CSE 3461: Introduction to Computer Networking and Internet Technologies

Packet Switching

Presentation G

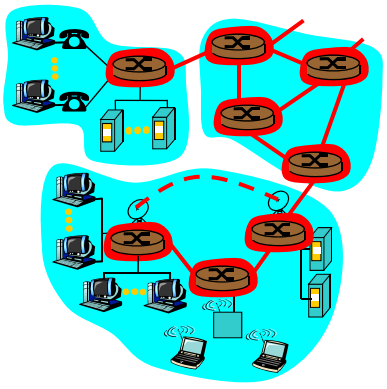
Study: 10.5, 10.6, 12.1, 12.2, 13.1, 13.2, 18.3, 18.4

Gojko Babić

10-09-2012

The Network Core

- mesh of interconnected routers
- *the* fundamental question: how is data transferred through net?
  - —circuit switching: dedicated circuit per call: telephone net
  - —packet-switching: data sent thru net in discrete "chunks"

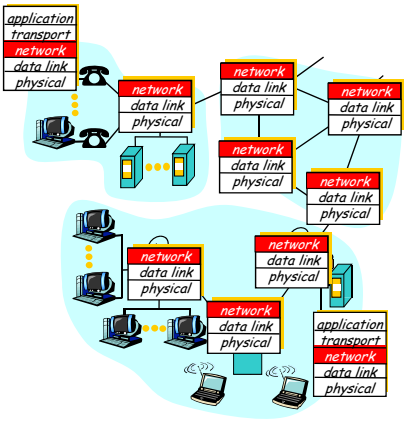


Network Layer Functions

- transport packet from sending to receiving hosts
- network layer protocols in *every* host, router

three important functions:

- *path determination:* route taken by packets from source to dest. *Routing algorithms*
- *switching:* move packets from router's input to appropriate router output
- *call setup:* some network architectures require router call setup along path before data flows



Network Core: Packet Switching

each end-end data stream divided into *packets*

- user A, B packets *share* network resources
- each packet uses full link bandwidth
- resources used *as needed*,

Bandwidth division into "pieces"
Dedicated allocation
Resource reservation

resource contention:

- *aggregate resource demand can exceed amount available*
- *congestion: packets queue, wait for link use*
- *store and forward: packets move one hop at a time*
  - *transmit over link*
  - *wait turn at next link*

Packet Switching: Basic Operation

- Data transmitted in small packets
  - — Typically 1000 octets (bytes)
  - — Longer messages split into series of packets
  - — Each packet contains a portion of user data plus some control information
- Control information
  - — Routing (addressing) information
- Packets are received, stored briefly (buffered) and past on to the next node
  - — Store and forward
- Packets sent one at a time through any network link

Use of Packets

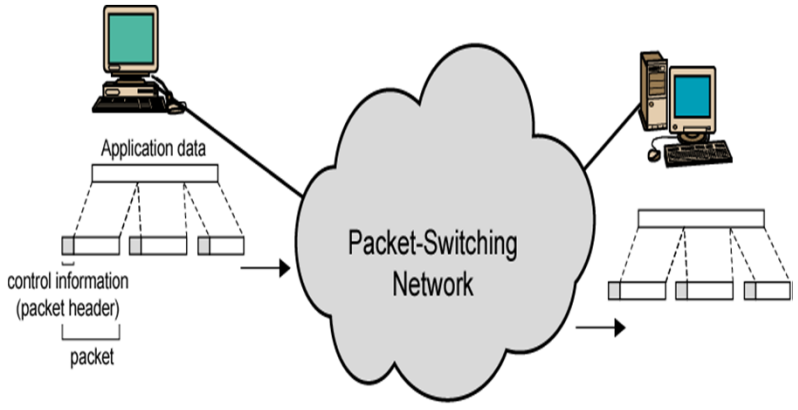
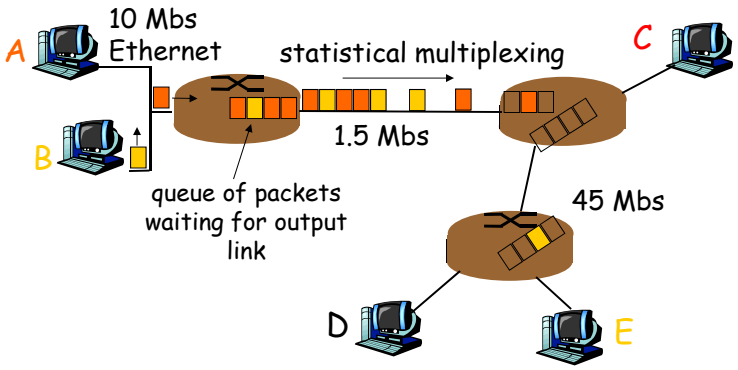


Figure 10.8

Network Core: Packet Switching

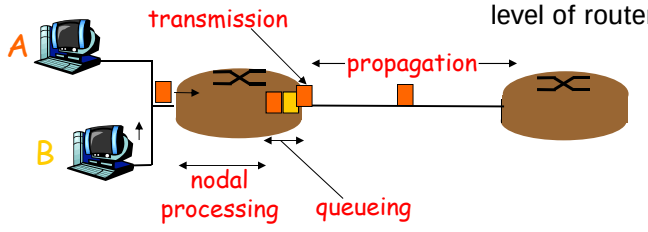


Delays in Packet-Switched Networks

packets experience delay on end-to-end path

- four sources of delay at each hop

- nodal processing:
  - check bit errors
  - determine output link
- queueing
  - time waiting at output link for transmission
  - depends on congestion level of router



Delays in Packet-Switched Networks

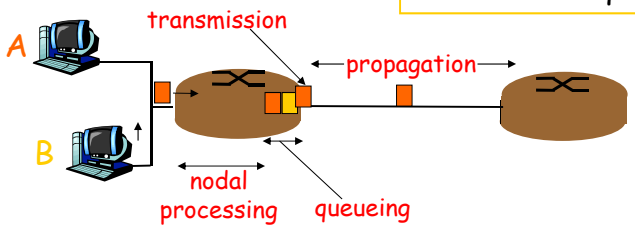
Transmission delay:

- C=link bandwidth (bps)
- m=packet length (bits)
- time to send bits into link = m/C

Propagation delay:

- d = length of physical link
- s = propagation speed in medium (~$2x10^8$ m/sec)
- propagation delay = d/s

Note: s and C are *very* different quantities!



Packet Switching: Advantages

- Line efficiency
  - — Single node to node link can be shared by many packets over time
  - — Packets queued and transmitted as fast as possible
- Data rate conversion
  - — Each station connects to the local node at its own speed
  - — Nodes buffer data if required to equalize rates
- Packets are accepted even when network is busy
  - — Delivery may slow down
  - — Priorities can be used
- Packets handled in two ways:
  - – Datagram
  - – Virtual-circuit

Datagram and Virtual-Circuit

- Datagram approach:
  - — Each packet treated independently
  - — Packets can take any practical route
  - — Packets may arrive out of order
  - — Packets may go missing
  - — Up to receiver to re-order packets and recover from missing packets
- Virtual-Circuit approach
  - — Preplanned route established before any packets sent
  - — Call request and call accept packets establish connection (handshake)
  - — Once connection established, each packet contains a virtual circuit identifier instead of destination address
  - — No routing decisions required for each packet
  - — Clear request to drop circuit

Virtual Circuits

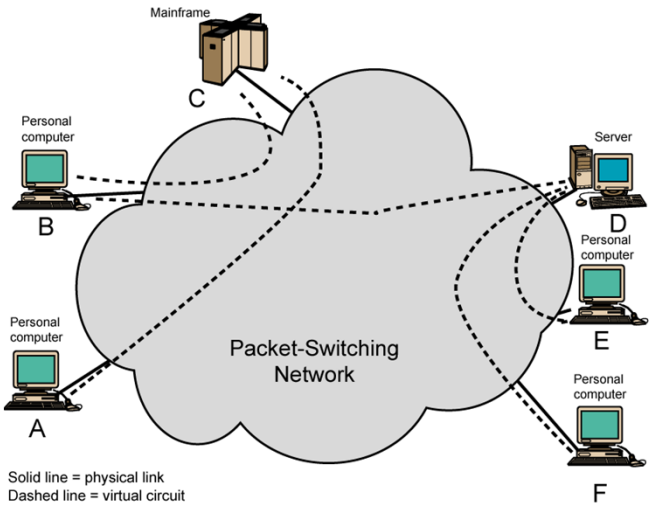
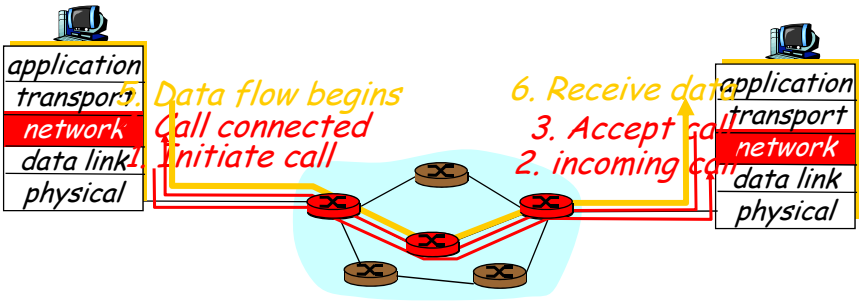


Figure 10.13

Virtual Circuits: Signaling Protocols

- used to setup, maintain & teardown VC
- used in ATM, frame-relay, X.25
- not used in today's Internet



Packet Switching: Virtual-Circuit Approach

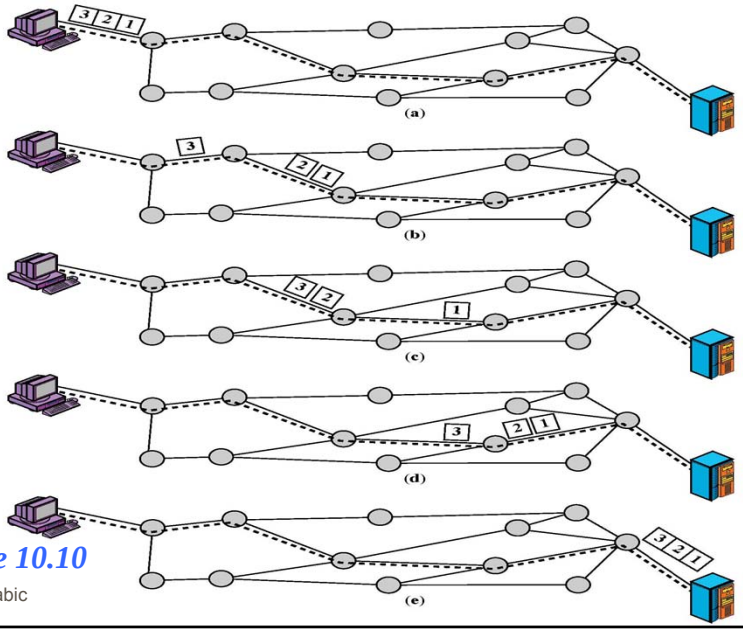
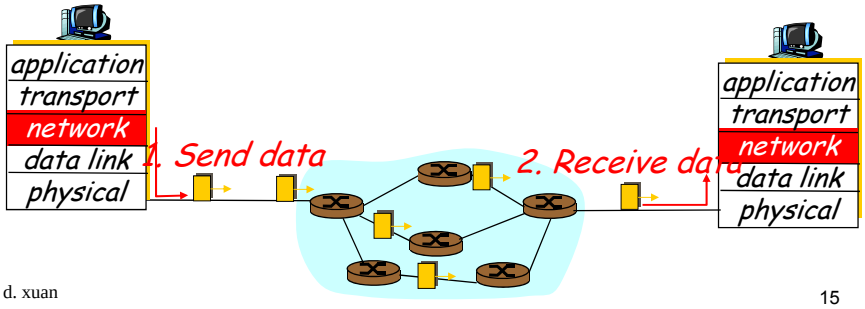


Figure 10.10

Datagram Networks: Internet Model

- no call setup at network layer
- routers: no state about end-to-end connections
  - no network-level concept of “connection”
- packets typically routed using destination host ID
  - packets between same source-dest pair may take different paths



d. xuan

15

Packet Switching: Datagram Approach

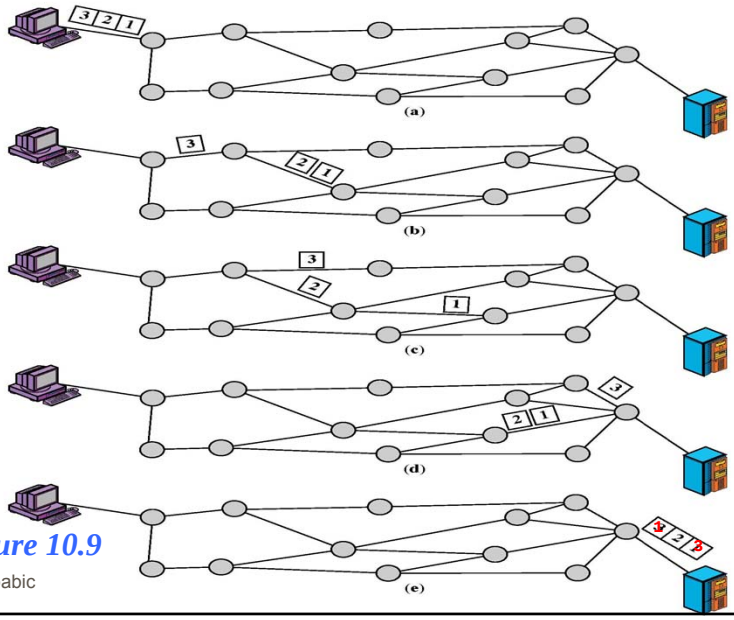


Figure 10.9

g. babic

16

Virtual Circuits vs. Datagram

- Virtual circuits
  - — Network can provide sequencing and error control
  - — Packets are forwarded more quickly
    - No routing decisions to make
  - — Less reliable
    - Loss of a node looses all circuits through that node
- Datagram
  - — No call setup phase
    - Better if few packets
  - — More flexible
    - Routing can be used to avoid congested parts of the network

Packet Size

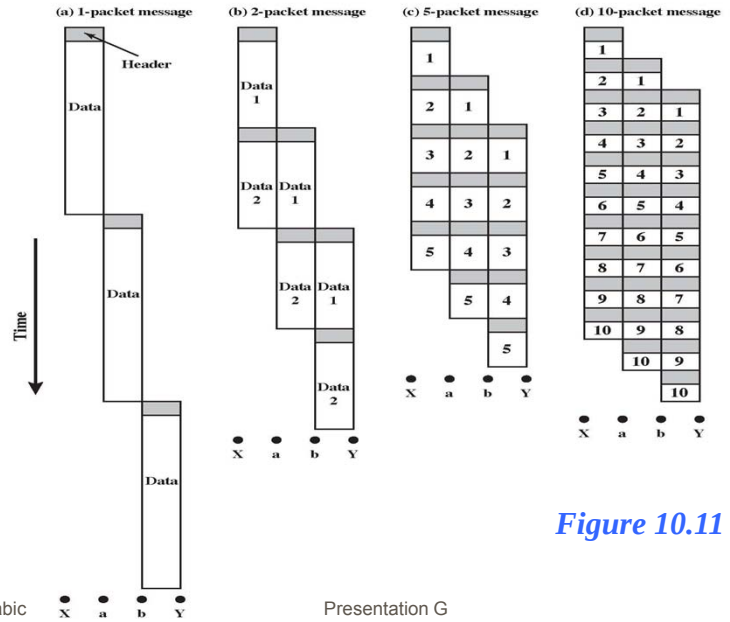
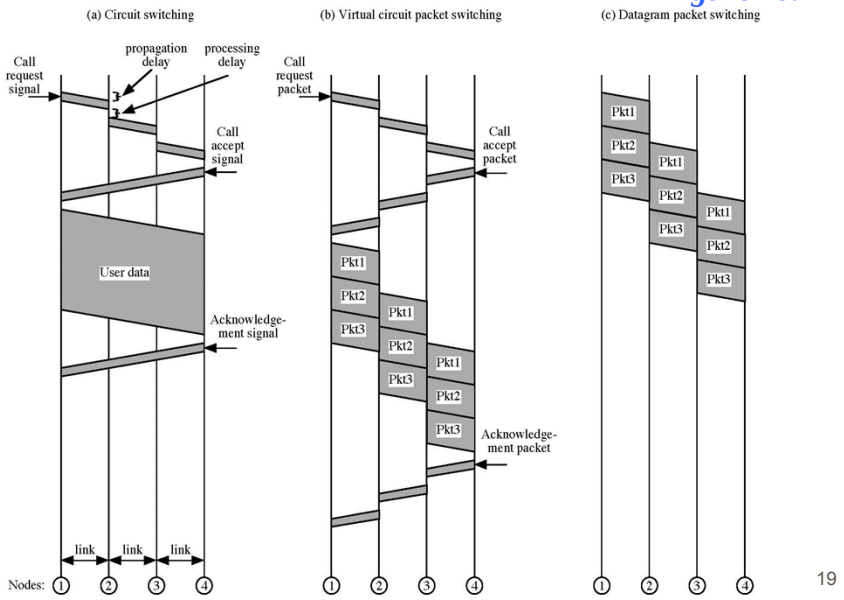


Figure 10.11

Circuit Switching vs. Packet Switching

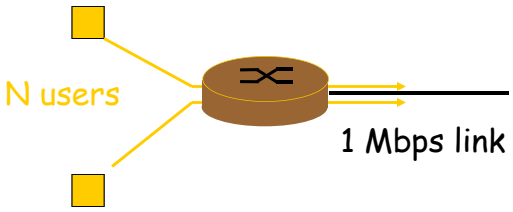
Figure 10.12



Packet Switching vs. Circuit Switching

Packet switching allows more users to use network!

- 1 Mbit link
- each user:
  - 100Kbps when "active"
  - active 10% of time
- circuit-switching:
  - 10 users
- packet switching:
  - with 35 users, probability > 10 active less than .0004



X.25 Protocol

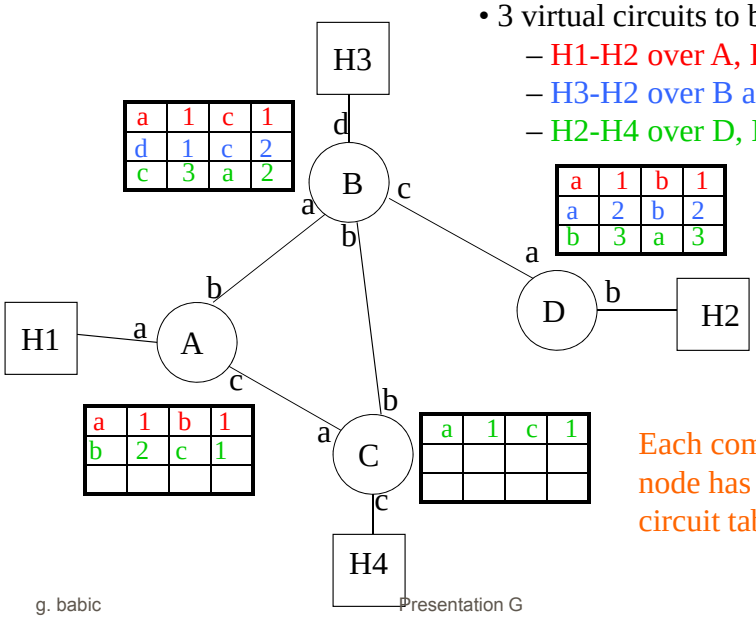
- Almost universal on virtual-circuit packet switched networks and packet switching in ISDN
- Defines three layers:
  - Physical
  - Link: Link Access Protocol Balance – LAPB (Subset of HDLC)
  - Packet: Virtual Circuit Service
- Virtual Circuit Service: Logical connection between two stations
- Specific route established through network for each connection
  - Internal virtual circuit
- Typically one to one relationship between external and internal virtual circuits
- Considerable overhead
- Not appropriate for modern digital systems with high reliability

X.25 Packets

- Call control packets:
  - Call Request packet includes: packet type indicator, destination and source address, and virtual circuit number
  - Call Accept packet includes: packet type indicator, and virtual circuit number
- Multiplexing of virtual circuits (data packets) at layer 3
- Layer 3 data packets include flow and error control
  - Data packet have send sequence numbers and receive sequence numbers similar as in data link layer, plus virtual circuit number, instead of destination address

X.25 Network: Connection Establishment

- 3 virtual circuits to be established:
  - H1-H2 over A, B and D
  - H3-H2 over B and D
  - H2-H4 over D, B, A and C



Each communication node has its virtual circuit table

IPv4 Datagram Format

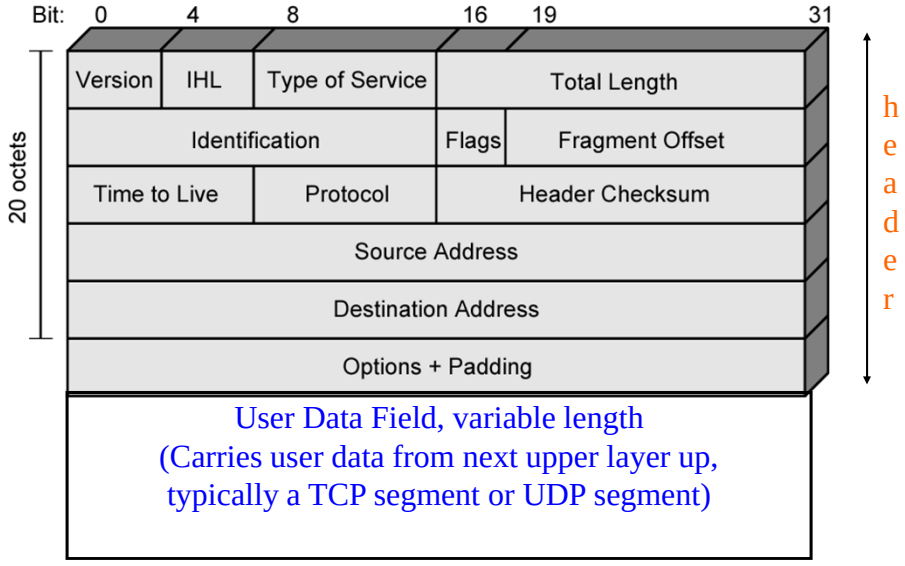


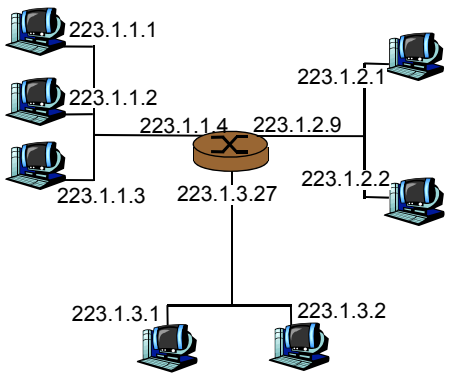
Figure 18.6 Max length of datagram 65,535 octets

IP Network: Design Issues

- Routing is based on the destination address:
  - — End systems and routers maintain routing tables that indicate next router to which datagram should be sent
    - Static routing
    - Dynamic routing: Flexible response to congestion and errors
    - Source routing: Source specifies (in *Options* field) route as sequential list of routers to be followed
    - Route recording and time-stamping (in *Options* field) by routes
- Datagram lifetime
- Fragmentation and re-assembly
- Error control
- Flow control

IP Addressing: Introduction

- IP address: 32-bit identifier for host, router *interface*
- *interface:* connection between host, router and physical link
  - — router's typically have multiple interfaces
  - — host may have multiple interfaces
  - — IP addresses associated with interface, not host, router
- Dotted decimal notation



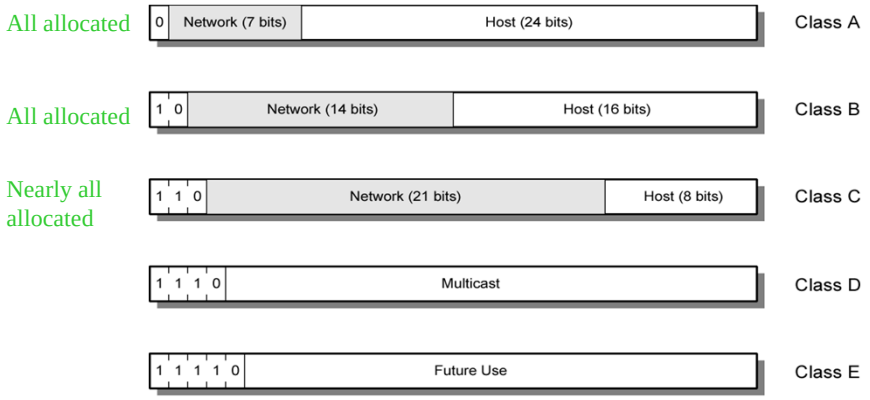
223.1.1.1 = 11011111 00000001 00000001 00000001

223 1 1 1

IP Addressing: Class-full Addressing

1111111 → reserved for loopback

Figure 18.7



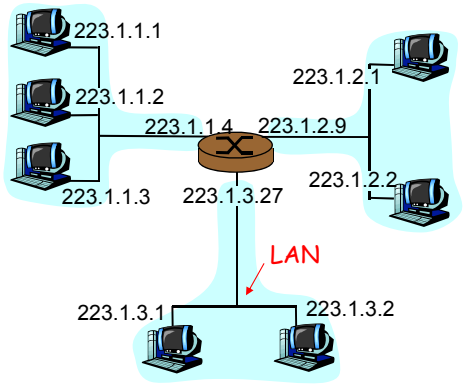
IP addressing: CIDR

- classful addressing:
  - inefficient use of address space, address space exhaustion
  - e.g., class B net allocated enough addresses for 65K hosts, even if only 2K hosts in that network
- CIDR: Classless InterDomain Routing
  - network portion of address of arbitrary length
  - address format: a.b.c.d/x, where x is # bits in network portion of address

200.23.16.0/23

IP Addressing

- IP address:
  - network part (high order bits)
  - host part (low order bits)
- *What's a network ?* (from IP address perspective)
  - device interfaces with same network part of IP address
  - can physically reach each other without intervening router



network consisting of 3 IP networks (for IP addresses starting with 223, first 24 bits are network address)

Getting Datagram from Source to Destination 1

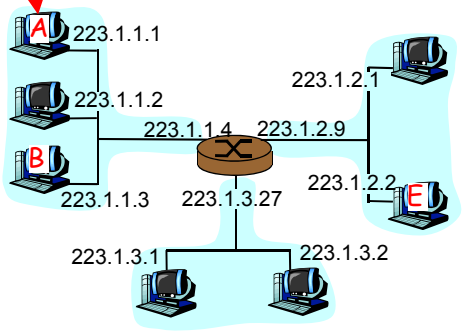
IP datagram:

misc fields	source IP addr	dest IP addr	data

- datagram remains unchanged, as it travels source to destination
- address fields of interest here

routing table in *A*

Dest. Net.	next router	Nhops
223.1.1		1
223.1.2	223.1.1.4	2
223.1.3	223.1.1.4	2



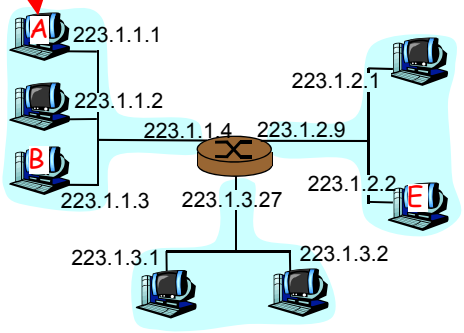
Getting Datagram from Source to Destination 2

misc fields	223.1.1.1	223.1.1.3	data

Starting at A, given IP datagram addressed to B:

- look up net. address of B
- find B is on same network as A
- link layer will send datagram directly to B inside link-layer frame, since B and A are directly connected

Dest. Net.	next router	Nhops
223.1.1		1
223.1.2	223.1.1.4	2
223.1.3	223.1.1.4	2



Getting Datagram from Source to Destination 3

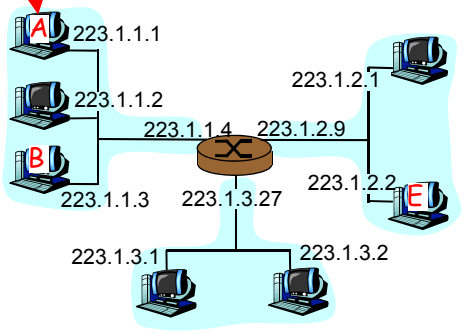
misc fields	223.1.1.1	223.1.2.3	data

Starting at A, dest. E:

- look up network address of E
- E on *different* network, i.e. A, E not directly attached
- routing table: next hop router to E is 223.1.1.4
- link layer sends datagram to router 223.1.1.4 inside link-layer frame
- datagram arrives at 223.1.1.4

continued.....

Dest. Net.	next router	Nhops
223.1.1		1
223.1.2	223.1.1.4	2
223.1.3	223.1.1.4	2



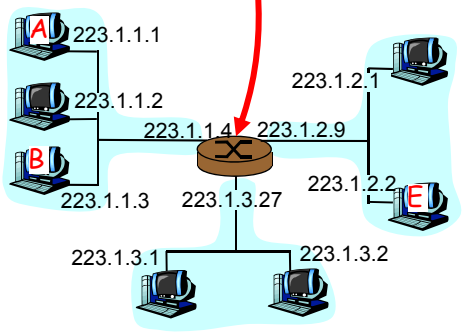
Getting Datagram from Source to Destination 4

misc fields	223.1.1.1	223.1.2.3	data

Arriving at 223.1.4, destined for 223.1.2.2

- look up network address of E
- E on *same* network as router's interface 223.1.2.9, i.e. router, E directly attached
- link layer sends datagram to 223.1.2.2 inside link-layer frame via interface 223.1.2.9
- datagram arrives at 223.1.2.2!!! (hooray!)

Dest. network	next router	Nhops	interface
223.1.1	-	1	223.1.1.4
223.1.2	-	1	223.1.2.9
223.1.3	-	1	223.1.3.27

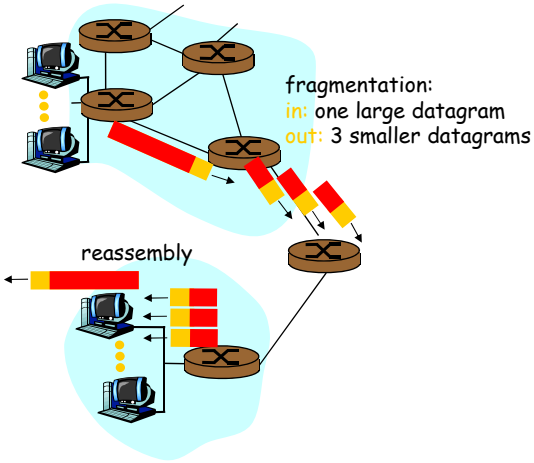


Datagram Lifetime & Type of Service

- Datagrams could loop indefinitely:
  - Consumes resources
- Datagram marked with lifetime:
  - *Time to Live* field in IP
  - Hop count
    - Decrement time to live on passing through each router
  - Time count
  - Once lifetime expires, datagram discarded (not forwarded)
- Type of Service filed:
  - Specify treatment of data unit during transmission through networks

IP Fragmentation & Reassembly

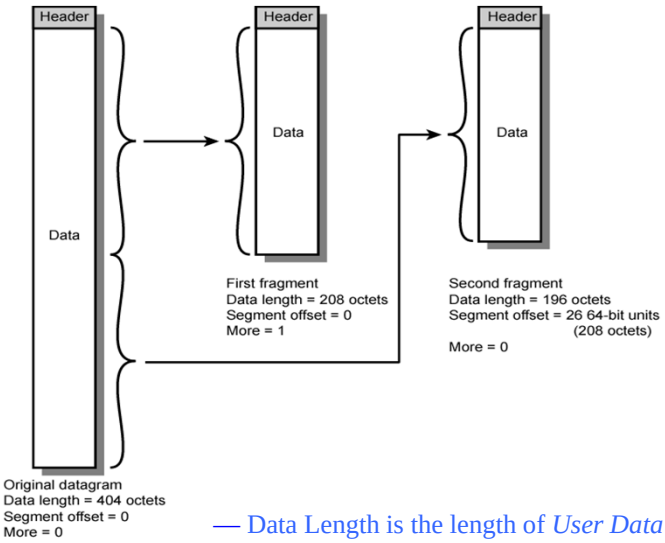
- network links have MTU (max.transfer size) - largest possible link-level frame.
  - different link types, different MTUs
- large IP datagram divided ("fragmented") within net
  - one datagram becomes several datagrams
  - "reassembled" only at final destination
  - IP header bits used to identify, order related fragments



Fragmentation and Re-assembly

- IP re-assembles at destination (resulting in packets getting smaller as data traverses internet), using:
  - Data unit ID identified by:
    - *Source Address* and *Destination Address*
    - *Protocol* layer generating data (e.g. TCP)
    - *Identification* supplied by that layer
  - *Fragment Offset*: position of fragment of user data in original datagram, in multiples of 64 bits (8 octets)
  - *More* bit: indicates that this is not the last fragment; also *Don't Fragment* bit
  - Re-assembly may fail if some fragments get lost; re-assembly time out assigned to first fragment to arrive.

Fragmentation Example



— Data Length is the length of *User Data Filed*

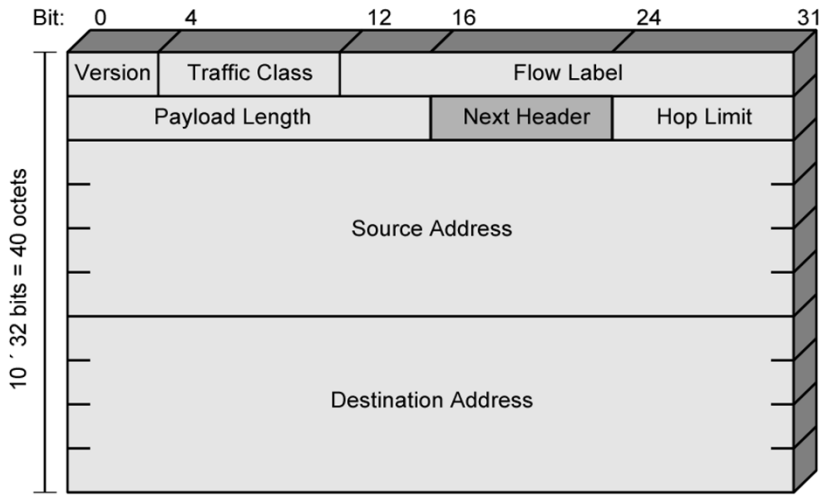
Error Control and Flow Control

- Error Control:
  - — Not guaranteed delivery
  - — Router should attempt (ICMP protocol used) to inform source if packet discarded, for time to live expiring
  - — Datagram identification needed
  - — Source may modify transmission strategy
  - — May inform high layer protocol
- Flow Control:
  - — Allows routers and/or stations to limit rate of incoming data
  - — Limited in connectionless systems
  - — Send flow control packets (ICMP used)
  - — Requesting reduced flow; again ICMP used
  - — **No flow control currently provided for in Internet**

ICMP – Internet Control Message Protocol

- IP protocol filed identifies ICMP
- Often considered as a part of IP layer
- Provides feedback from the network:
  - —destination (network, host, or protocol) unreachable or unknown
  - — time to live expiring
  - — parameter problem
  - — fragmentation needed but *Don't Fragment* bit set
  - — source quench
- Can be used by the host to obtain certain information:
  - — echo request and echo replay (ping program)
  - — timestamp request and timestamp replay

IPv6 Header Format



Routing in Packet Switching Networks

- Complex, crucial aspect of packet switched networks
- Characteristics required:
  - Correctness
  - Simplicity
  - Robustness
  - Stability
  - Fairness
  - Optimality
  - Efficiency
- Routing Strategies:
  - Fixed
  - Flooding
  - Random
  - Adaptive

Elements of Routing Techniques

- Performance criteria:
  - minimize number of hops
  - minimize delay
  - maximize throughput
  - minimize cost
- Decision time:
  - datagram → each packet
  - virtual circuit → only call request packet
- Decision place
  - distributed → made by each node
  - centralized → made by central node
  - source → made by originating node

Elements of Routing Techniques (continued)

- Network information source:
  - — Distributed routing → each node makes decisions
    - Nodes use local knowledge
    - May collect information from adjacent nodes
    - May collect information from all nodes on a potential route
  - — Central routing → one central node makes decisions
    - One node collects information for all nodes
- Update timing:
  - — Fixed - never updated
  - — After major load changes
  - — After topology changes
  - — Regular updates

Example of Network for Fixed Routing

- Each link is assigned its cost, that is a base for routing decisions
- Link costs in different directions may be different
- Can have link value (i.e. link cost) inversely proportional to capacity
- Define cost of path between two nodes as sum of costs of links traversed

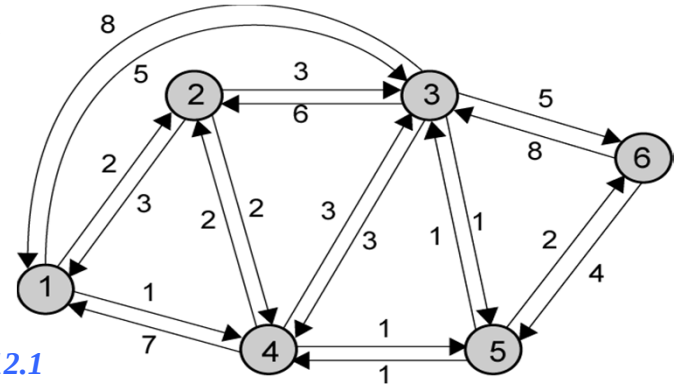


Figure 12.1

Fixed Routing

- Least cost algorithm, for each node pair, finds a path with the least cost
- Single permanent route for each source to destination pair
- Route fixed, at least until a change in network topology

CENTRAL ROUTING DIRECTORY

To Node \ From Node	1	2	3	4	5	6
1	—	1	5	2	4	5
2	2	—	5	2	4	5
3	4	3	—	5	3	5
4	4	4	5	—	4	5
5	4	4	5	5	—	5
6	4	4	5	5	6	—

Figure 12.2

Node 1 Directory

Destination	Next Node
2	2
3	4
4	4
5	4
6	4

Node 2 Directory

Destination	Next Node
1	1
3	3
4	4
5	4
6	4

Node 3 Directory

Destination	Next Node
1	5
2	5
4	5
5	5
6	5

Node 4 Directory

Destination	Next Node
1	2
2	2
3	5
5	5
6	5

Node 5 Directory

Destination	Next Node
1	4
2	4
3	3
4	4
6	6

Node 6 Directory

Destination	Next Node
1	5
2	5
3	5
4	5
5	5

Flooding

- No network info required
- Incoming packets retransmitted on every link except incoming link
- Eventually a number of copies will arrive at destination
- Each packet is uniquely numbered so duplicates can be discarded
- Nodes can remember packets already forwarded to keep network load in bounds
- Can include a hop count in packets
- Property of flooding
  - All possible routes are tried, thus it is very robust
  - At least one packet will have taken minimum hop count route, thus it can be used to set up virtual circuit
  - All nodes are visited, thus useful to distribute information (e.g. routing)

Flooding Example

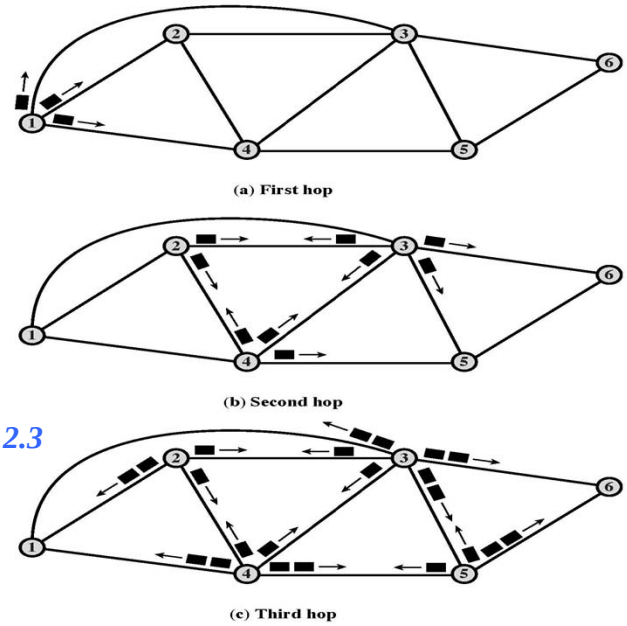


Figure 12.3

Random Routing

- Node selects one outgoing path for retransmission of incoming packet
- Selection can be random or round robin
- Can select outgoing path based on probability calculation
- No network info needed
- Resulting route is typically not least cost nor minimum hops
- But far less traffic than flooding

Adaptive Routing

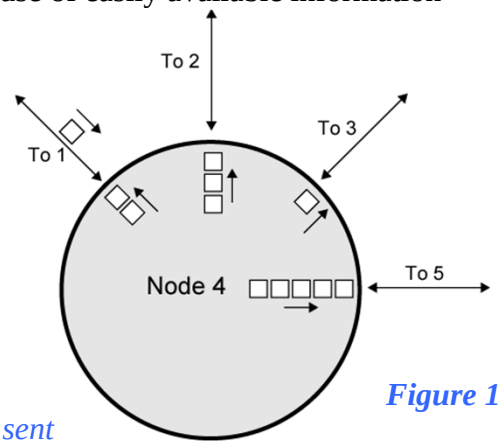
- Used by almost all packet switching networks
- Routing decisions change as conditions on the network change
  - —Link (or node) failure
  - —Congestion, i.e. change in traffic load
- Requires information about network
- Decisions more complex
- Tradeoff between quality of network information and overhead
- Classification based on information sources
  - —Local (isolated)
  - —Adjacent nodes
  - —All nodes
- Reacting too quickly can cause oscillation
- Reacting too slowly can make irrelevant

Isolated Adaptive Routing

- Route to outgoing link with shortest queue
- Can include bias for each destination
- Rarely used - do not make use of easily available information

Node 4's Bias Table for Destination 6

Next Node	Bias
1	9
2	6
3	3
5	0

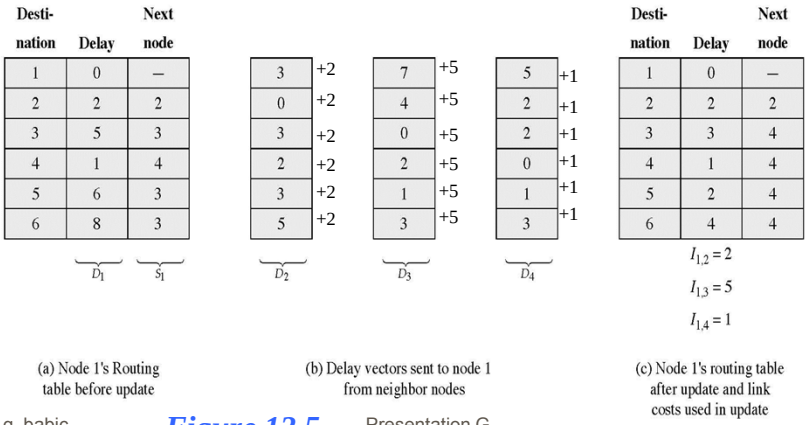


Packet arriving with node 6 as destination will be sent through link 3; min (Q + B) = 4

Figure 12.4

First Generation ARPANET Routing

- **Node makes routing decisions based on "Next node" column in its routing table**
- **Periodically, neighbor nodes exchange "Delay" columns**
- **Every node calculates delays on each of its links based on current queue lengths**
- **Having received new "Delay" columns, a node updates its routing table**



(a) Node 1's Routing table before update

(b) Delay vectors sent to node 1 from neighbor nodes

(c) Node 1's routing table after update and link costs used in update

ARPANET Routing

- Distributed adaptive
- First generation ARPANET routing:
  - Estimated delay used as performance criterion
  - Node exchanges delay vector with its neighbors
  - Update routing table based on incoming information
  - Doesn't consider line speed, just queue length
  - Queue length is not necessary a good measurement of delay
  - Responds slowly to congestion
- Second generation ARPANET routing:
  - Delay measured directly (time-stamped packets)
  - If there are any significant changes in delay, the information is sent to all other nodes using flooding
  - Each node maintains an estimate of delay on every network link
  - Good under light and medium loads
  - Under heavy loads, oscillation may occur

What is Congestion?

Congestion:

- informally: “too many sources sending too much data too fast for *network* to handle”
- different from flow control!
- manifestations:
  - —lost packets (buffer overflow at routers)
  - —long delays (queueing in router buffers)
- a top-10 problem!

What is Congestion?

- Congestion occurs when the number of packets being transmitted through the network approaches the packet handling capacity of the network
- Congestion control aims to keep number of packets below level at which performance falls off dramatically
- Generally 80% line utilization is critical
- Data network is a network of queues
  - — Packets arriving are stored at input buffers
  - — Routing decision made
  - — Packet moves to output buffer
  - — Packets queued for output transmitted as fast as possible
  - — If packets arrive too fast to be routed, or to be output, buffers will fill and congestion starts occurring
  - — finite queues mean data may be lost

Interaction of Queues

- Node to node flow control can propagate congestion through network

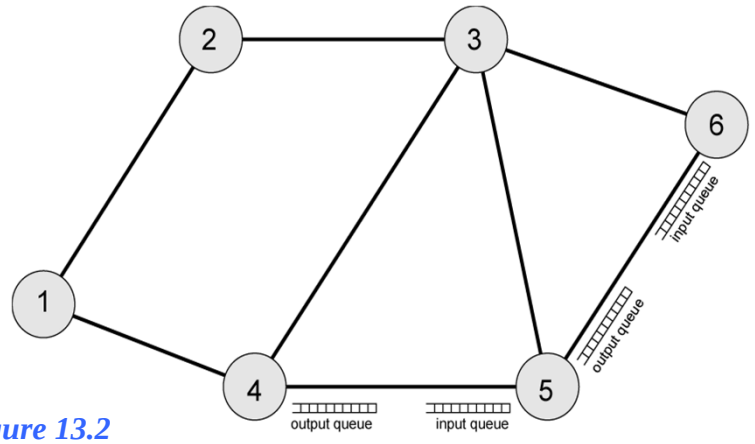


Figure 13.2

Effects of Congestion – No Control

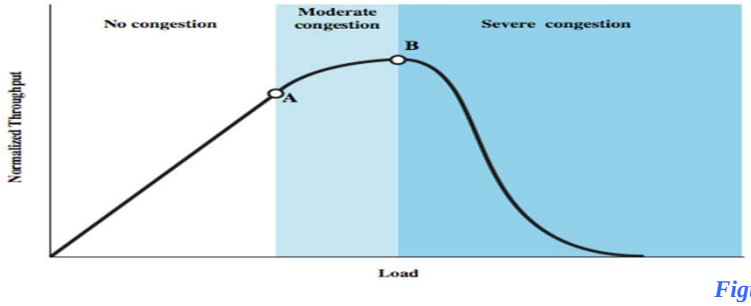
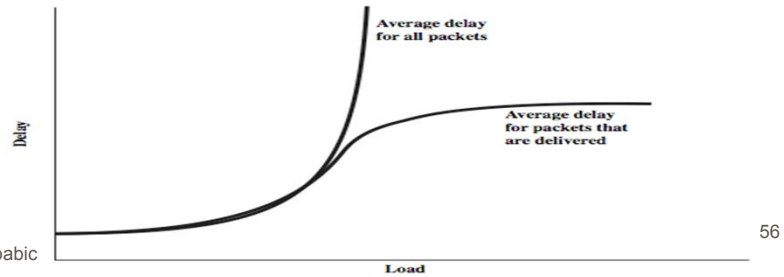
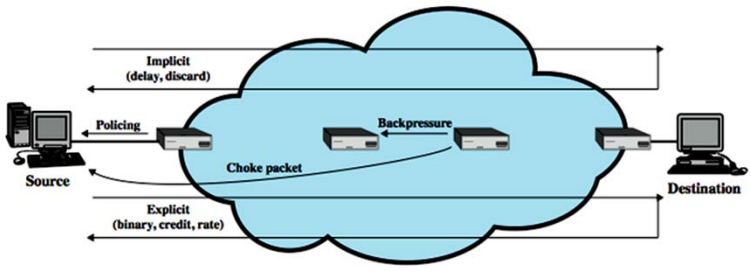


Figure 13.4



Mechanisms for Congestion Control



- backpressure
- choke packets
- implicit congestion signaling
- explicit congestion signaling

Backpressure & Choke Packet

- Backpressure
  - —if node becomes congested it can slow down or halt flow of packets from other nodes and other nodes have to apply control on outgoing packet rates
    - o propagates back to source
  - —used in connection oriented networks that allow hop-by-hop flow control (e.g. X.25)
- Choke Packet
  - — generated at congested node and sent back to source node
  - — ICMP Source Quench packet
    - o from router or destination end system and source cuts back until it no longer receives source quench messages
    - o message is issued for every discarded packet

Implicit & Explicit Congestion Signaling

- Implicit congestion signaling:
  - With network congestion transmission delay increases and packets may be discarded
  - Source can detect congestion and reduce flow
  - Responsibility of end systems
  - Effective on connectionless (datagram) networks
- Explicit congestion signaling:
  - Network alerts end systems of increasing congestion and end systems take steps to reduce offered load
  - Backward: congestion avoidance notification in opposite direction to packet required
  - Forward: congestion avoidance notification in same direction as packet required