


18-345 – Fall 08

Lecture 18

**Network Layer:
Routing and Addressing**


Peter Steenkiste

Readings: Chapter 7




Overview Network Layer

- Network layer
- Packet network topology
- Distance vector routing
- Link state routing
- Internet addressing
- Internet protocol
- Routing in the Internet
- Putting things together



Specialized Routing

- Flooding
 - Useful in starting up network
 - Useful in propagating information to all nodes
- Deflection Routing
 - Fixed, preset routing procedure
 - No route synthesis
 - Take alternate path if primary path blocked
 - Not common in networks (backplane technology)




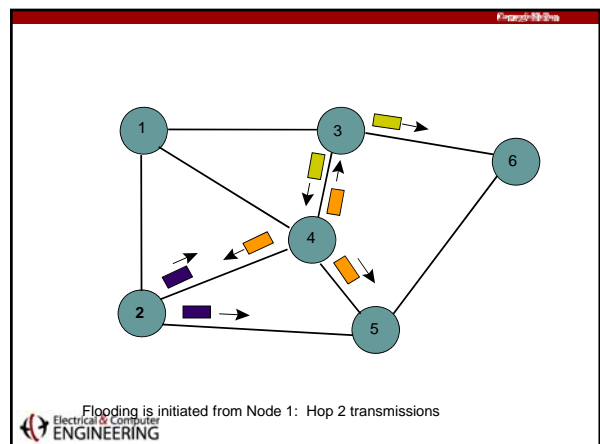
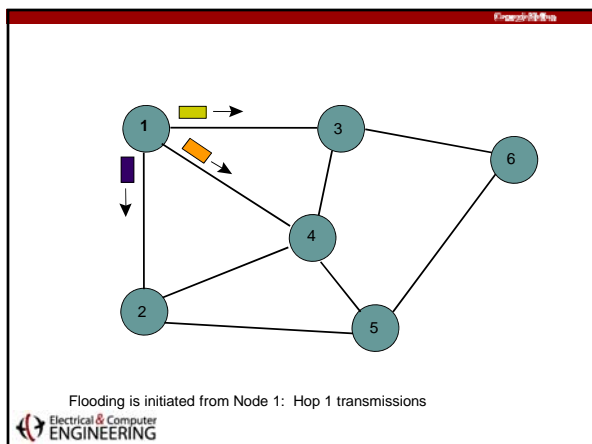
Flooding

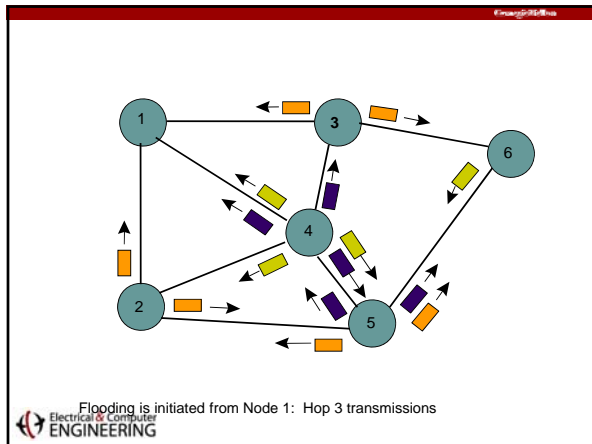
Send a packet to all nodes in a network

- No routing tables available
- Need to broadcast packet to all nodes
 - E.g. to propagate link state information

Approach

- Send packet on all ports **except one where it arrived**
- Exponential growth in packet transmissions



Limited Flooding - Options

- Time-to-Live field in each packet limits number of hops to certain diameter
- Each switch adds its ID before flooding; discards repeats
- Source puts sequence number in each packet; switches records source address and sequence number and discards repeats
 - Safest and most efficient solution
 - Requires routers to keep some state

Shortest Path routing Approaches

Distance Vector Protocols (previous lecture)

- Neighbors exchange list of distances to destinations
- Best next-hop determined for each destination
- Ford-Fulkerson (distributed) shortest path algorithm

Link State Protocols

- Link state information (link up/down?) flooded to all routers
- Routers have complete topology information
- Shortest path (& hence next hop) calculated
- Dijkstra (centralized) shortest path algorithm

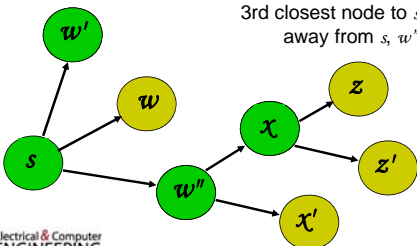
Link-State Algorithm

- Basic idea: two step procedure
 1. Each source node gets a map of all nodes and link metrics (link state) of the entire network
 2. Find the shortest path on the map from the source node to all destination nodes
- Broadcast of link-state information
 - Every node i in the network broadcasts to every other node in the network:
 - ID's of its neighbors: $N_i = \text{set of neighbors of } i$
 - Distances to its neighbors: $\{C_{ij} \mid j \in N_i\}$
 - Flooding is a popular method of broadcasting packets

Dijkstra Algorithm: Finding shortest paths in order

Find shortest paths from source s to all other destinations

Closest node to s is 1 hop away
 2nd closest node to s is 1 hop away from s or w'
 3rd closest node to s is 1 hop away from s , w' , or x



Dijkstra's algorithm

- N : set of nodes for which shortest path already found
- Initialization: (*Start with source node s*)
 - $N = \{s\}$, $D_s = 0$, "s is distance zero from itself"
 - $D_j = C_{sj}$ for all $j \neq s$, distances of directly-connected neighbors
- Step A: (*Find next closest node i*)
 - Find $i \notin N$ such that
 - $D_i = \min D_j$ for $j \notin N$
 - Add i to N
 - If N contains all the nodes, stop
- Step B: (*update minimum costs*)
 - For each node $j \notin N$
 - $D_j = \min (D_j, D_i + C_{ij})$ ← Minimum distance from s to j through node i in N
 - Go to Step A

Execution of Dijkstra's algorithm

Iteration	N	D ₂	D ₃	D ₄	D ₅	D ₆
Initial	{1}	3	2 ✓	5	∞	∞
1	{1,3}	3 ✓	2	4	∞	3
2	{1,2,3}	3	2	4	7	3 ✓
3	{1,2,3,6}	3	2	4 ✓	5	3
4	{1,2,3,4,6}	3	2	4	5 ✓	3
5	{1,2,3,4,5,6}	3	2	4	5	3

Electrical & Computer ENGINEERING

Shortest Paths in Dijkstra's Algorithm

Electrical & Computer ENGINEERING

Reaction to Failure

- If a link fails,
 - Router sets link distance to infinity & floods the network with an update packet
 - All routers immediately update their link database & recalculate their shortest paths
 - Recovery very quick
- But watch out for old update messages
 - Add time stamp or sequence # to each update message
 - Check whether each received update message is new
 - If new, add it to database and broadcast
 - If older, send update message on arriving link

Electrical & Computer ENGINEERING

Why is Link State Better?

- Fast, loopless convergence
- Support for precise metrics, and multiple metrics if necessary (throughput, delay, cost, reliability)
- Support for multiple paths to a destination
 - algorithm can be modified to find best two paths
- But ... not as scalable!

Electrical & Computer ENGINEERING

Source Routing

- Source host selects path that is to be followed by a packet
 - Strict: sequence of nodes in path inserted into header
 - Loose: subsequence of nodes in path specified
- Intermediate switches read next-hop address and remove address
- Source host needs link state information or access to a route server
- Source routing allows the host to control the paths that its information traverses in the network
- Potentially the means for customers to select what service providers they use

Electrical & Computer ENGINEERING

Example

Electrical & Computer ENGINEERING

Non-Hierarchical Addresses and Routing

4 hosts in each network

• No relationship between addresses & routing proximity

• Routing tables require 16 entries each

Electrical & Computer ENGINEERING

Hierarchical Addresses and Routing

Network id = 00

Network id = 10

• Prefix of address indicates network where host is attached

• Routing tables require 4 entries each

Electrical & Computer ENGINEERING

Flat vs Hierarchical Routing

- Flat Routing
 - All routers are peers
 - Does not scale
- Hierarchical Routing
 - Partitioning: Domains, autonomous systems, areas...
 - Some routers part of routing backbone
 - Some routers only communicate within an area
 - Efficient because it matches typical traffic flow patterns
 - Scales

Electrical & Computer ENGINEERING

The Internet Protocol

Reading: Section 8.1 and 8.2

Electrical & Computer ENGINEERING

IETF Request For Comments - RFC's

- New protocols presented through *Internet Drafts (IDs)*
- An ID can become an RFC, which subsequently can be:
 - updated or obsolete (New RFC number issued)
- IAB maintains a list of RFCs for the protocol suite
- Protocol State
 - Standard, Draft Standard, Proposed Standard
 - Experimental, Informational, Historic
- Protocol Status
 - Required, Recommended, Elective, Limited Use, Not Recommended
- Freely Available: <http://www.ietf.org>

Electrical & Computer ENGINEERING

Internet Standards

When a protocol is standardized it is assigned a *Standard Number (STD)*, which reference associated RFCs

Very important Internet standards:

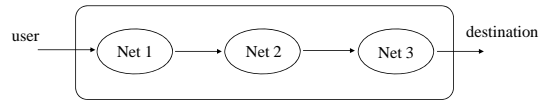
- STD 1 -- Internet Official Protocol Standards
 - state and status of each Internet protocol or standard
 - issued by the IAB approximately quarterly.
- STD 2 -- Assigned Internet Numbers
- STD 3 -- Host Requirements
- STD 4 -- Router Requirements
 - requirements for IPv4 Internet gateway (router) software
 - RFC 1812 -- Requirements for IPv4 Routers.

Electrical & Computer ENGINEERING

Internet Protocol

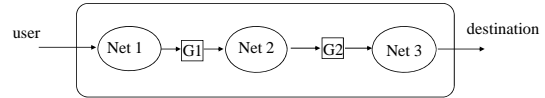
- Unreliable, best effort, connectionless packet delivery
 - motivated by adaptability to failure of network elements
 - packets may be lost, out of order, or even duplicated
 - higher layer protocols must deal with these, if necessary
- STD number 5, which also includes:
 - Internet Control Message Protocol (ICMP)
 - Internet Group Management Protocol (IGMP)

Goal: Collect diverse networks into coordinated whole



Approach:

- Universal set of machine identifiers (32 bit addresses)
- Network independent interfaces: Internet Gateways
- Route packets using destination network, not host



Gateway -> Router

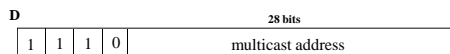
Original Internet Addressing

- RFC 1166
- Each host on Internet has unique 32 bit Internet address that is used in all communications with that host
- Each address has two parts: *netid* and *hostid*
- *netid* unique & administered by Internet Network Info. Center
- Facilitates routing
- A separate address is required for each physical connection of a host to a network; "multi-homed" hosts
- Dotted Decimal Notation:
int1.int2.int3.int4 where intj = integer value of jth octet

IP address classes

- A**
- | | | | | |
|--------|-------|---------|--|-------------------------|
| 7 bits | | 24 bits | | network with many hosts |
| 0 | netid | hostid | | |
- 126 networks with up to 16 million hosts 1.0.0.0 to 127.255.255.255
- B**
- | | | | | |
|---------|---|---------|--------|-------------------------|
| 14 bits | | 16 bits | | network with many hosts |
| 1 | 0 | netid | hostid | |
- 16,382 networks with up to 64,000 hosts 128.0.0.0 to 191.255.255.255
- C**
- | | | | | |
|---------|---|---|--------|------------------------|
| 22 bits | | | 8 bits | network with few hosts |
| 1 | 1 | 0 | netid | |
- 2 million networks with up to 254 hosts 192.0.0.0 to 223.255.255.255
 - More than half class B nets have less than 50 hosts!

IP address classes (contd.)

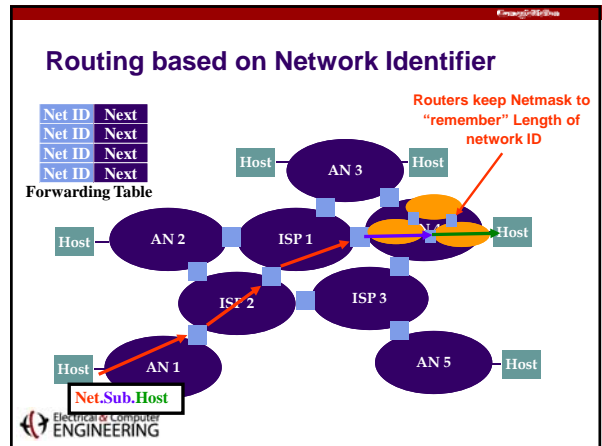
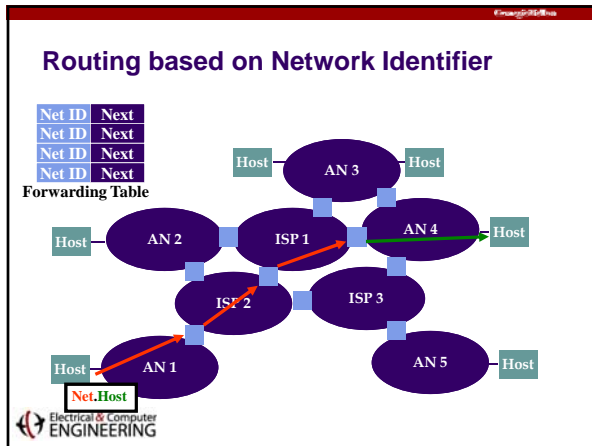


- Up to 250 million multicast groups at the same time 224.0.0.0 to 239.255.255.255
- Permanent group addresses
 - All systems in LAN; All routers in LAN;
 - All OSPF routers on LAN; All designated OSPF routers on a LAN
- Temporary groups addresses created as needed
- Special multicast routers

Subnet Addressing

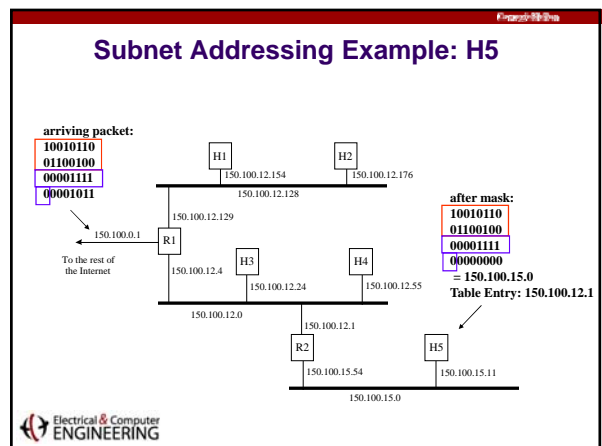
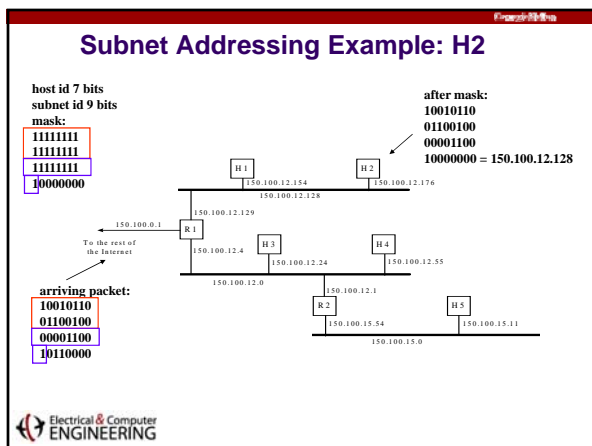
- Subnet addressing introduces another hierarchical level
- Transparent to remote networks
- Simplifies management of multiplicity of LANs
- Routers use masking used to find subnet number





- ### Internet Routing
- Direct routing*
- destination host in same physical network
 - encapsulate IP packet in frame of physical network
- Indirect routing*
- destination host not in same physical network
 - IP packet sent to IP gateway (attached to a physical network)
 - Indirect routing required if destination host is in a different subnet of the same physical network
- Routing Tables*
- Destination IP address to route to next gateway
- Electrical & Computer ENGINEERING

- ### IP Packet Delivery Modes
- Unicast: packet to a single destination
 - Broadcast:
 - limited broadcast: 255.255.255.255; all hosts on subnet
 - used in BOOTP
 - network directed broadcast: netid.111...11; not subnetted
 - routers broadcast packet to all nodes in network;
 - used in ARP
 - subnet directed broadcast: {netid}{subnetid}{1...11}
 - broadcast done by router that receives packet into subnet
 - all-subnets directed broadcast: netid.111...11; subnets defined
 - all hosts in all subnets in network
 - not desirable, can lead to broadcast storm
- Electrical & Computer ENGINEERING



Subnet Addressing Example: Routing Table

Destination	Next Hop	Interface
127.0.0.1	127.0.0.1	lo0
Default	150.100.15.54	emd0
150.100.15.0	150.100.15.11	emd0

Electrical & Computer ENGINEERING

IP Address Exhaustion Problem

- Class B too large; Class C too small
 - Rate of class B allocation implied exhaustion in 1994
 - IPv6 long-term solution, needed short-term solution:
- New allocation policy introduced in 1990 (RFC 2050)
 - Class A & B only assigned for clearly demonstrated need
 - Consecutive blocks of class C assigned (up to 64 blocks)
 - Looks like a subnet within the enclosing class B
 - Lower half of class C space assigned to regional authorities
 - Upper half of class C unassigned & unallocated
- Private IP Addresses (Intranets)
 - Class A, B, C ranges set aside for use within organizations
 - These addresses not forwarded by Internet

Electrical & Computer ENGINEERING

CIDR Addressing: Variable Length Network ID

- Length of network address is variable - specified using netmask
 - Generalization of subnetting
- Can make the address space just large enough, e.g.
 - Can merge a group of adjacent class C addresses to form a larger network address, or ...
 - Can break up a class B address into multiple network addresses
- Must now have the netmask to identify the network id.
 - Address class bits no longer useful
 - Netmask maintained by routers

Electrical & Computer ENGINEERING

CIDR Example

- ISP is allocated 8 class C chunks, 200.10.0.0 to 200.10.7.255
 - Allocation uses 3 bits of class C space
 - Remaining 21 bits are network number, written as 200.10.0.0/21
- Replaces 8 class C routing entries with 1 combined entry
 - Routing protocols carry prefix with destination network address
- ISPs get block of addresses from Regional Internet Registries (RIRs)
 - ARIN (North America, Southern Africa), APNIC (Asia-Pacific), RIPE (Europe, Northern Africa), LACNIC (South America)

Electrical & Computer ENGINEERING

CIDR Addressing: ISP-based Allocation

- Service provider hands out network addresses to its customers from a large block assigned to it.
 - To senders, the customers look like subnets in the ISP
- Routing entries for the ISPs customers can be merged in many cases
 - Packets must be forwarded to that ISP, which forwards them to the customer
 - Only ISP has to distinguish between its customers

Electrical & Computer ENGINEERING

CIDR Address Allocation: Example

Single route entry:
128.5/16

Separate route entries for each customer:
128.5.010/19
128.5.110/19
128.5.011/19

ISP 1: 128.5.X.X
ISP 2: 128.5.36.X

Customer 1: 128.5.010xxxxx.X
Customer 2: 128.5.110xxxxx.X
Customer 3: 128.5.011xxxxx.X

Electrical & Computer ENGINEERING

Shortcomings of CIDR

- CIDR does not help with the addresses that were assigned before CIDR introduction
 - E.g. 128.2 for CMU
- Many exceptions to CIDR addresses.
 - Many customers subscribe with several ISPs for redundancy – network address???
 - Example: 45 Mbs primary ISP; 5 Mbs two backup ISPs
 - Customer receives a block of addresses and then moves to a different ISP – keep addresses?
 - These exceptions require adding that network as a separate route entry in the forwarding tables
- Requires longest prefix match table look up

Route Lookup with CIDR: Longest Prefix Match

- With CIDR there can be multiple matches when looking up an address.
 - E.g. customer belongs to multiple ISPs
- Solution: lookup is based on longest prefix match.
 - If there are multiple matches in the lookup, the match with the most bits wins
- Complicates route lookup!

ISP 10110110 -> ISP 1
 My Entry 10110110 010 -> ISP 2
 10110110 010 0100011

Host Routing Table Example

Destination	Gateway	Genmask	Iface
128.2.209.100	0.0.0.0	255.255.255.255	eth0
128.2.0.0	0.0.0.0	255.255.0.0	eth0
127.0.0.0	0.0.0.0	255.0.0.0	lo
0.0.0.0	128.2.254.36	0.0.0.0	eth0

- Host 128.2.209.100 when plugged into CS ethernet
- Dest 128.2.209.100 → routing to same machine
- Dest 128.2.0.0 → other hosts on same ethernet
- Dest 127.0.0.0 → special loopback address
- Dest 0.0.0.0 → default route to rest of Internet
 - Main CS router: gigrouter.net.cs.cmu.edu (128.2.254.36)