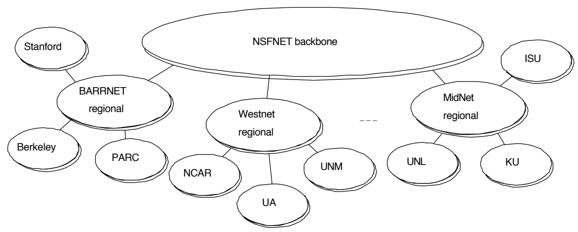
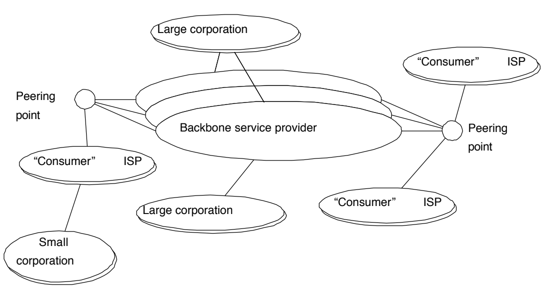
Internet Structure - Past



Internet Structure - Today



Subnetting

- Add another level to address/routing hierarchy: subnet
- Subnet masks define variable partition of host part
- Subnets visible only within site

Network number	Host number

Class B address

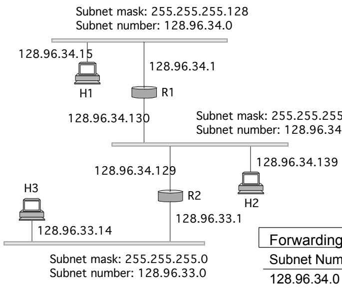
111111111111111111111111	00000000

Subnet mask (255.255.255.0)

Network number	Subnet ID	Host ID

Subnetted address

Subnet Example



Forwarding table at router R1

Subnet Number	Subnet Mask	Next Hop
128.96.34.0	255.255.255.128	interface 0
128.96.34.128	255.255.255.128	interface 1
128.96.33.0	255.255.255.0	R2

Forwarding Algorithm

```
D = destination IP address
for each entry (SubnetNum, SubnetMask, NextHop)
   D1 = SubnetMask & D
   if D1 = SubnetNum
      if NextHop is an interface
         deliver datagram directly to D
      else
         deliver datagram to NextHop
```

- Use a default router if nothing matches
- Not necessary for all 1s in subnet mask to be contiguous
- Can put multiple subnets on one physical network
- Subnets not visible from the rest of the Internet

Supernetting

- Assign block of contiguous network numbers to nearby networks
- Called CIDR: Classless Inter-Domain Routing
- Represent blocks with a single pair

  `(first_network_address, count)`
- Restrict block sizes to powers of 2
- Use a bit mask (CIDR mask) to identify block size
- All routers must understand CIDR addressing

Route Propagation

- Know a smarter router
  - hosts know local router
  - local routers know site routers
  - site routers know core router
  - core routers know everything
- Autonomous System (AS)
  - corresponds to an administrative domain
  - examples: University, company, backbone network
  - assign each AS a 16-bit number
- Two-level route propagation hierarchy
  - interior gateway protocol (each AS selects its own)
  - exterior gateway protocol (Internet-wide standard)

Popular Interior Gateway Protocols

- RIP: Route Information Protocol
  - developed for XNS
  - distributed with Unix
  - distance-vector algorithm
  - based on hop-count
- OSPF: Open Shortest Path First
  - recent Internet standard
  - uses link-state algorithm
  - supports load balancing
  - supports authentication

EGP: Exterior Gateway Protocol

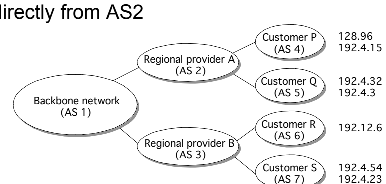
- Overview
  - designed for tree-structured Internet
  - concerned with reachability, not optimal routes
- Protocol messages
  - neighbor acquisition: one router requests that another be its peer; peers exchange reachability information
  - neighbor reachability: one router periodically tests if the another is still reachable; exchange HELLO/ACK messages; uses a k-out-of-n rule
  - routing updates: peers periodically exchange their routing tables (distance-vector)

BGP-4: Border Gateway Protocol

- AS Types
  - stub AS: has a single connection to one other AS
    - carries local traffic only
  - multihomed AS: has connections to more than one AS
    - refuses to carry transit traffic
  - transit AS: has connections to more than one AS
    - carries both transit and local traffic
- Each AS has:
  - one or more border routers
  - one BGP speaker that advertises:
    - local networks
    - other reachable networks (transit AS only)
    - gives path information

BGP Example

- Speaker for AS2 advertises reachability to P and Q
  - network 128.96, 192.4.153, 192.4.32, and 192.4.3, can be reached directly from AS2



- Speaker for backbone advertises
  - networks 128.96, 192.4.153, 192.4.32, and 192.4.3 can be reached along the path (AS1, AS2).
- Speaker can cancel previously advertised paths

Multicast routing

- Multicast - list of sender and receiver not known
- Multicast within LANs is simple because we can use the underlying multicast capabilities of Ethernet
- Internet multicast implemented on top of a collection of networks that support broadcast by extending the routers
- Hosts join multicast groups using Internet Group Management Protocol (IGMP)
- How receivers and senders agree on a specific multicast address is orthogonal to routing issues
  - SDP – Session description protocol
  - SAP – Session announcement protocol
- Problem: Create multicast tree for the routers

Link state multicast

- Each router monitors its lan for multicast packets
- Use this information to build shortest-path multicast tree
- May have to maintain information about each group (many multicast groups can co-exist at the same time)
  - Usually caches these trees

Distance Vector Multicast

- Two steps
  - broadcast mechanism to forward packets to all the networks
  - Pruning mechanism to remove networks that are not currently participating
- Reverse-Path Broadcast (RPB)
  - Routers forward packets along all the outgoing links (except ones that route towards to source)
- Reverse-Path Multicast (RPM)
  - Propagate “no members of G here” back to source

Protocol Independent Multicast (PIM)

- Define operating modes
  - Sparse mode: If few routers are interested in this multicast
  - Dense mode: When most routers want this stream
- Rendezvous point - RP
  - Somehow choose RP
  - Use RP to forward requests to join and prune multicast groups
- Creates source-specific tree or shared tree

Problem – debugging multicast topology

- Suppose – multicast transmission from Berkeley to ND, the receiver is not receiving it. How do you debug it?
- Unicast tools link ping and traceroute do not work because we want to get the whole multicast topology; not if one host can get multicast
  - Just because Stanford is receiving this stream is no help to debug why it is not working for ND

Approaches

- Receiver to Source direction
  - Multicast routing information is used to discover the tree topology
  - Need to know session identities
- Source to receiver
  - Don't need the identities of receivers
  - Multicast forwarding information is used to get the tree
- SNMP based approach
  - Simple Network Management Protocol
  - Each router maintains information. Query all routers to get routing info.

Approaches (cont.)

- Use other mechanisms (such as RTCP – Real time Transport Control Protocol – part of RTP Realtime Transport Protocol)
- RTCP sends announcements periodically and use that to discover topology
  - RTCP is unreliable

Peering and Transits

- Thousands of ISPs. ISPs connect using transit providers and backbone providers to route packets
- Decisions are made on business goals and $$$
- Peering does not give access to other peering points, I.e. peering is non-transitive
- No explicit service level agreement (SLA)
- Peering can be cheaper
  - For example, Notre Dame can peer with Ameritech and ATT to transfer mutual traffic (from DSL and Cable customers)
  - Lower latency to preferred ISPs

Notre Dame to Saint Marys

- traceroute www.saintmarys.edu
  - traceroute to www.saintmarys.edu (147.53.8.10), 30 hops max, 40 byte packets
  - 1 eafs-e06.gw.nd.edu (129.74.250.1) 0.664 ms 0.469 ms 0.450 ms
  - 2 c245-e01.gw.nd.edu (129.74.245.14) 0.301 ms 0.574 ms 0.345 ms
  - 3 monk-fe00.gw.nd.edu (129.74.45.4) 1.046 ms 0.918 ms 0.823 ms
  - 4 klimek-i00.gw.nd.edu (129.74.248.102) 4.784 ms 4.569 ms 4.688 ms
  - 5 mren-m10-lsd6509.startap.net (206.220.240.86) 4.863 ms 5.884 ms 6.659 ms
  - 6 chin-mren-ge.abilene.ucaid.edu (198.32.11.97) 5.234 ms 4.512 ms 4.879 ms
  - 7 iplsng-chinng.abilene.ucaid.edu (198.32.8.77) 15.137 ms 22.735 ms 8.524 ms
  - 8 ul-abilene.indiana.gigapop.net (192.12.206.250) 8.584 ms 9.009 ms 8.814 ms
  - 9 ihets-gw-1-ge15-0.ind.net (157.91.6.37) 8.458 ms 8.581 ms 8.823 ms
  - 10 sbn-fa0-0.ind.net (199.8.76.73) 9.256 ms 8.826 ms 8.638 ms
  - 11 stmarys-edu-T1.ind.net (199.8.73.110) 30.135 ms 26.131 ms 25.682 ms
  - 12 * * smcswitch.saintmarys.edu (147.53.1.1) 31.876 ms !X

Reasons why you don’t peer

- No explicit SLA
- Use cold-potato algorithm to offset traffic costs
  - Carry traffic in your local network as much as possible rather than use an optimal (possibly more expensive transit route)
  - Transit points use hot potato algorithm, dumping the packets as soon as possible to the back bone (even if it was not optimal)
- Don’t want to help potential competitors
  - Ameritech would want your friends to move to Ameritech so that you all can get faster traffic, not peer with AT&T so that you can enjoy the benefit