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

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

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

Virtual Circuits: Signaling Protocols

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

Data flow begins
Call connected
Initiate call

2. incoming call
3. Accept call
4. Call connected
5. Data flow begins
6. Receive data
Packet Switching: Virtual-Circuit Approach

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
Packet Switching: Datagram Approach

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 0.0004

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

User Data Field, variable length
(Carries user data from next upper layer up, typically a TCP segment or UDP segment)

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

![Diagram of IP addressing](image)

```
223.1.1.1 = 11011111 00000001 00000001 00000001
223 1 1 1
```

IP Addressing: Class-full Addressing

- 1111111 → reserved for loopback

<table>
<thead>
<tr>
<th>Class</th>
<th>Network (bits)</th>
<th>Host (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>B</td>
<td>1 0</td>
<td>16</td>
</tr>
<tr>
<td>C</td>
<td>1 1 0</td>
<td>8</td>
</tr>
<tr>
<td>D</td>
<td>1 1 1 0</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1 1 1 1 0</td>
<td></td>
</tr>
</tbody>
</table>

*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

```
11001000 00010111 00010000 00000000
```

```
200.23.16.0/23
```

IP Addressing

- **IP address:**
  - network part (high order bits)
  - host part (low order bits)

- **What’s a network?**
  - device interfaces with same network part of IP address
  - can physically reach each other without intervening router

```
223.1.1.1
223.1.1.2
223.1.1.3
223.1.1.4
223.1.2.1
223.1.2.2
223.1.2.9
223.1.3.1
223.1.3.2
223.1.3.27
```

```
LAN
```

```
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:

<table>
<thead>
<tr>
<th>misc fields</th>
<th>source IP addr</th>
<th>dest IP addr</th>
<th>data</th>
</tr>
</thead>
</table>
- datagram remains unchanged, as it travels source to destination
- address fields of interest here

Routing table in A

<table>
<thead>
<tr>
<th>Dest. Net.</th>
<th>next router</th>
<th>Nhops</th>
</tr>
</thead>
<tbody>
<tr>
<td>223.1.1</td>
<td>223.1.1.4</td>
<td>2</td>
</tr>
</tbody>
</table>

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

Getting Datagram from Source to Destination 2

<table>
<thead>
<tr>
<th>misc fields</th>
<th>223.1.1.1</th>
<th>223.1.1.3</th>
<th>data</th>
</tr>
</thead>
</table>

Routing table in A

<table>
<thead>
<tr>
<th>Dest. Net.</th>
<th>next router</th>
<th>Nhops</th>
</tr>
</thead>
<tbody>
<tr>
<td>223.1.1</td>
<td>223.1.1.4</td>
<td>2</td>
</tr>
</tbody>
</table>

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
Getting Datagram from Source to Destination 3

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

Getting Datagram from Source to Destination 4

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!)
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: in: one large datagram out: 3 smaller datagrams*

*reassembly*
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

- 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
- Can be used by the host to obtain certain information:
  - echo request and echo replay (ping program)
  - timestamp request and timestamp replay
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

![Diagram of network](image)

**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

<table>
<thead>
<tr>
<th>CENTRAL ROUTING DIRECTORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>From Node</td>
</tr>
<tr>
<td>To Node</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

**Figure 12.2**

- **Node Directory** table for each node:
  - **Node 1 Directory**
  - **Node 2 Directory**
  - **Node 3 Directory**
  - **Node 4 Directory**
  - **Node 5 Directory**
  - **Node 6 Directory**
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

(a) First hop

(b) Second hop

(c) Third hop

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

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

**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

<table>
<thead>
<tr>
<th>Destination</th>
<th>Delay</th>
<th>Next node</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>3</td>
</tr>
</tbody>
</table>

\[ +2 \quad +5 \quad +1 \]
\[ +2 \quad +5 \quad +1 \]
\[ +2 \quad +5 \quad +1 \]
\[ +2 \quad +5 \quad +1 \]
\[ +2 \quad +5 \quad +1 \]

(a) Node \( Y \)'s Routing table before update

Figure 12.4

Figure 12.5
**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.

**Studying Assignment: 13.1, 13.2**

Interaction of Queues

- Node to node flow control can propagate congestion through network.

*Figure 13.2*
Congestion Control in Packet Switching

- Can use flow control
  - Node to node
  - End to end
- Send control packets from congested node to some or all source nodes
  - Requires additional traffic during congestion
- Rely on routing information, e.g. link delay information in ARPANET
  - But delay may vary too quickly to be used for congestion control
- End to end probe packets to measure delay
  - Adds to overhead
- Add congestion info to packets as they cross nodes
  - Either backwards or forwards
Frame Relay

- Designed to be more efficient than X.25
- Developed before ATM and larger installed base than ATM
- ATM now of more interest on high speed networks
- Frame relay can be considered as a type of packet switching
- Call control carried in separate logical connection
- Multiplexing and switching at layer 2
  - Eliminates one layer of processing
- No hop by hop error or flow control
  - Increased reliability makes this less of a problem
- End to end flow and error control (if used) are done by higher layer
- Streamlined communications process
  - Lower delay
  - Higher throughput
Data Frame Format

<table>
<thead>
<tr>
<th>Flag</th>
<th>Address</th>
<th>Information</th>
<th>FCS</th>
<th>Flag</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;-1-&gt;</td>
<td>&lt;-4-&gt;</td>
<td>Variable &lt;-2-&gt;</td>
<td>&lt;-1-&gt;</td>
<td></td>
</tr>
</tbody>
</table>

(a) Frame format

8 7 6 5 4 3 2 1

Upper DLCI  CR  EA 0
Lower DLCI  FECN  BECN  DE  EA 1

(b) Address field - 2 octets (default)

8 7 6 5 4 3 2 1

Upper DLCI  CR  EA 0
DLCT  FECN  BECN  DE  EA 0

Lower DLCI or DL-CORE control  DC  EA 1

(c) Address field - 3 octets

8 7 6 5 4 3 2 1

Upper DLCI  CR  EA 0
DLCT  FECN  BECN  DE  EA 0

Lower DLCI or DL-CORE control  DC  EA 1

(d) Address field - 4 octets

8 7 6 5 4 3 2 1

Upper DLCI  CR  EA 0

EL  Address field extension bit
C3  Command sequence bit
FECN  Forward explicit congestion notification
BECN  Backward explicit congestion notification
DLCT  Data link connection identifier
DL-CORE control indicator
DE  Discard eligibility

Data Frame

- One frame type: only for user data and no control frames
- No sequence numbers
  - No flow nor error control
- End to end functionality, i.e. transfer of information between ends
- LAPF (Link Access Procedure for Frame Mode)
  - Frame delimiting, alignment and transparency
  - Frame multiplexing and demultiplexing using addressing field
  - Ensure frame is neither too long nor short
  - Detection of transmission errors
  - Frames in error simple discarded
  - Congestion control functions
Asynchronous Transfer Mode - ATM

- ATM uses an approach sometimes referred as cell switching
- ATM cell switching can be considered as type of packet switching
  - Transfer of data in discrete chunks
  - Multiple logical connections over single physical interface
- In ATM flow on each logical connection is in fixed sized packets called cells
- Minimal error and flow control
  - Reduced overhead
- Data rates (physical layer) 25.6Mbps to 622.08Mbps
- ATM cells
  - 5 octet header
  - 48 octet information field
  - Small cells reduce queuing delay for high priority cells
  - Small cells can be switched more efficiently
  - Easier to implement switching of small cells in hardware

ATM Cell Format

CLP – Cell Loss Priority
Circuit Switch Elements

Figure 10.4

g. babic

FDM of Three Voice-band Signals

(a) Spectrum of voice channel

(b) Spectrum of voice channel modulated on 64kHz

(c) Spectrum of composite signals using subscribers at 64kHz, 68KHz, and 72 KHz

Figure 8.5

g. babic
X.25 Data Packets

- User data passes to X.25 level 3
- X.25 appends control information
  - Header: identifies virtual circuit and provides sequence numbers for flow and error control
- X.25 packet passed down to LAPB entity
- LAPB appends further control information

Network Subject to Oscillations
Input and Output Queues at Node

Figure 13.1