IP ROUTING AND ADDRESSING

Tawfiq Khan
TCOM610
George Mason University

Internet History

- Started in 1960s by ARPA: Advanced Research Projects Agency of DoD
- Small number of sites linked by 56 kb/s lines
- ARPANET: connect research, academic, and government
  - Statically provisioned
  - AUP: no commercial use
  - Scalability issues began to appear
- Decommissioned in 1989

Internet History - NSFNet

- Started in 1986 by NSF
- Core backbone with regional networks and peers
- Link speeds evolve to T1 (1.5Mbps) then T3 (45 Mbps)
- Connect campuses and research organization (SCC)
- Routing protocols based around tree Structure
- Start the public peering project:
  - Federal Internet eXchange (FIX) points on East & West Coasts
  - Commercial Internet eXchange (CIX) on West Coast
  - Commercial ISP emerges
- Decommissioned in 1995 – end of NSFNet, beginning of HTML/HTTP/Mosaic

Internet Today

- Many National and multi-national Backbone
- Some tier-1: Verizon/MCI, Sprint, ATT/BellSouth, Level3, ATDN, Qwest, CW/Savvis, Abovenet, Teleglobe etc
- Distributed architecture with scalable, dynamic routing protocols
- OC48 to multiple OC192 Speed (10Gbps), 10GE
- Internet: network of networks interconnected at NAP and private peering

NAP and Direct Interconnect

- NAP: Network Access Point, where networks interconnect and exchange traffic
- Initially on 100 Mbps FDDI ring, or ATM switch; Later FE and GE
- Faced scalability and capacity problem; Become performance bottleneck
- New direct interconnection: direct links between two networks
- Very popular among commercial ISPs

NSFNet and vBNS

- NSF contracted MCI for 5 yrs high performance and high speed VBNS network
- Connect research university and NSF SCC
- Focus on high performance computing users
  - Evolved from OC-3 to OC48 speeds
  - Connects to commodity Internet at NAPs
  - Early deployment of advanced services
  - Original funding ended in 2000, but still in existence
  - Evolve into secure Internet access and VPN platform

Internet2 and Abilene

- Internet2: membership over 200 universities
- Provides interconnection architecture for research university networks
- Foster growth of advanced Internet application and networking protocols
- Facilitated leading edge capacity of test bed networks
- Projects: HOPI, E2EPI, Middleware, Security, SIP etc.
- http://www.internet2.edu/
- **“Internet2 develops and deploys advanced network applications and technologies for research and higher education, accelerating the creation of tomorrow's Internet“**

Internet Growth

- Growth of the Internet has had an impact on many architectural issues
- Routing has becoming increasingly challenging
  - Number of autonomous organizations
  - Structure of organizations
    - multiple sites and multiple homed
- How to measure Internet size?
  - Domain name registration, web servers, Internet host address
  - Backbone connection, BGP routes, AS number

Acronym

- IANA: allocation and assignment of various numeric such as address family, protocol number, port etc.
- IR: Internet registry for domain name, AS
- ARIN: IP number and ASN allocation for America
- IRR: Internet routing registry, where you register your routes, Autonomous System number, network administrators and routing policies

Basic Internet Routing

- All hosts and router interfaces are uniquely identified by 32 bits IPv4 (or 128 bites IPv6) address
- 32 bit address may be divided into network part and host part
- Most hosts are default routed (single default route out) and directly connected to switches
- Servers may have multiple connections with multiple defaults or one HSRP default out
- How about return traffic? All communication requires two-way reach ability
  - Router need to announce host routes to outside world

DNS

- DNS is the directory for Internet
  - Access Internet services by name instead of address
- Allows the separation of service and IP address
- If DNS is down or is un-resolvable, people won't be able to locate the service
  - Internet yellow page
- Every services start with a DNS name
- DNS service need to be redundant, reliable, fast and secure
  - Sustain DOS/DDOS attacks

Internet Routing

- Routing
  - Take destination address and do longest prefix match (BGP route with next-hop)
  - Do recursive look-up on next-hop using IGP routing table
  - Find the directly connected interface to forward traffic
  - Do ARP to get Layer 2 address for that IP
  - Forward the packet with layer 2 encapsulation
  - Analogy: Post-mail among different countries without ship and cars (no airplane)

Destination-base hop-by-hop

- Routing is done on routers with Layer 3 IP information (sip, sport, dip, dport, proto)
- Typically most routing are done solely based on destination IP address
- Special case: policy based routing, source routing, tunnel etc
- Internet routing is hop-by-hop
  - Each router perform its decision independently
- As packets travel over the Internet, its physical/layer2 address changes in each hop, but the layer3 And above remain the same

Hop-by-hop routing

- Routers have to work together to make sure a packet will be eventually delivered to its destination, regardless of many intermediate router's independent decision
  - Through routing protocol standards
- Different types of routing protocol
  - Static vs dynamic, link state vs distance vector
  - Reliable, loop-free, fast convergence protocols
  - IGP (OSPF, ISIS, EIGRP), BGP4, MPLS
- Two packets from the same host to the same destination may travel different routes
  - Equal-cost multiple paths, load-balancing
- Internet may not be reliable: packets may get out of order or get lost
  - Rely on layer 4 or higher for reordering or retransmission
- Use ping to verify reach ability and traceroute to see the packet path

Internet Routing Verification

- DNS lookup
- Ping
- Traceroute
- Route server
- Internet looking glass

IP ADDRESSING

Basic Addressing

- Notation
  - Address is 32 bits long
  - Representations: 2179825081, 0x81ED7DB9, or 129.237.125.185
  - Comprised of both Network and Host number
- Originally network part was first 8 bits
  - 1981 – Classful routing – RFC 791
  - 5 Classes: Class A, Class B, Class C, Class D, Class E

Classful Addressing

- Class A
  - 1.0.0.0/8 through 126.0.0.0/8 (1st octet starts with 0)
  - Originally for very large organizations
  - Now being divided for use by service providers
- Class B
  - 128.0.0.0/16 through 191.255.0.0/16 (1st octet starts with 10)
  - Originally for midsize organizations such as universities and regional networks

Classful Addressing

- Class C
  - 192.0.0.0/24 through 223.255.255.0/24 (1st octets starts with 110)
  - Originally for small organizations
- Class D
  - 224.0.0.0/24 through 239.255.255.0/24 (1st octets starts with 111)
  - Reserved for multicast services
- Class E
  - 240.0.0.0/24 through 255.255.255.0/24
  - Reserved for experimental use

Address Mask

- Address Masks are used to distinguish the bits for network and for host
- Notation: 1 bit to indicate network part
  - 4294967280, FFFFFFF0 or 255.255.255.240, or /28
  - Refer to the inner cover of the textbook
- Example: 10.0.0.0/8, mask: 255.0.0.0 (10 part is the network, rest is for hosts)

Subnetting

- Dividing a single class A, B or C into smaller sub-networks
  - Allowed more effective utilization of Class A and B address spaces
- Network size – /8s and /16s had an unrealistic number of host addresses
- Dividing Class C address space into
  - 2 subnets with 128 hosts (/25, 255.255.255.128)
  - 4 subnets with 64 hosts (/26, 255.255.255.192)
  - 8 subnets with 32 hosts (/27, 255.255.255.224)
  - …
  - 64 subnets with 4 hosts (/30, 255.255.255.252)
- Now whole 32 bits are divided into: network, subnet, and host portions
- Early-day router does not carry prefix info

Sub-netting Example

- You are allocated 192.168.200.0/24, How would you subnet to support three networks having 50 hosts each?
- 2 bits creates 4 subnets (we have to use the zero net)
  - 192.168.200.0/26
  - 192.168.200.64/26
  - 192.168.200.128/26
  - 192.168.200.192/26
- Each subnet has 6 bits for hosts - 62 hosts
- If you have 192.168.200.0/24 and you need to support 3 networks
  - net-a has 75 hosts
  - net-b has 30 hosts
  - net-c has 30 hosts
  - How can you do that?
- Use mask conversion table to quickly convert network mask and the number of hosts covered

Variable Length Subnet Masks (VLSM)

- Organizations were locked into a fixed size subnet for each classful address
- Allowed the pie to be divided differently
- Routing protocols had to be modified to carry prefix information
- Best match needed - multiple matching routes
- Subnets need to be sized appropriately for the particular network size
- Class C variable sub-netting example
  - One subnet with 128 hosts
  - Two subnets with 64 hosts
  - 192.214.11.x: 192.214.11.0/25 for 128 hosts, 192.214.11.128/26 and 192.214.11.192/26 for 64 hosts

VLSM Example

- You have 192.168.200.0/24, and you need to support 3 networks
  - net-a has 75 hosts
  - net-b has 30 hosts
  - net-c has 30 hosts
- What will the subnets be? 6 bits for host to support 62 hosts (excluding all 1s and all 0s), 5 bits for 126 hosts
- First divide 192.168.200.0/24 into two nets
  - 192.168.200.0/25 - supporting 126 hosts
  - 192.168.200.128/25 - supporting 126 hosts
- Then divide second net into two nets
  - 192.168.200.128/26 - supporting 62 hosts
  - 192.168.200.192/26 - supporting 62 hosts

IP Address Space Depletion

- Only 32 bits of address space
- Class boundaries were easy to understand and decode but they were inefficient
  - /16s were going fast
  - /24s were filling up the routing tables
- Multiple techniques deployed to address the issue
  - 1985 – Subnetting – RFC 950
  - 1987 – VLSM – RFC 1009
  - 1992 – CIDR – RFC 1518
  - NAT: RFC1918 private address inside and public address outside
  - IPv6 – new 128 bit address
- Need flexibility to take advantage of the remaining available IPv4 address space

CIDR

- Classless Inter-Domain Routing
  - 1993 - RFCs 1517, 1518, 1519, 1520
  - No classful routes - everything must have a prefix and a mask
  - Support route aggregation and hierarchical Internet architecture
- All prefixes accompanied by mask: 198.32.0.0/13 (synonymous to” 198.32.0.0 255.248.0.0) and no octet boundary
- Like VLSM, but stronger and more flexible – allowed address space to be allocated and distributed differently, efficiently and hierarchically across organizations
- Three prerequisites
  - Prefix in routing protocols
  - Longest match based forwarding
  - Topological significance addresses assignment

CIDR Issues

- Routing and aggregating with CIDR
  - New routing protocols needed
  - New behavior needed at routers (longest match)
  - Potential loop if most specific route is not followed
- Implications
  - Multi-homed customers reduce efficiency
  - Need to pin up aggregate and point to Null0
  - Aggregates require non-portable address space inherited from provider
  - Renumbering is required if changing providers
- CIDR provides for routing table efficiency by allowing fewer routes to be advertised upstream

CIDR

- Supernet: 198.32.0.0/16 to 198.32.1.0/24 (shorter prefix length)
- Aggregate: lump all continuous more specific into one supernet (8 contiguous /24 into /21)
- More Specific: relative to summary or aggregate: 198.32.0.0/20 is more specific of 192.32.0.0/16
- ARIN gets one 208.0.0.0/8 block, allocates 208.2.0.0/16 to one ISP, that ISP further divides that into various /21 for tier2 ISP, and /24 for smaller customers. The Internet will only see one /16 advertisement from that ISP (ISP still see multiple prefix including supernet and more specifics in their own routing table)

CIDR Examples

- List the /24s in CIDR block 204.24.104/21
  - Note that there are 3 bits available to subnet, and therefore we can do 8 /24s
  - 204.24.104.0/24
  - 204.24.105.0/24
  - ...
  - 204.24.110.0/24
  - 204.24.111.0/24
- List the /21s defined by 152.13.160.0/19
  - Note that 2 bits are available and therefore we will be able to have 4 /21s
  - The stride of a /21 is 8 (binary 1000)
  - 152.13.160.0/21
  - 152.13.168.0/21
  - 152.13.176.0/21
  - 152.13.184.0/21

Longest Match

- Following most specific route in routing
- Example
  - 10.0.0.0/8 Port 1
  - 10.1.0.0/16 Port 2
  - 10.1.1.0/24 Port 3
- Where do we send 10.1.0.1?
- Where do we send 10.1.1.1?
- Where do we send 10.2.1.1?
- Answer:
- Port 2
- Port 3
- Port 1

Aggregation

- Identifying aggregates
- Example
  - 152.13.160.0/21
  - 152.13.168.0/21
  - 152.13.176.0/21
  - 152.13.184.0/21
  - All of these share the same leading 19 bits
- Look at the bits that don’t change (101 in 3rd octet)
- Need to have all specific in the aggregate – only aggregate your own address space
- You get: 152.13.160.0/19

Route Aggregation

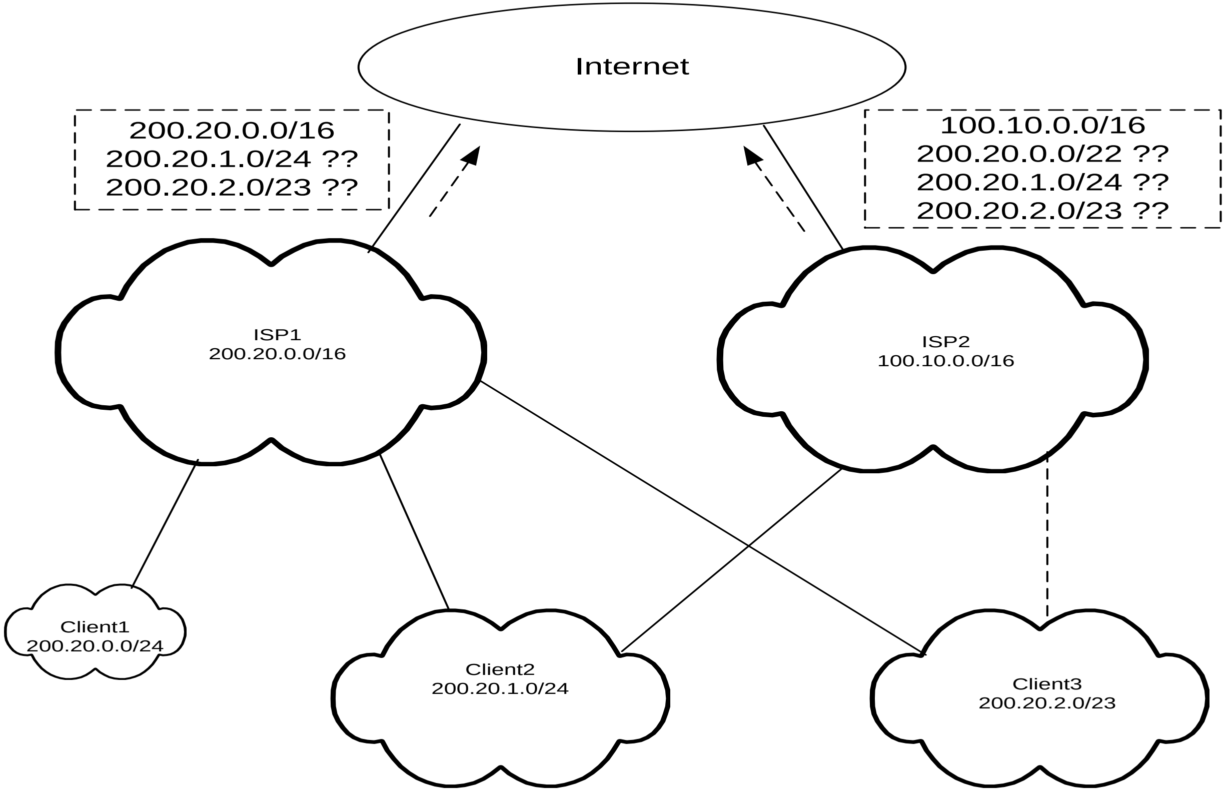
- Must carefully handle the less specific routes of an aggregate
  - If specific destination becomes unavailable and a default route exists, a routing loop can result
  - Particularly problematic in the case of multi-homed networks with address space from a single provider
- Address space is typically inherited from providers, particularly in the case of enterprise customers and smaller providers
  - Providers need to carefully advertise routes for multi-homed customers
- Addresses may be obtained from different providers
  - Will not support redundancy if routes only advertised with single provider
  - Multiple-homed connection typically result in advertisement of aggregate and more specific
- Addresses may be obtained independently of providers, if large enough
  - Will support redundancy but requires specific routes in global routing tables

Aggregation Example

- Aggregate the following prefixes
  - 150.3.82.0/24, … 0101 0010 | …
  - 150.3.83.0/24, … 0101 0011 | …
  - 150.3.84.0/24, … 0101 0100 | …
  - 150.3.85.0/24, … 0101 0101 | …
- Answer
  - 150.3.82.0/23, … 0101 001 | 0 …
  - 150.3.84.0/23, … 0101 010 | 0 …
  - Can you aggregate further? (/22)

Aggregation Example

- What prefixes do you need to aggregate it out to a /21?
  - 150.3.80.0/23, … 0101 000 | 0 …
  - 150.3.86.0/23, … 0101 011 | 0 …
- Answer
  - From 150.3.80.0/23, … 0101 0 | 000 …
  - To 150.3.87.0/23, … 0101 0 | 111 …


Internet
200.20.0.0/16
200.20.1.0/24 ??
200.20.2.0/23 ??
100.10.0.0/16
200.20.0.0/22 ??
200.20.1.0/24 ??
200.20.2.0/23 ??
ISP1
200.20.0.0/16
ISP2
100.10.0.0/16
Client1
200.20.0.0/24
Client2
200.20.1.0/24
Client3
200.20.2.0/23