtual Private Networks (VPNs)
- IP multicast
- Managed Security and Firewall

ISP Service Pricing

- Highly variable pricing
  - Often due to available infrastructure
  - Package features also vary
- Bundled packages for multiple sites often advantageous because of service consistency
- Usually billed based on combined 95th usage of all ports of higher direction of inbound or outbound
  - May require per port charge for small traffic volume
  - Avoid per port charge by minimum committed usage
  - Prices / MBPS has dropped significantly over last a few years (300 Mbps => 10 Mbps/month)

SLA

- Service Level Agreements (SLAs) or Service Level Guarantees (SLGs)
- Specify performance and availability levels
  - Details of specification and penalties
  - Monitoring and reporting responsibilities
  - If SLA is violated, customer may get service credit
- Typically include measures such as packet loss, delay, availability, maintenance/outage notification requirements
  - Typically: North-America < 55 ms delay, < 0.03% packet loss, 99.9% delivery rate or availability

Provider Selection

- Physical network topology
  - Network bottlenecks and subscription ratios
  - Level of network and element redundancy
  - Peerings with other networks (Are they Tier1?)
- Distances
- Traffic agreements
- Physical Connections – Need network map with connections
- Consistent and adequate bandwidth for traffic paths, with redundancy
- High speed backbone required
- High speed links to other providers needed (peering)

ISP Bottlenecks

- Price, Network Connectivity, and Network Redundancy
- Networks defined by the weakest links
  - Over-subscription of critical resources
    - 5:1 ratio a typical target, but 8:1 or 10:1 is acceptable
  - Small tail circuits to POPs or customers
    - Critical resources (e.g., server facilities) need to be located on high bandwidth infrastructure
- ISP connectivity: Tier1/Tier2 – how many AS away from major networks and their customers?
- ISP's Monitoring extremely important

Network Redundancy

- Network Redundancy
  - Redundancy necessary to enable alternate routes despite link failures and direct or NAP peering problems
  - Multiple POP connections and peerings desirable
  - Usually provided via global redundancy, rather than per site
- Sparing plan important
  - On-site critical component supply
  - Manage the spares: spare ratio
  - Depends on components and MTBF (Mean Time Between Failures)
- Sparing often done in concert with vendor
  - Increases MTTR (Mean Time To Repair) but often acceptable and economical

Distance and Traffic Agreement

- Router hops are significant but link speeds, network technologies, and geography are also important
- Number of autonomous systems traversed
  - Affected by direct peerings and NAP peerings
  - Reasonable measure of performance since losses and delays often happen at boundaries
- Larger ISPs choose peerings carefully for direct and NAP interconnections
  - smaller ISPs are often customers of large ISPs and pay for transit
- Usually aggregate capacity and selected peering policy information available for review by customers

Transit vs Peering

- Transit
  - Use your ISP to get to other networks – not just ISP's customers, but whole internet
  - Receive full Internet route, partial routes plus default or default only from ISP
  - Pay ISP monthly based on 95th of the total Inbound or Outbound traffic volume
  - Usually cold-potato routing from ISP and not necessarily consistent routing
  - Send traffic to your ISP with traffic destined to anywhere in Internet
- Peering
  - Only receive other ISP's customer routes ONLY
  - May not use other network to transit to other networks
  - Two flavors: Settlement-free peering or paid peering ($$)
  - Usually require consistent route announcement, traffic volume and ratio requirements, hot-potato routing, NOC support and similar network size

Transit and Peering

- Category of Networks
  - Content networks: Yahoo, Microsoft, Google
  - Eyeball networks: MSOs (Comcast, Cox, TWC etc), RBOCs (Verizon, SBC, Southwest Bell, USWest etc), AOL, MS-MSN
  - Based on classic client-server model, but in new Peer2Peer domain, everyone is client AND server
  - New broadband changes the landscape and the categories of the networks
- Peering Requirements
  - Network Size and Geographical Distributions: OC48 US backbone with at least 15 major cities
  - Traffic Volume: at least 1 Gbps traffic exchange
  - Traffic Ratio: usually 1:2 or 1:1.5 (balanced in/out traffic)
  - Operations: 7x24 NOC, AUP, Anti-SPAM etc

Verizon Business SF Peering Requirements

- Geographic Scope: 50% of its regions (25 states)
- Traffic Exchange Ratio: < 1.8:1
- Backbone Capacity: OC192 in US, Oc48 in EU, OC12 in AS
- Traffic Volume: 1.5 Gbps in US, 150 Mbps in EU, 30 Mbps in AS
- Transit Autonomous Systems: 1500 Unique AS in US, 100 Unique AS in EU, and 10 Unique AS in AS
- Operations Reqs: 7x24 NOC, cold potato routing, consistent route announcements, redundancy for failure, response for security breach or abuse

Peering

- History of peering
  - Initial a few separate networks interconnected in some public peering points
  - Later NSF MAE public peering facility
  - Due to rapid growth, public peering become bottlenecks of network performance
  - Then the beginning of direct or private peering
- Public Peering
  - Locations: MAE-EAST, MAE-WEST, PAC-BELL, Ameriwest
  - Peer with shared media such as FDDI ring, FE/GE Ethernet
  - Easy to peer with all parties in the facility
  - Still important internationally, but not in USA
- Private Peering
  - Two networks connects with dedicated link in multiple locations
  - Exchange network’s customer routes over dedicated link
  - For better performance and usually settlement-free

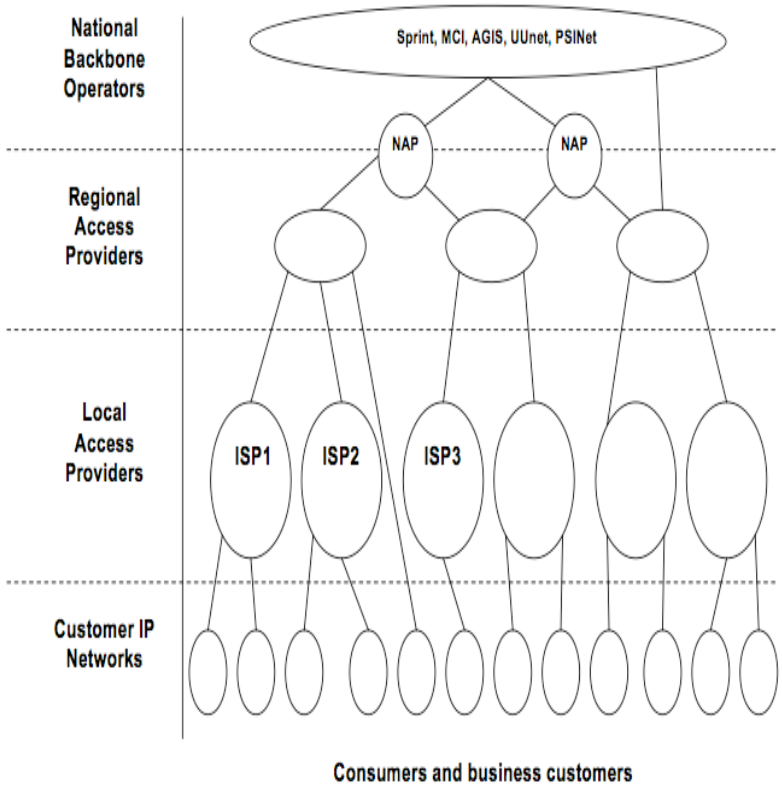
Peering

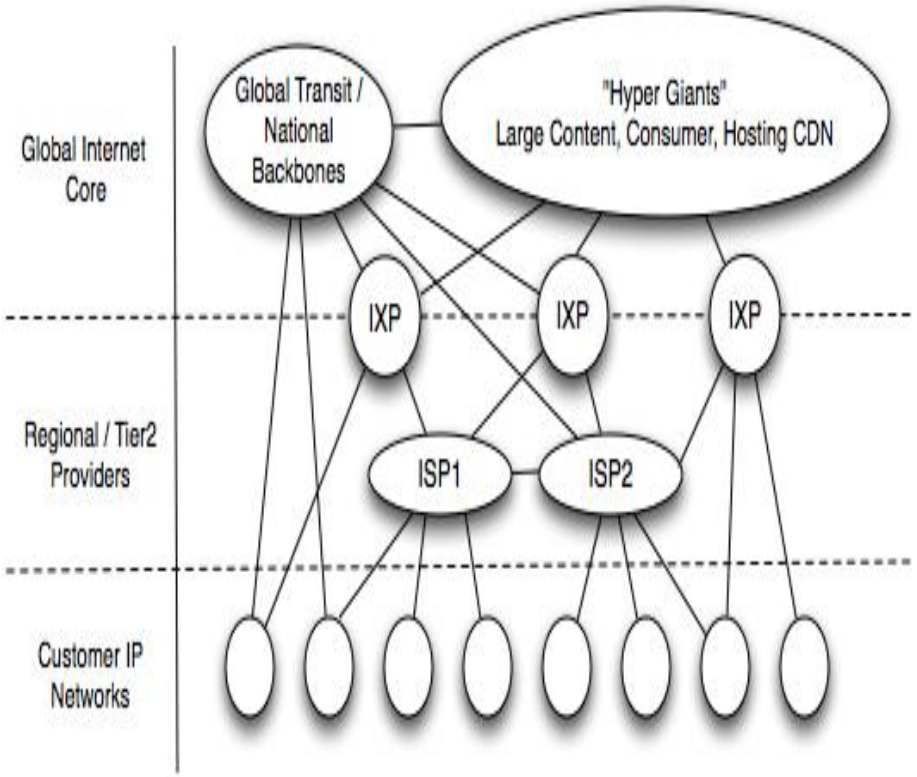
- US primary peering locations
  - Eight locations: Seattle, San Jose, Los Angle, Dallas, Chicago, New York, Washington DC, and Atlanta
- Equinix and PAIX Facilities
  - All Tier1 and major networks in the same building
- Tier1
  - Networks that are transit-free and default-free
  - Historically US-centric criteria
  - Includes: att/SBC, Sprint, Verizon/Uunet, Level3, Qwest, ATDN, and possibly Savvis, Verio, Global Crossing
  - Directly peer with each other to get whole Internet routes
  - It will evolve as Asia-Pacific and Europe become more important
- Tier2
  - At least buy transit from one of the tier1 ISP
  - May be combination of settlement-free peering, paid peering and limited transit
  - MSO, RBOC, Regional/New Networks,

Peering Evolution

- Dan Golding: Peering Evolution
  - http://www.nanog.org/mtg-0302/golding.html
- Bill Norton: Peering Ecosystems
  - http://www.nanog.org/mtg-0405/pdf/norton.pdf

Internet Architecture Evolution





ISP Service BGP Community

- RFC1998 – BGP Community Application
- ISP may use BGP community to implement their customer BGP policy
  - No need to customize routing policy per customer
  - Standardize ISP policy implementation
  - No operator involved if customers want to implement different policy
  - More granulates on BGP policy
    - Could be per prefix rather than per AS
  - Facilitate the easy implementation of highly sophisticated routing policy
    - Move away from simple model of ISP routing policy

ISP BGP Community

- Typical application for BGP community
  - Change local-pref for customer routes
    - Lower or raise your route's local-pref within ISP
    - May or may not allow customer to lower than peer route's local-pref
  - Suppress route announcement to other networks
    - Don't want specific peer of ISP to see its route
    - Customer controlled partial transit or paid peering
  - Control AS-PATH prepend
    - Could be done on per peer base and for limited subset of prefix
  - Customer triggered remote blackhole
    - Drop attack traffic at ISP border
    - Enable by customer at real-time

Example Policy (AS1000)

- Change local-preference value
  - default: customer local-pref=300, peer local-pref=100
  - 1000:250 Set local preference of routes to 250
  - 1000:350 Set local preference of routes to 350
- Prepend to ISP's peers
  - 65001:ASN Prepend once to peer ASN
  - 65001:0 Prepend once to all peers
  - 65002:ASN Prepend twice to ASN
  - 65002:0 Prepend twice to all peers
  - 65003:ASN Prepend thrice to ASN
  - 65003:0 Prepend thrice to all peers

Example Policy (AS1000)

- Suppress/Leak routes to peers:
  - 65000:0 Suppress route to all peers
  - 65000:ASN Suppress route to ASN
  - 65111:ASN Leak route to ASN (that was otherwise suppressed)
- Customer Triggered Blackhole
  - 1000:9999 Blackhole the prefix – must be more specific than /27

Example Implementation

```
route-map FR-20001-1 permit 10
 match ip address prefix-list FR-20001-1
 match community LOCALPREF250
 set local-preference 250
 set com-list OURCOMS delete
 set community 1000:11000 additive
!
route-map FR-20001-1 permit 20
 match ip address prefix-list FR-20001-1
 match community LOCALPREF350
 set local-preference 350
 set com-list OURCOMS delete
 set community 1000:11000 additive
!
route-map FR-20001-1 permit 30
 match ip address prefix-list FR-20001-1
 set local-preference 300
 set com-list OURCOMS delete
 set community 1000:11000 additive
!
ip community-list expanded LOCALPREF250 permit _1000:250_
ip community-list expanded LOCALPREF350 permit _1000:350_
ip community-list expanded OURCOMS permit _1000:.*_
```

Exercise

- Implement AS-PREPEND ISP community policy using Cisco IOS
- Implement community-based selective route-leaking BGP policy
- Implement customer triggered blackhole BGP policy

Troubleshooting BGP

- Download Cisco BGP troubleshooting guide from NANOG site: http://www.nanog.org/mtg-0310/pdf/smith.pdf
- This Cisco guides provide lots of examples and detailed cases
- First on bringing up BGP session
  - Configured with same parameters? (AS, IP Address, BGP version etc)
  - Have direct IP connectivity to other side?
  - If peered with loopback, ebgp-multihop?
  - eBGP vs iBGP? TTL?
  - Is there layer 2 MTU issue?
  - Is there ACL or firewall in between?
  - MD5 password match?
  - Show ip bgp sum | show log
  - Show tcp all brief
  - Debug ip bgp | debug ip tcp transactions
  - Debug ip bgp update ACL#

Troubleshooting BGP

- Then on route announcements
  - Originate your routes in BGP?
  - Redistribute routes into BGP?
  - Check your BGP table and IP routing table
    - Sh ip route x.x.x.x, sh ip bgp x.x.x.x
  - Check your policy and BGP announcements
    - Show ip bgp neighbor x.x.x.x received-routes | advertised-routes | routes
  - Only best BGP routes are exchanged
  - Configure send-community? (no sent by default)
  - Clear BGP session (soft-reconfig or route-refresh)
  - Check syntax on your policy
    - Regex, route-map, multiple match, default deny etc

Missing Routes

- Customer see partial routes – must be IBGP full mesh problem
  - Better to use peer-group and RR
- Inconsistent route selection: MED problem – cause inconsistent route selection, not converging
  - Deterministic MED – Order all paths from the same AS first
  - BGP best-path compare-router-id to ignore time-dependency
- Route not marked for not-synchronized; will not use unless synchronized
- Persistent route oscillation problem: next hop must NOT learn from BGP so that you have to advertise your next-hop reachability through IGP;
- Maximum-paths tells router to reset next-hop to itself – may cause problem in confederation

Route-view Servers

- http://www.routeviews.org
- To obtain real-time information about the global routing system
  - Peer with all major providers
  - Free read access to all peer's routing tables
  - Historic archive of all BGP routes and events
    - Become major data resources for BGP academic study
  - Several BGP analysis tools based on route-view data
    - CIDR report, BGP dynamics, Skitter etc
  - Important tool when troubleshooting how other network see your routes
- Other provider-only route servers with limited commands
- Major route servers listed under http://www.traceroute.org and http://www.bgp4.as

Looking Glass

- Similar to route-view, to provide user capability to verify how other networks route
- Allow you to troubleshoot without calling your ISP
- Typically implemented by providers to facilitate customer or peers routing troubleshooting
  - Usually in different locations of their networks
- Run different commands
  - Ping, traceroute, show ip bgp
- Best source: http://www.traceroute.org
- Use Cisco troubleshooting guide in NANOG archive!