Communication Networks: Technology & Protocols

Varaiya - Walrand

University of California

Berkeley

Course Objectives

- Explain Main Technology and Protocols
- Discuss Important Trends

Course Organization

- Three Days of Lectures
- Exercises to Consolidate Understanding
- Day 1: Protocols and TCP/IP (1)
- Day 2: TCP/IP (2), LANs, ATM, C/S
- Day 3: QoS, Optical, Switches, Wireless

<

Table of Contents (1/2)

Day One

Day Two

<

Table of Contents (2/2)

- Day Three

<

Overview

- Web
- Voice over IP
- Terminology: LAN, MAN, WAN
- History

<

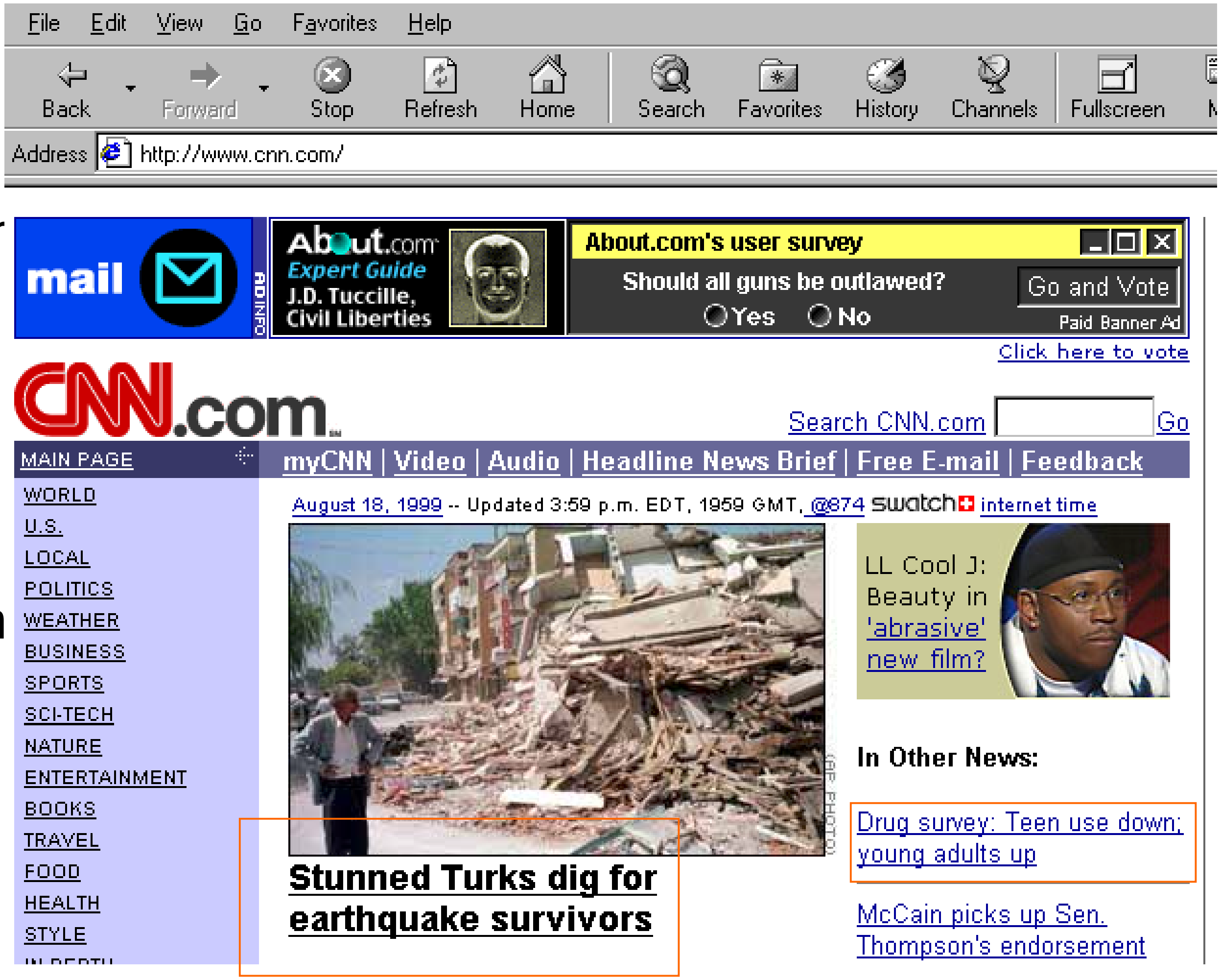
Web

- Example
- Locating Resource: DNS
- Connection
- End-to-end
- Packets
- Bits
- Points to remember

Web: Example

- Click -> get page
- page from local or remote computer
- link: http://cnn.com

  specifies
  - protocol (http)
  - location (cnn.com



Address http://www.cnn.com/

About.com Expert Guide J.D. Tuccille, Civil Liberties

About.com's user survey
Should all guns be outlawed?
○ Yes ○ No
Go and Vote
Paid Banner Ad
Click here to vote

Search CNN.com Go

myCNN | Video | Audio | Headline News Brief | Free E-mail | Feedback

WORLD
U.S.
LOCAL
POLITICS
WEATHER
BUSINESS
SPORTS
SCI-TECH
NATURE
ENTERTAINMENT
BOOKS
TRAVEL
FOOD
HEALTH
STYLE

August 18, 1999 -- Updated 3:59 p.m. EDT, 1959 GMT, @874 swatch internet time

Stunned Turks dig for earthquake survivors

LL Cool J: Beauty in 'abrasive' new film?

In Other News:

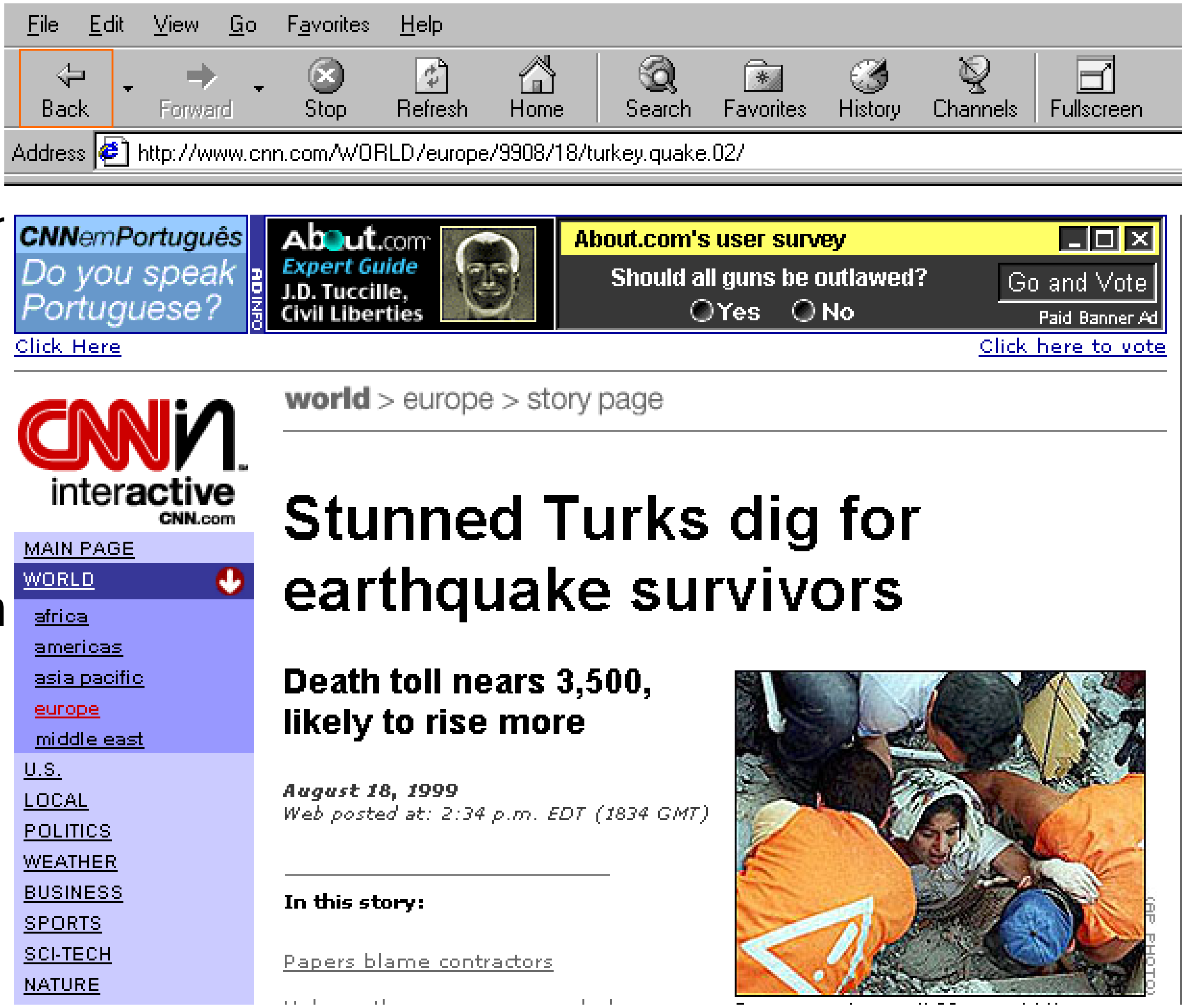
Drug survey: Teen use down; young adults up

McCain picks up Sen. Thompson's endorsement

Web: Example

- Click -> get page
- page from local or remote computer
- link: http://cnn.com

  specifies
  - protocol (http)
  - location (cnn.com



Address: http://www.cnn.com/WORLD/europe/9908/18/turkey.quake.02/

world > europe > story page

Stunned Turks dig for earthquake survivors

Death toll nears 3,500, likely to rise more

August 18, 1999
Web posted at: 2:34 p.m. EDT (1834 GMT)

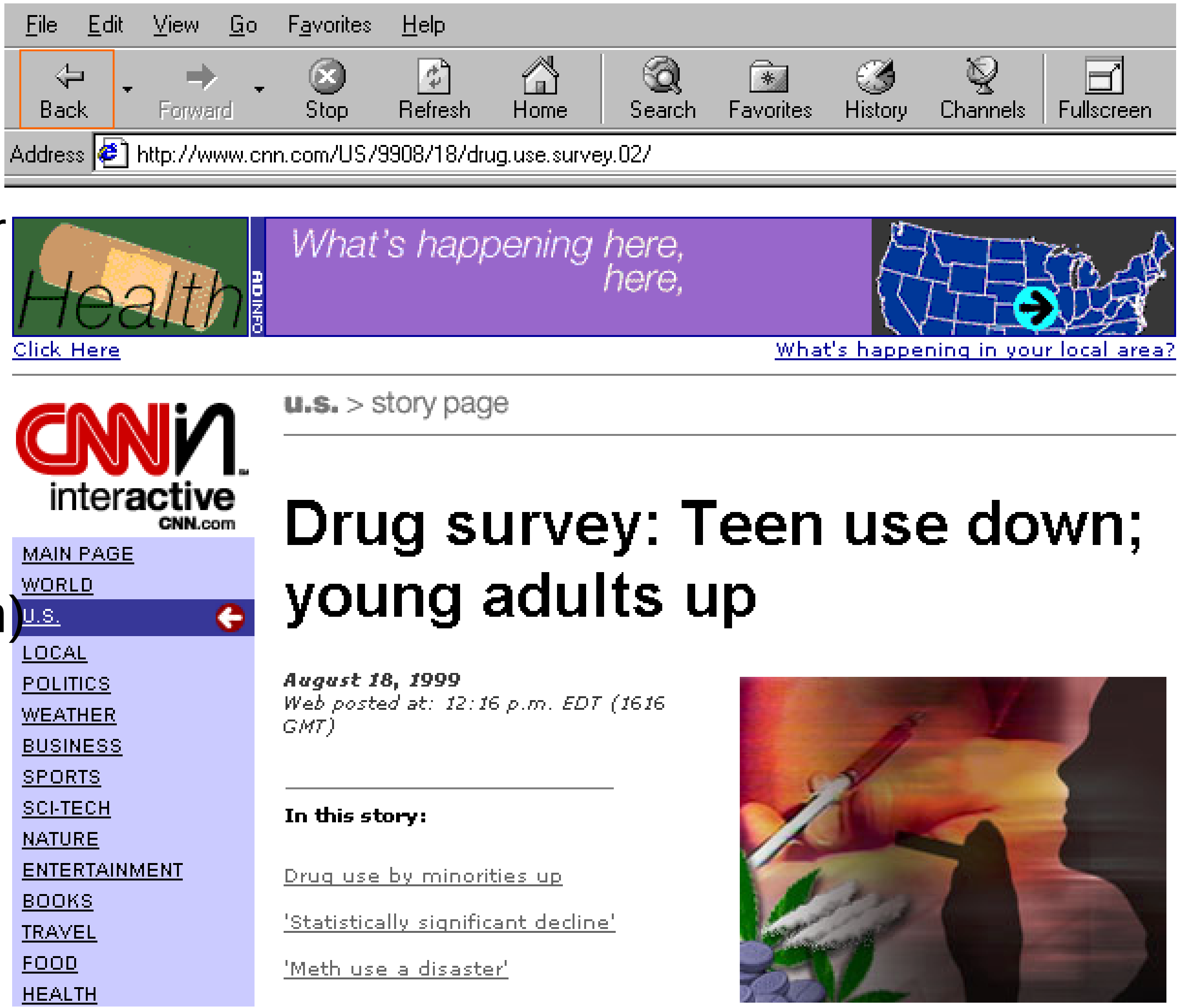
In this story:

Papers blame contractors

Web: Example

- Click -> get page
- page from local or remote computer
- link: http://cnn.com

  specifies
  - protocol (http)
  - location (cnn.com)



Address: http://www.cnn.com/US/9908/18/drug.use.survey.02/

u.s. > story page

Drug survey: Teen use down; young adults up

August 18, 1999
Web posted at: 12:16 p.m. EDT (1616 GMT)

In this story:

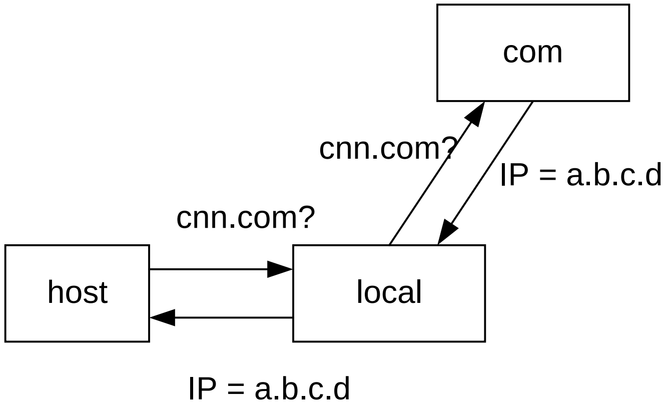
Drug use by minorities up

'Statistically significant decline'

'Meth use a disaster'

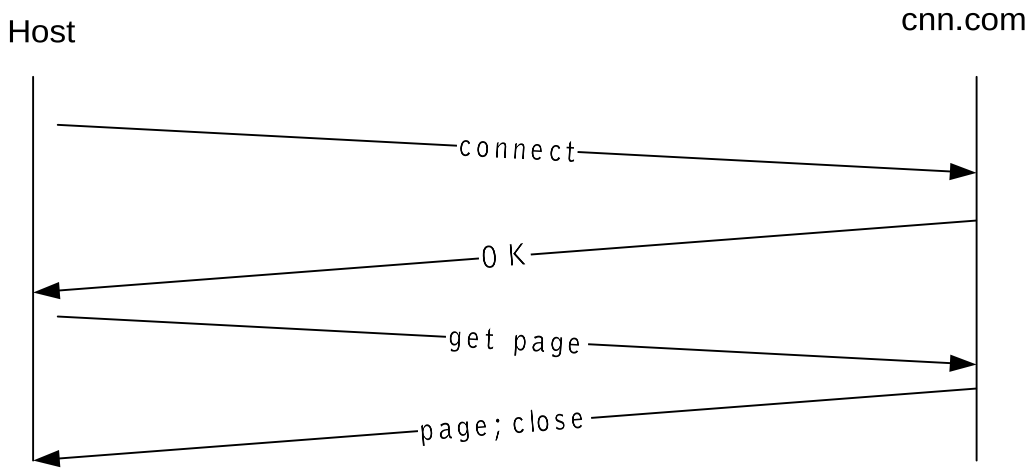
Web: Locating Resource

- cnn.com is the **name** of a computer (and, implicitly, of a file in that computer)
- To find the address, the application uses a hierarchical directory service called the **Domain Name System**



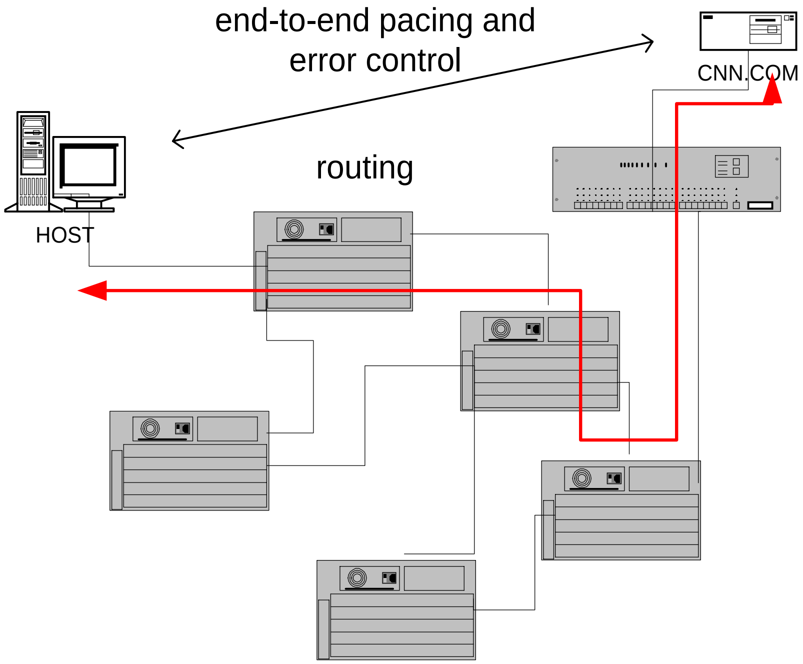
Web: Connection

- The protocol (http) sets up a **connection** between the host and cnn.com to transfer the page
- The connection transfers the page as a byte stream, without errors: **pacing + error control**



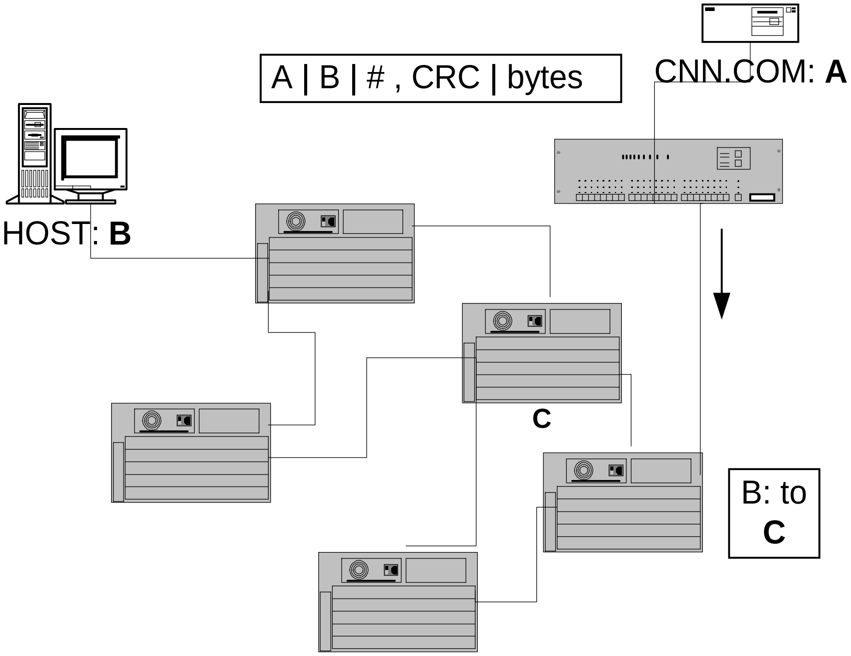
Web: End-to-end

- The byte stream flows from end to end across many links and switches: **routing (+ addressing)**
- That stream is regulated and controlled by both ends: **retransmission** of erroneous or missing bytes; **pacing**



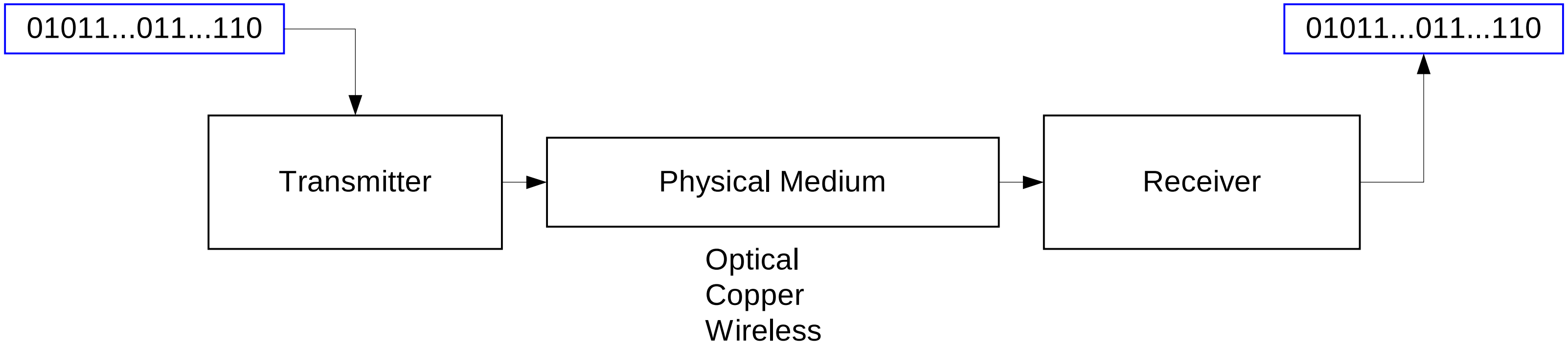
Web: Packets

- The network transports bytes grouped into packets
- The packets are "self-contained" and routers handle them one by one
- The end hosts worry about errors and pacing: Destination sends ACKs Source checks losses



Web: Bits

- Equipment in each node sends the packets as a string of bits
- That equipment is not aware of the meaning of the bits

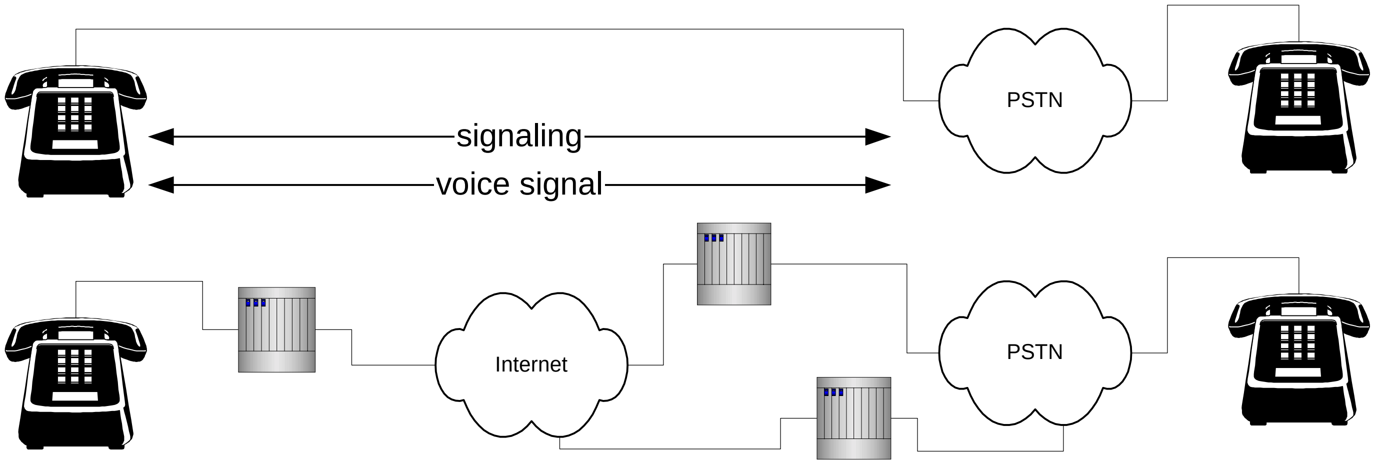


Web: Points to remember

- Separation of tasks
  - send bits on a link: transmitter/receiver [clock, modulation,...]
  - send packet on each hop [framing, error detection,...]
  - send packet end to end [addressing, routing]
  - pace transmissions [detect congestion]
  - retransmit erroneous or missing packets [acks, timeout]
  - find destination address from name [DNS]
- Scalability
  - routers don’t know about connections
  - names and addresses are hierarchical

Voice Over IP: General Operations

- Gateways must reproduce the signaling and voice
- Signaling: dialed digits, dial tone, ringing
- Voice: Packetize and absorb delay jitter
- Note: Routing problem

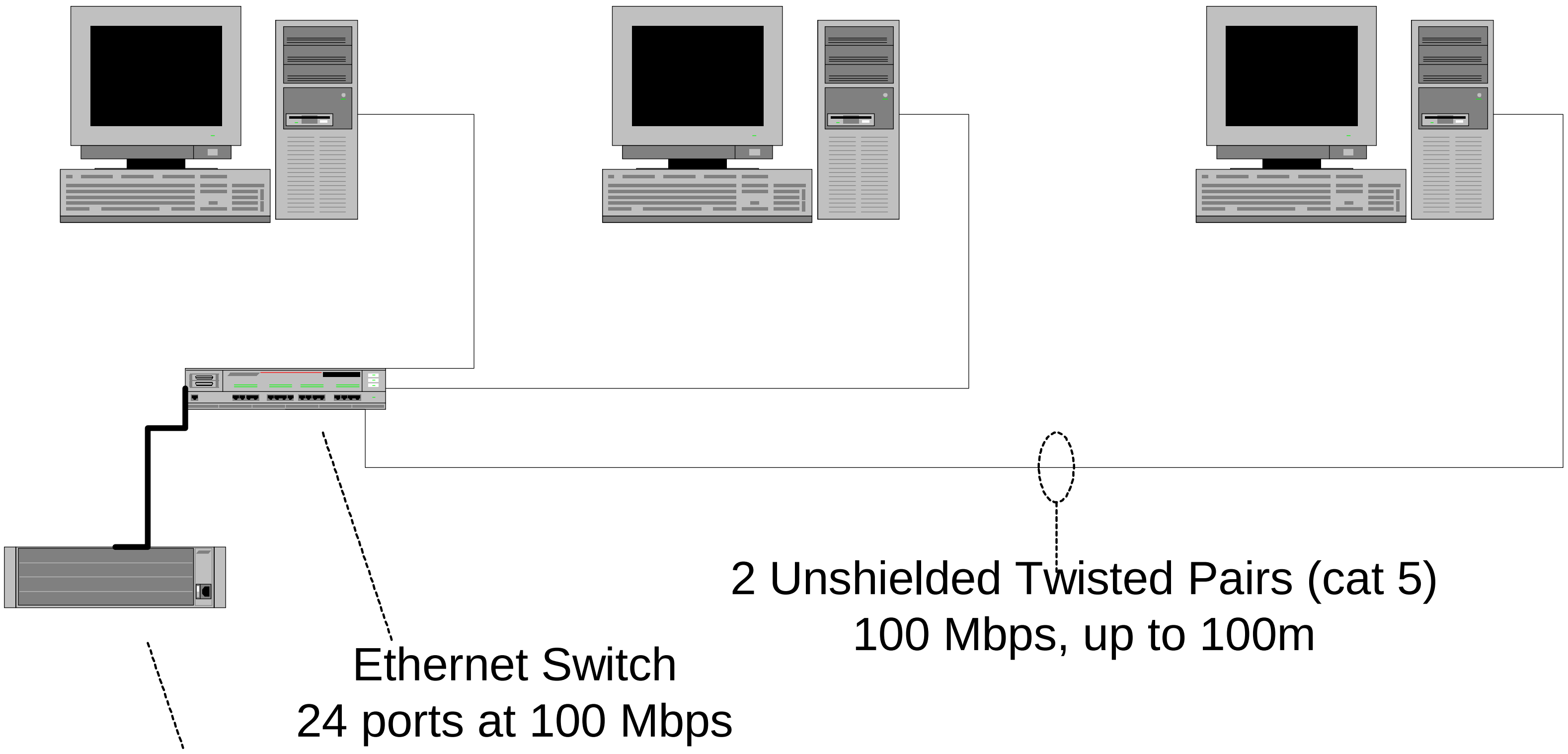


<

Terminology

- LAN
- MAN
- WAN

LAN

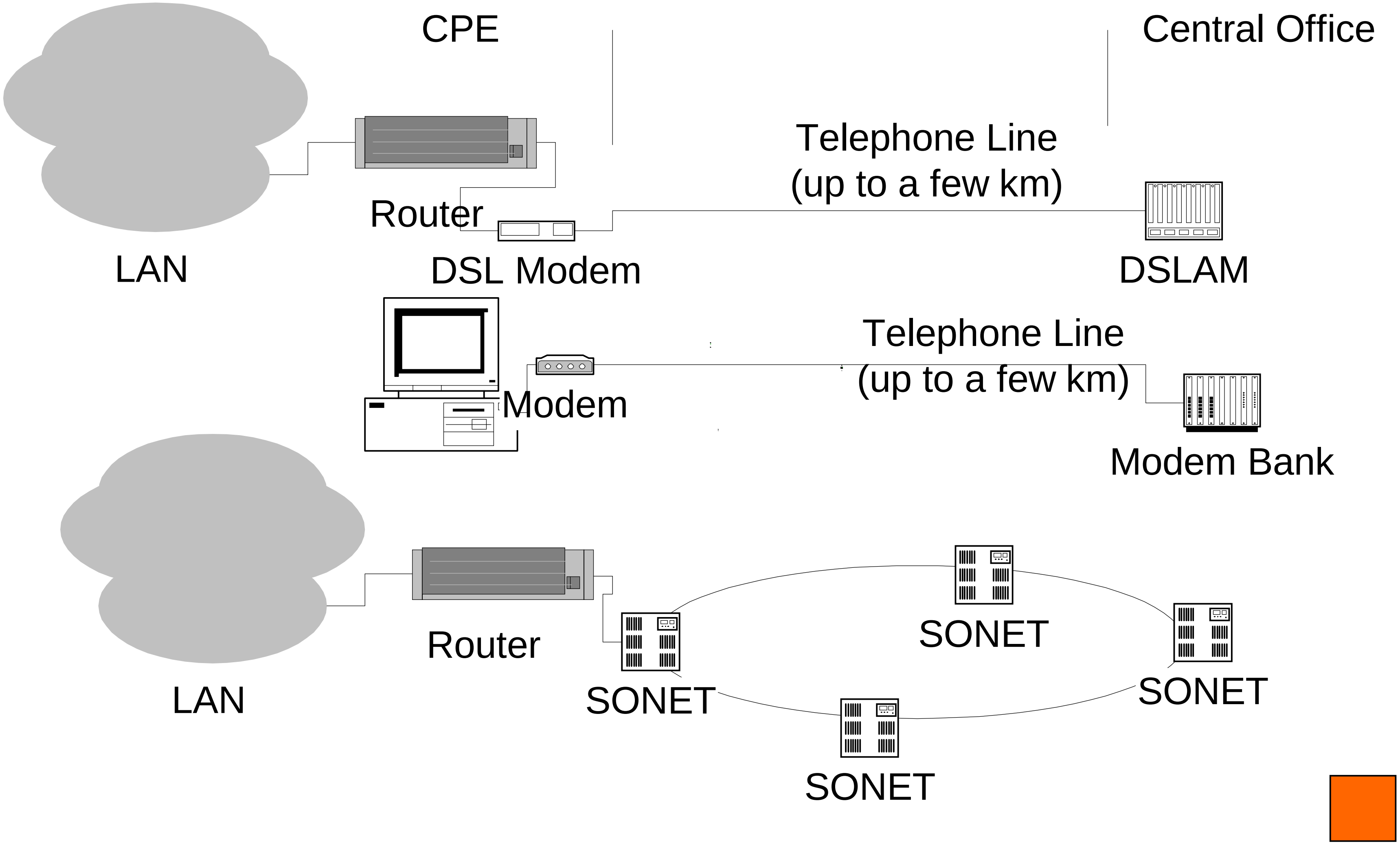


2 Unshielded Twisted Pairs (cat 5)
100 Mbps, up to 100m

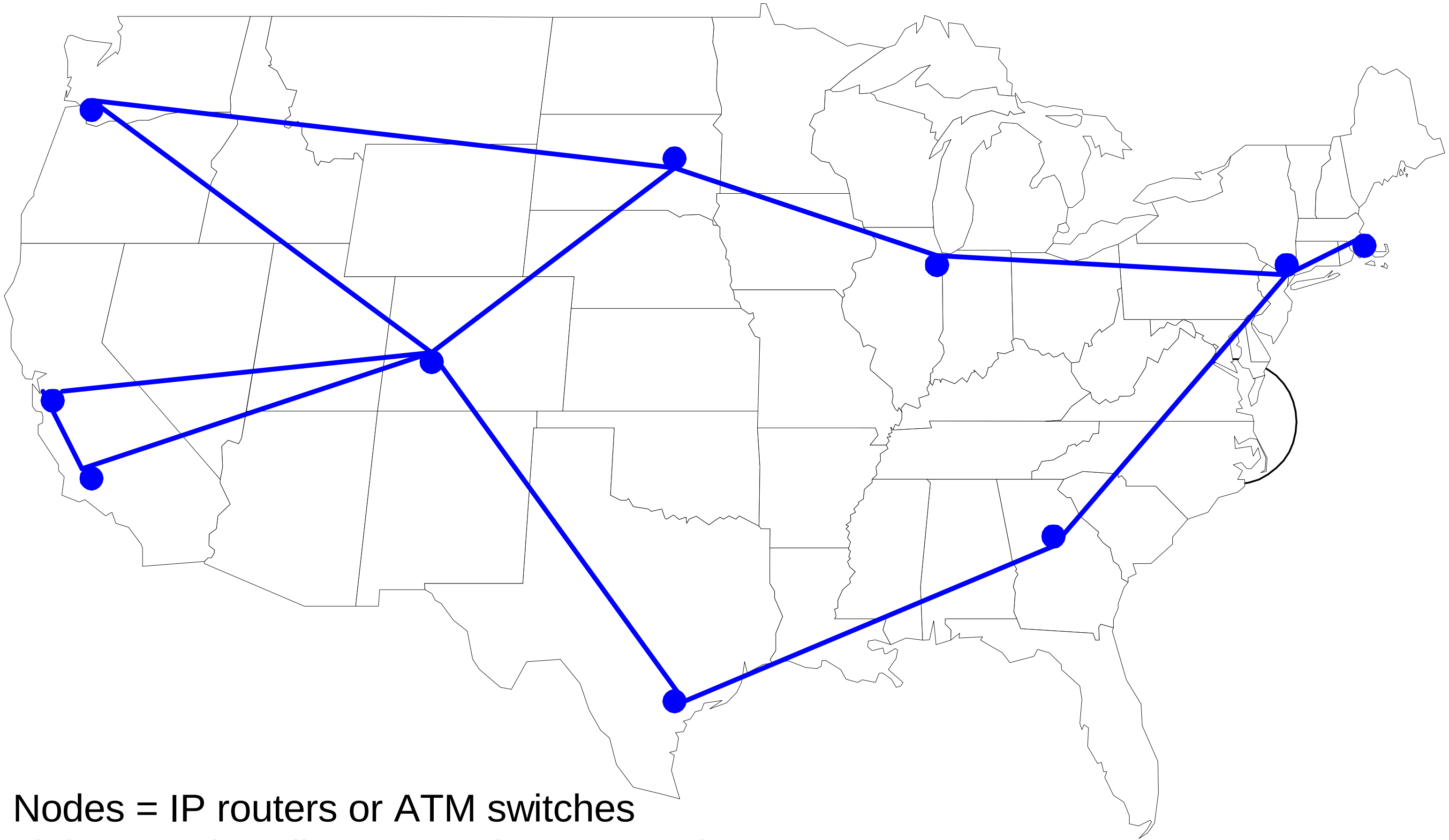
Ethernet Switch
24 ports at 100 Mbps

Router
(or Gigabit Switch)

MAN


CPE
Central Office
Telephone Line
(up to a few km)
Router
LAN
DSL Modem
DSLAM
Telephone Line
(up to a few km)
Modem
Modem Bank
LAN
Router
SONET
SONET
SONET
SONET

WAN



Nodes = IP routers or ATM switches

Links = Optical fibers: 155Mbps - 2.48Gbps

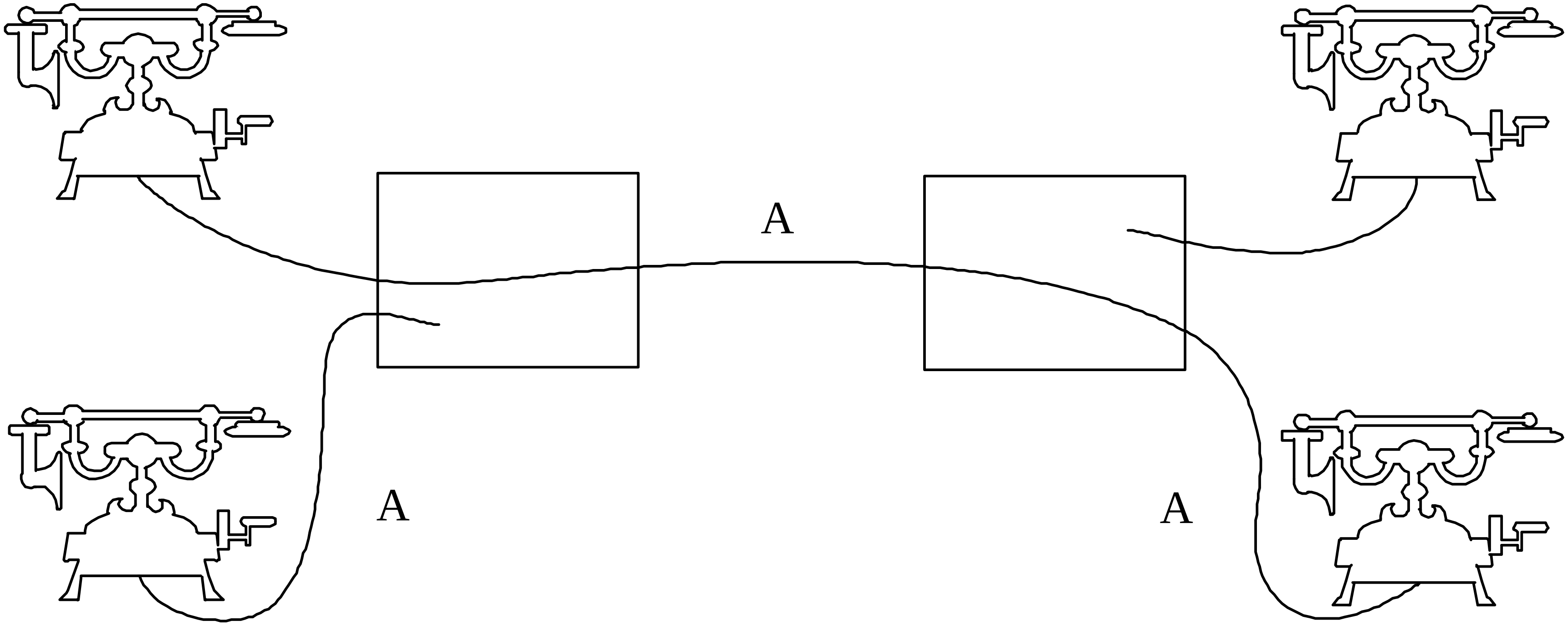
with WDM: one fiber = multiple (up to 128) bit streams

<

History

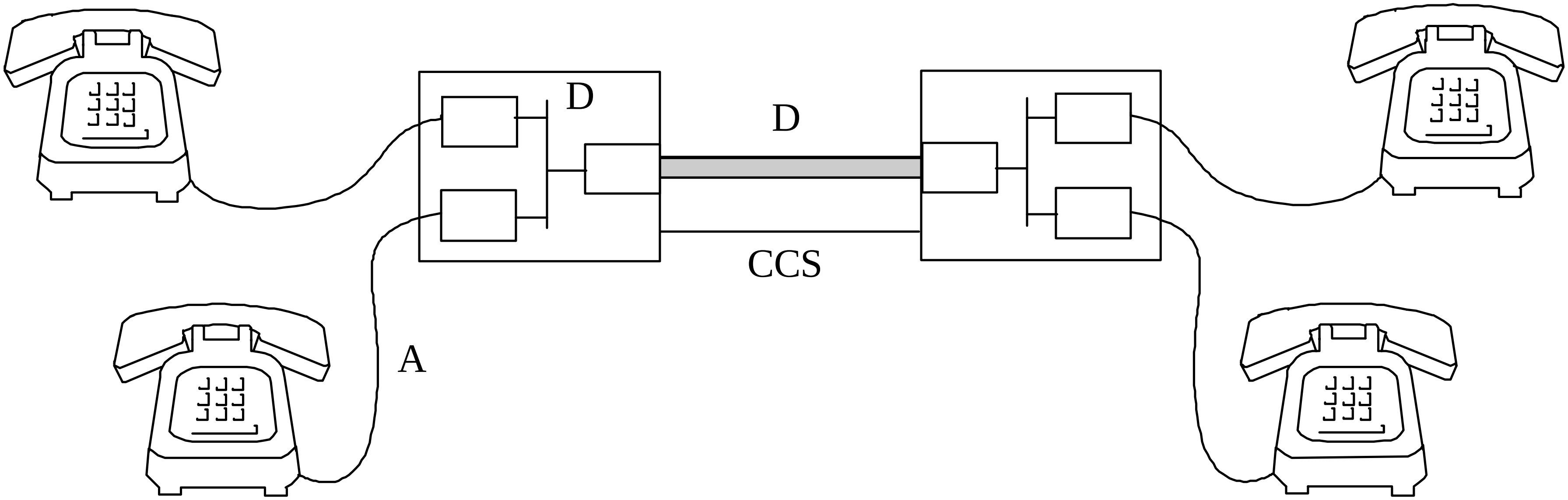
- Telephone
- Computers
- CATV
- Main Innovations

History: Telephone: 1880



A = Analog transmissions
Switching of “circuits” by operators
Signaling by conversation between users and operators

History : Telephone: 1988



Transmissions are analog (A) or digital (D)
Electronic circuit-switching
Signaling by data network (CCS)

History : Telephone: DCS

Medium	Signal	No. Voice Circuits	Rate (Mbps)
T1 paired cable	DS-1	24	1.5
T1C paired cable	DS-1C	48	3.1
T2 paired cable	DS-2	96	6.1
T3 coax, radio, fiber	DS-3	672	45.0
Coax, waveguide, radio, fiber	DS-4	4032	274.0

Notes:

- Rate of DS-1 =1.544 Mbps > 24x64kbps = 1.536 Mbps
- Rate of DS-3 = 45.0 Mbps > 28x(rate of DS-1) = 43.232 Mbps

Extra control bits are needed to accommodate differences in rates

History : Telephone: STS

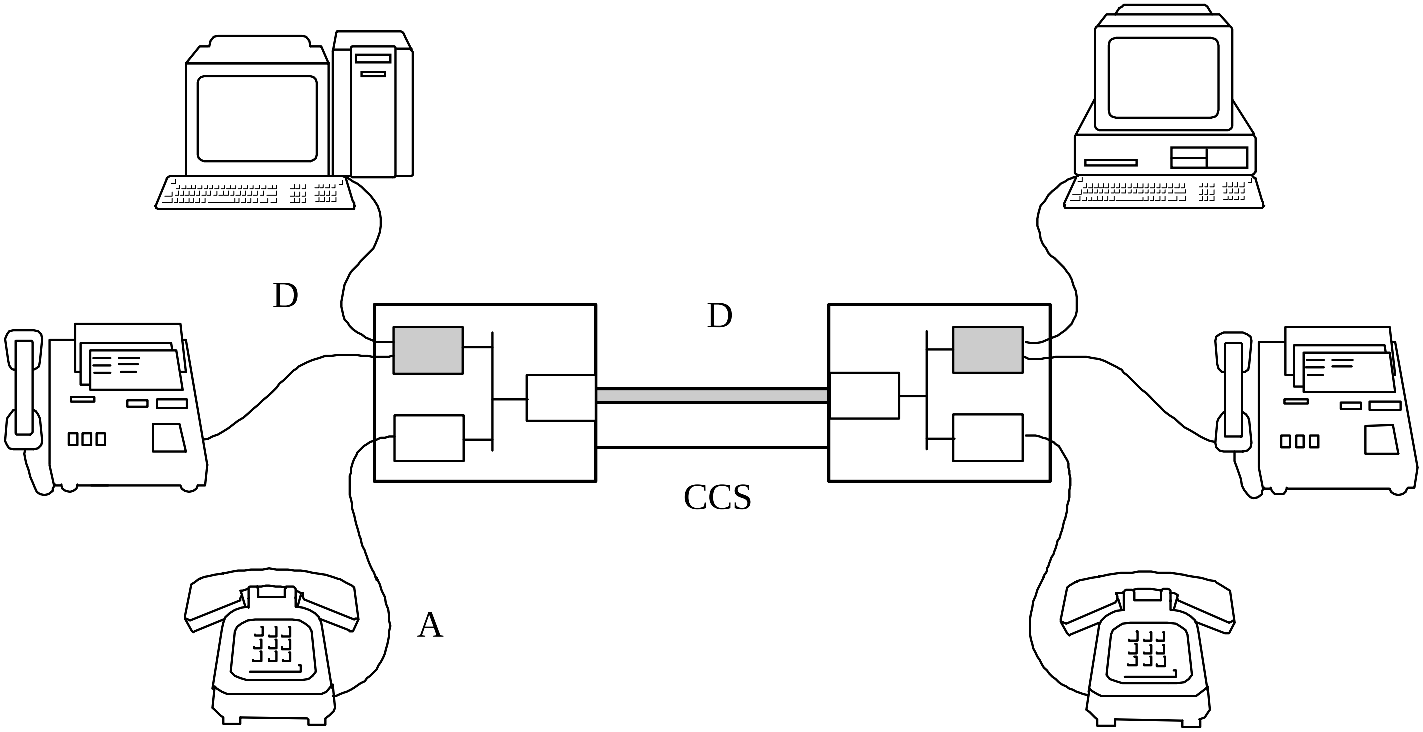
Carrier	Signal	Rate (Mbps)
OC-1	STS-1	51.840
OC-3	STS-3	155.520
OC-12	STS-12	622.080
OC-48	STS-48	2488.320
OC-192	STS-192	9853.280

T1.2

Notes:

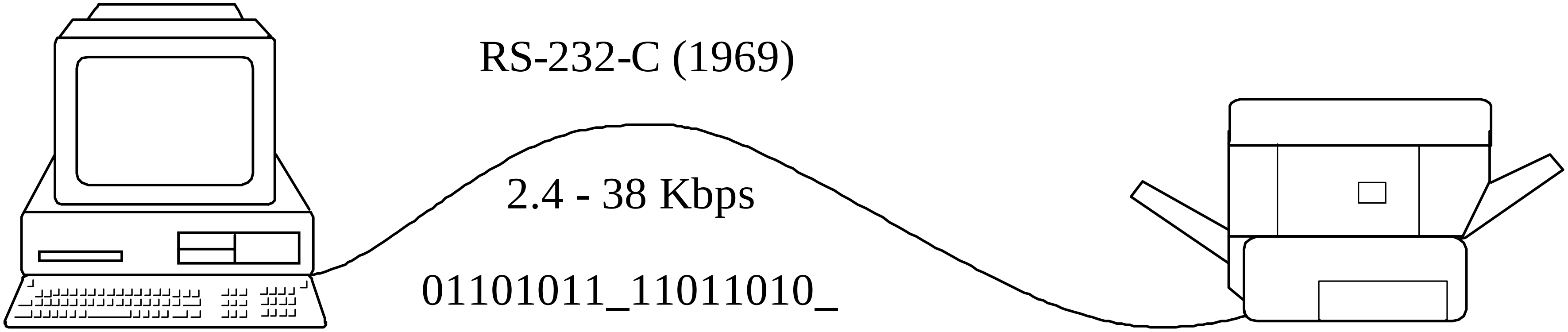
- Rate of STS-192 = 192x(rate of STS-1)
- No extra control bits are needed for multiplexing

History : Telephone: ISDN



The digital transmission is available to users
Basic service: 2B + D (B = 64kbps, D = 16kbps)

History : Computers: RS232 (serial line)



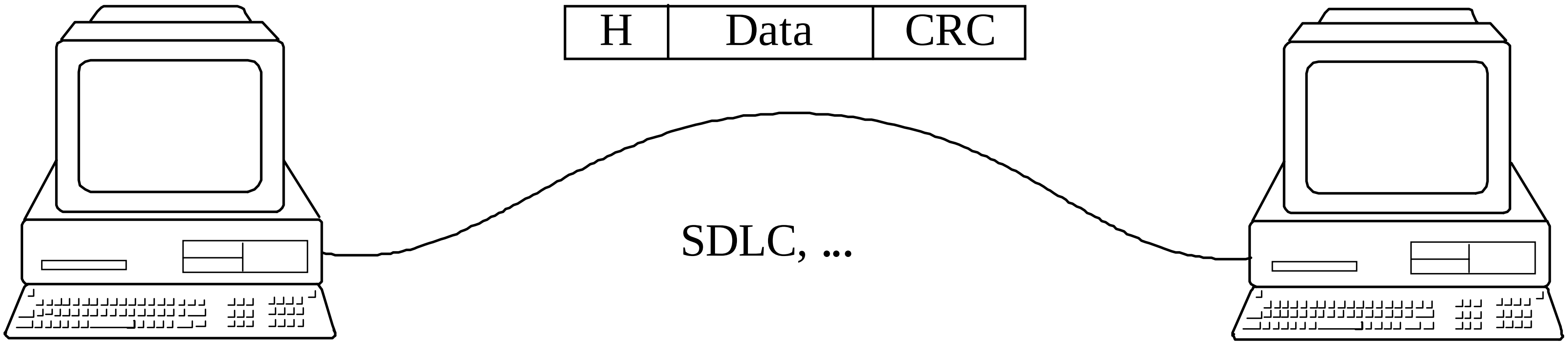
RS-232-C (1969)

2.4 - 38 Kbps

01101011_11011010_

Transmission of one 8-bit character at a time

History : Computers: SDLC



H	Data	CRC

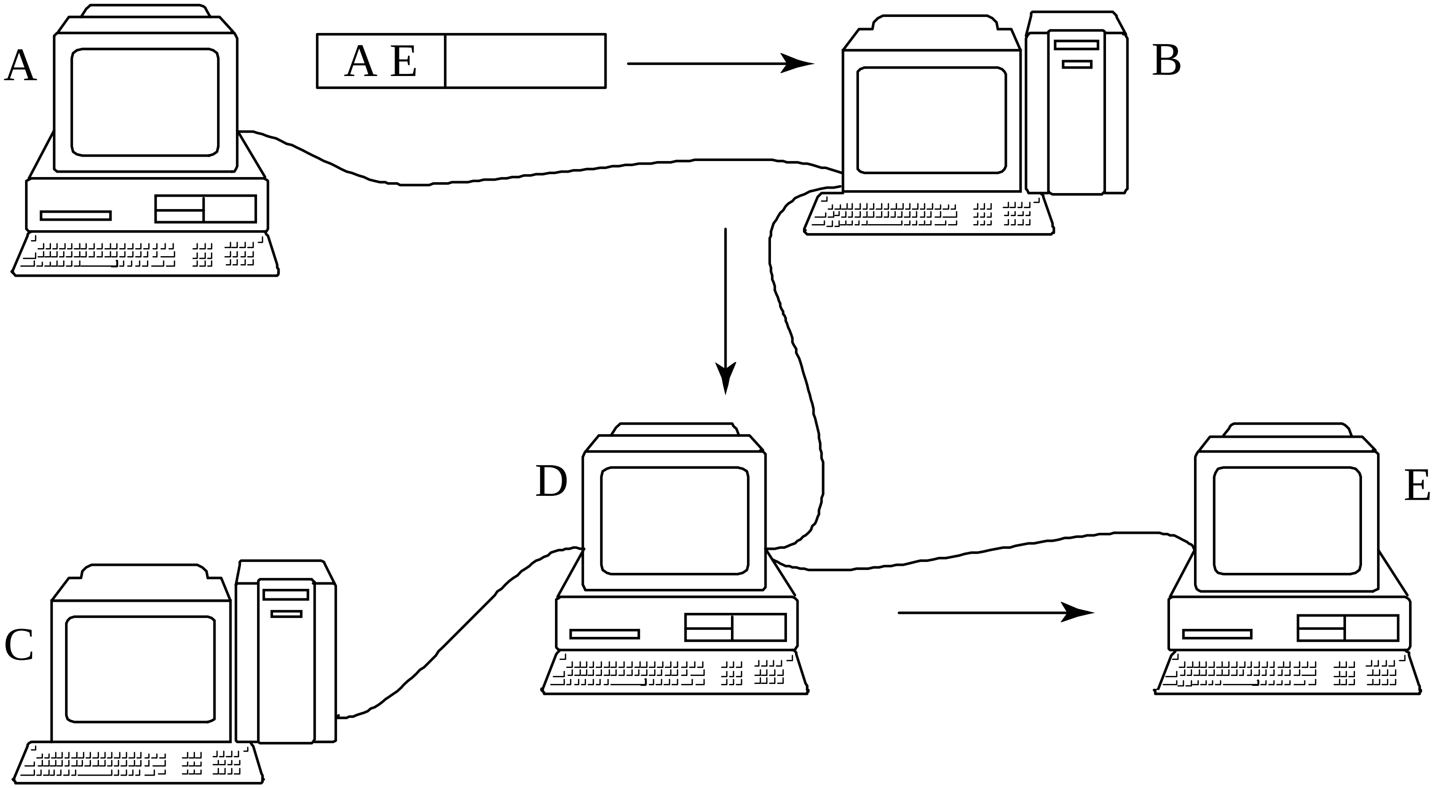
SDLC, ...

One “PACKET” at a time

Packet is framed to be identifiable by receiver

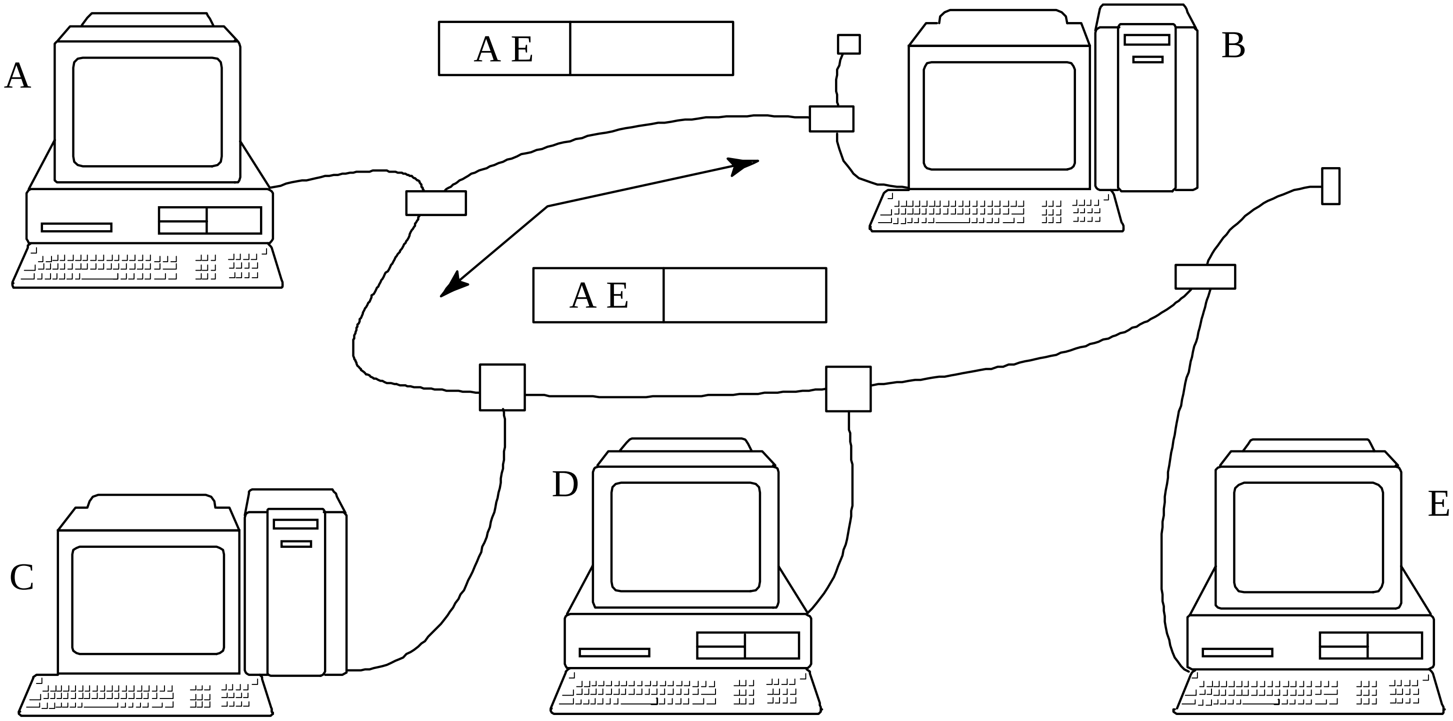
Control characters are added: header, CRC

History : Computers: Store-and-forward



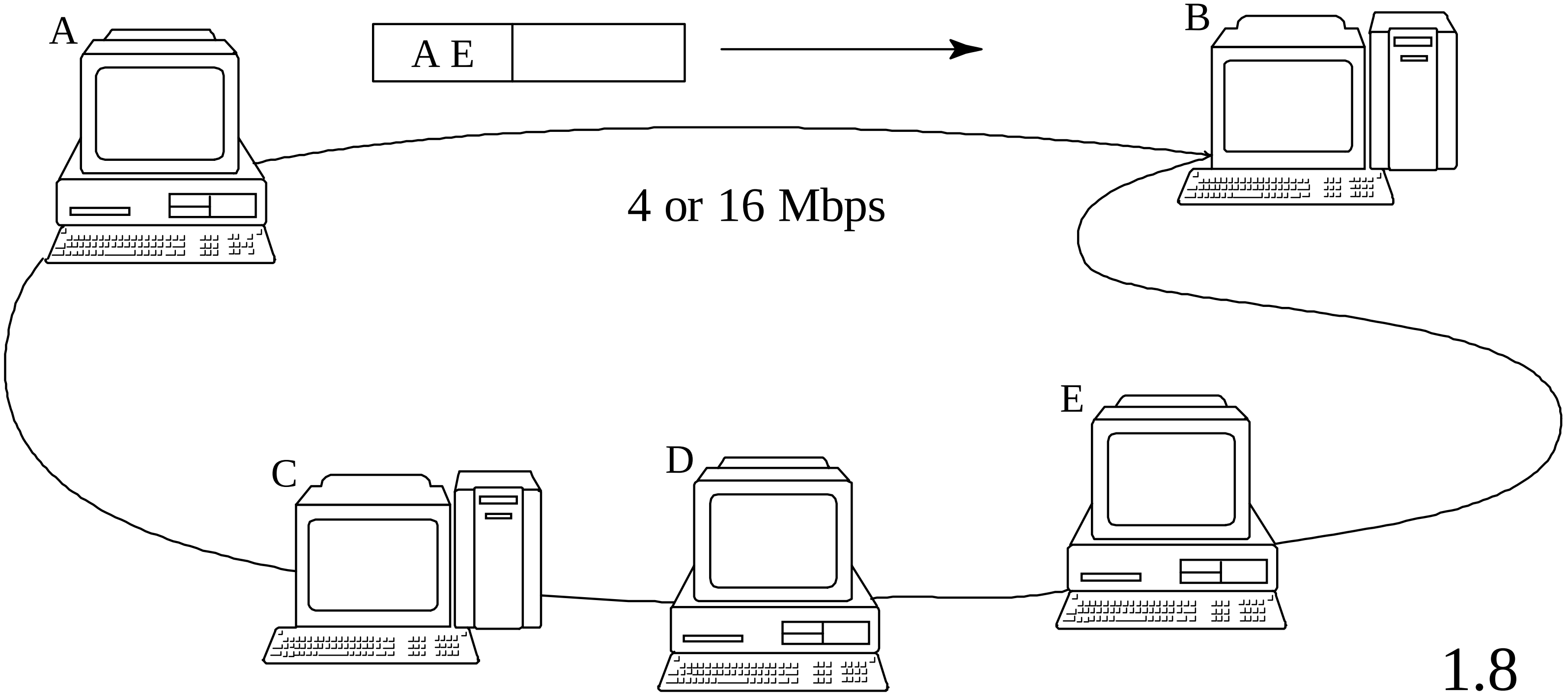
Packet contains source and destination “addresses”
Forwarding decision made by every computer

History : Computers: Ethernet



In original Ethernet (1974), computers broadcast packets
If collision, wait random time and try again

History : Computers: Token Ring

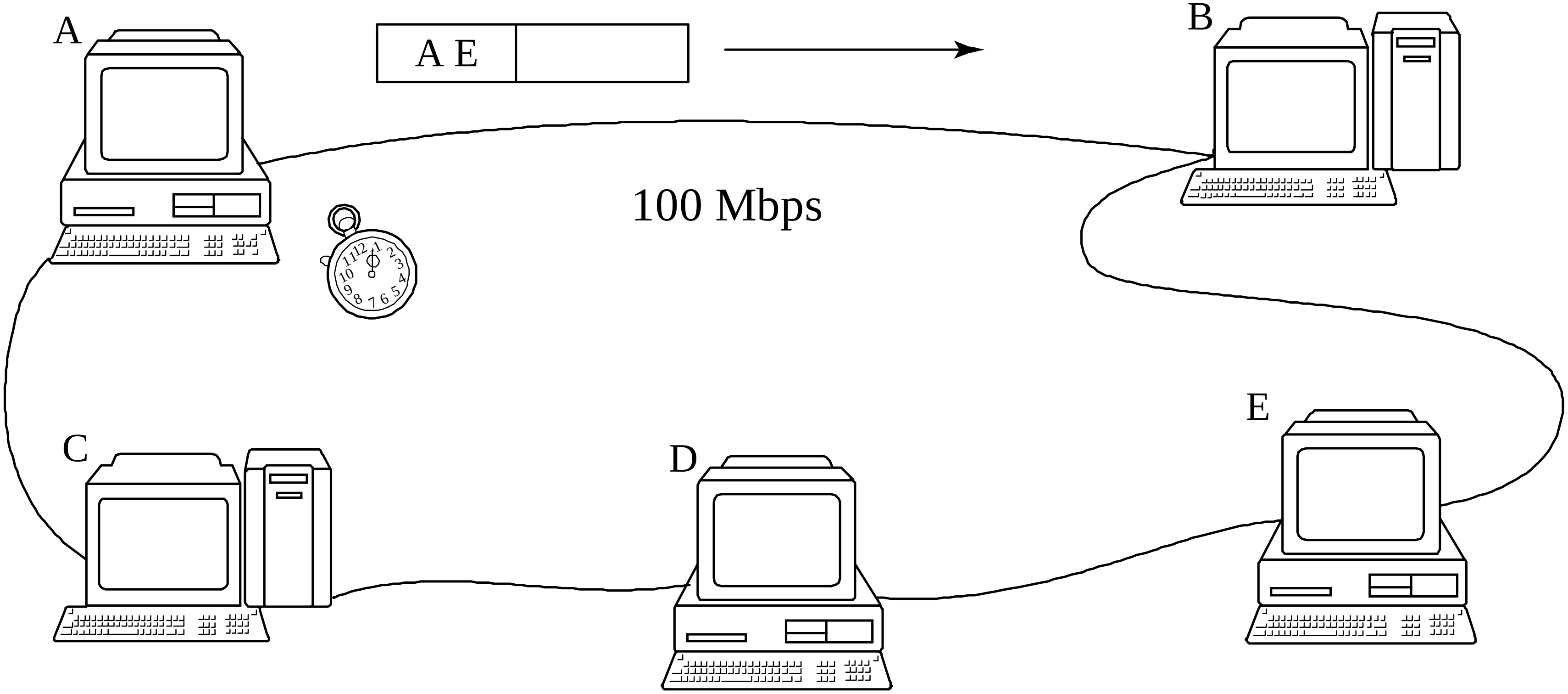


IBM (1980)

Ring, not store-and-forward

Access control by token-passing

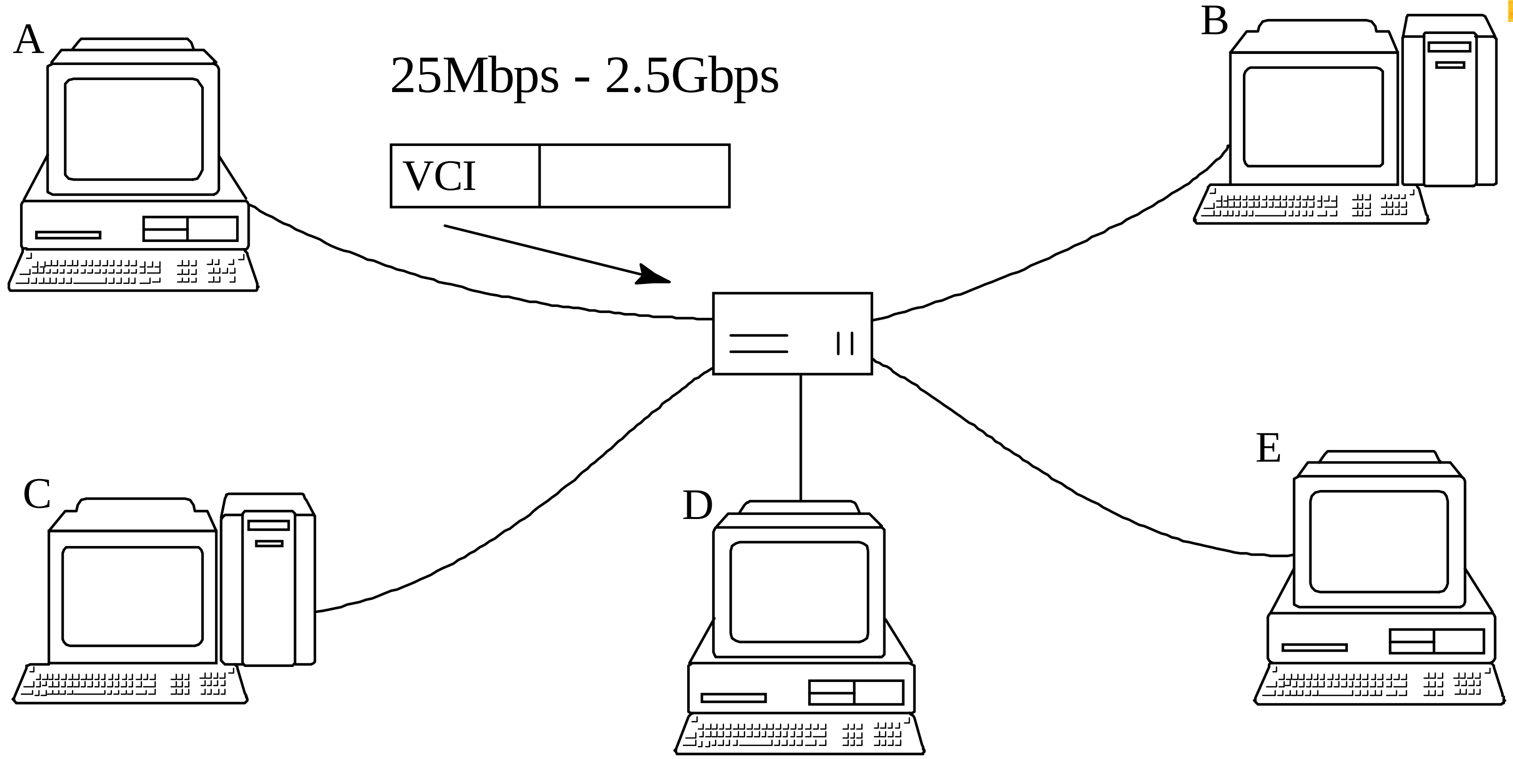
History : Computers: FDDI



Access control by timed token passing mechanism

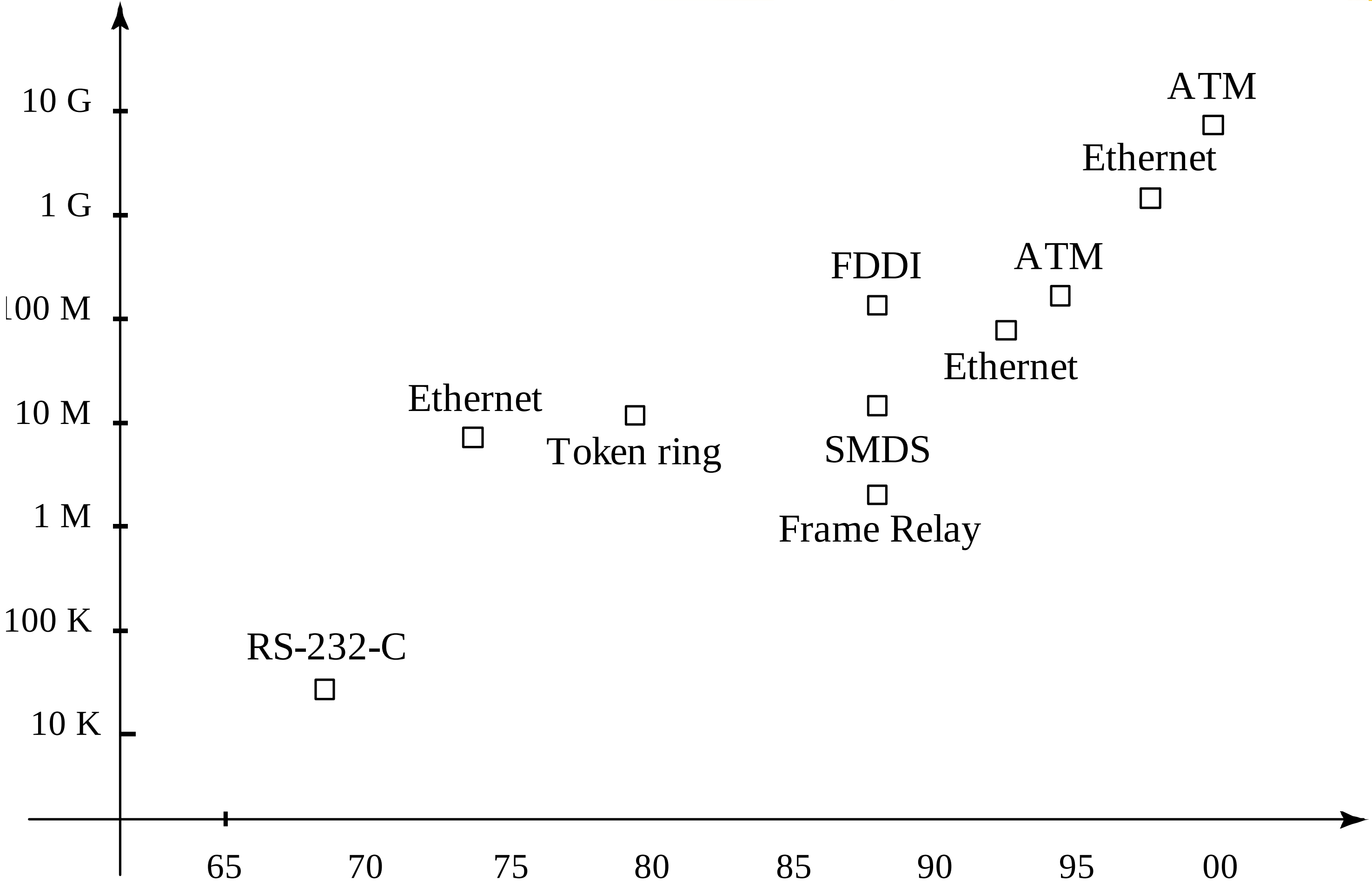
Bounds the media access time

History : Computers: ATM

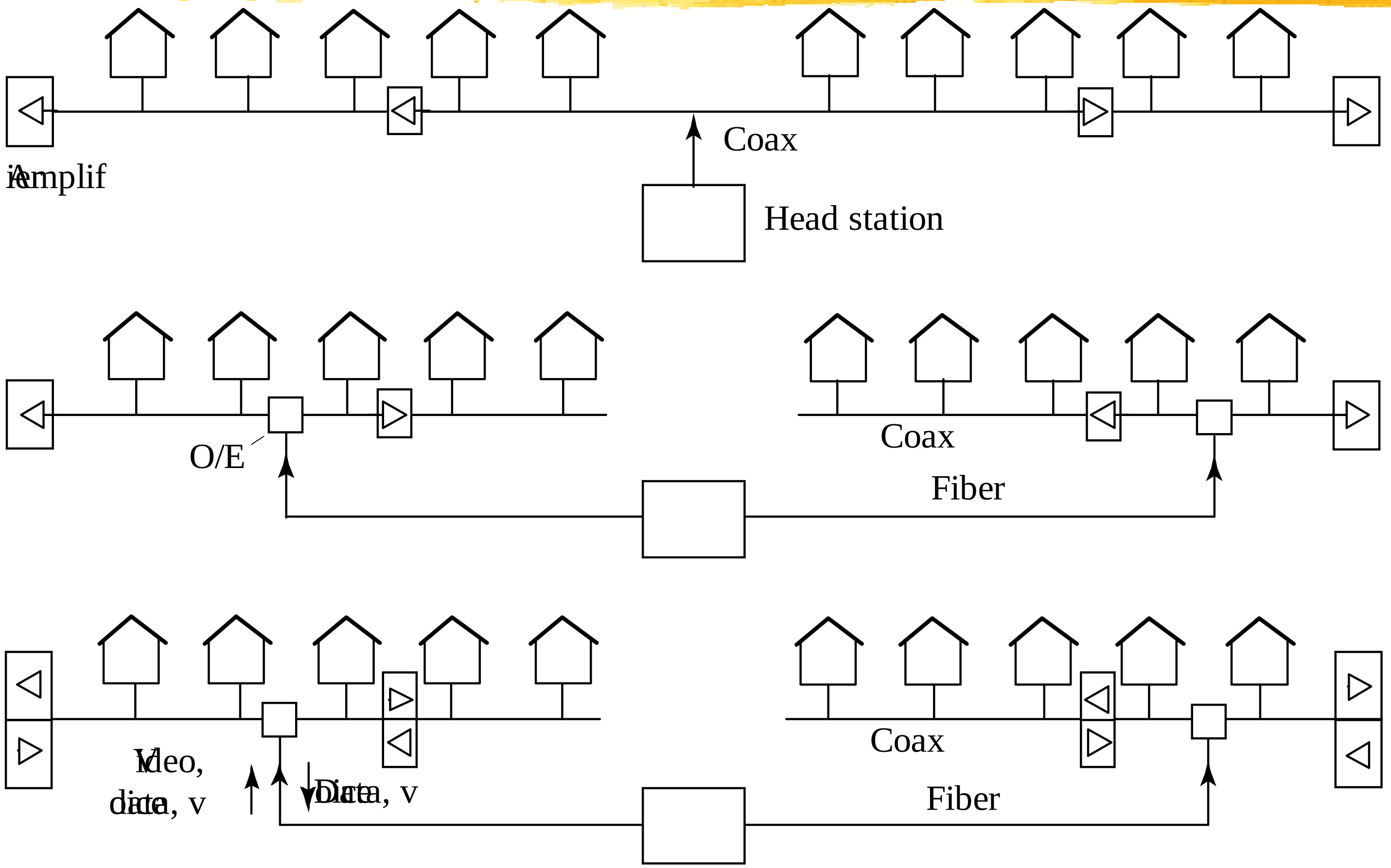


53-byte cells include a virtual circuit identifier
Many different classes of service are supported
Topologies other than stars are possible

History : Computers: Speeds



History : CATV



1. Analog broadcast; 2. Digital broadcast; 3. Digital two-way

History : Main Innovations

Telephone

Circuit switching
Separation control/voice
ISDN
Optical Links - SONET

Computer

Packet switching
Multiple Access
Layered Architecture
Internetworking
Integrated Services
ATM

CATV

Digitization/Compression
Fiber to the curb
Two-way links
Service integration

Wireless

Radio, TV broadcast
Cellular telephones
Wireless LANs
Cellular packets
Service integration

<

Technology: Links & Switches

- Links
- Switches

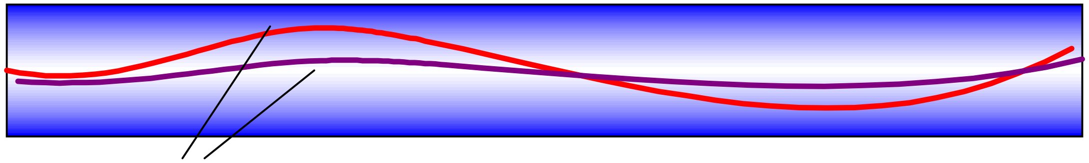
Links

- Transmits R bps over L kms with BER
- Examples:

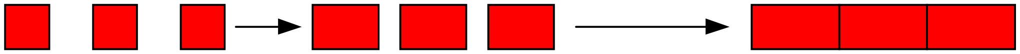
Link	R	L	BER	Note
Optical				
MMF	100Mbps	2km	E-12	E.g.: Fast Ethernet
SMF	10Gbps	100km	E-12	with dispersion compensation
Wireless				
	2Mbps	10m	E-8	E.g.: wireless Ethernet
	30kbps	5km	E-5	E.g.: GSM
UTP	100Mbps	100m	E-10	E.g.: Fast Ethernet
Cable	1Gbps	1km	E-10	E.g: CATV

Links (cont.)

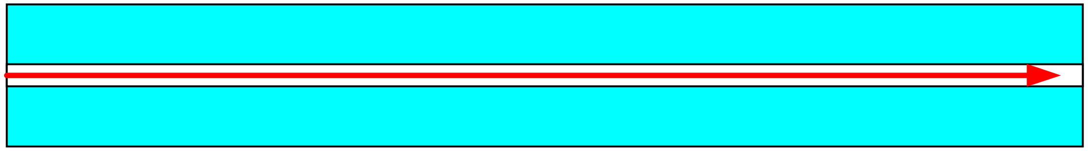
MMF:



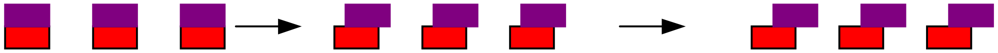
Different modes => Different speeds along fiber
=> modal dispersion



SMF:

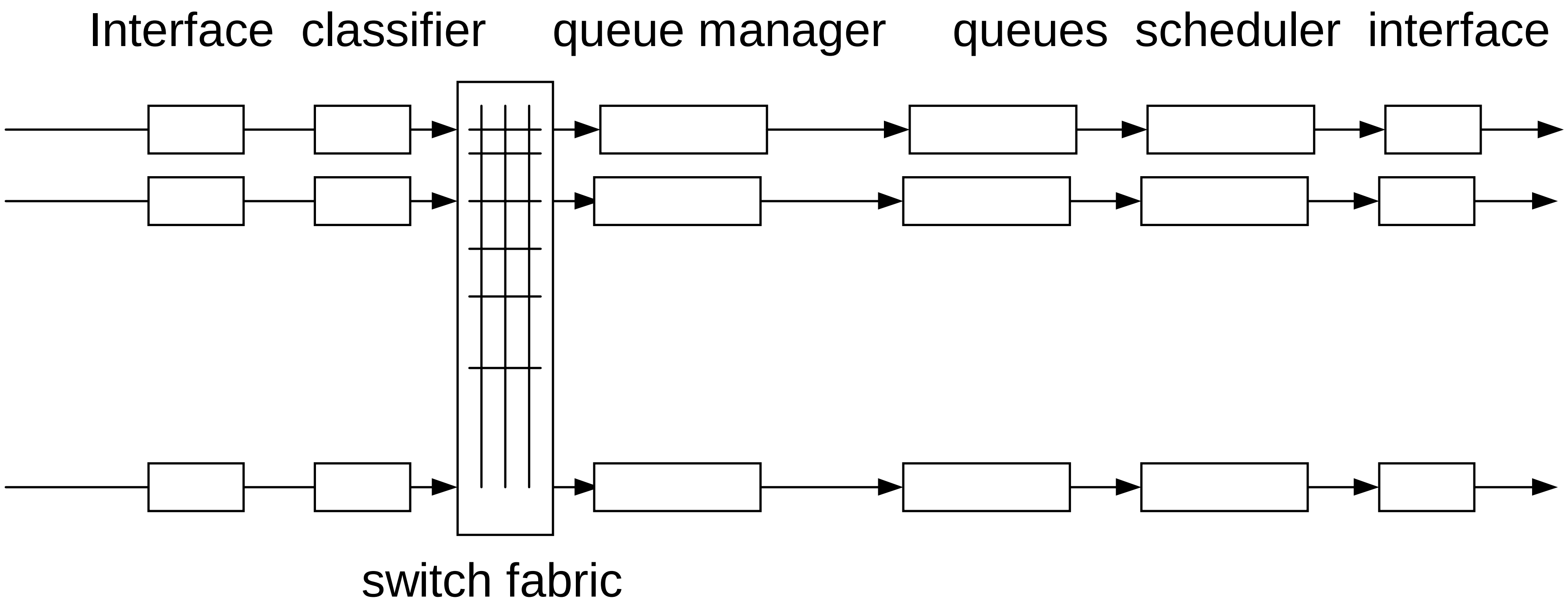


Single mode. However, small material dispersion



Attenuation: Limits L
Dispersion: Limits R×L

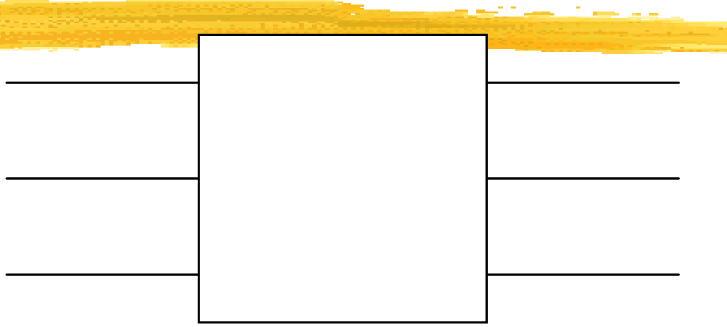
Switches



Interface classifier queue manager queues scheduler interface

switch fabric

Switches (cont.)



Characteristics

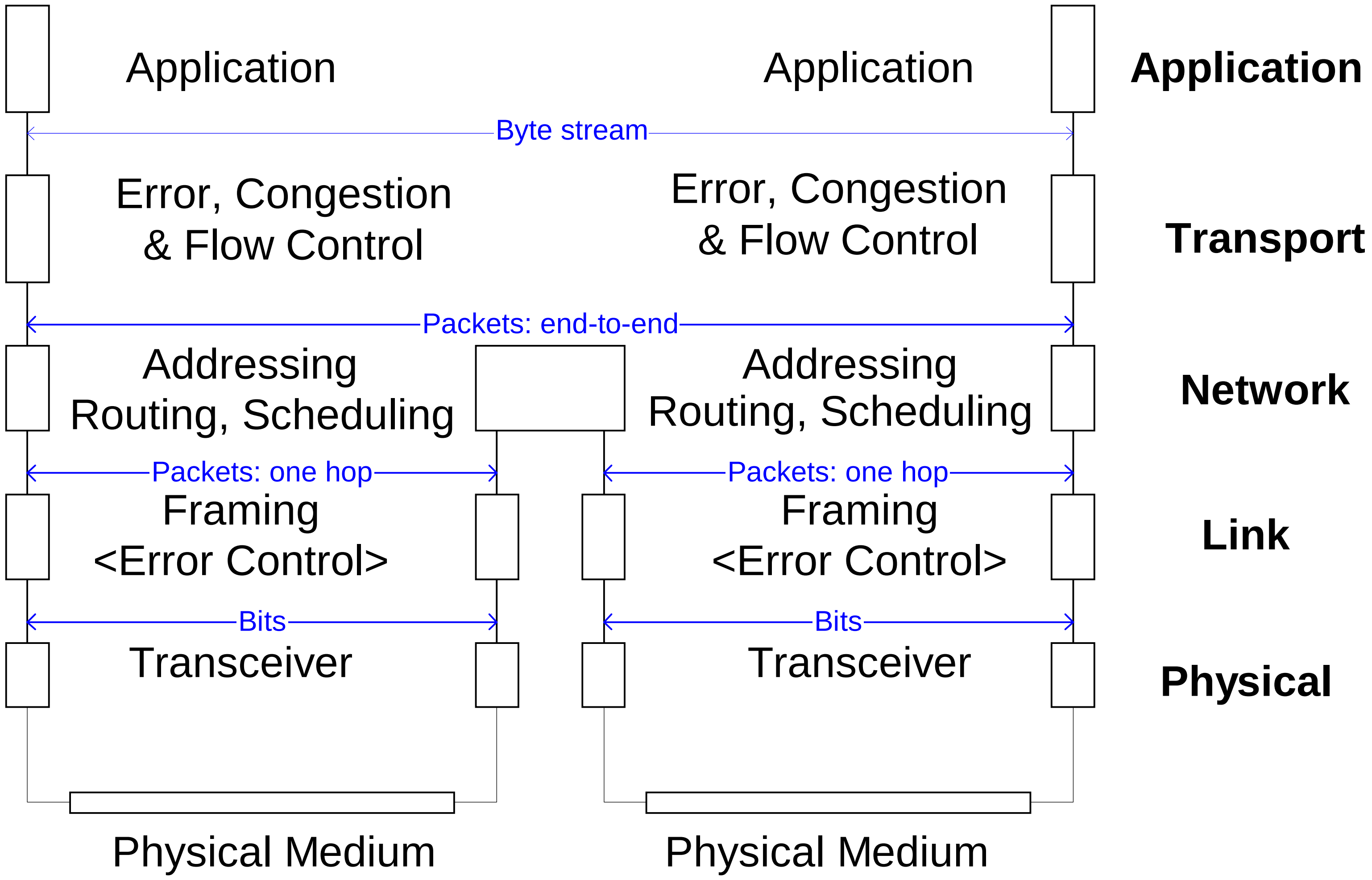
- Interfaces. Examples:
  100Mbps Ethernet; GE
  OC-12 SONET
  T1, T3, ...
- Number of interfaces
- Throughput (packets/s, bps)
- CoS, QoS, ...
- Protocols. Examples:
  Flow control
  Link Aggregation
  Spanning Tree
  OSPF, ...
  SNMP

< Mechanisms

- Overview
- Transceiver
- Framing
- Error Control

- Scheduling
- Flow Control
- Congestion Control
- Addressing
- Routing

Mechanisms: Overview

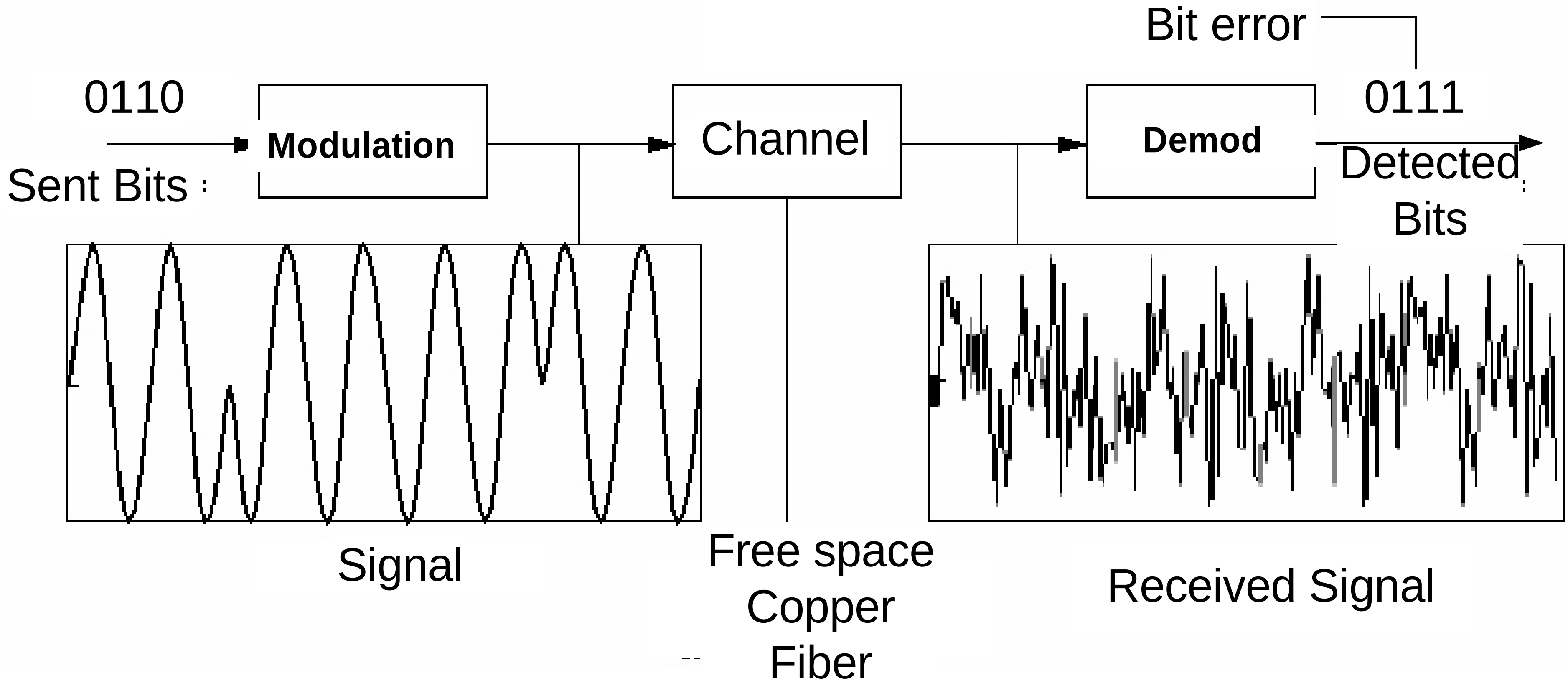


<

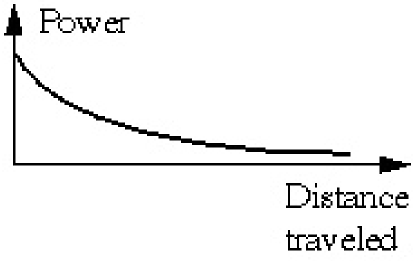
Transceiver:

- Bits -> Channel -> Bits
- Impairements
- Characteristics

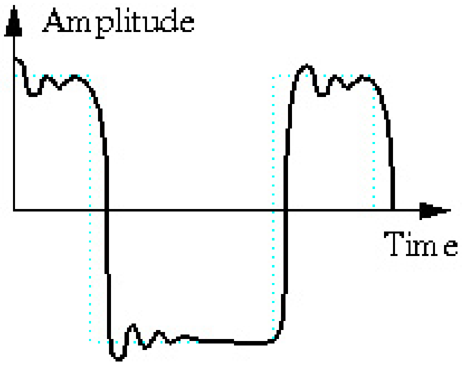
Transceiver: Bits --> Channel --> Bits



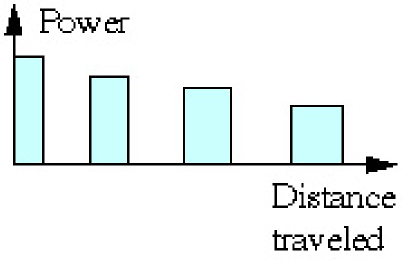
Transceiver : Impairments



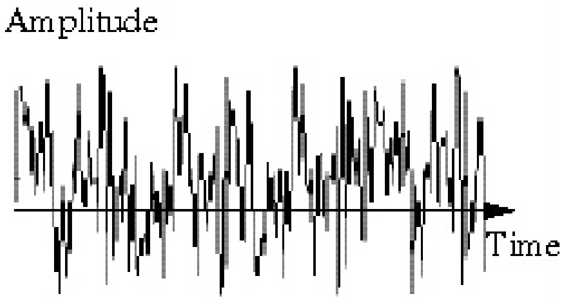
Attenuation



Distortion



Dispersion



Noise

Transceiver : Characteristics

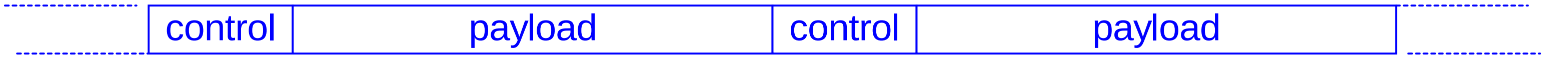
- Bit rate
- Error Rate
- Distance
- Examples (again)

Link	R	L	BER	Note
Optical				
MMF	100Mbps	2km	E-12	E.g.: Fast Ethernet
SMF	10Gbps	100km	E-12	with dispersion compensation
Wireless				
	2Mbps	10m	E-8	E.g.: wireless Ethernet
	30kbps	5km	E-5	E.g.: GSM
UTP	100Mbps	100m	E-10	E.g.: Fast Ethernet
Cable	1Gbps	1km	E-10	E.g: CATV

Framing

TDM:

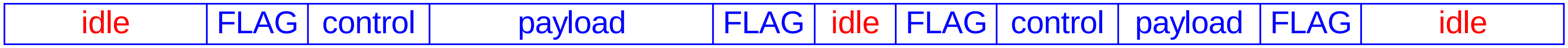
- periodic frames with constant size
- frames may be divided into channels
- examples: T1, T3, E1,..., SONET/SDH



control | payload | control | payload

Packets:

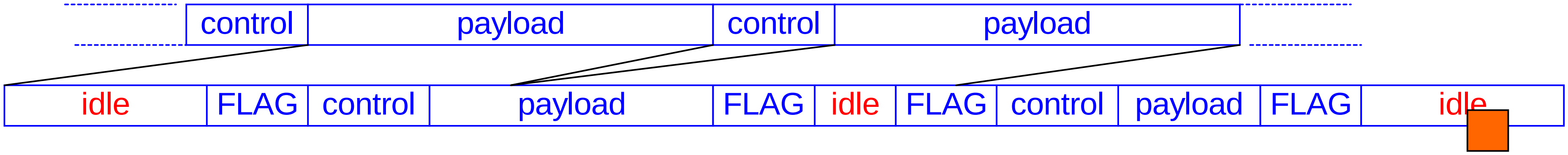
- asynchronous frames with constant or variable size
- frames have identification field
- examples: ATM, Ethernet, ...



idle | FLAG | control | payload | FLAG | idle | FLAG | control | payload | FLAG | idle

Packets over TDM

- asynchronous frames with constant or variable size in TDM palyload
- examples: PoS (packets over SONET)



control | payload | control | payload

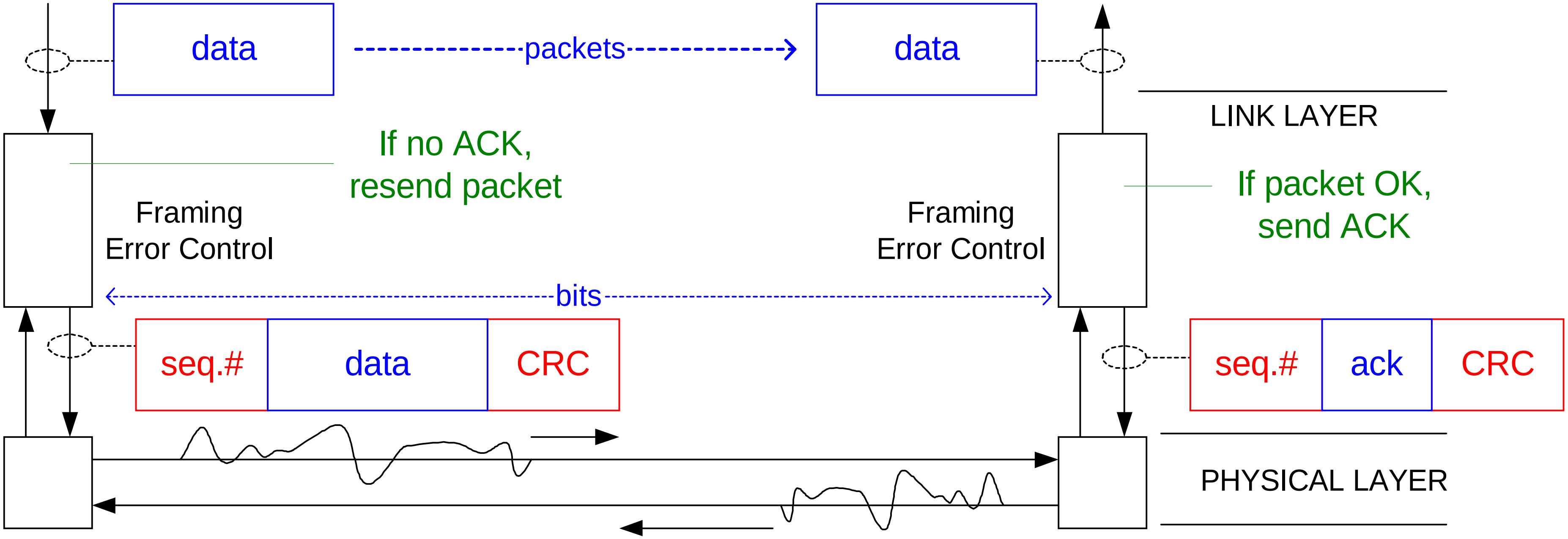
idle | FLAG | control | payload | FLAG | idle | FLAG | control | payload | FLAG | idle

<

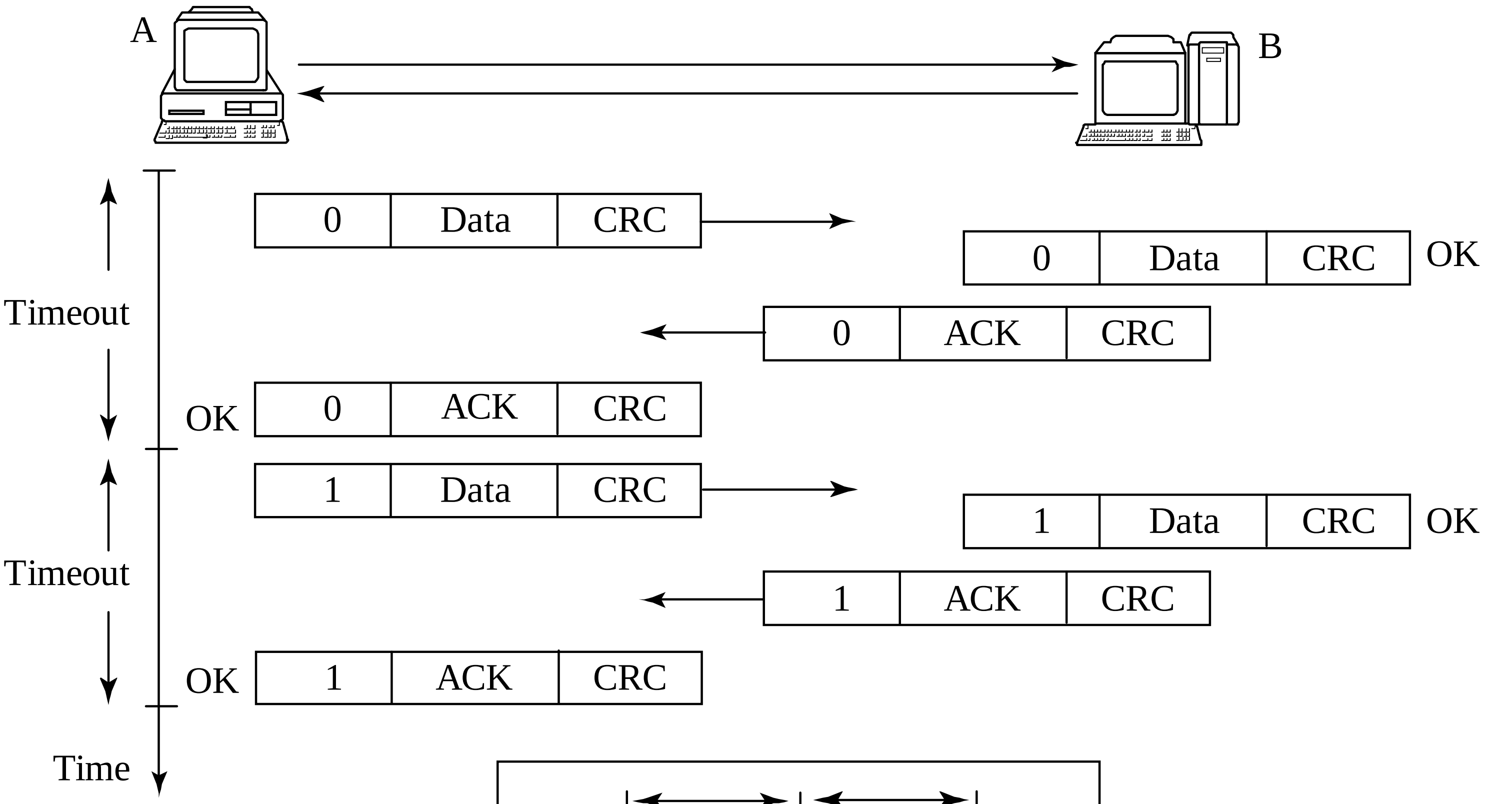
Error Control

- Example
- ABP - when no error
- ABP - when errors
- ABP - Efficiency
- Go Back N
- Go Back N - Efficiency / **Error Control Codes**

Error Control: Example



Error Control: ABP - when no error



A

B

Timeout

OK

Timeout

OK

Time

0	Data	CRC

0	Data	CRC

OK

0	ACK	CRC

0	ACK	CRC

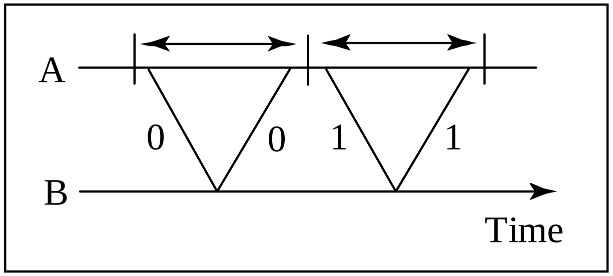
1	Data	CRC

1	Data	CRC

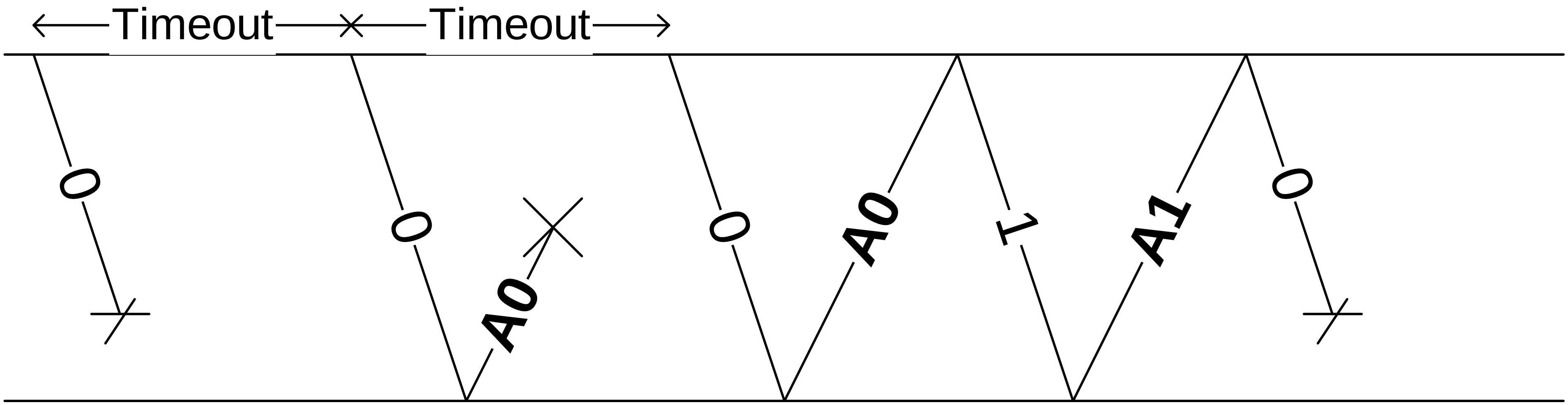
OK

1	ACK	CRC

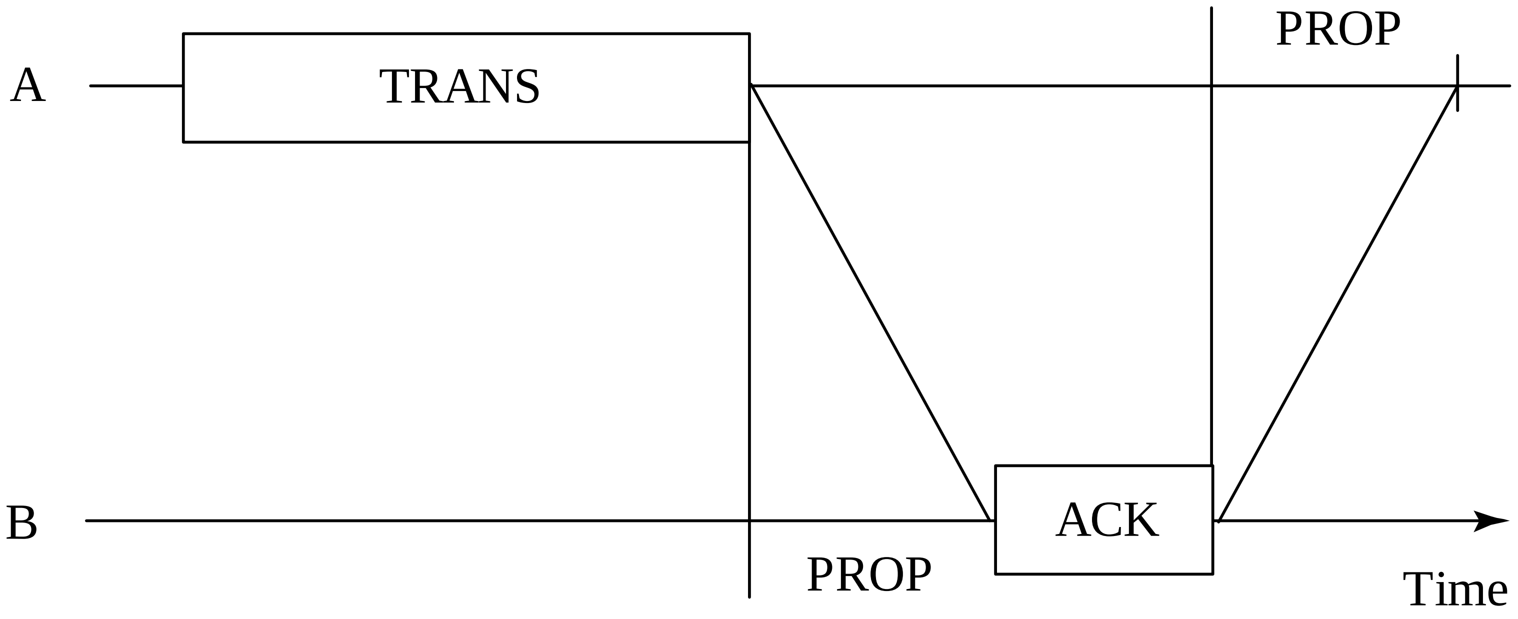
1	ACK	CRC



Error Control: ABP - when errors

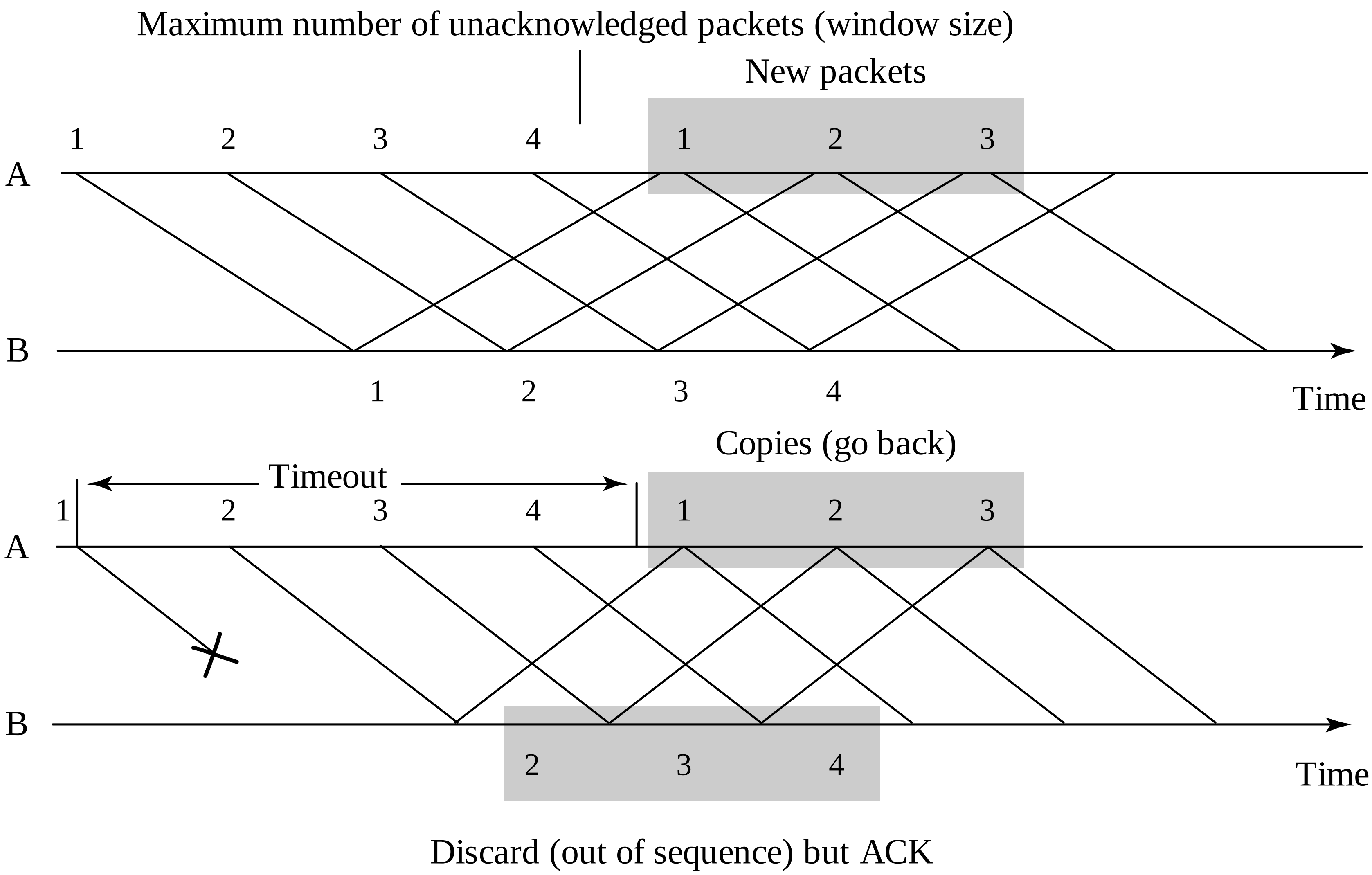


Error Control: ABP - Efficiency



$$\text{Efficiency} = \frac{\text{TRANS}}{\text{TRANS} + \text{ACK} + 2\text{PROP}}$$

Error Control: Go Back N



Maximum number of unacknowledged packets (window size)

New packets

Time

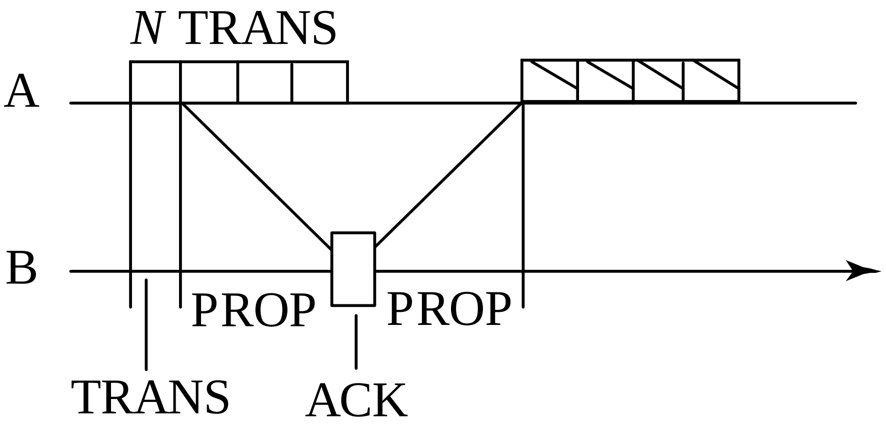
Copies (go back)

Timeout

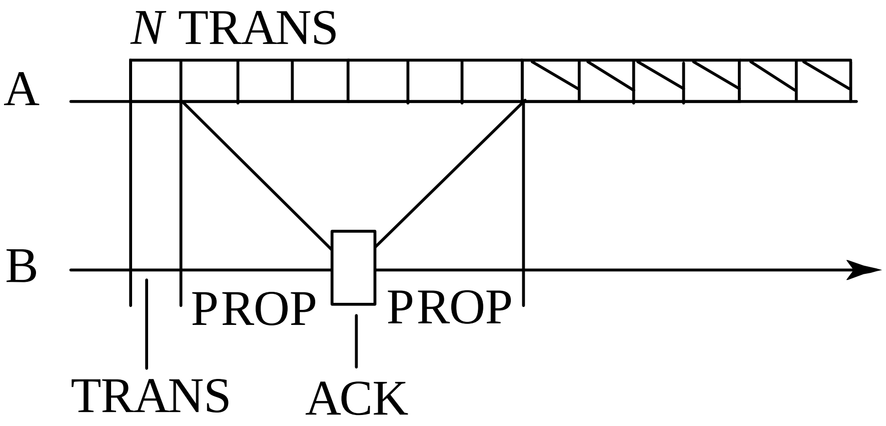
Time

Discard (out of sequence) but ACK

Error Control: Go Back N - Efficiency



Efficiency < 100%



Efficiency = 100%

Error Control: Error Control Codes

Error detection with odd parity:

0100100**1** → 010**1**00**1**

Parity bit

Transmission error

Error detected because parity is now even

Error correction with replication:

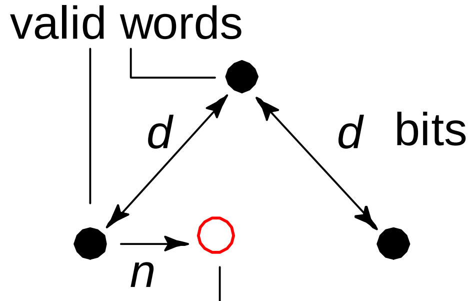
000 000 000 → 00**1** 000 000

Replication

Transmission error

Error can be corrected by majority decision on each bit

Properties of codes:



Erroneous word
- can be detected if n < d
- can be corrected if n < d/2

<

Scheduling

- Objectives
- TDM
- Statistical Multiplexing
- SM: Priority
- SM: DRR

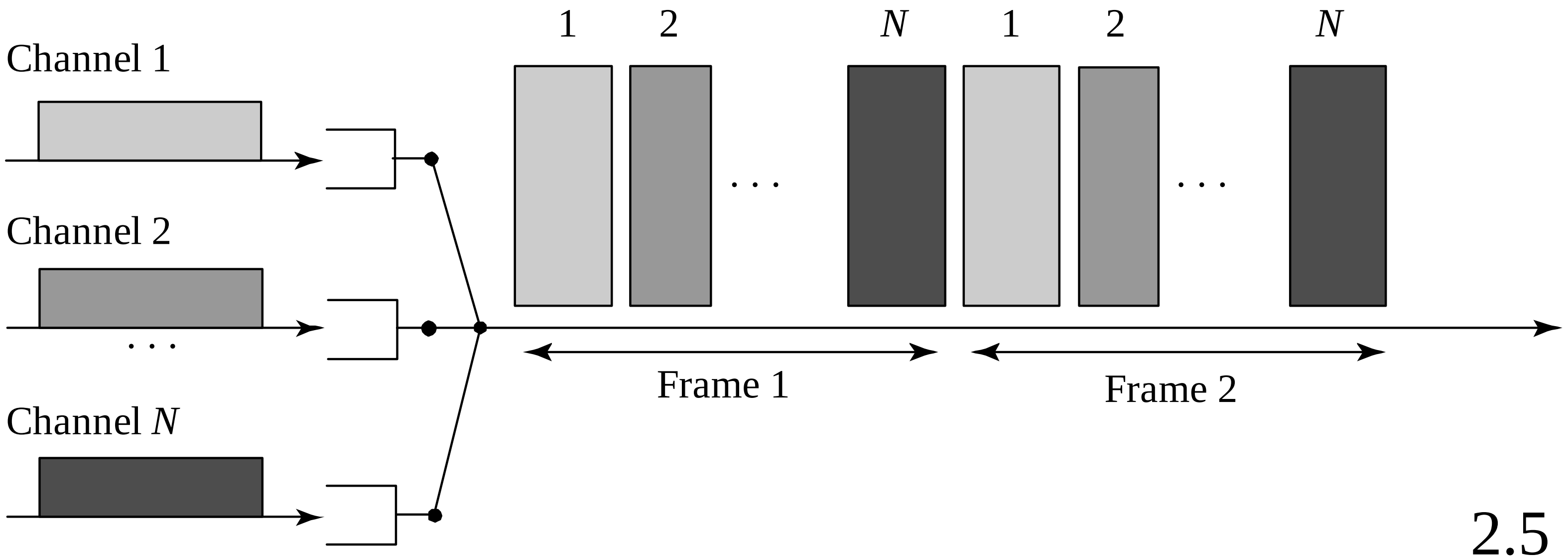
Scheduling: Objectives

Allocate the transmission rate of the output port to packets streams

Three possibilities:

- All packets treated equally
- Different classes of service: CoS
- Different qualities of service: QoS

Scheduling: Time Division multiplexing

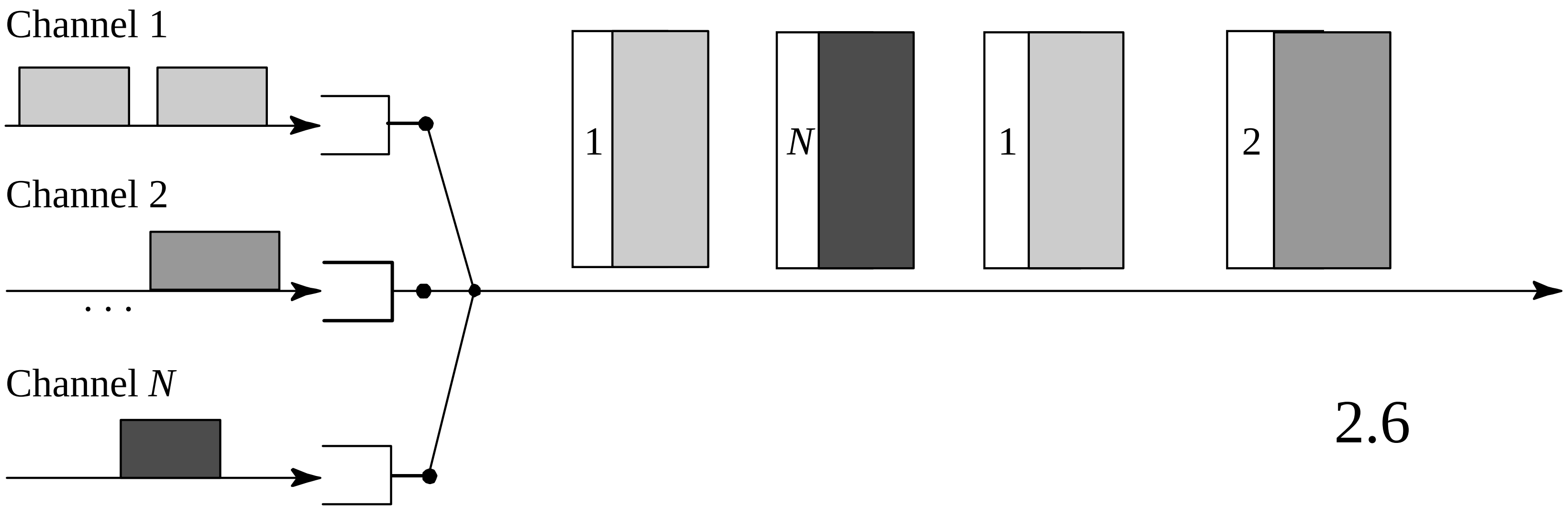


2.5

Transmission rate allocated into fixed fractions to channels

Note: Unused bandwidth cannot be made available to other channels

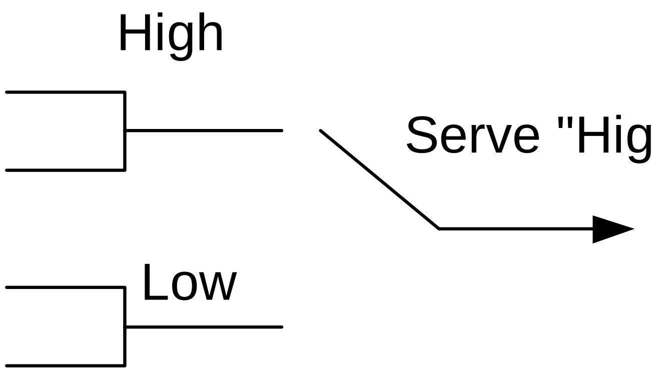
Scheduling: Statistical Multplexing



2.6

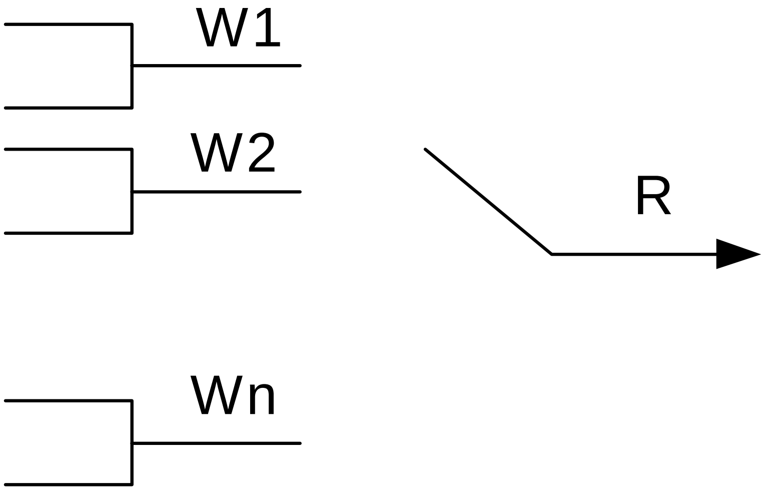
Transmission allocated “when needed” [here: First come, first served]
Note: Overhead to identify the channel in packet [may already be there]

Scheduling: Priority



- Lower delays for “High”
- Delays for “Low” larger than under FCFS
- Must make sure that “Low” is not shut out

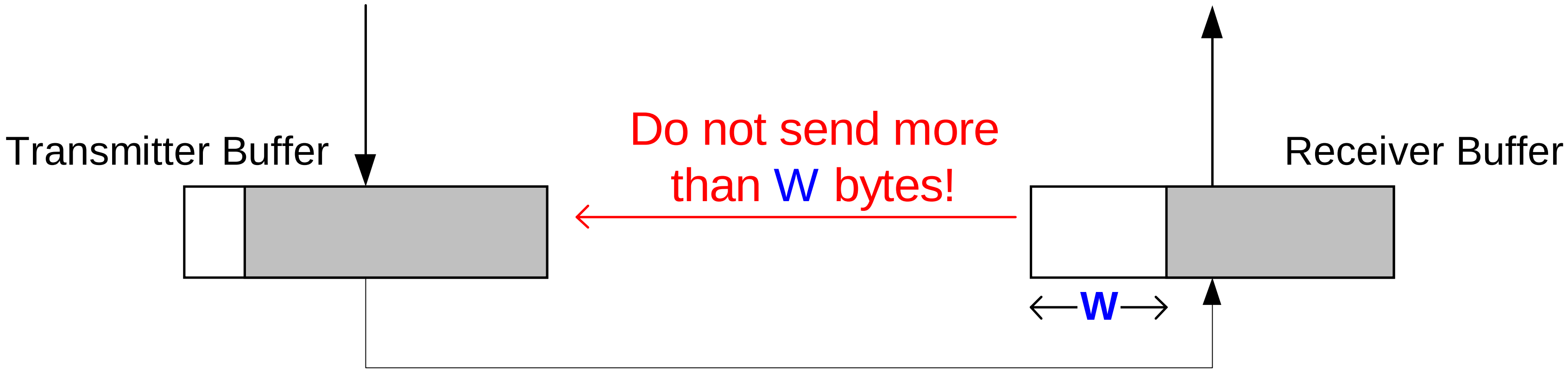
Scheduling: Deficit Round Robin



```
Initialize: $D1 = D2 = … = Dn = 0$
For $i = 1, …, n, 1, …, n, ….$
  Serve up to $Wi - Di$ bytes from queue i [skip to i + 1 when empty]
  Complete packet with x bytes, $Di = x$
```

Queue i gets at least RWi/(W1 + … + Wn); reuse available bandwidth

Flow Control



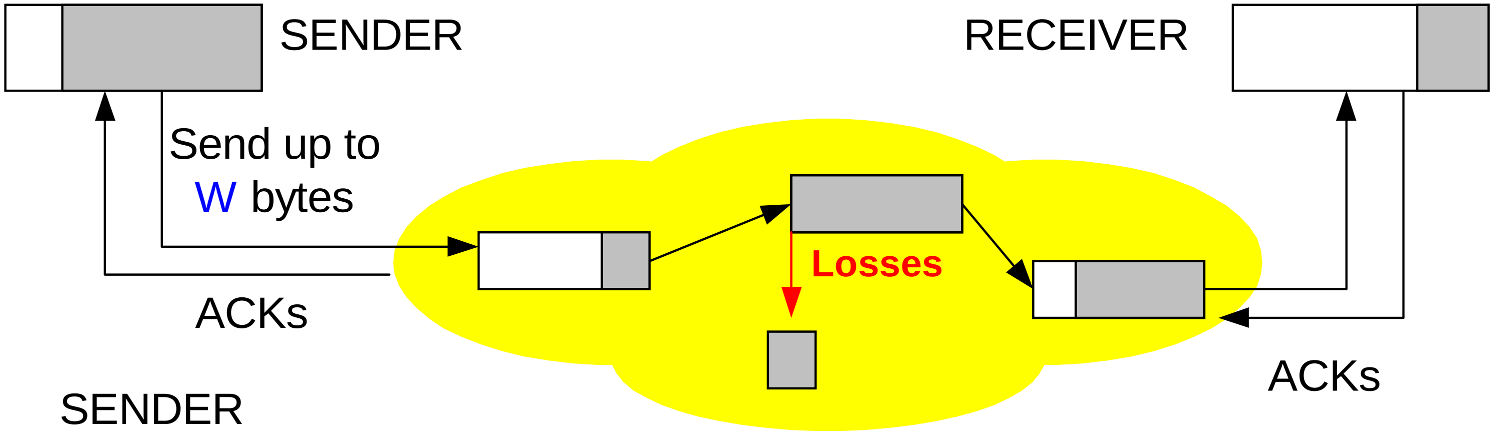
W = receiver advertised window

<

Congestion Control

- End-to-end
- Open Loop
- Explicit switch feedback

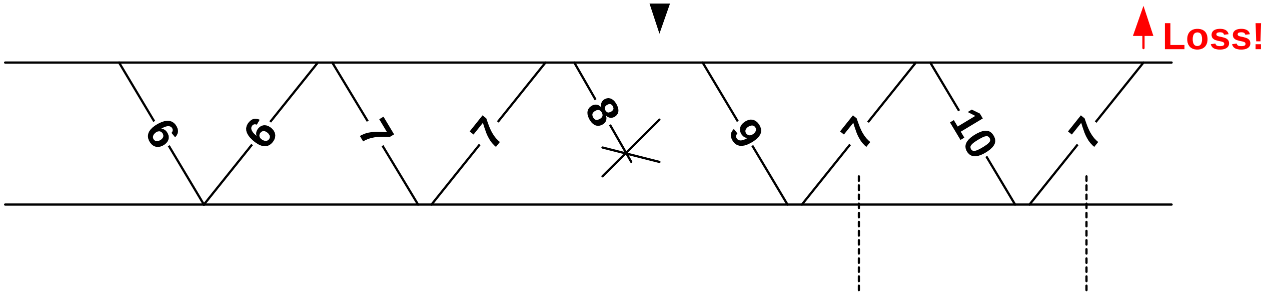
Congestion Control: End-to-end



SENDER

- reduces W when it detects losses
- increases W when it does not detect losses

How to detect losses: duplicated ACKS or timeout



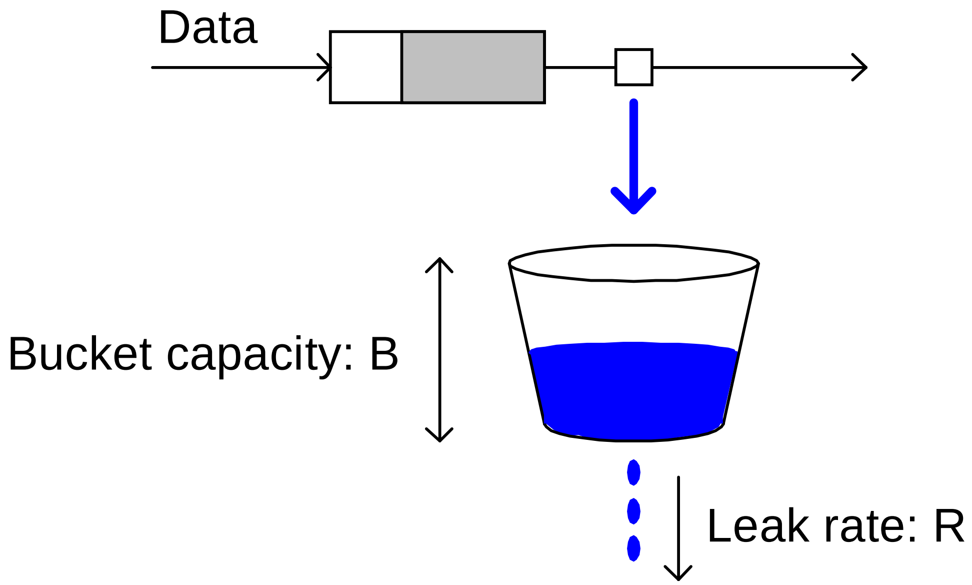
ACK# = last packet
received in sequence

Congestion Control: Open Loop (1/2)

Leaky bucket controller

Add one unit of fluid in bucket per byte transmitted

Can transmit packet only if bucket does not overflow



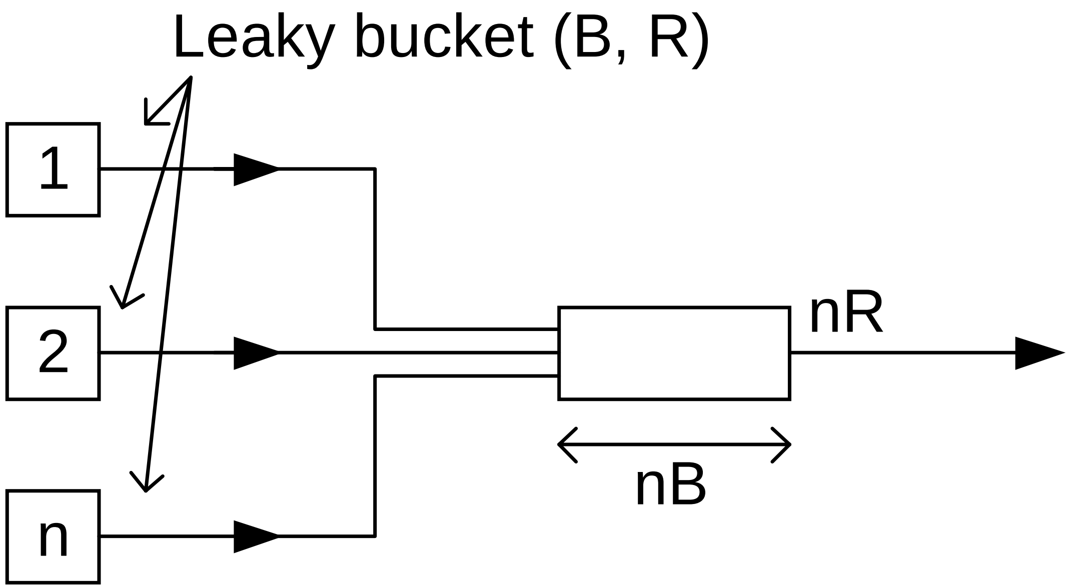
Note: Shaping (delay packets until there is space in bucket)

Marking (mark “excess” packets)

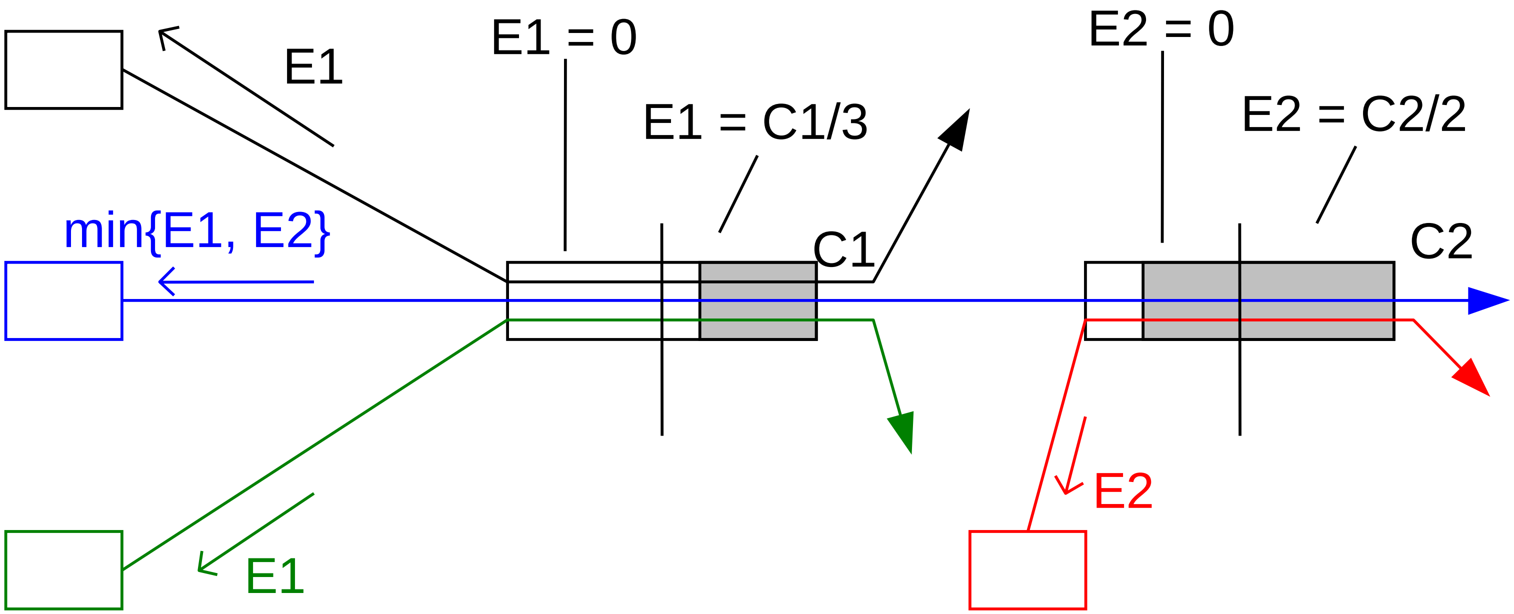
Police (discard excess packets)

Congestion Control: Open Loop (2/2)

Regulate all the flows to avoid congestion:



Congestion Control: Explicit Feedback



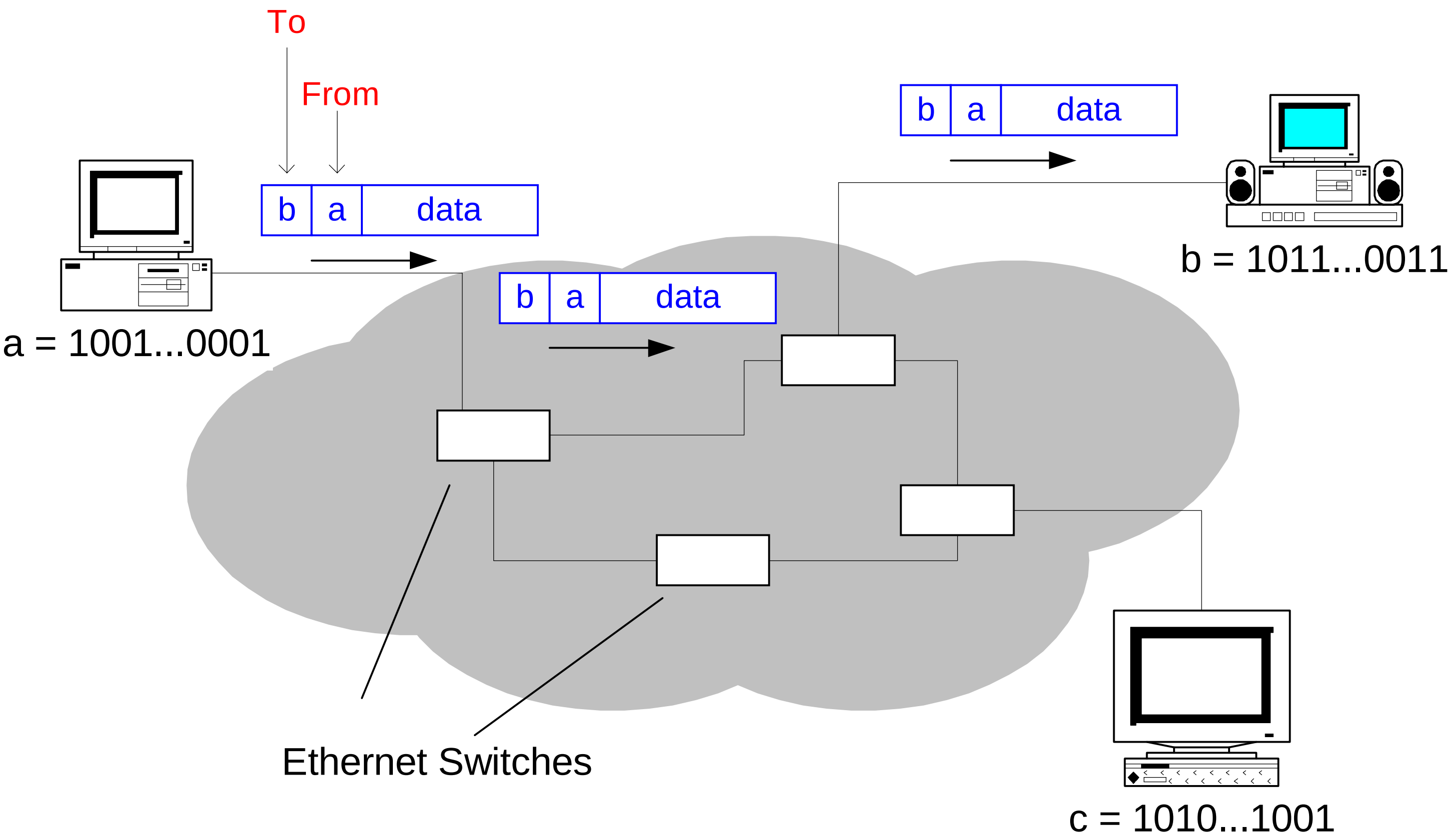
< Addressing

- Local
- Global: Internetworking

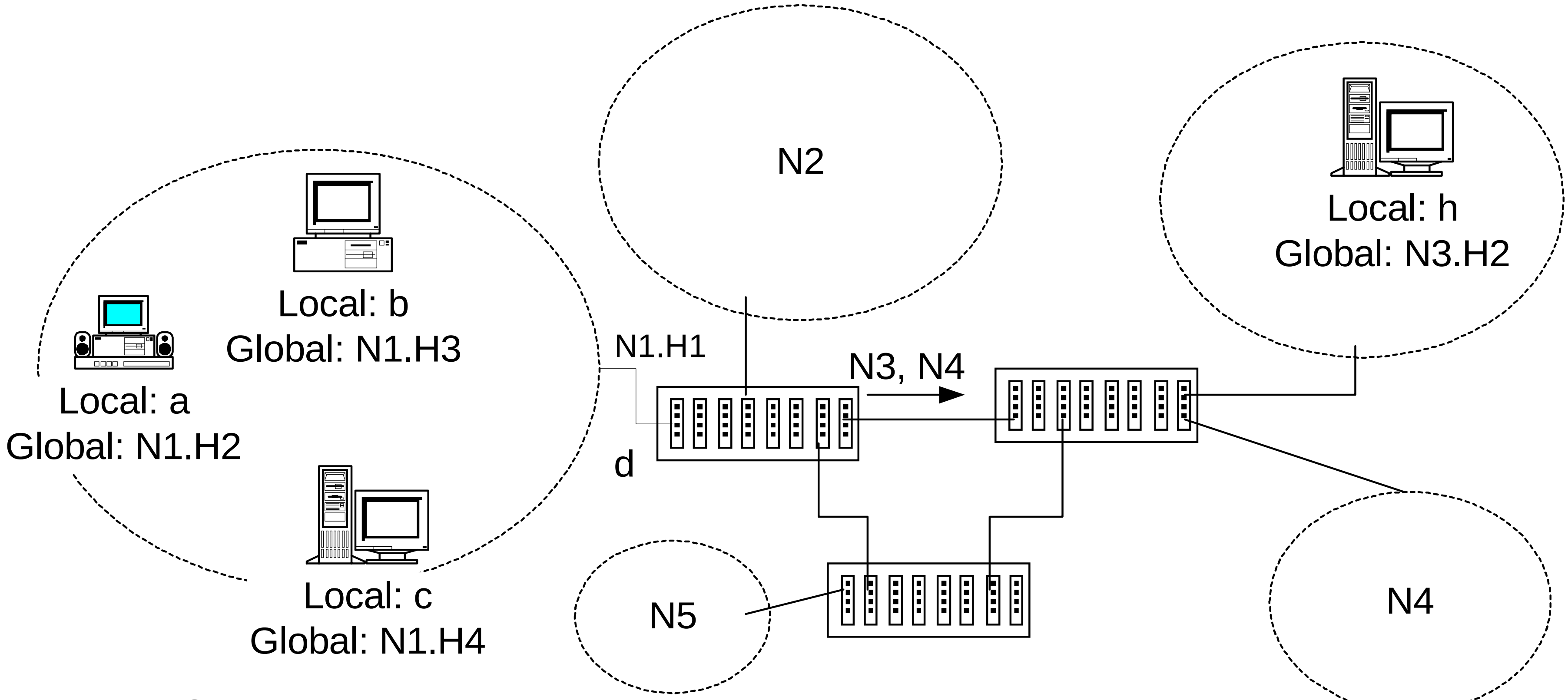
Addressing: Local

ETHERNET:

Every Ethernet NIC has unique 48-bit address



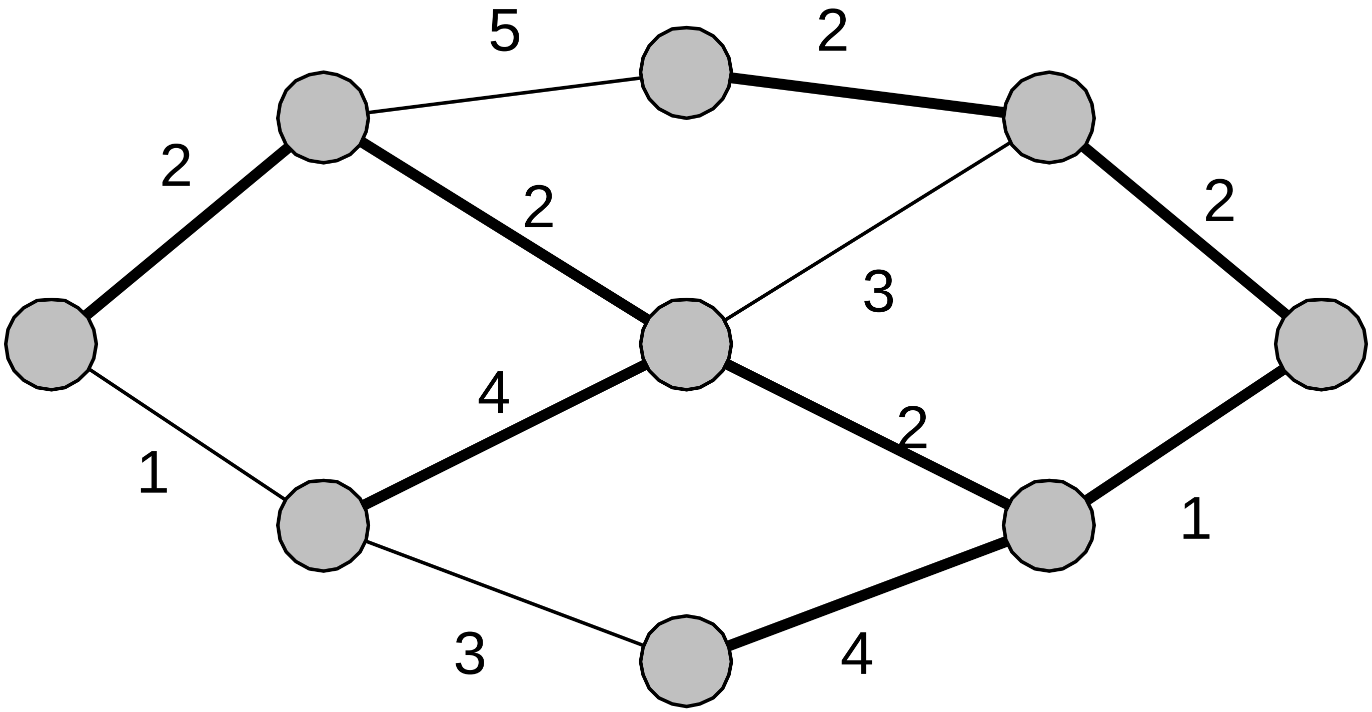
Addressing : Internetworking



N1.H2:

- To send to N1.H3: N1 => Local, use Ethernet [To b | From a | Data]
- To send to N3.H2: N2 => Not local => Send to "Gateway" N1.H1 [d]

≤ Routing



Minimum cost path

Cost = measure of delay

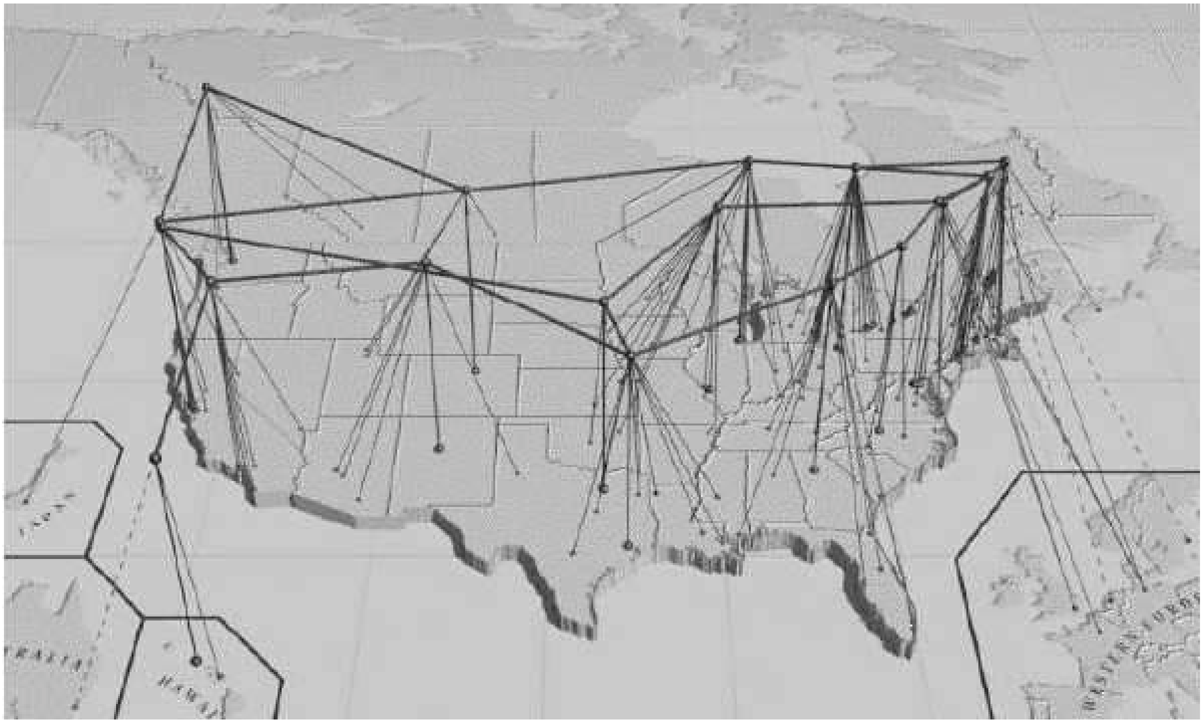
Nodes exchange information to discover these best paths

<

Internet & TCP/IP Networks

- The Internet
- TCP/IP Protocols: Overview
- IP
- Network Examples
- TCP
- Extensions
- CDN

The Internet



[4.1]

ANS backbone (Advanced Networks & Services, a Worldcom company), one of many Internet backbones

≤ The Internet (cont.)

1962-64: Leonard Kleinrock (MIT 1961-1967), Paul Baran (RAND 1962-1965), and Donald Davies independently develops the idea of **distributed, packet-switching** networks.

ARPANET goes online in 1969 (UCLA-SRI).

Bob Kahn and Vint Cerf develop the basic ideas of the **Internet** in 1973.

In 1974 BBN opens the first public packet-switched network - Telenet.

TCP/IP (Transmission Control Protocol and Internet Protocol) is established as the standard for ARPANET in 1982.

1987: the number of network hosts breaks 10,000.

1989: the number of hosts breaks 100,000.

≤ The Internet (cont.)

Tim Berners-Lee develops the **World Wide Web**. CERN releases the first Web server in 1991.

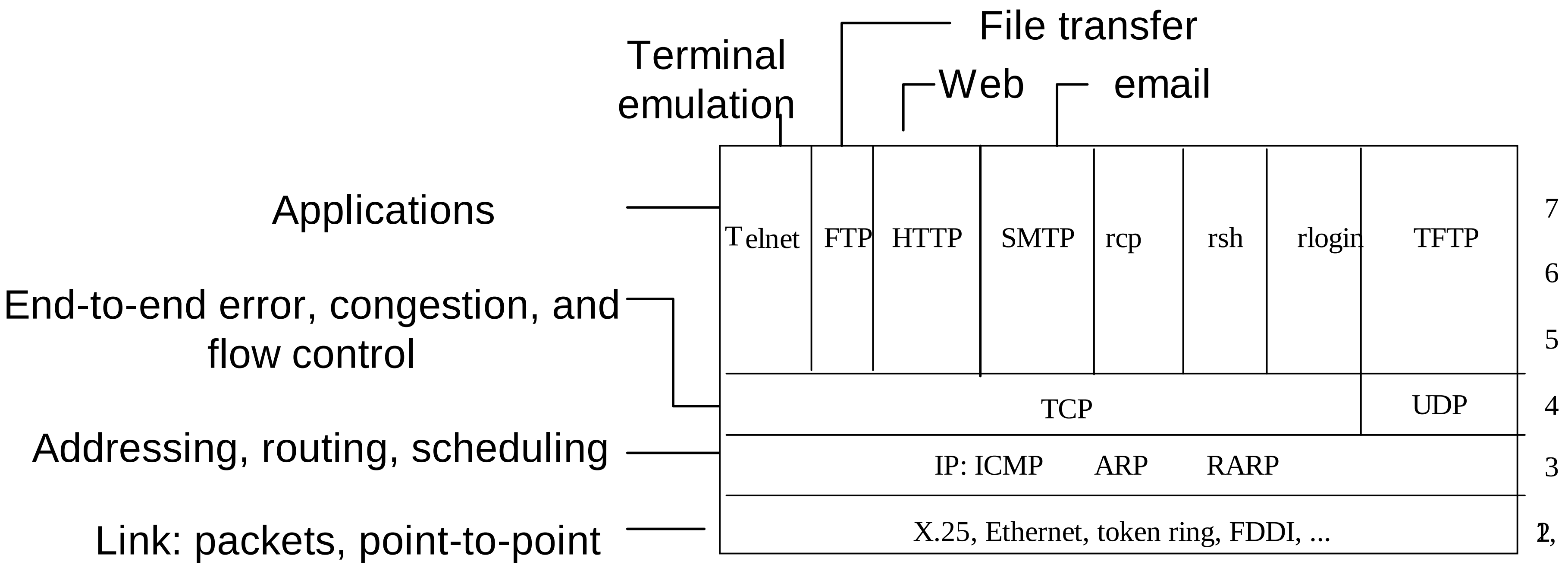
1992: the number of hosts breaks 1,000,000.

The World Wide Web sports a growth rate of 341,634% in service traffic in its third year, 1993.

The main U.S. Internet backbone traffic begins routing through commercial providers as NSFNET reverts to a research network in 1994.

The Internet 1996 World Exposition is the first World's Fair to be held on the internet.

TCP/IP Protocols: overview



Terminal emulation

File transfer

Web

email

Applications

End-to-end error, congestion, and flow control

Addressing, routing, scheduling

Link: packets, point-to-point

Telnet FTP HTTP SMTP rcp rsh rlogin TFTP

TCP UDP

IP: ICMP ARP RARP

X.25, Ethernet, token ring, FDDI, ...

7

6

5

4

3

1, 2

[4.2]

<

IP: Internet Protocol

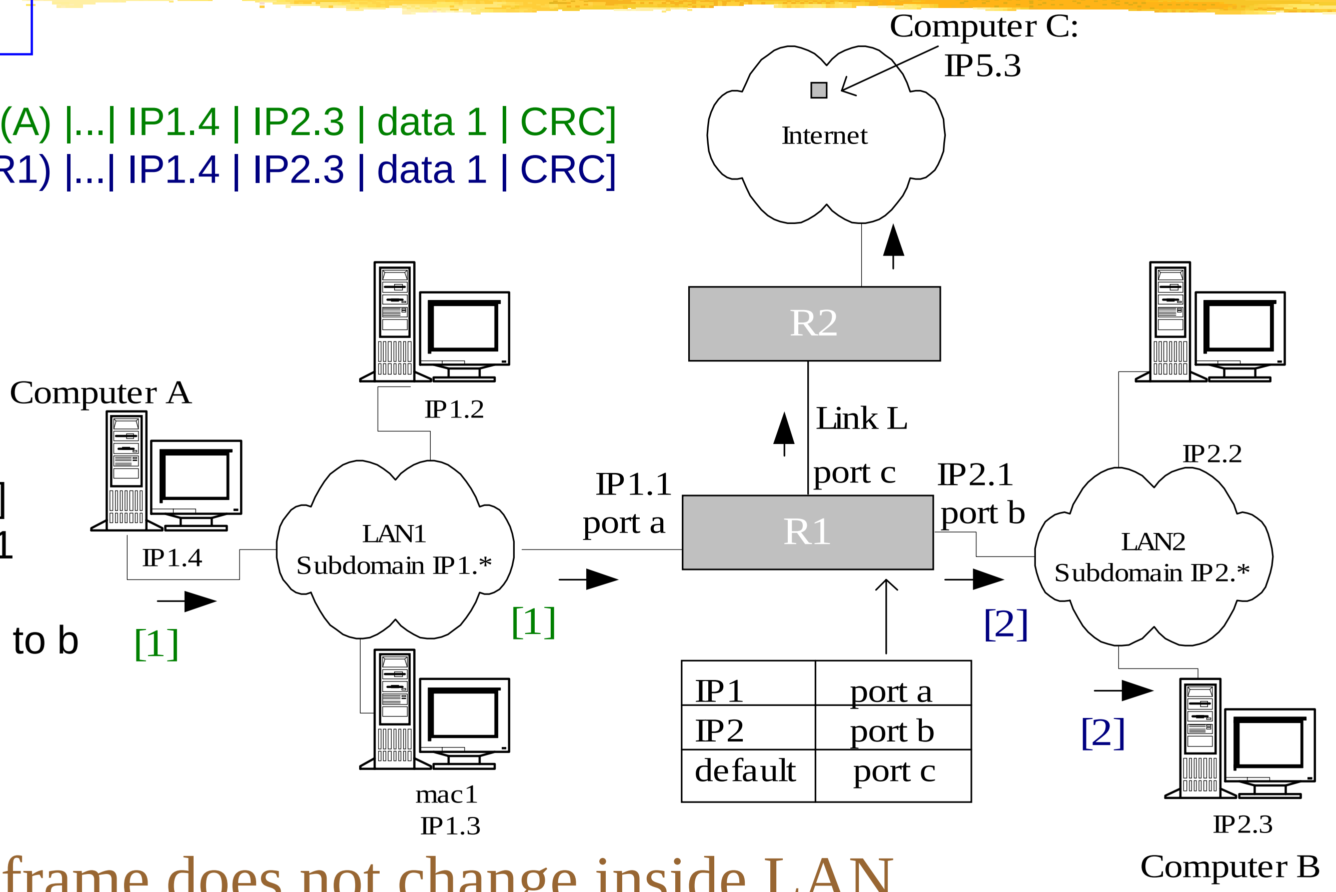
- Forwarding
- Header
- Addressing
- Fragmentation
- OSPF

IP: Forwarding

A: send [data 1] to B

[1] = [mac(a, R1) | mac(A) |...| IP1.4 | IP2.3 | data 1 | CRC]
[2] = [mac(B) | mac(b, R1) |...| IP1.4 | IP2.3 | data 1 | CRC]

1. Learn IP2.3 (DNS)
2. [IP1.4 | IP2.3 | data 1]
3. IP2 is not IP1 => to R1
4. A sends [1] to R1
5. R1 sees B is attached to b
6. R1 learns mac(B)
7. R1 sends [2] to B

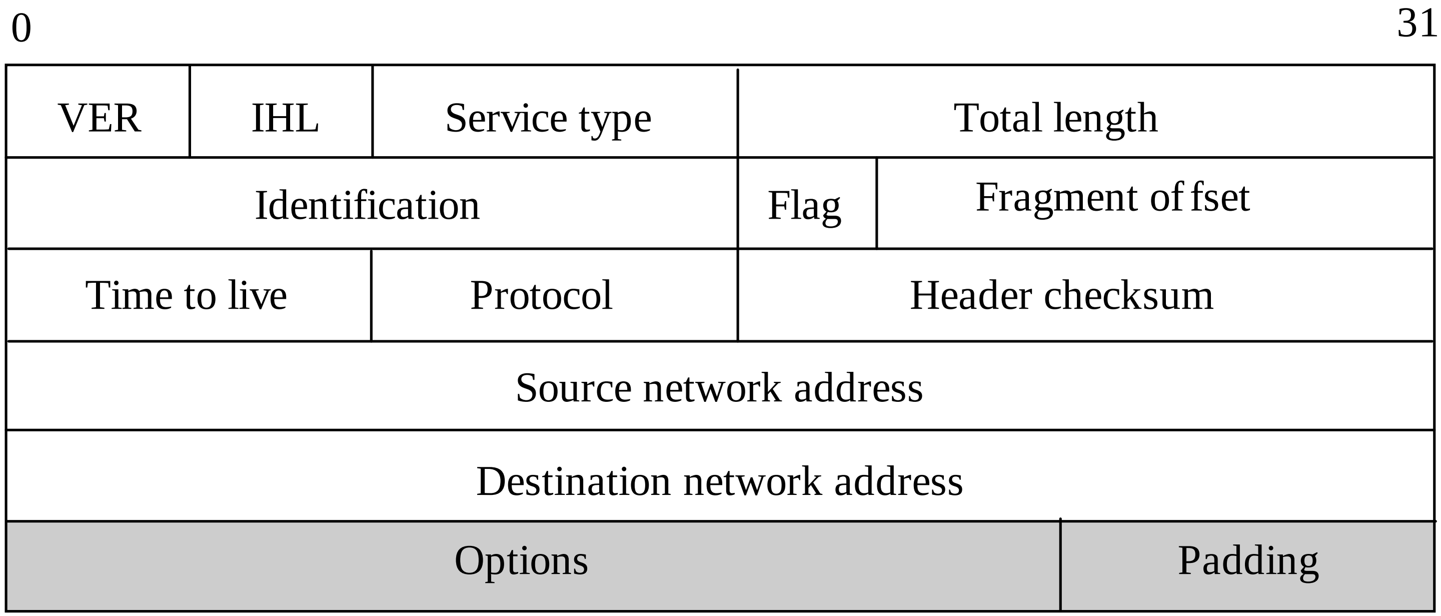


IP1	port a
IP2	port b
default	port c

Notes: Ethernet frame does not change inside LAN
IP packet never changes
Hierarchical addresses to simplify routing tables

[4.3]

IP : Header



0 31

VER	IHL	Service type	Total length
Identification		Flag	Fragment offset
Time to live	Protocol	Header checksum	
Source network address			
Destination network address			
Options			Padding

IP : Addressing (32 bits in IPv4)

Class-based: [Net | Host]

A: (0 | 7) | (24) : 128 networks with 16M hosts each
B: (10 | 16) | (16) : 64k networks with 64k hosts each
C: (110 | 21) | (8) : 2M networks with 256 hosts each
D: (1110 | 28) : Multicast

Subnet: [Net | Subnet | Host]

Subnet mask determined subnet
E.g., B = [128.32.156.14 with M = 255.255.255.0]
=> subnet = 128.32.156.*, host = 14 on that subnet
Assume A = [128.32.134.28, M = 255.255.255.0]
To send from A to B, A knows B is on different subnet
=> A sends to default gateway

IP : Addressing (cont.)

Classeless (CIDR): More efficient use of addresses

Example:

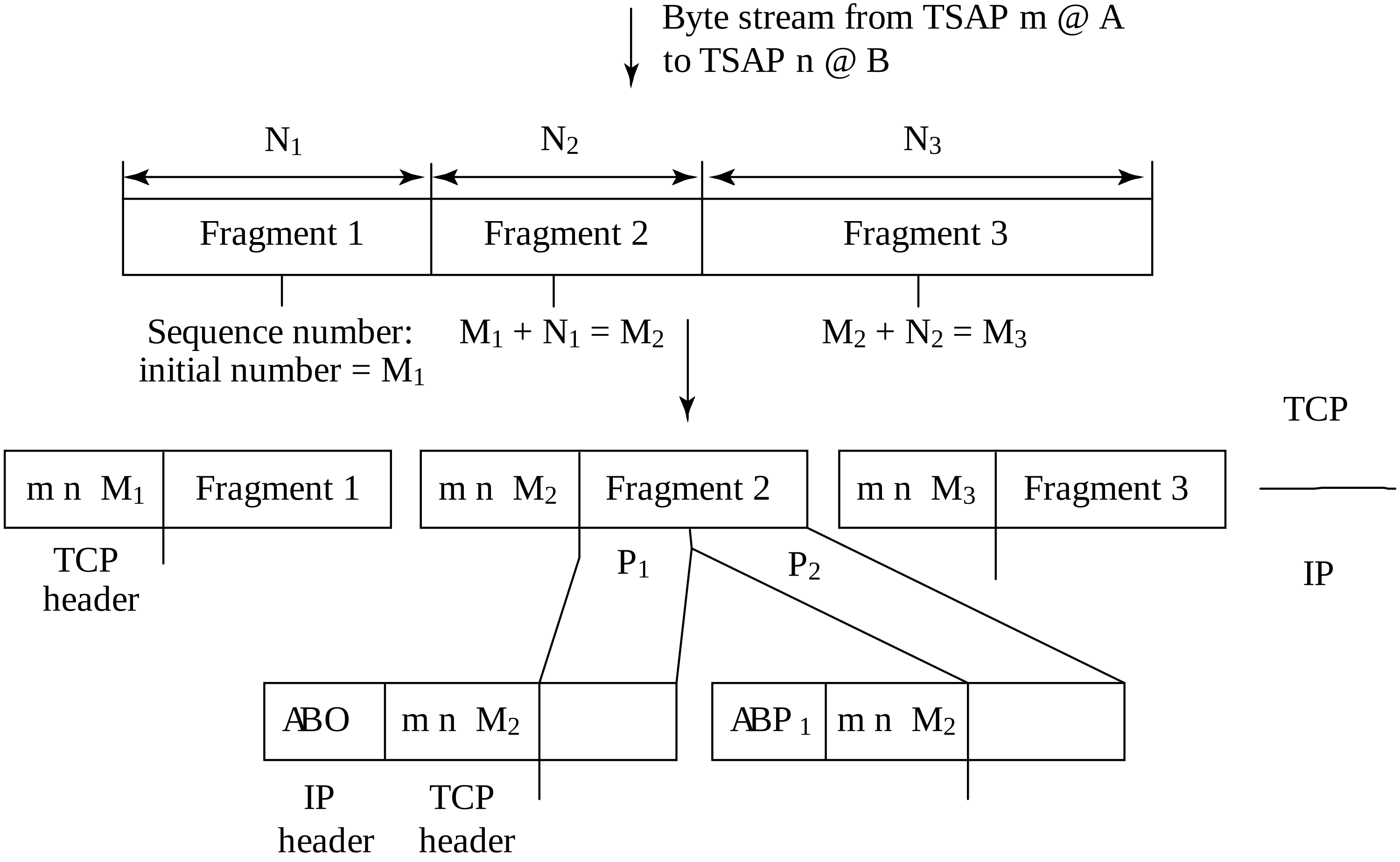
0010	-> port 1
00110	-> port 2
00111	-> port 3

Routing by longest prefix match

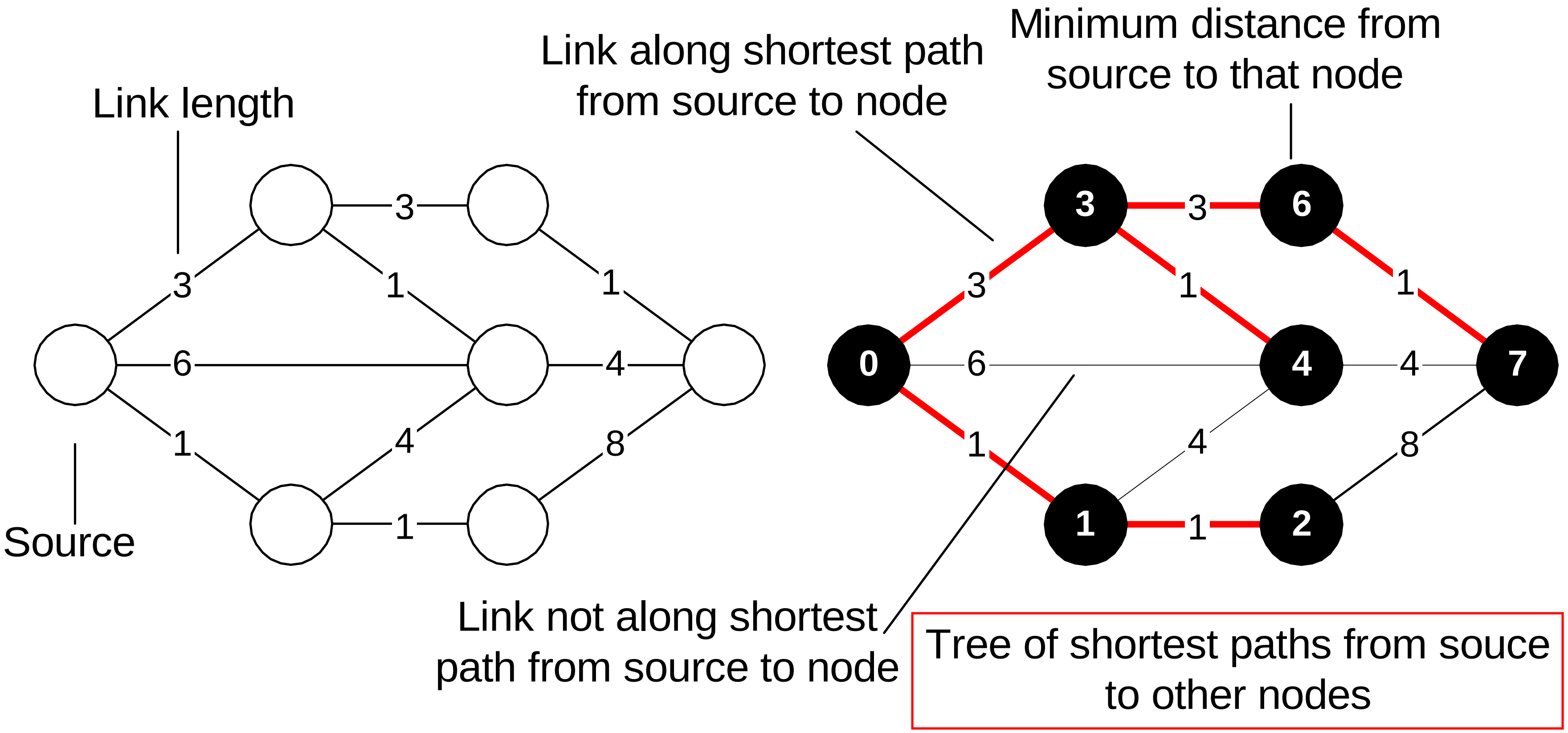
Notes:

- DHCP: Addresses are temporary and assigned from pool
- Mobile IP: Leave forwarding address with a home agent that intercepts packets for mobile host

IP : Fragmentation

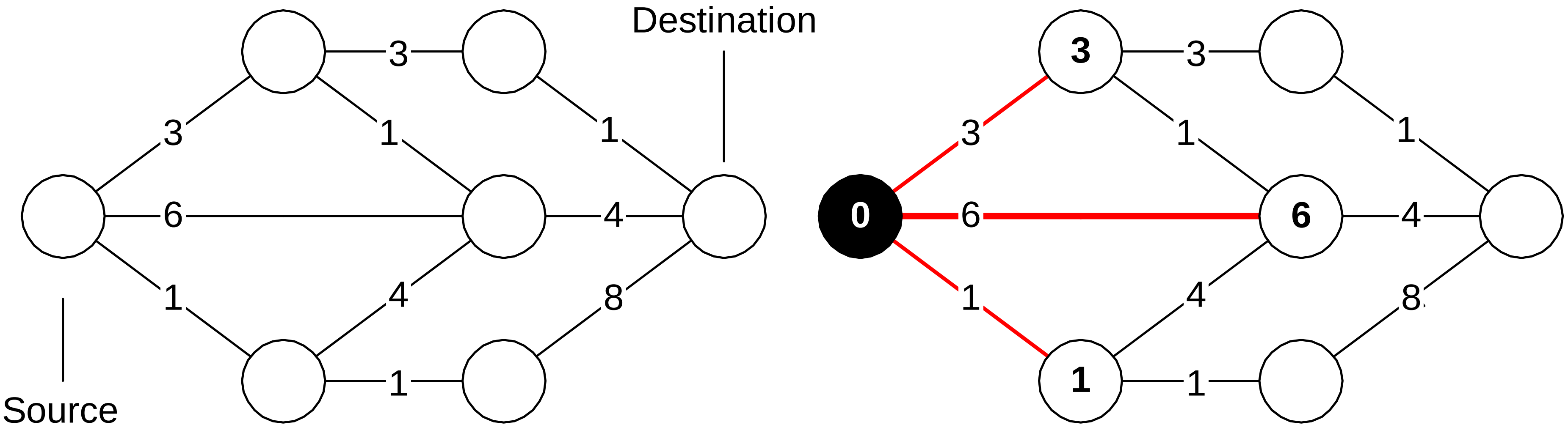


IP : OSPF (1/8)


Link length
Source
3
6
1
3
1
4
1
1
4
8
Link along shortest path from source to node
Minimum distance from source to that node
0
1
2
3
4
6
7
3
6
1
3
1
4
1
1
4
8
Link not along shortest path from source to node

Tree of shortest paths from souce to other nodes

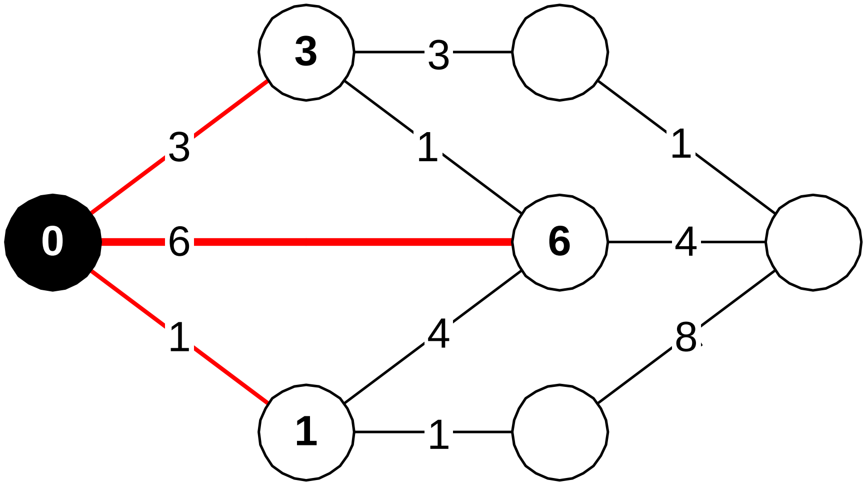
IP : OSPF (2/8)

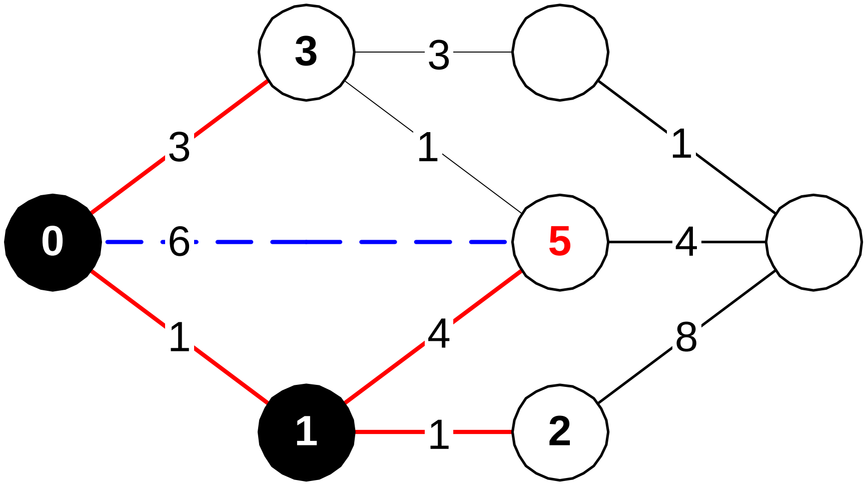


Algorithm:

- Initial label = infinity for all nodes, 0 for source; nodes unmarked
- Pick umarked node with smallest label [here: source]
- Update its children and mark it; mark/unmark shortest path links
  [Here: three nodes are updated (3, 6, 1), their links are marked.]

IP : OSPF (3/8)

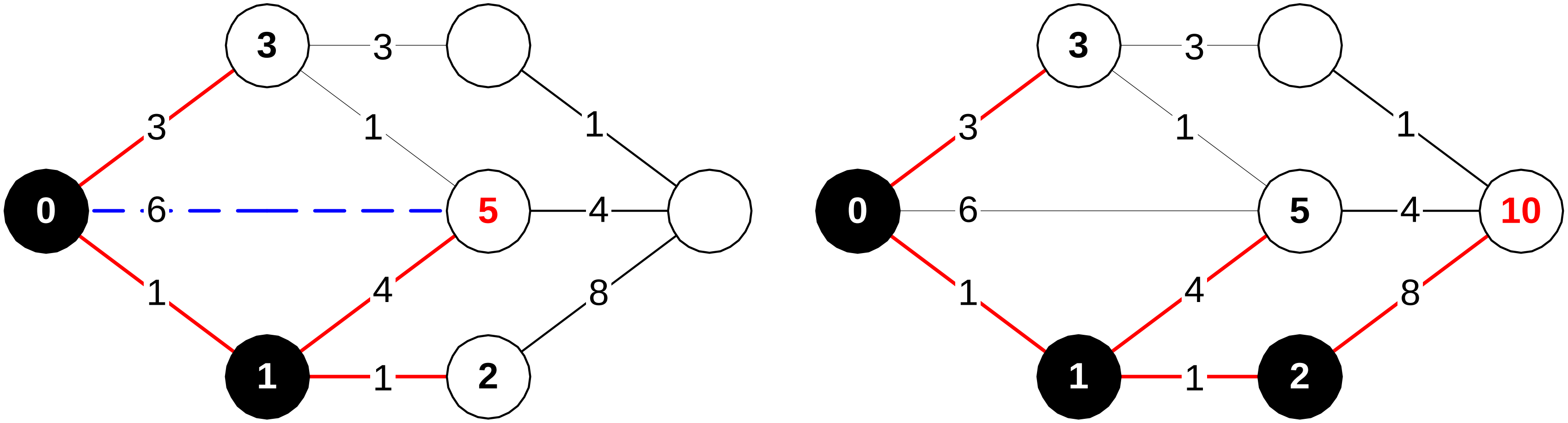




Algorithm:

- Initial label = infinity for all nodes, 0 for source; source marked
- Pick umarked node with smallest label [node with label 1]
- Update its children and mark it; mark/unmark shortest path links
  [nodes with label 5: umark link from source, mark link 1-5]

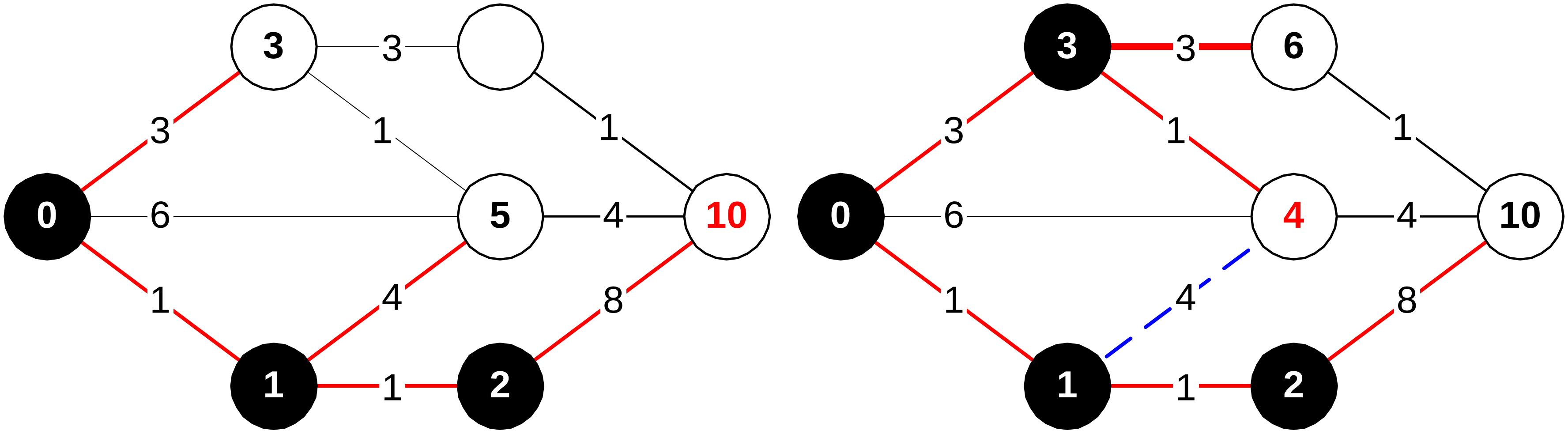
IP : OSPF (4/8)



Algorithm:

- Initial label = infinity for all nodes, 0 for source; source marked
- Pick umarked node with smallest label [node with label 2]
- Update its children and mark it; mark/unmark shortest path links [label 10, mark links 2-10]

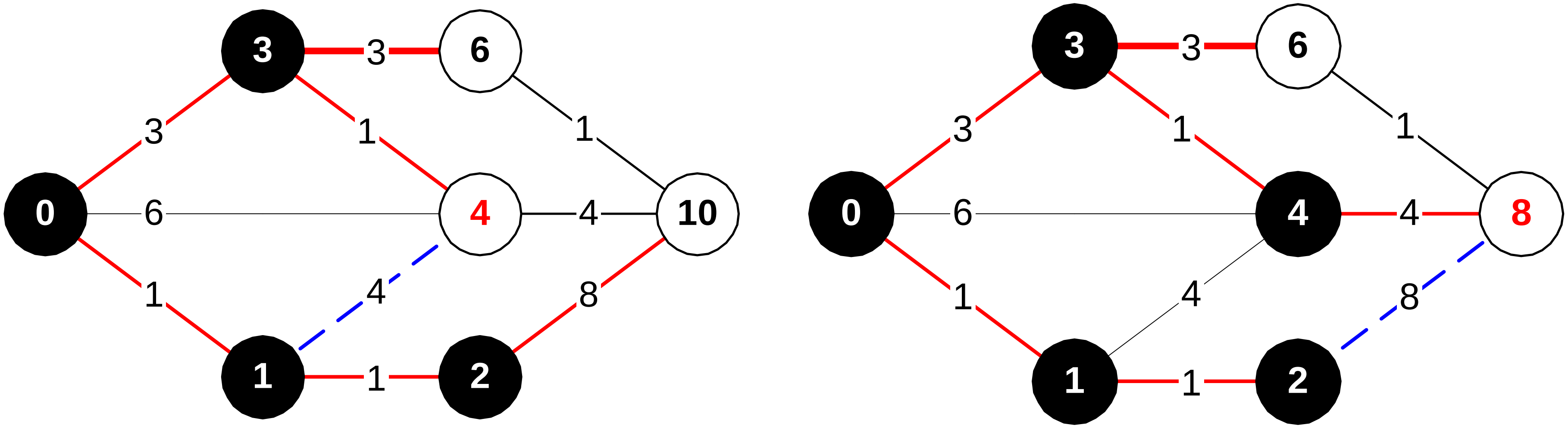
IP : OSPF (5/8)



Algorithm:

- Initial label = infinity for all nodes, 0 for source; source marked
- Pick umarked node with smallest label [node with label 3]
- Update its children and mark it; mark/unmark shortest path links [update node from label 5 to 4, umark old link, mark new one]

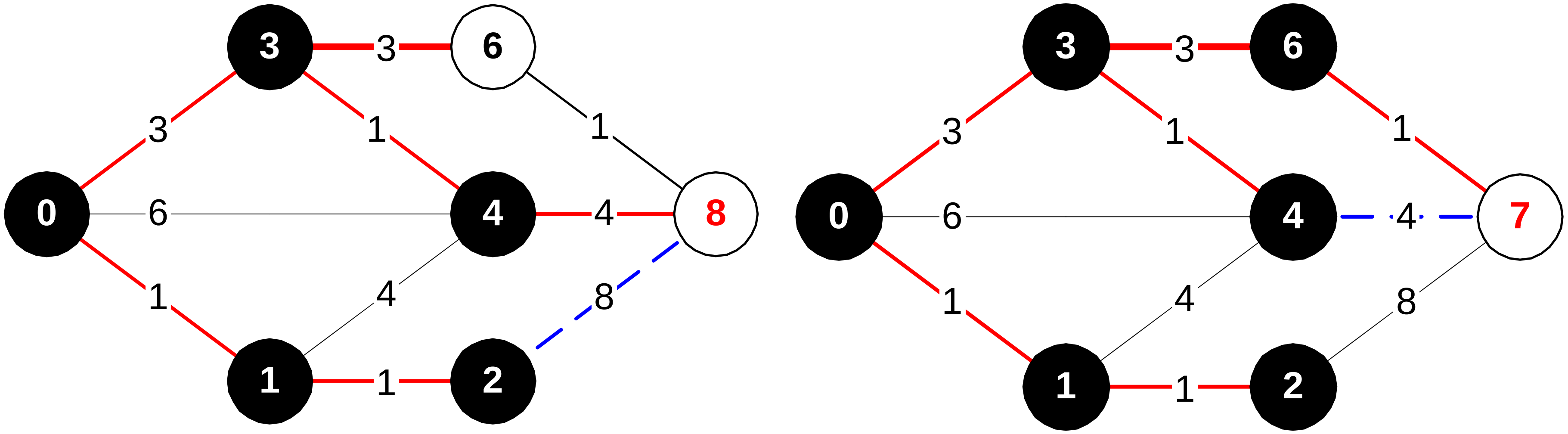
IP : OSPF (6/8)



Algorithm:

- Initial label = infinity for all nodes, 0 for source; source marked
- Pick umarked node with smallest label [node with label 4]
- Update its children and mark it; mark/unmark shortest path links [update label from 10 to 8, unmark/mark link]

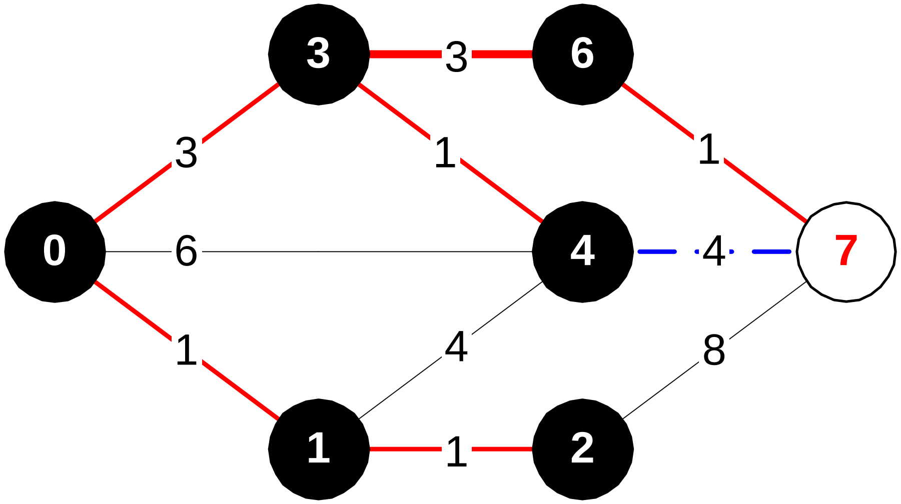
IP : OSPF (7/8)

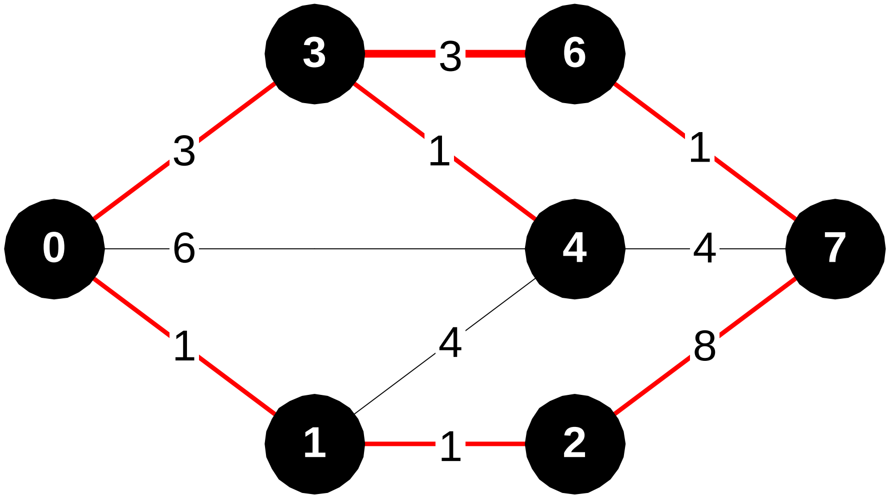


Algorithm:

- Initial label = infinity for all nodes, 0 for source; source marked
- Pick umarked node with smallest label
- Update its children and mark it; mark/unmark shortest path links

IP : OSPF (8/8)





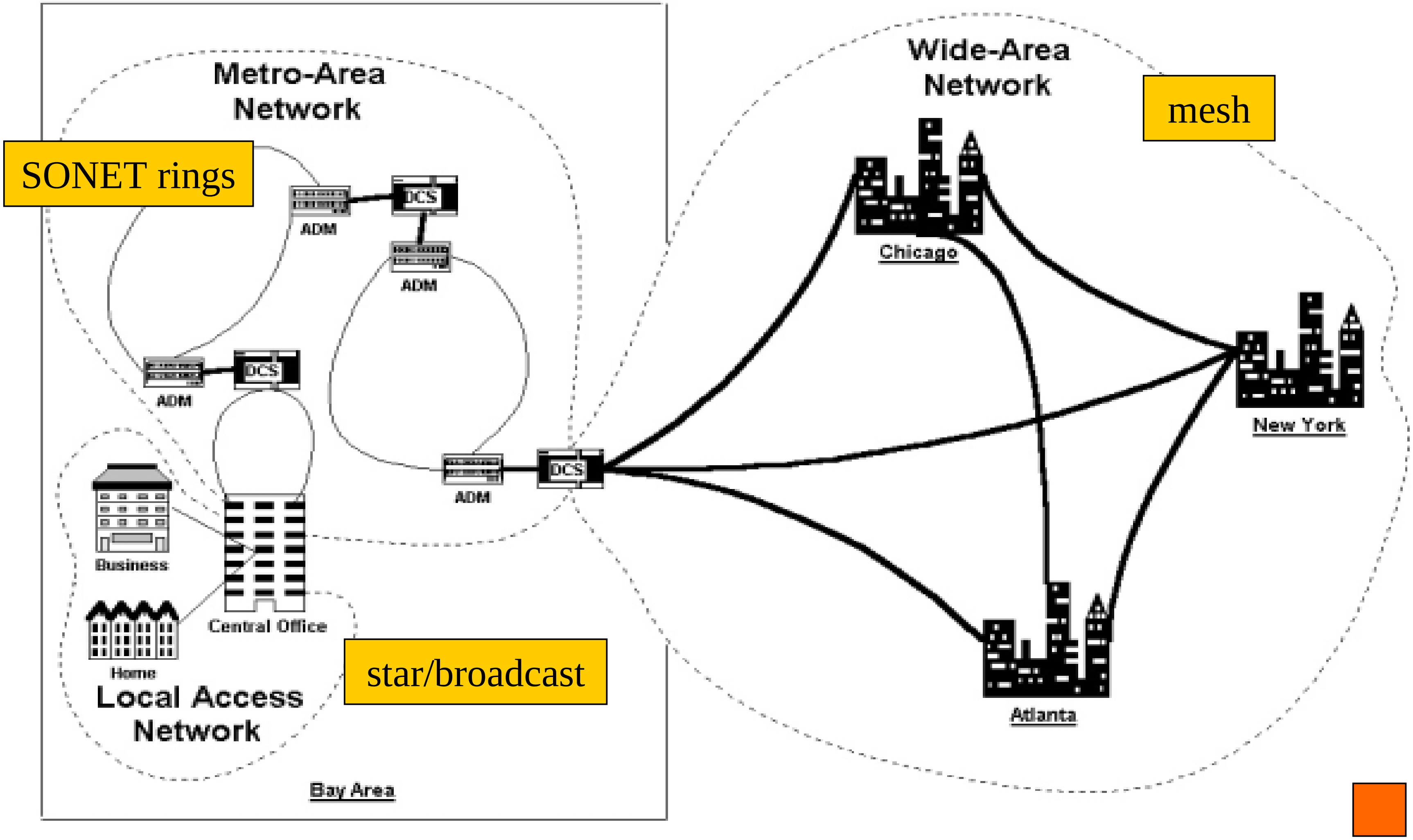
Algorithm:

- Initial label = infinity for all nodes, 0 for source; source marked
- Pick umarked node with smallest label
- Update its children and mark it; mark/unmark shortest path links

Network Examples

- Hierarchy
- Backbone, NAP, regional Network
- Public IXP or NAP in US
- Major US Data Centers
- Regional, ISP, Local Access Networks
- Bay Area backbone structure
- Links from SFO to major data centers
- Links from Palo Alto to major data centers
- Traceroute from Concentric customer

E.g. Internet hierarchy

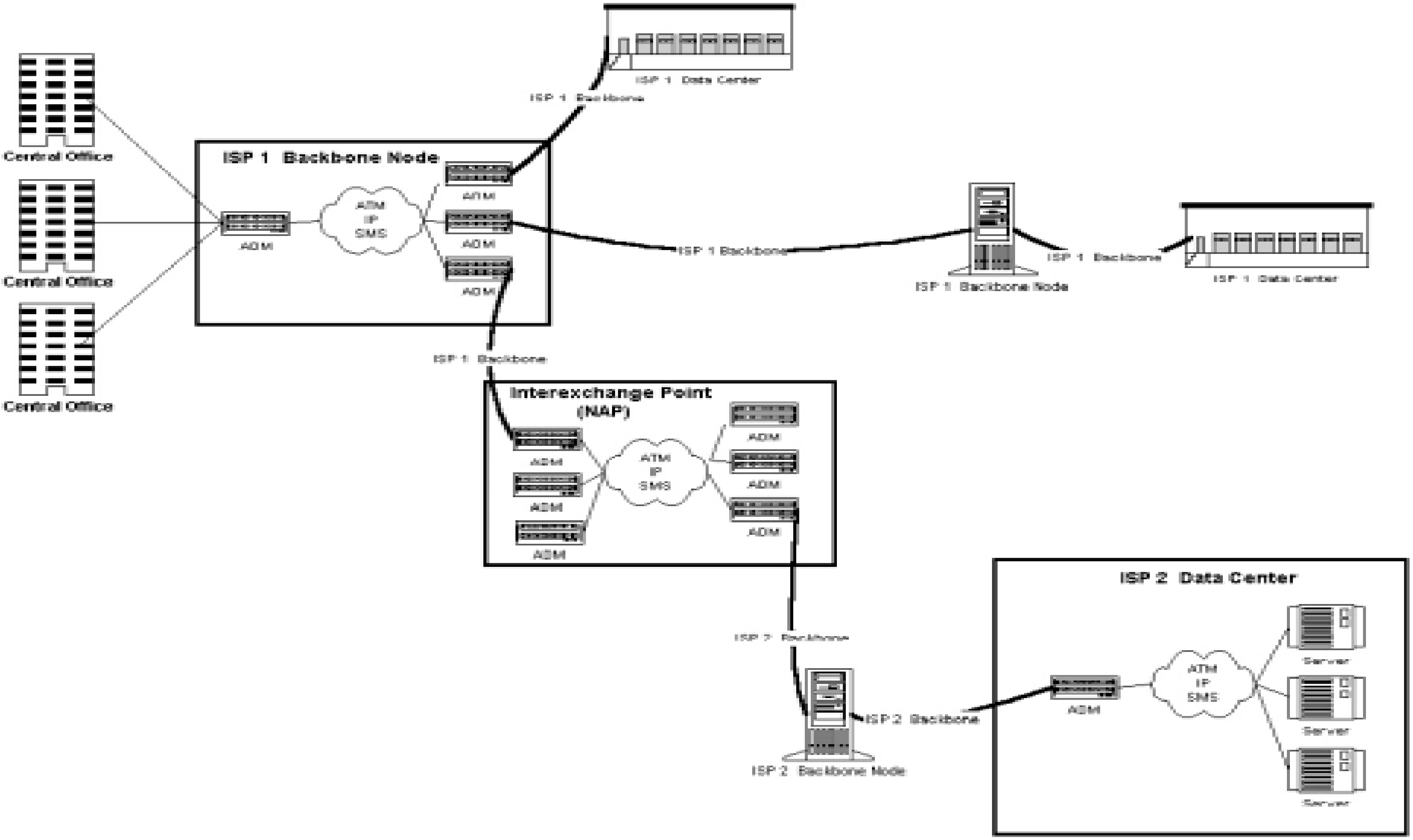


SONET rings

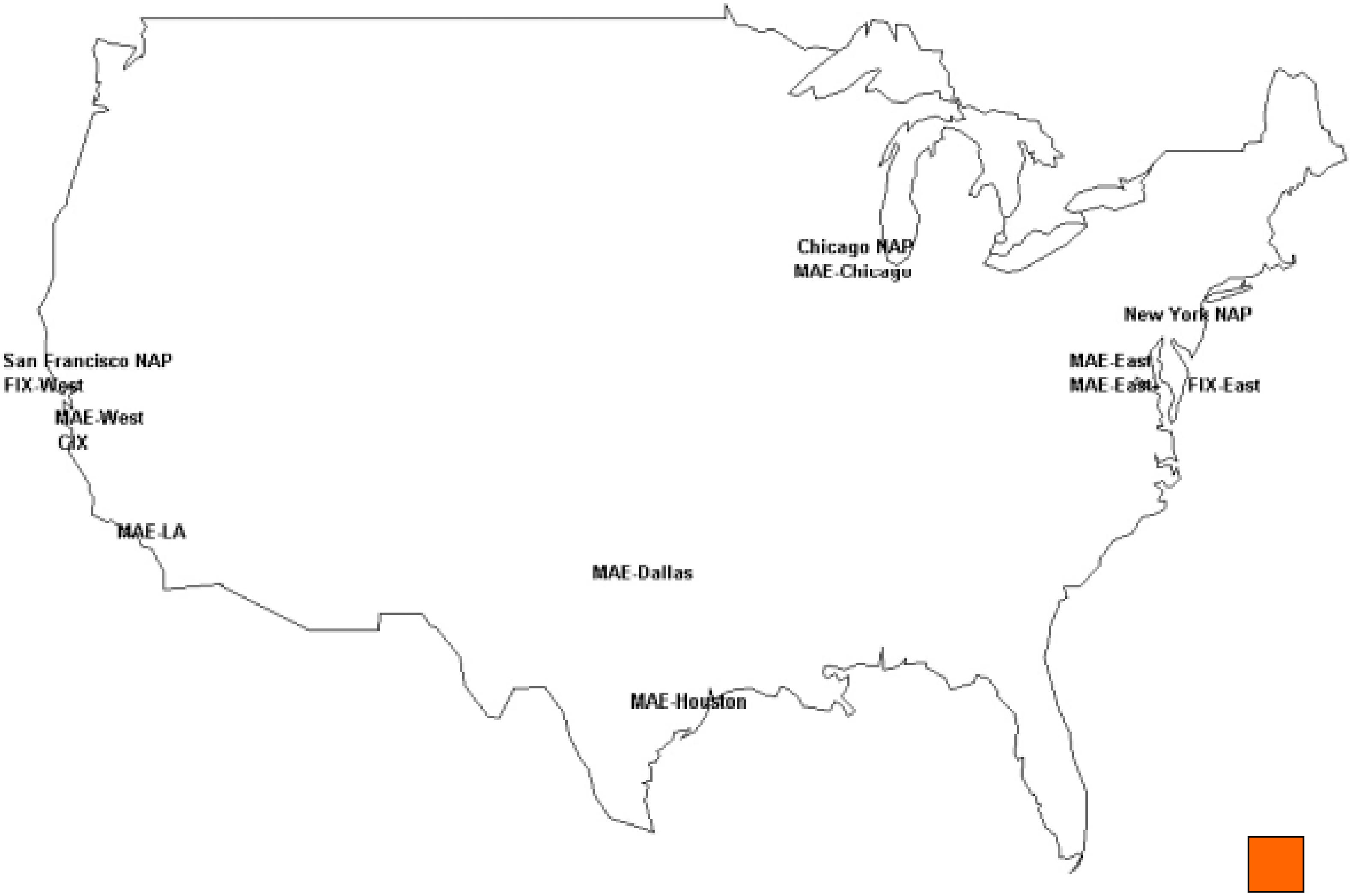
mesh

star/broadcast

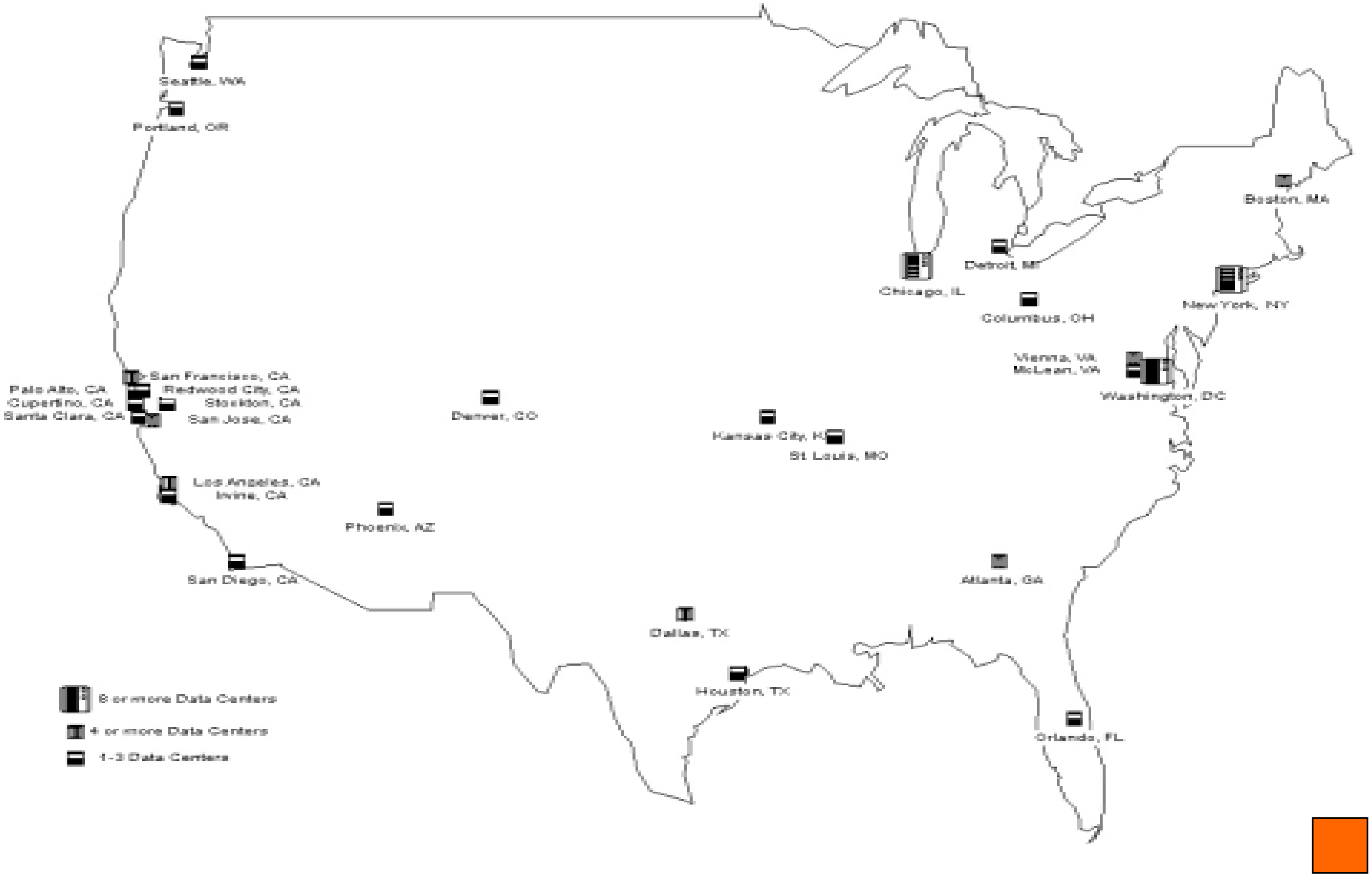
E.g. Backbone, NAP, regional network



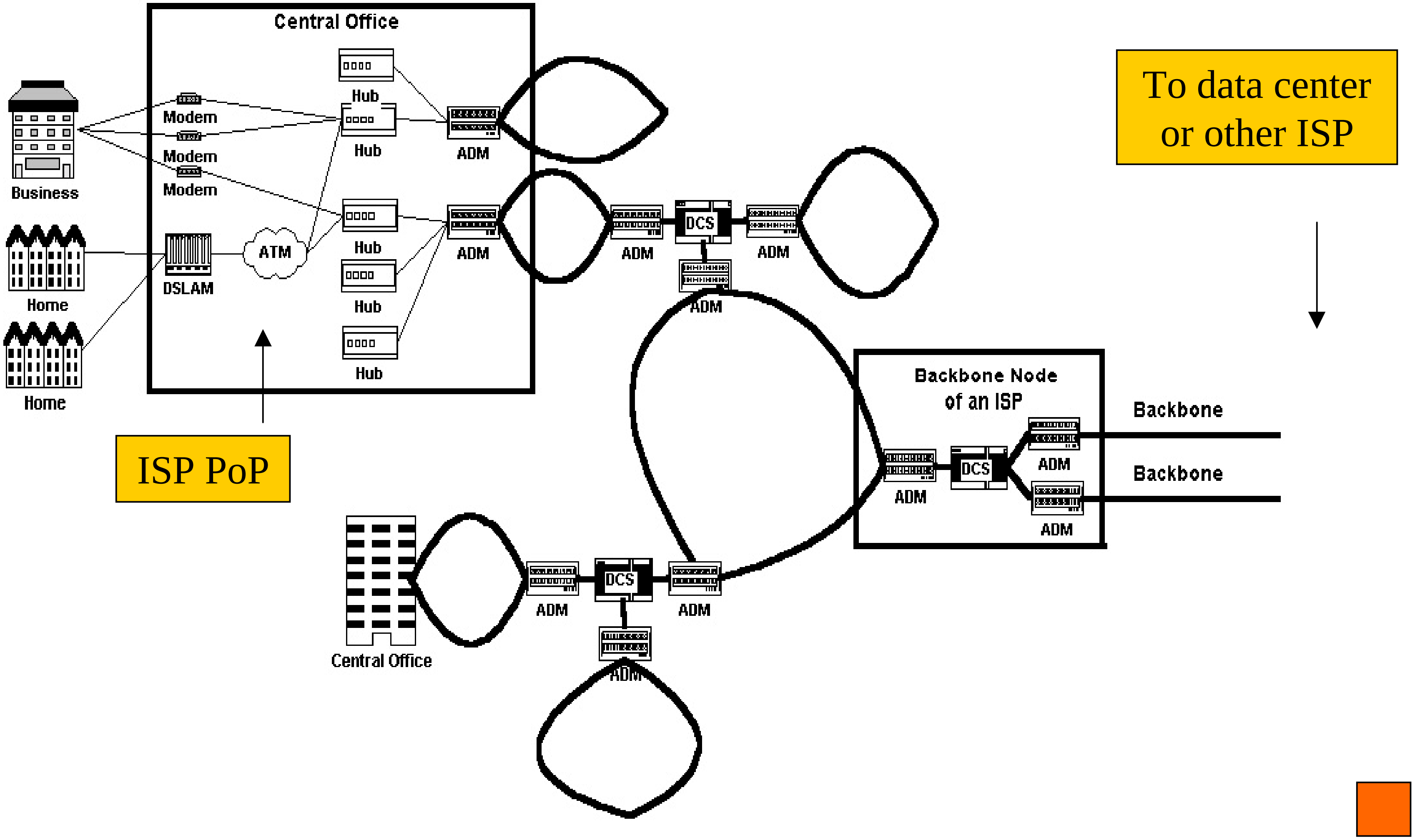
E.g. Public IXP or NAP in US



E.g. Major US data centers



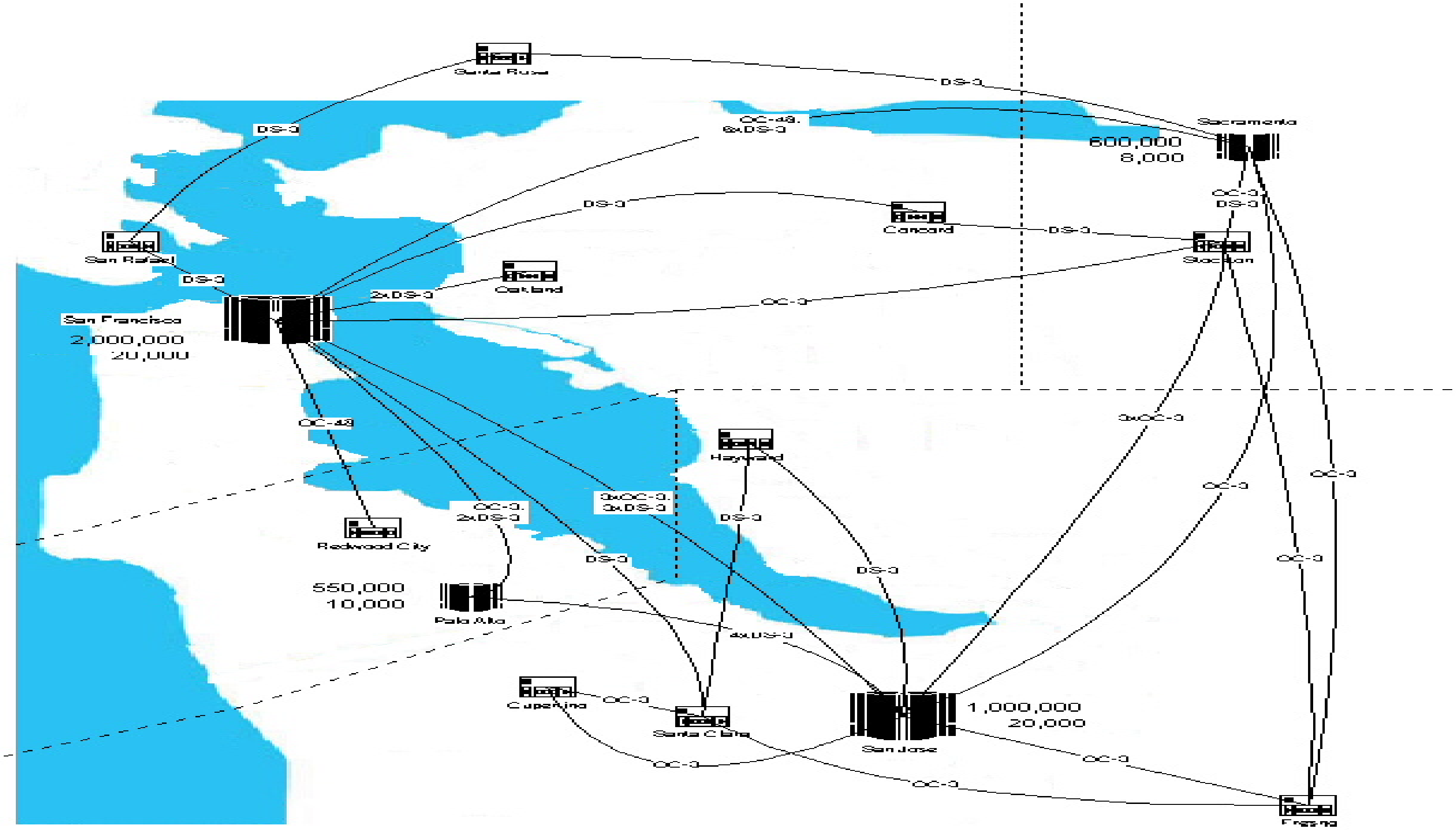
E.g. Regional, ISP, local access networks



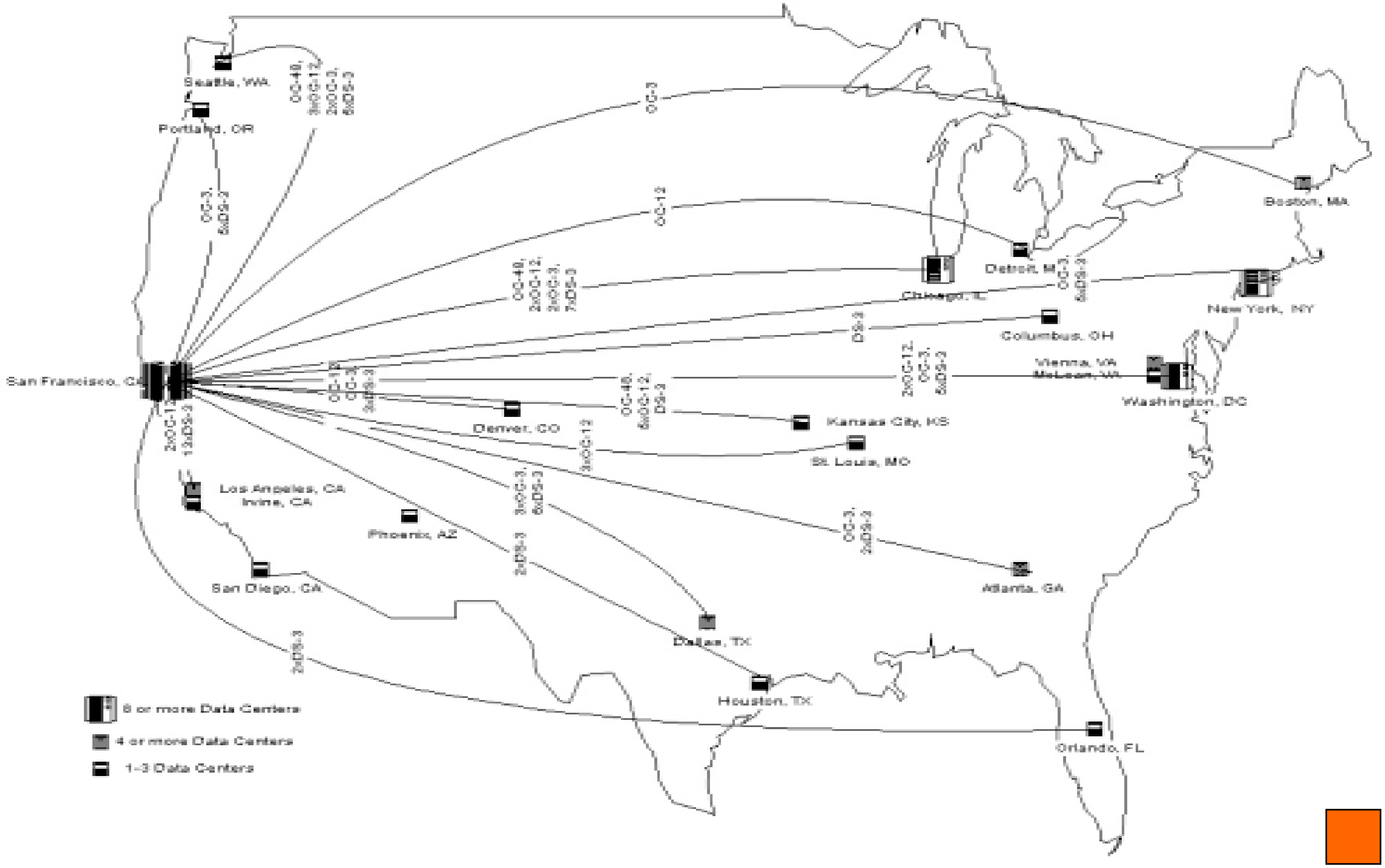
To data center or other ISP

ISP PoP

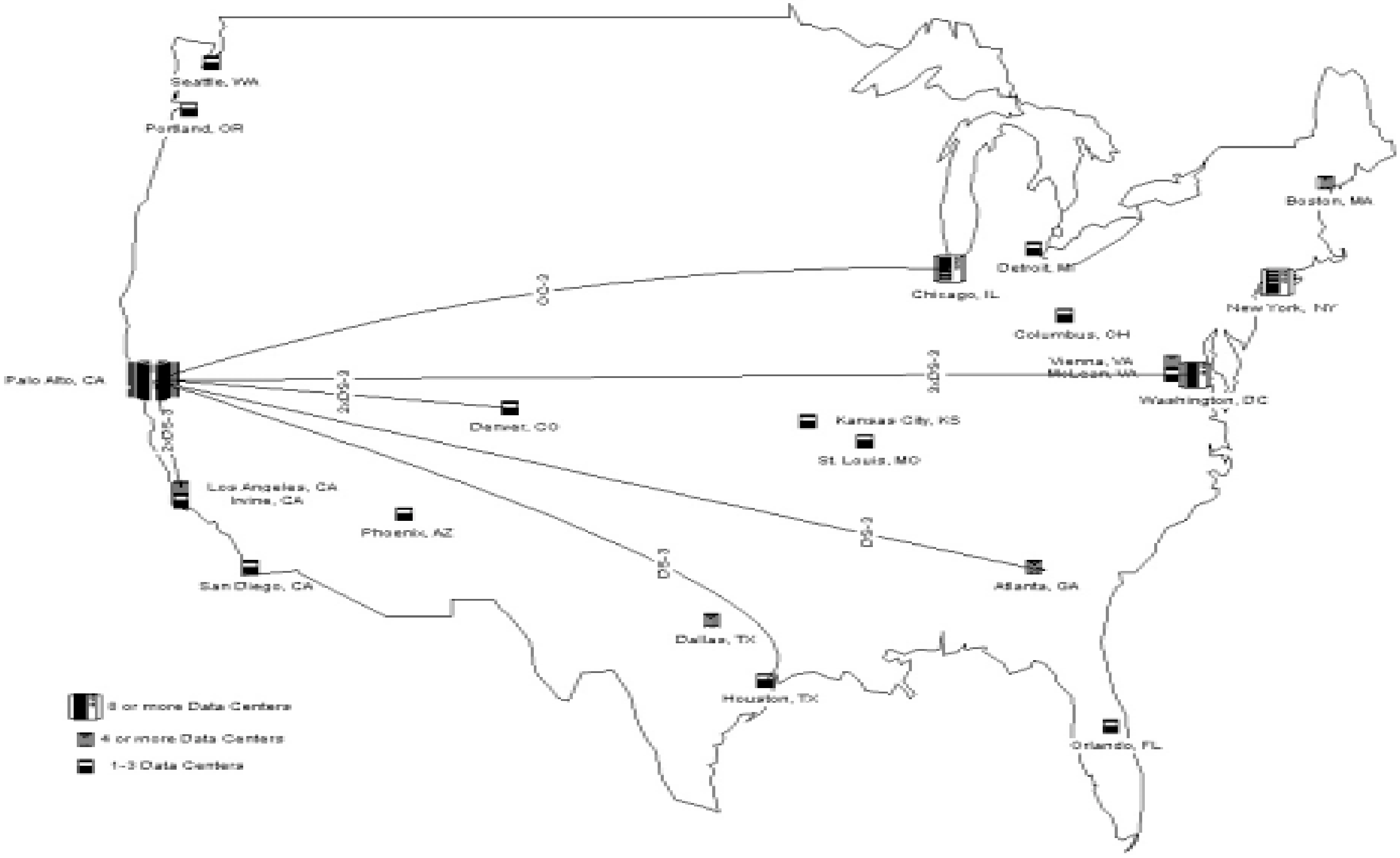
E.g. Bay Area backbone structure



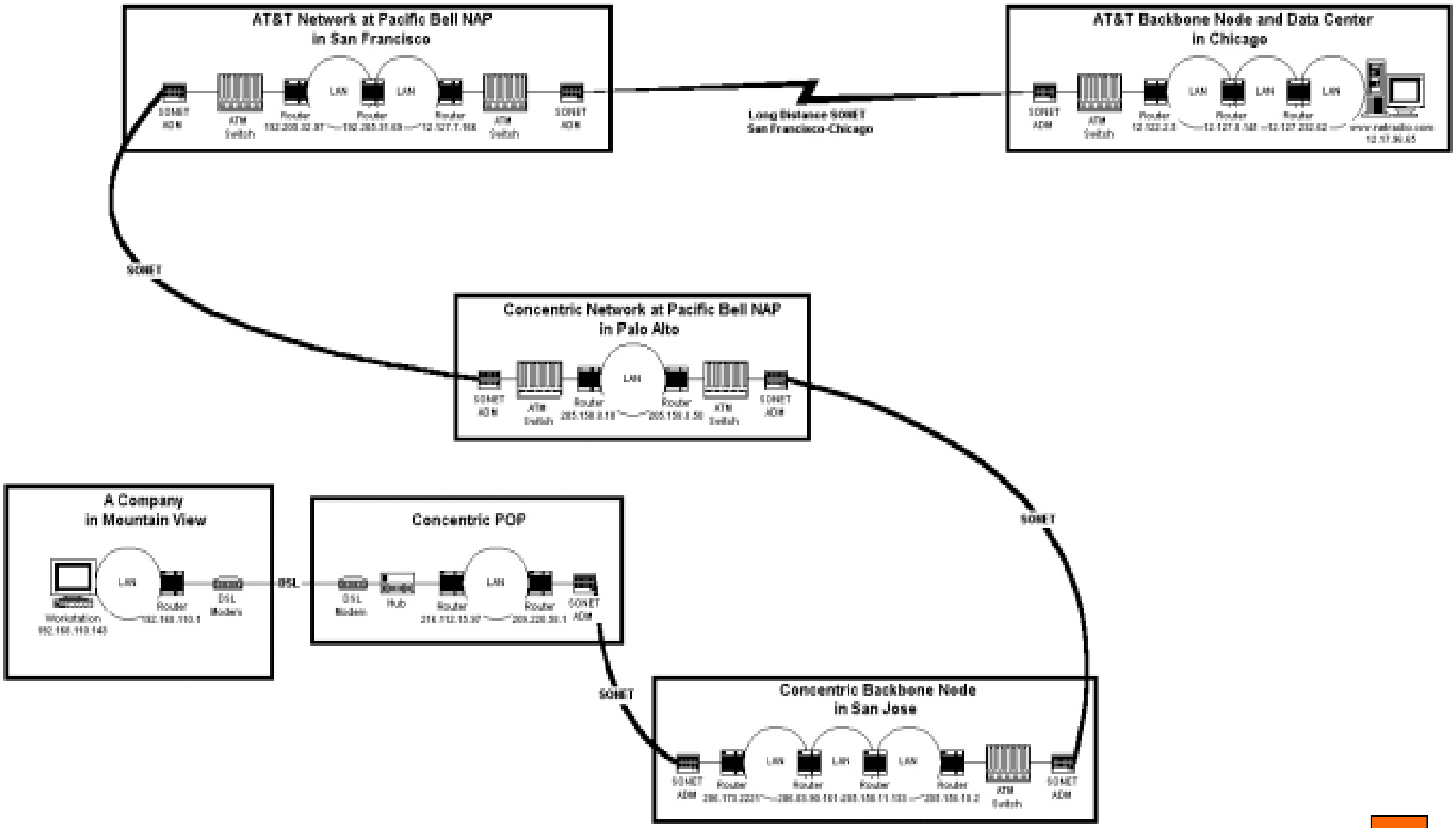
E.g. Links from SFO to major data centers



E.g. Links from Palo Alto to major data centers



E.g. Traceroute from Concentric customer

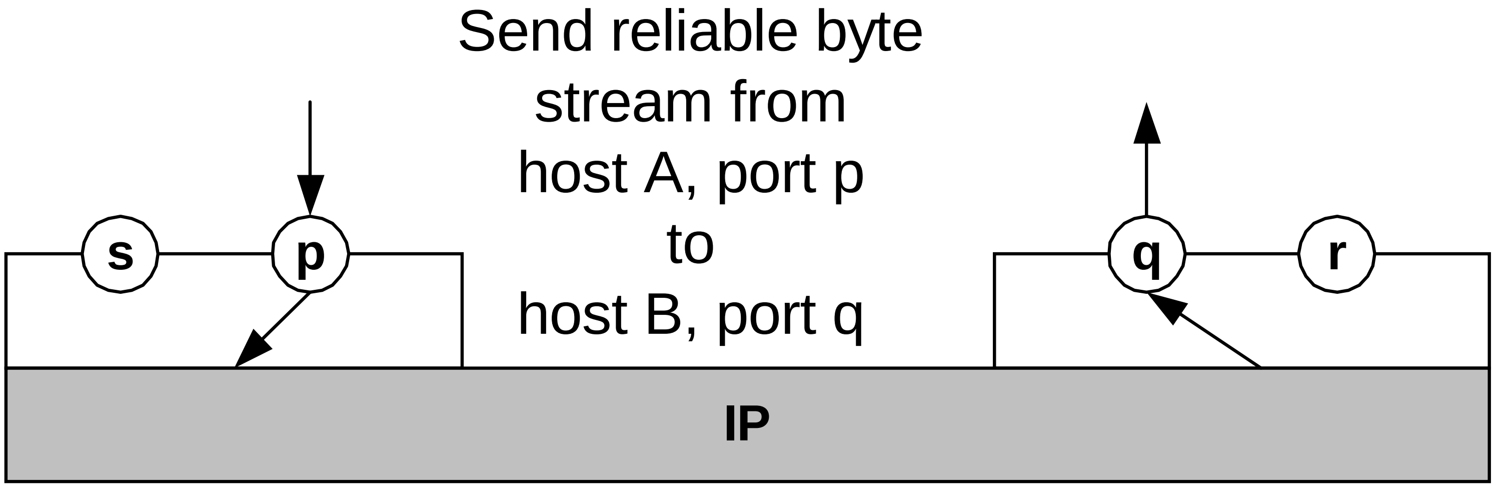


<

TCP: Transmission Control Protocol

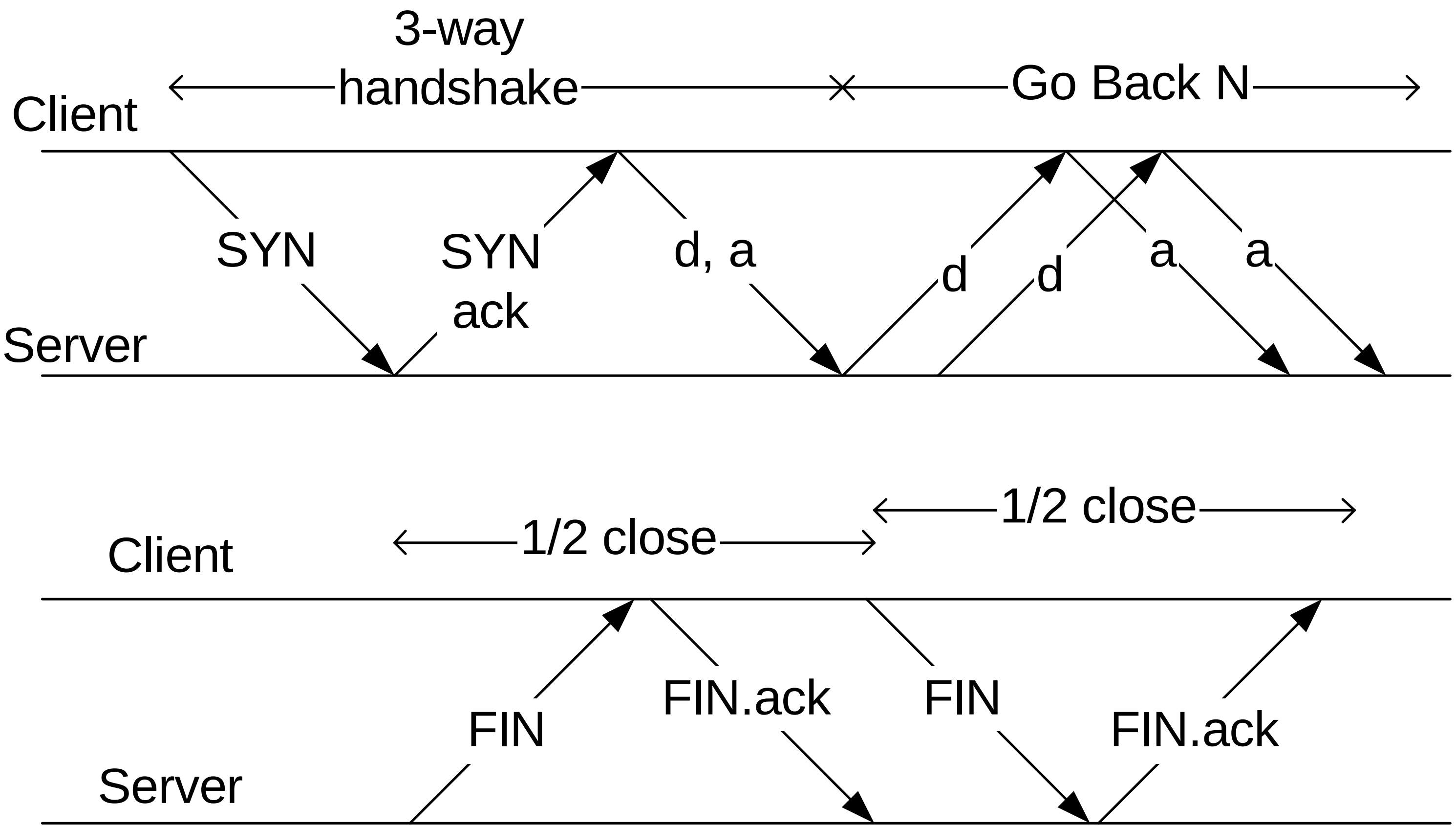
- Services
- Stages of connection
- Header
- Window Size
- Window Size Justification
- RED
- ECN

TCP: Services

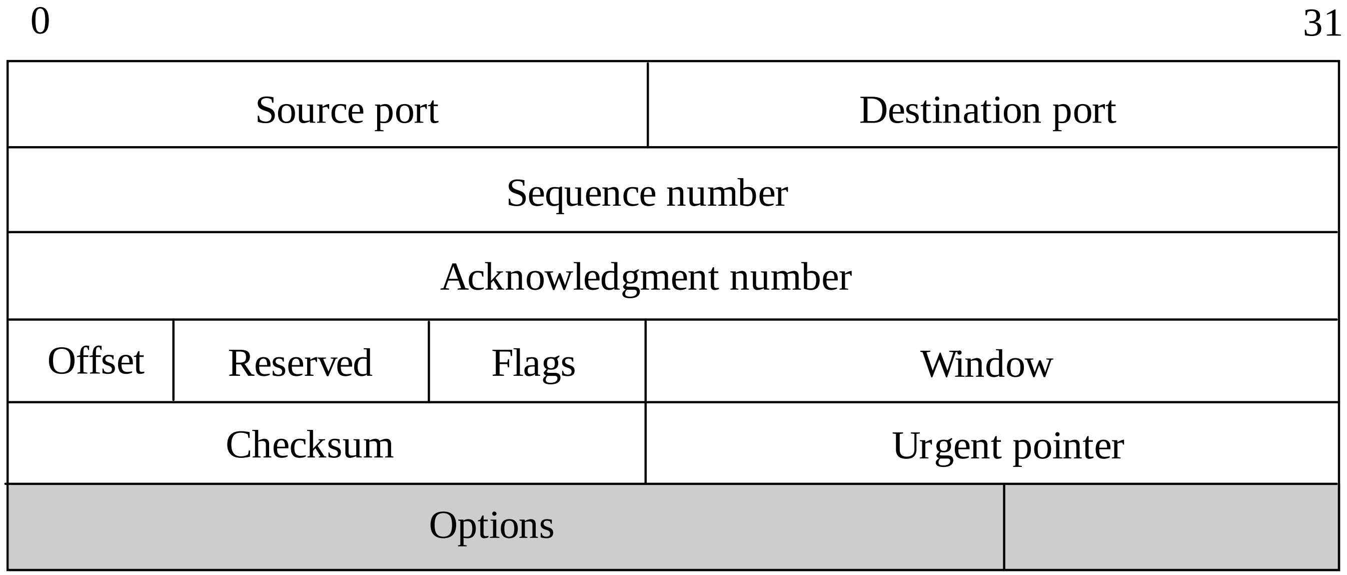


Send reliable byte stream from host A, port p to host B, port q

TCP: Typical stages of connection



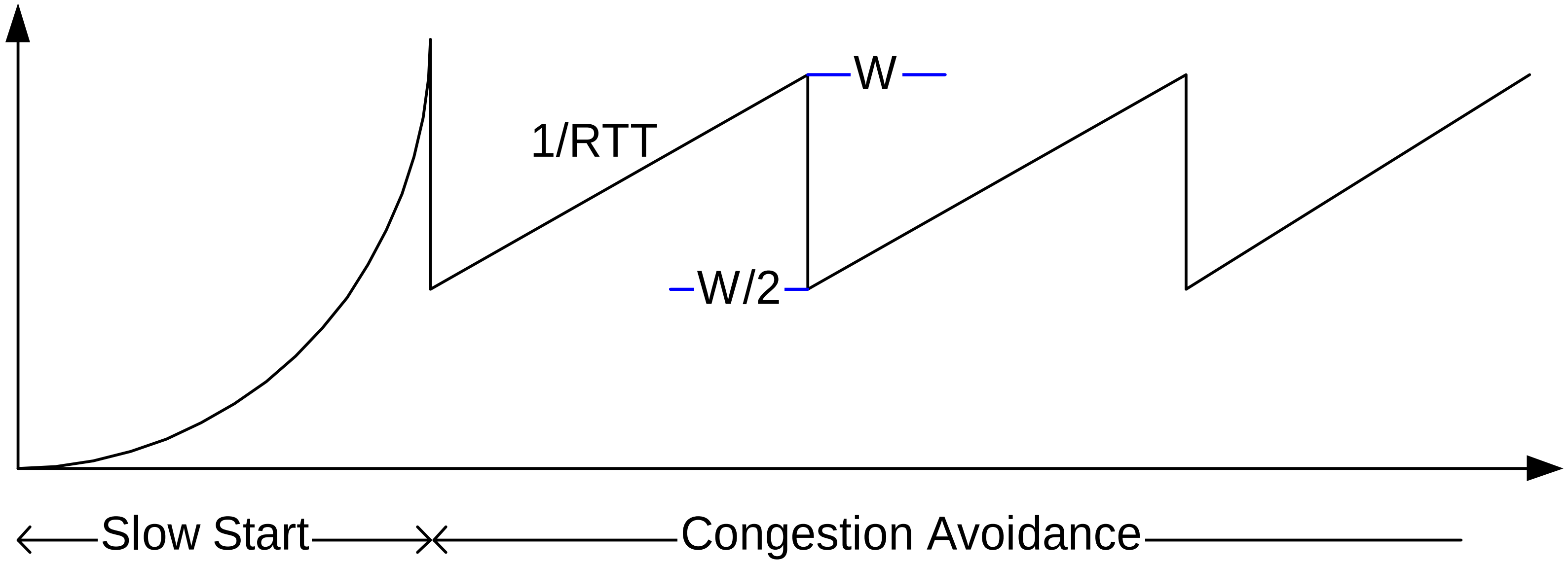
TCP: Header



Source port			Destination port
Sequence number			
Acknowledgment number			
Offset	Reserved	Flags	Window
Checksum			Urgent pointer
Options			

FLAGS: URG, ACK, PSH, RST, SYN, FIN
Offset: Where data starts
Window = receiver advertised window
Options: negotiate Maximum Segment Size ...

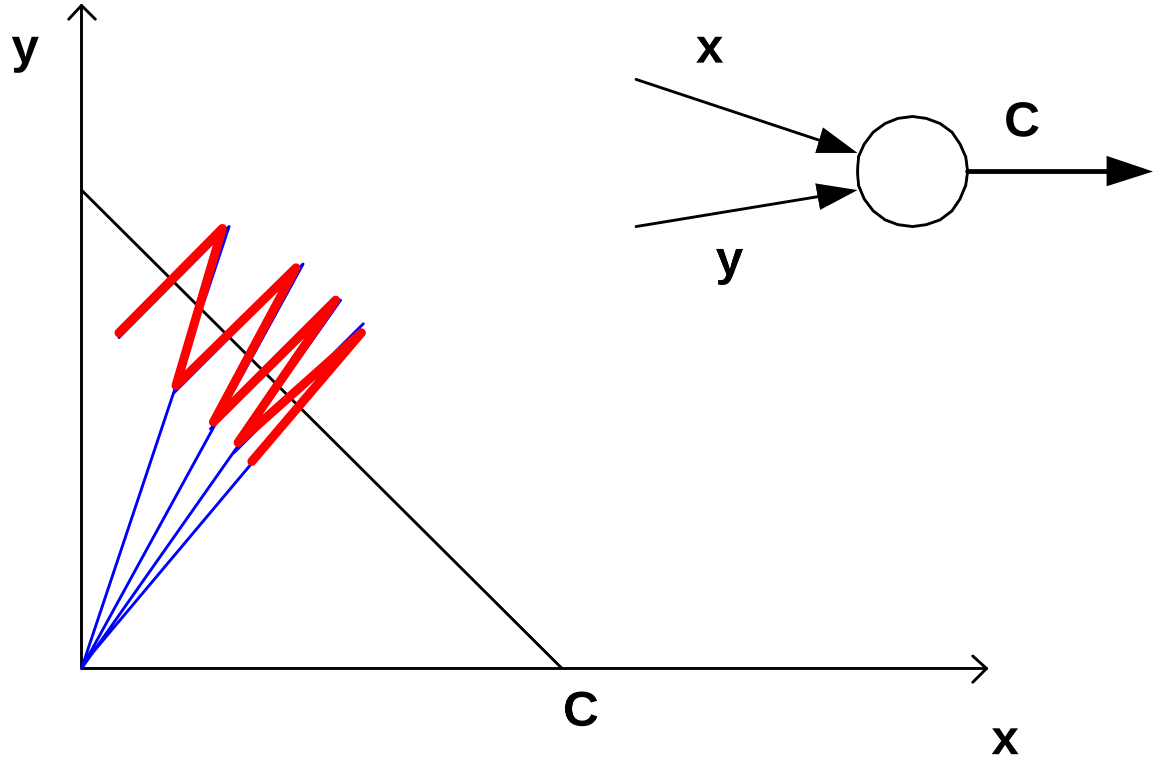
TCP: Window Size



Algorithm:

- Duplicated ACKs => W := W/2 with fast recovery
- Timeout => W = 1, wait T seconds, slow start
- [Double T if repeated timeouts]

TCP: Window Size justification



Additive increase - Multiplicative decrease:
Converges (under ideal assumptions) to fair share

TCP: Random Early Detection

Objectives:

- **Eliminate synchronized reductions in windows by sources**
- **Avoid batch losses by one connection**

Algorithm:

Drop randomly, with probability p(Qav) where

- **Qav = recent average queue length**
- **p(x) increases from 0 to 0.1 as x increases from L to H**

TCP: Explicit Congestion Notification

Objectives:
Avoid dropping packets in router

Algorithm:

- **Router indicates congestion by marking packet instead of dropping it**
- **Receiver sends back the marks in the ACKs**
- **Sender reacts to marks as it would to dup ACKs**

≤

Extensions

- DiffServ
- MPLS
- IPv6
- Ipsec
- Multicast

Extensions: DiffServ

Differentiated Services: Classify with ToS field in IP header

Precise definition of class of service (CoS) is still not adopted.
Some proposals:

- Expedited Forwarding + Best effort (everything else)
- In/Out of profile
- Controlled load with leaky buckets and admission control

Extensions: MPLS

Multiprotocol Label Switching:

Label defines path + QoS (or CoS)
Labels can be stacked

[IPH | MPLS1 | … | MPLSn | IP packet]
Routers use MPLS1, may push a new label or pop label.

Extensions: IPv6

Main objectives of IPv6

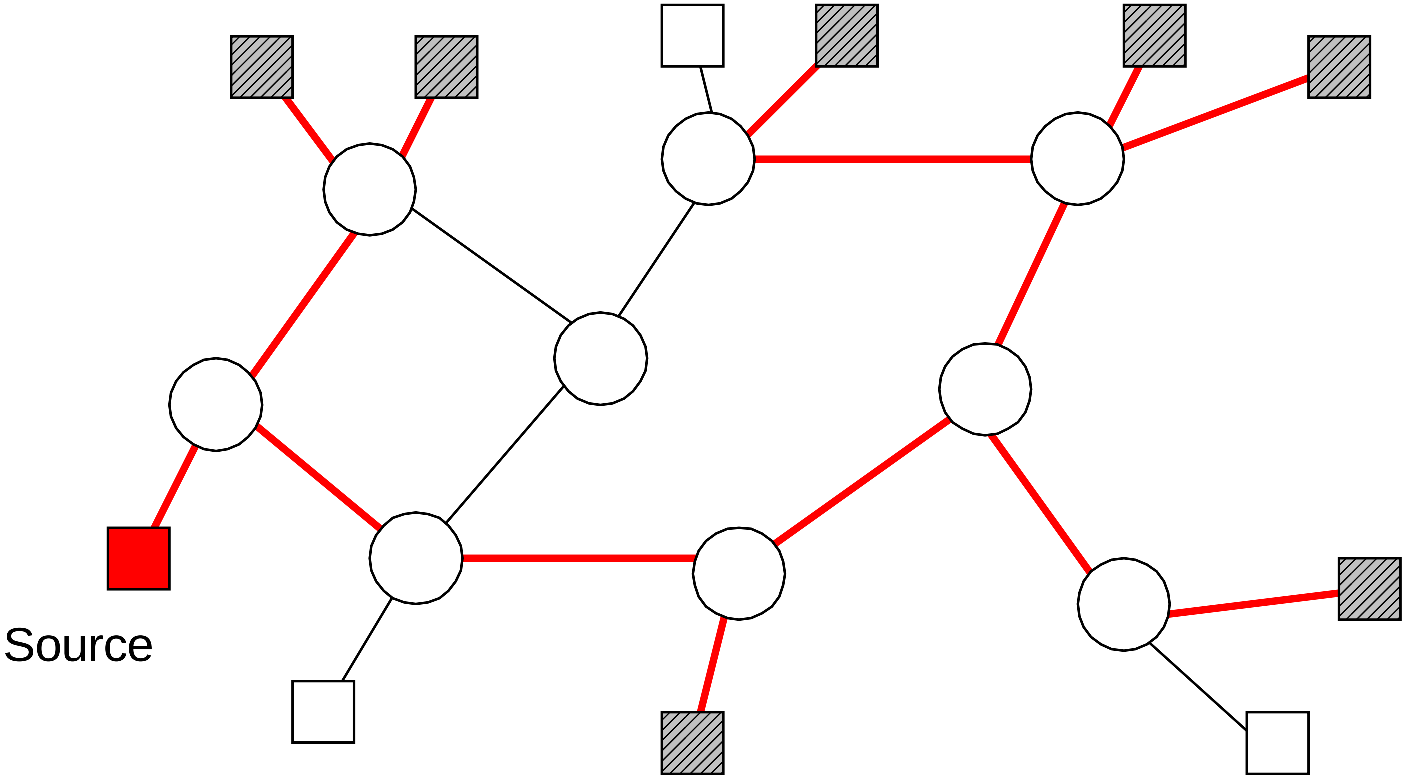
Extend address space: 128 bits instead of 32
Simplify header through optional headers only when needed
May add security at IP level
Normally, fragmentation is done at the source

Extensions: Ipsec

Objective of IPsec:
Secure end-to-end connection

Connection:
Source, destination, protocol, ID
Encryption with DES (40-bit, 56-bit, 2x56-bit)

Extensions: Multicast



At most one copy per link

Join and leave multicast group without informing source

Web Content Delivery Networks

- Motivations
- What is a CDN?
- How does a CDN work?
- Akamai Network
- Benefits
- CDN and Layer 4 switching
- CDN and Caching: Akamai
- References

(Prepared by Yogesh Bhumralkar)

CDN: Motivations (1/2)

- Congestion in the Internet.
- Web Servers sometimes become overloaded due to too many people trying to access their content.
- Communicating directly with the actual servers involves longer delays.
- Caches don’t provide enough control over what data is actually served by them.

CDN: Motivations (2/2)

- Need protection against flash crowds - when content becomes extremely popular over short term. Example: Starr Report or Star Wars trailer.
- Want to distribute content based on geographic location. Consider the following CNN example:
  - Want more servers on east coast serving New York sports related information
  - More west coast servers for Bay Area political news

CDN: What is it?

- Network of content servers deployed throughout the Internet available on a subscription basis to publishers.
- Web publishers use these to store their high-demand or rich content (ie, certain portions of their web site).
- Support for delivery of many content types (e.g, HTML, graphics, streaming media, etc.)
- Brings content closer to end-users but no changes required at end-hosts.

CDN: How does it work? (1/3)

- Preparation:
  - Web publishers decide on the portions of their web site they want to be served by the CDNs.
    - Use CDNs for images or rich content.
    - Most web pages: 70% objects
  - CDN companies provide web content distributors with the software tools to modify their HTML code.
  - The URL’s pointing to these objects on the publishers server are then modified so that the content can now be served from the CDN servers.

CDN: How does it work? (2/3)

- Monitoring/Routing:
  - Some kind of probing algorithms used to monitor state of network - traffic conditions, load on servers, and location of users.
  - generate network map incorporating this information - maps updated frequently to ensure the most current view of the network.
  - CDN develops its own “routing tables to direct the user to the fastest location.”

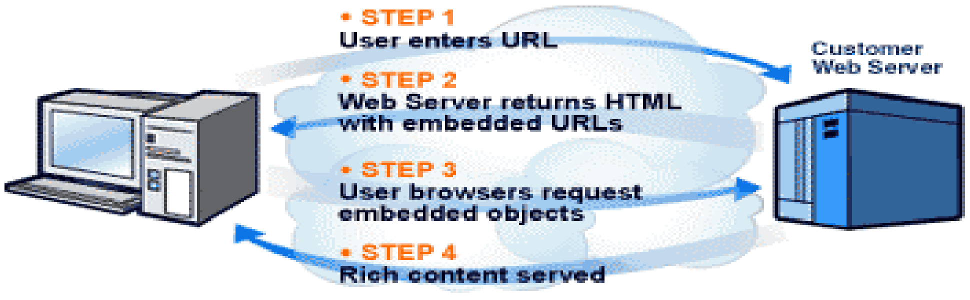
CDN: How does it work? (3/3)

- Delivery:
  - Data to be served by CDNs is pre-loaded onto the servers.
  - CDNs take care of migration of data to the appropriate servers.
  - Users retrieve modified HTML pages from the original server, with references to objects pointing to the CDN.
  - Content is served from the best server.

CDN: Akamai Networks

(pictures obtained from http://www.akamai.com)

Figure 1: "Internet Content Delivery Without FreeFlow"



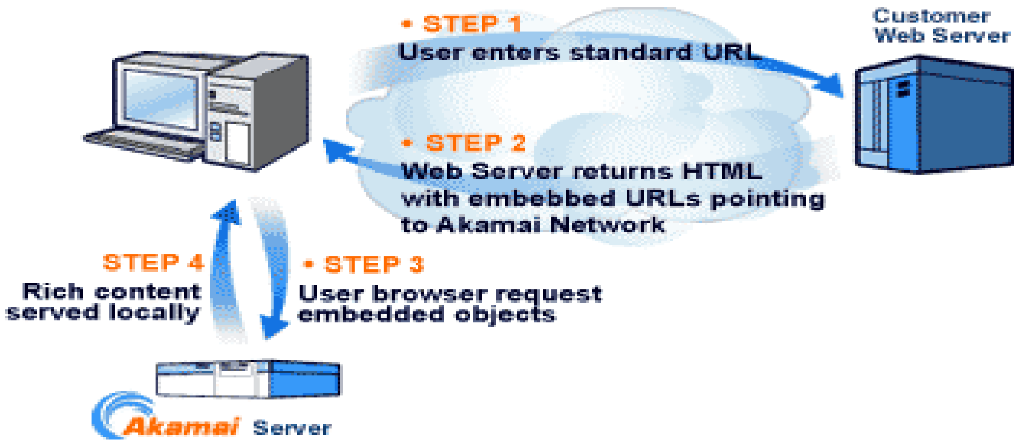
• STEP 1
User enters URL

• STEP 2
Web Server returns HTML with embedded URLs

• STEP 3
User browsers request embedded objects

• STEP 4
Rich content served

Figure 2: "Internet Content Delivery With FreeFlow"



• STEP 1
User enters standard URL

• STEP 2
Web Server returns HTML with embebbed URLs pointing to Akamai Network

• STEP 3
User browser request embedded objects

STEP 4
Rich content served locally

CDN: Benefits

- Highly scalable:
  - As the demand for a document increases, the number of servers serving that document also increases.
  - Ensure that no content server is overloaded by requests.
- Fault Tolerant: guarantee 100% uptime
- High speed connections from content servers to the Internet: Sandpiper - 100Mbps.

CDN and Layer 4 Switching (1/2)

- What is Layer 4 switching?
  - Switch employs the information contained in the transport header to assist in switching traffic.
  - Layer 4 info - port numbers to identify applications (port 80 for HTTP, 20/21 for FTP, etc.)
- Switch keeps track of established sessions to individual servers
  - use Destination IP address + destination port + Source IP address + source port for session identification

CDN and Layer 4 Switching (2/2)

- Switch performs Load Balancing:
  - Multiple servers assigned the same virtual ip address.
  - switch maintains information on server loads.
  - traffic load-balancing done based on specified criteria (e.g., least connections, round robin, etc.)
  - Maintain session management information:
    - ensure that all packets within a session are forwarded to the same server
    - Ex: eShopping sessions: 2 connections - persistent HTTP for shopping cart and SSL for purchases within cart.

CDN and Caching: Akamai

- Akamai servers currently located alongside many ISP caches.
- Content requested from Akamai's web customers is directly served by Akamai servers.
- Cache Interface Protocol: Akamai and Cisco
  - enable third-party caches to store content currently carried on the Akamai network and report back on the performance (ie, number of hits) to web site owners through Akamai's content delivery services.

CDN: References

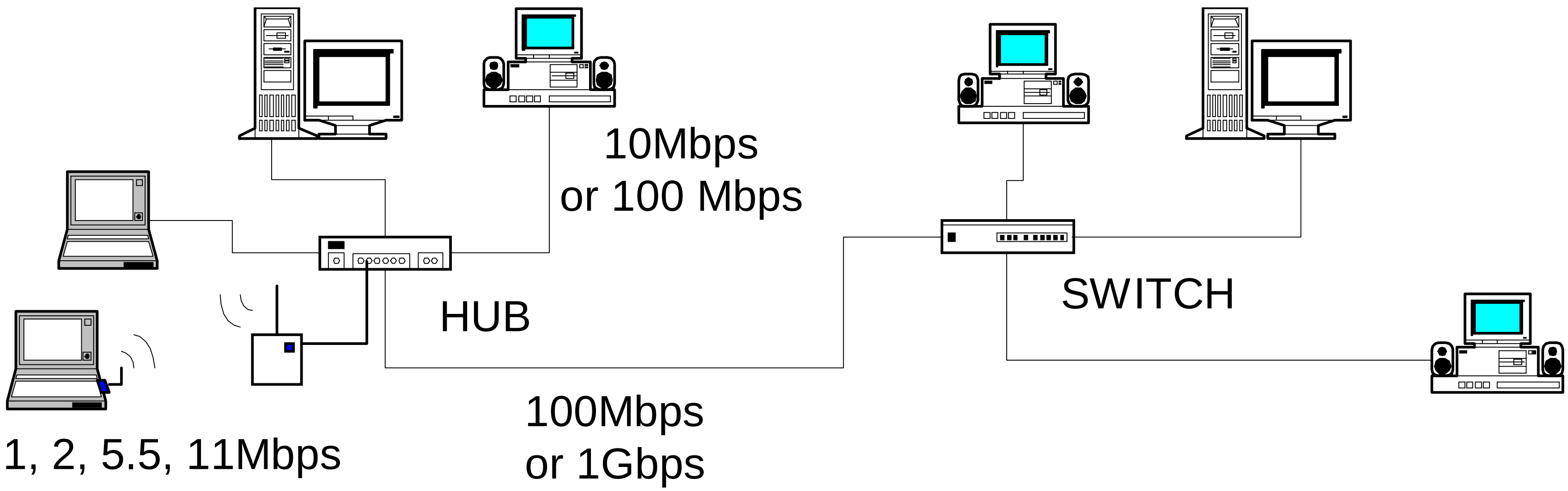
- Akamai Networks: http://www.akamai.com
- Sandpiper Networks: http://www.sandpiper.com

<

LANs: Ethernet

- Overview
- IEEE 802.1-3 Protocols: LLC/MAC/PHY
- Physical Layer
- MAC: Protocol, Frames, ARP, VLAN, Link Aggregation
- LLC
- Routing: Learning bridges, spanning tree

Ethernet: Overview



10Mbps
or 100 Mbps

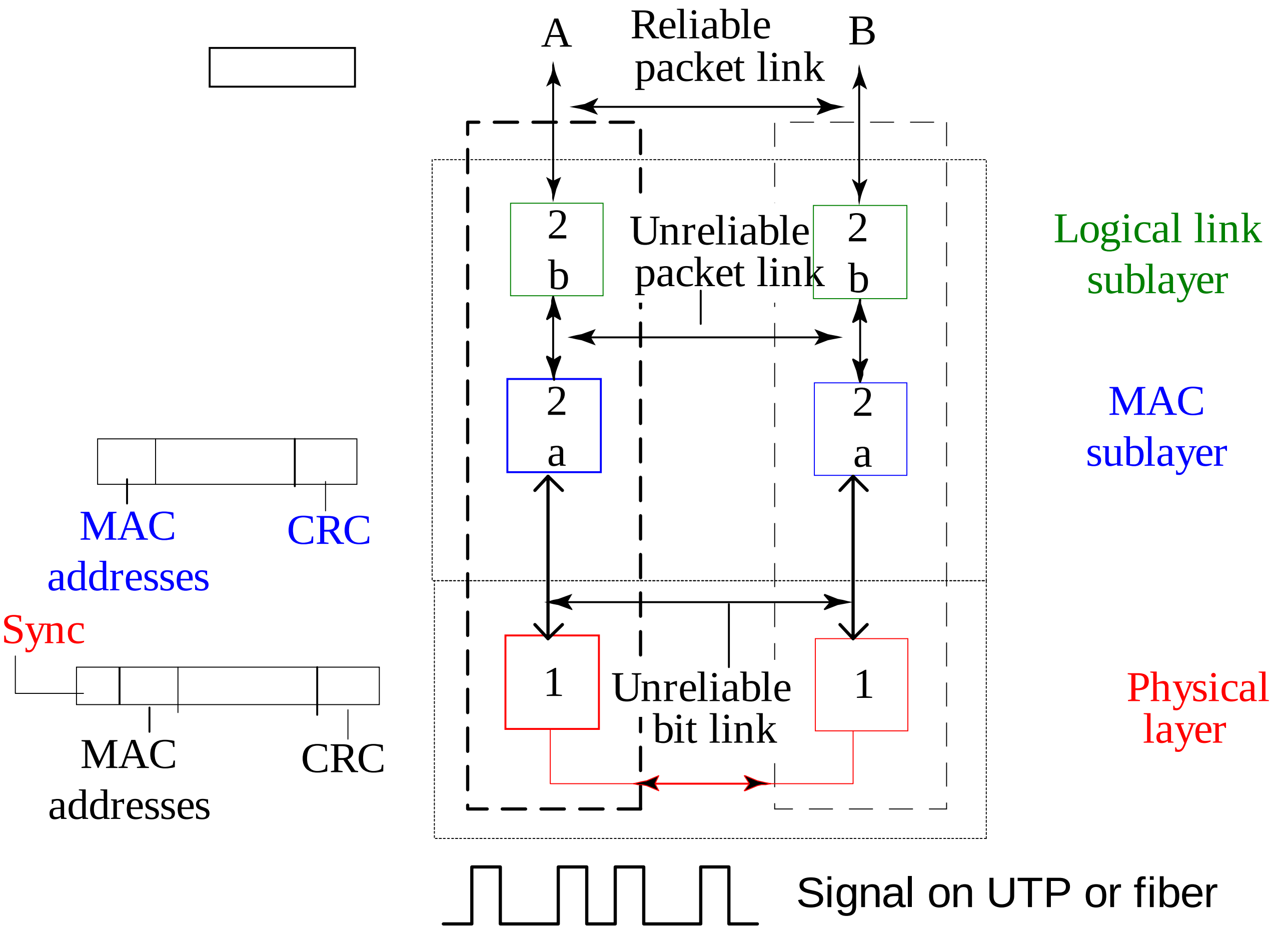
HUB

SWITCH

1, 2, 5.5, 11Mbps

100Mbps
or 1Gbps

Ethernet: IEEE 802.1-3 Protocols



Signal on UTP or fiber

Ethernet: Physical Layer

UTP

unshielded twisted pair

up to 110m

Fiber

100Mbps: 2000m

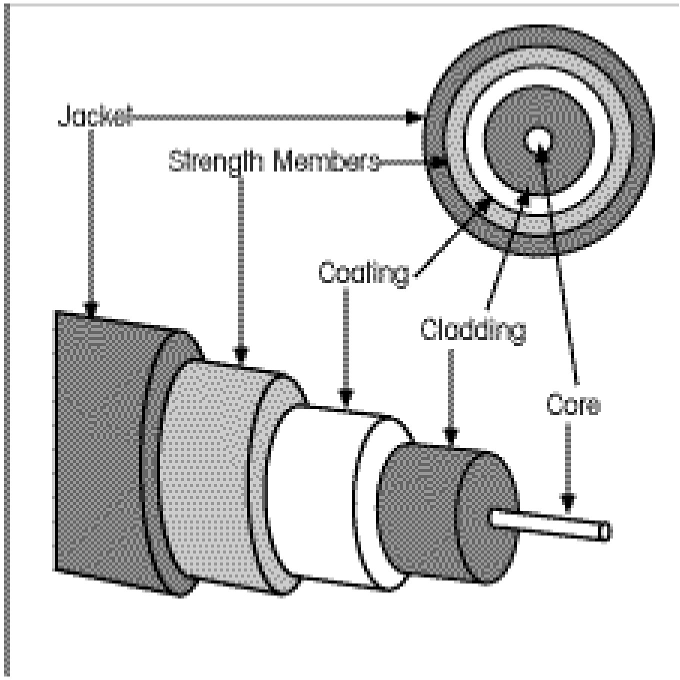
Gbps: 220m, 500m, 5000m

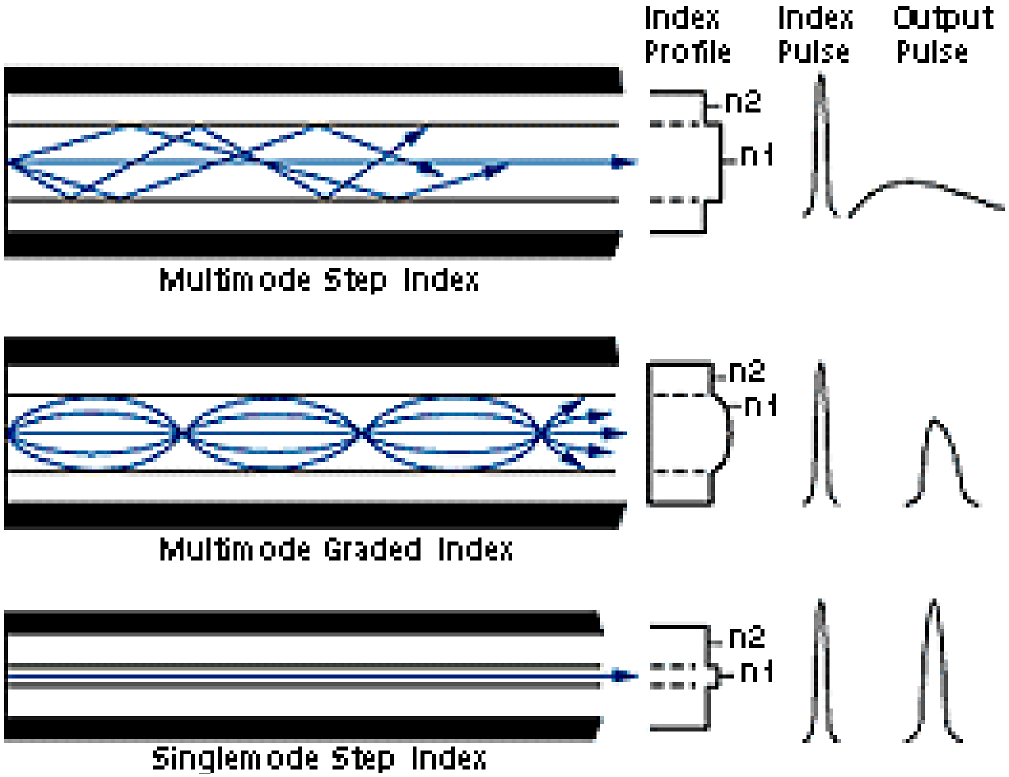
Wireless

2.4GHz DSSS: 1Mbps, 2Mbps, 5.5Mbps, and 11Mbps

25m - 200m

Ethernet: Physical Layer (cont.)



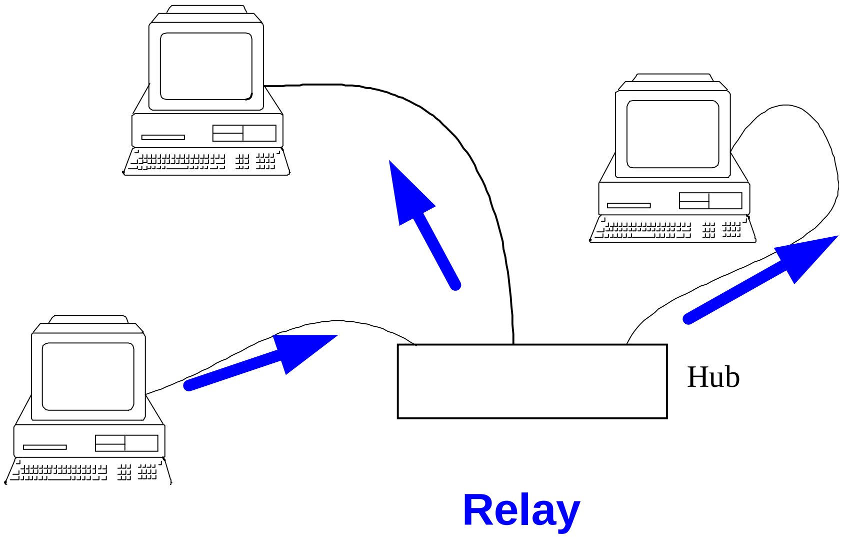


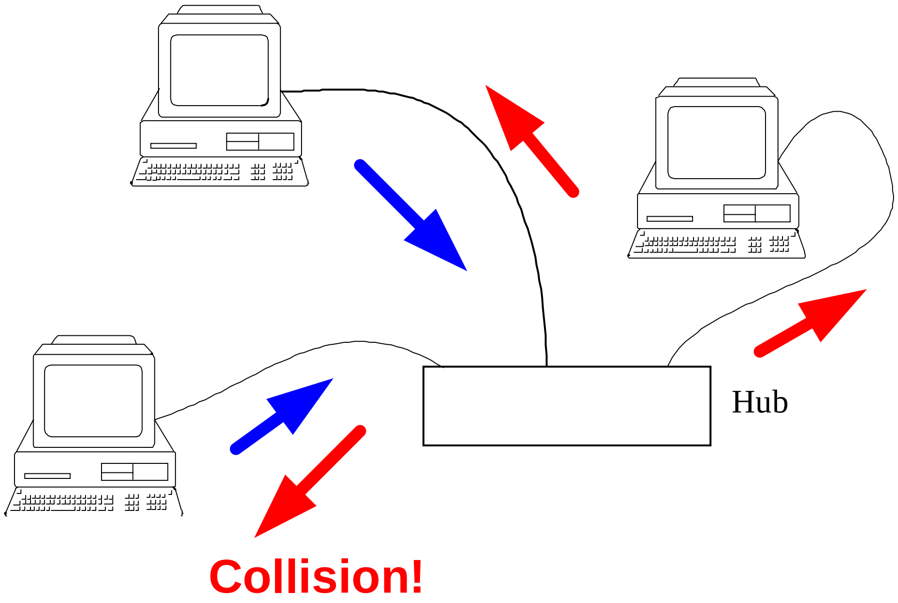
Ethernet: MAC

- Hub
- Switch
- Frame
- ARP
- VLAN
- Link Aggregation

Ethernet MAC: Hub

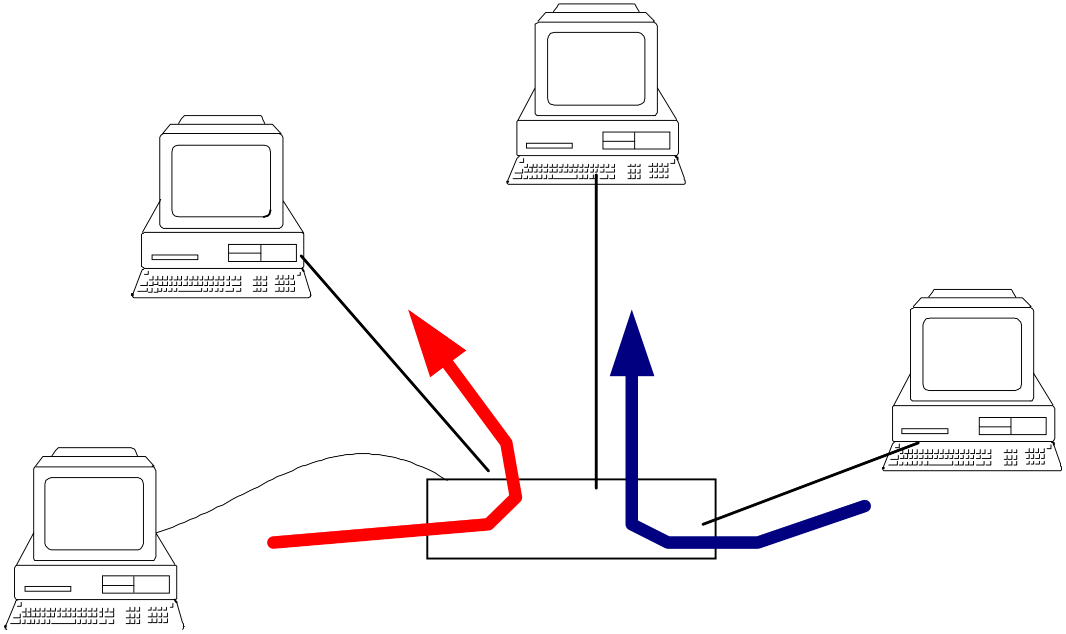
Single collision domain





After a collision, stop for a random time
wait for a random time, then try again.

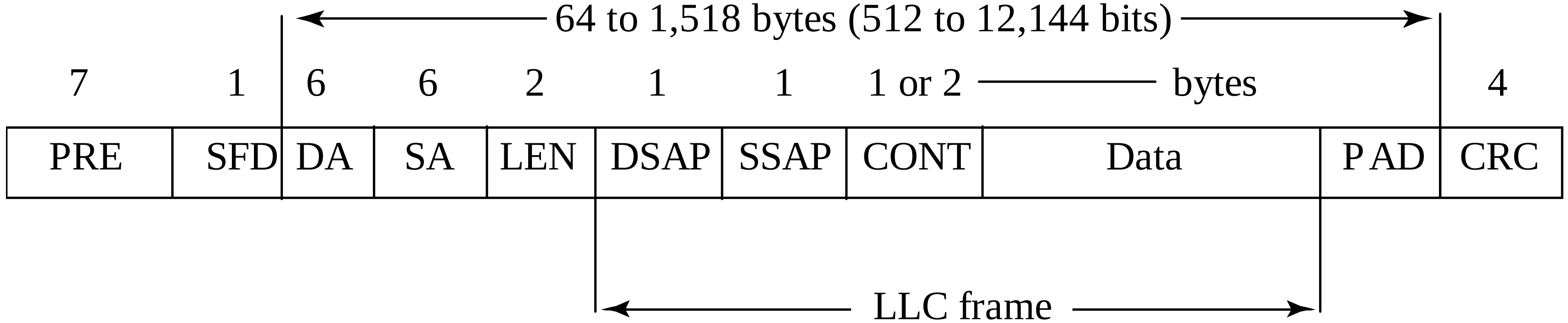
Ethernet MAC: Switch



Switch

Parallel transmissions are possible with a switch.
No collisions.

Ethernet MAC: Frame

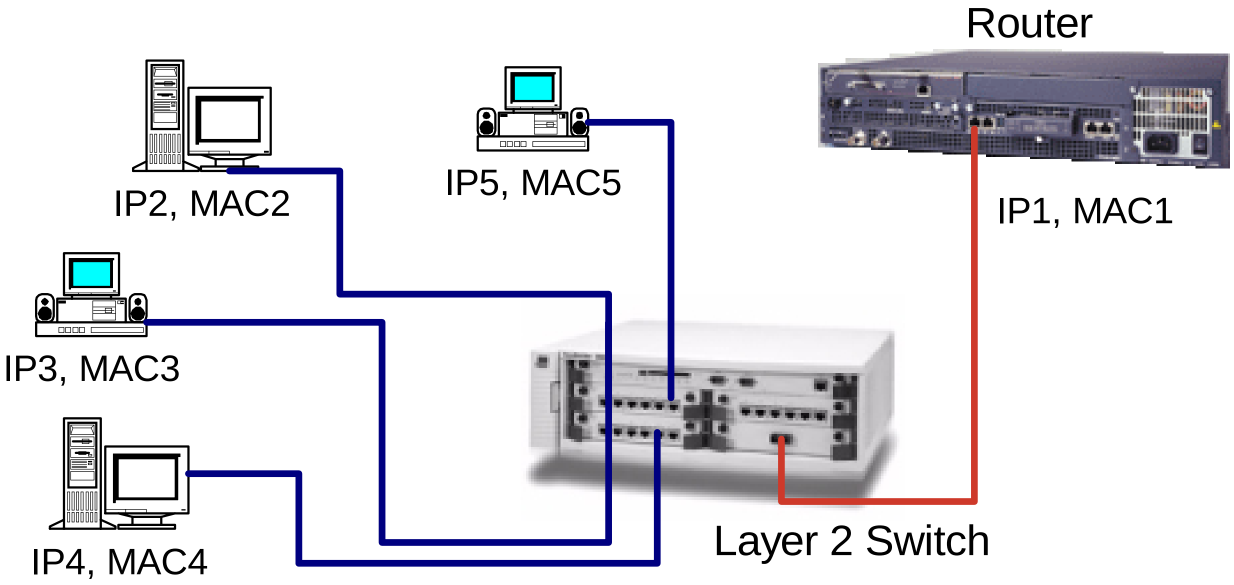


64 to 1,518 bytes (512 to 12,144 bits)

7	1	6	6	2	1	1	1 or 2	bytes		4
PRE	SFD	DA	SA	LEN	DSAP	SSAP	CONT	Data	P AD	CRC

LLC frame

Ethernet MAC: Address Resolution Protocol



To send to IP5, any other node must learn MAC5.

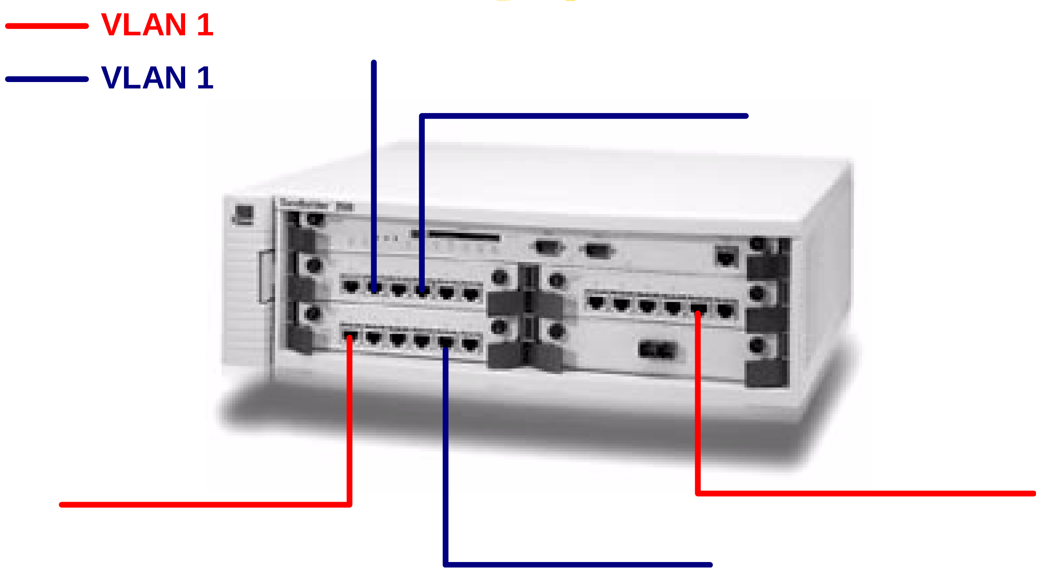
To learn MAC5, IP2 send broadcast message:

[To All, From MAC2, ARP: IP5?]

When it sees this message, IP5 replies:

[To MAC2, From MAC5, ARP: I am IP5]

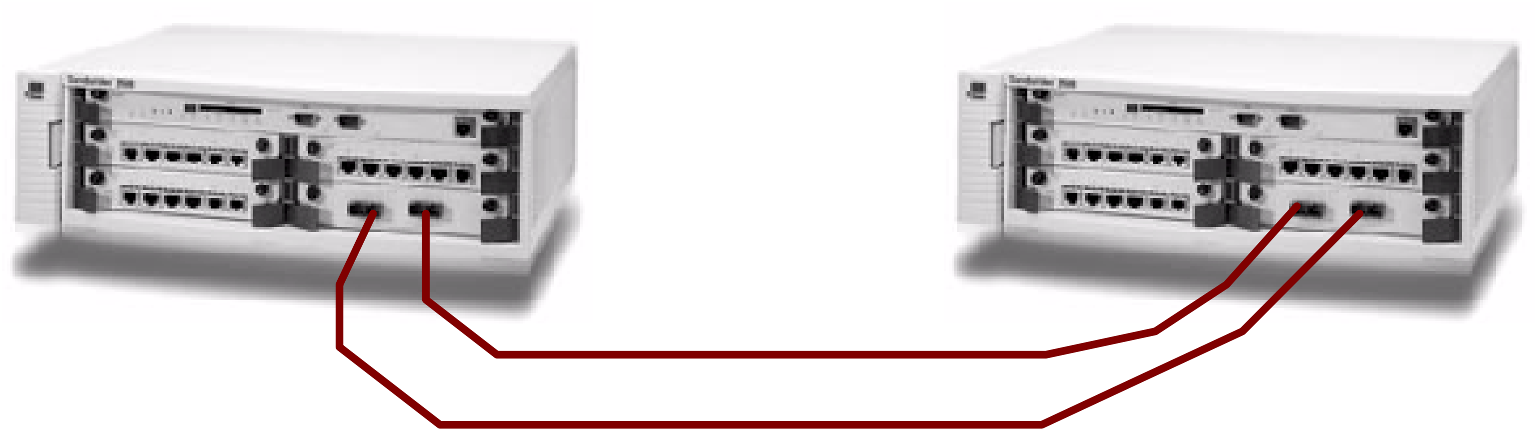
Ethernet MAC: VLAN



4-byte VLAN tag (IEEE802.1Q) inserted after Ethernet header

Broadcasts limited to specific VLAN
Limits scope of ARP packets => Security
(prevents hijacking of connections, can force firewall, …)

Ethernet MAC: Link Aggregation



Two links instead of one

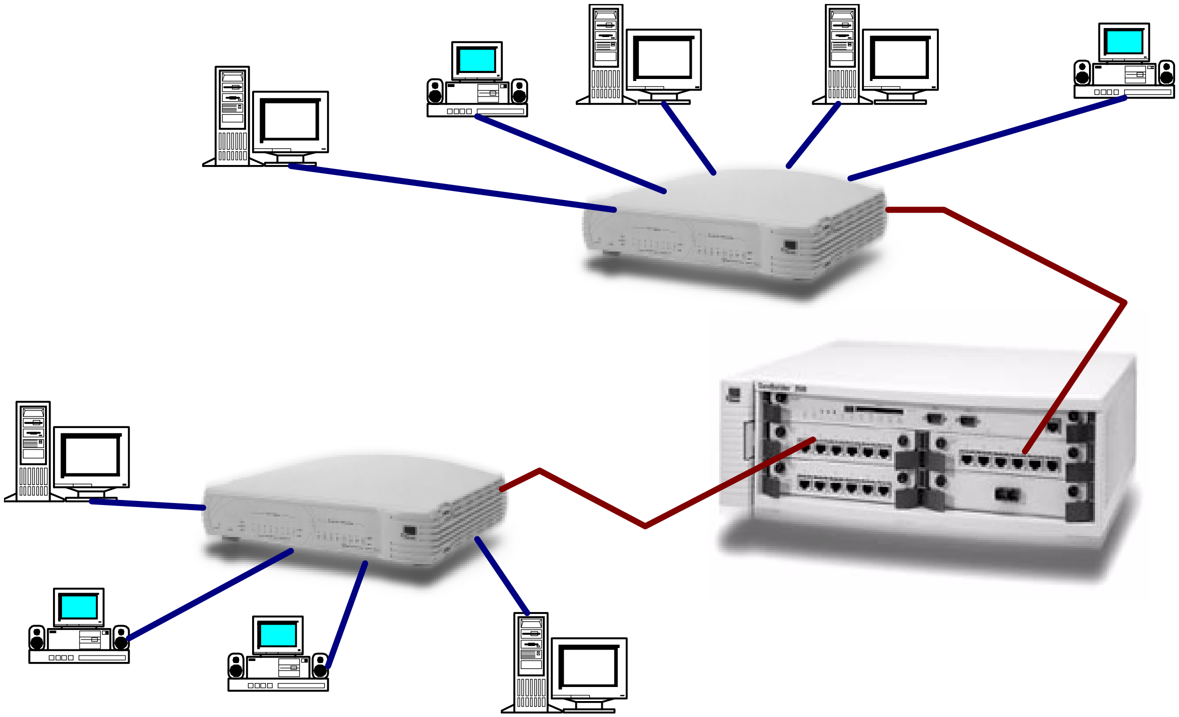
When both links are OK, they are used in parallel (aggregation)

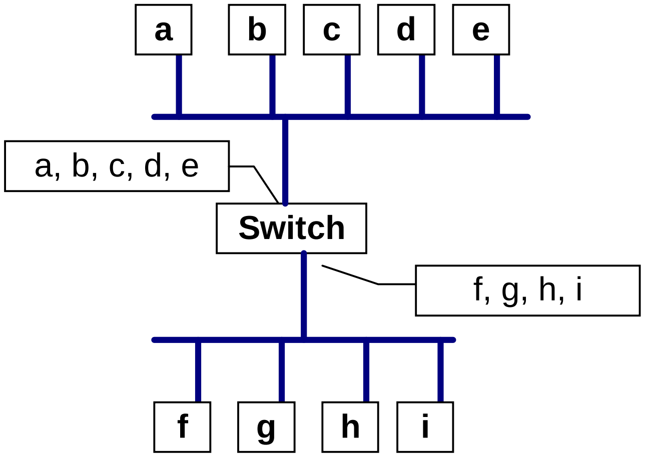
When a link fails, switch sends traffic to other link

Ethernet : LLC

- Connection-oriented or connectionless services
- Acknowledged of unacknowledged
- Multiplexing

Ethernet : Routing: Learning

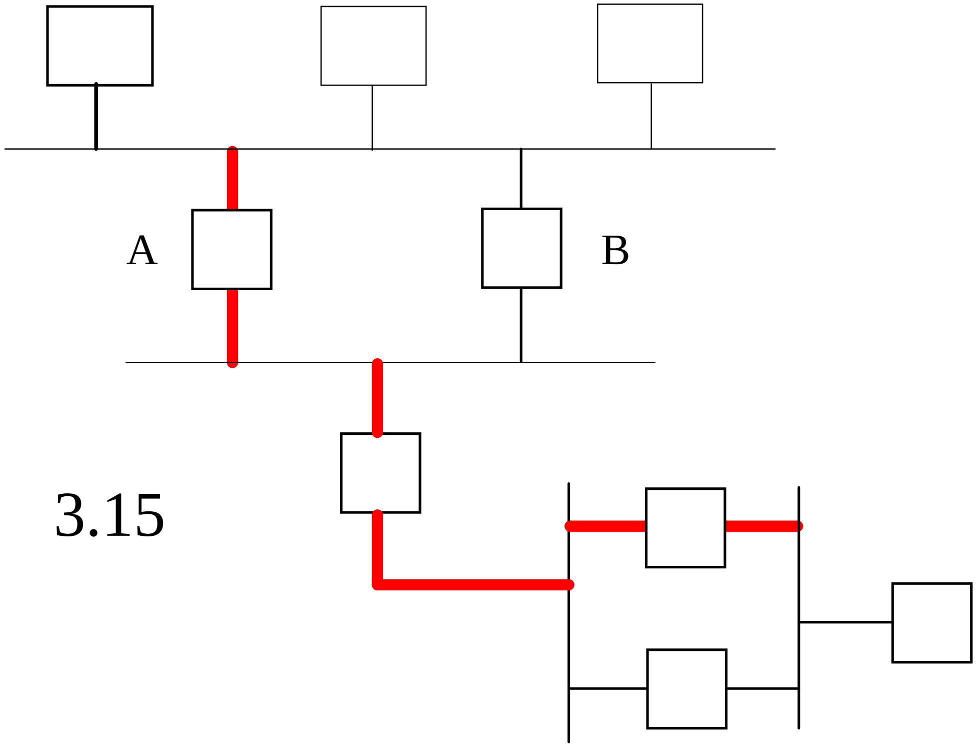




Switch learns the MACs attached to a port by watching the source addresses of packets that arrive at that port.

The switch then knows not to forward arriving packets destined to same port.

Ethernet : Routing: Spanning Tree



3.15

To avoid multiple copies and loops, nodes construct a spanning tree.
To construct the tree, they exchange messages of the form
[x | r | d]
where x = MAC of sender
r = MAC of assumed root
d = distance from r to x.
Each node stores the best message received so far. Best means with a smallest r and d to that r.

When it has found a better r, a node only relays those messages while adding 1 to the value of d.

<

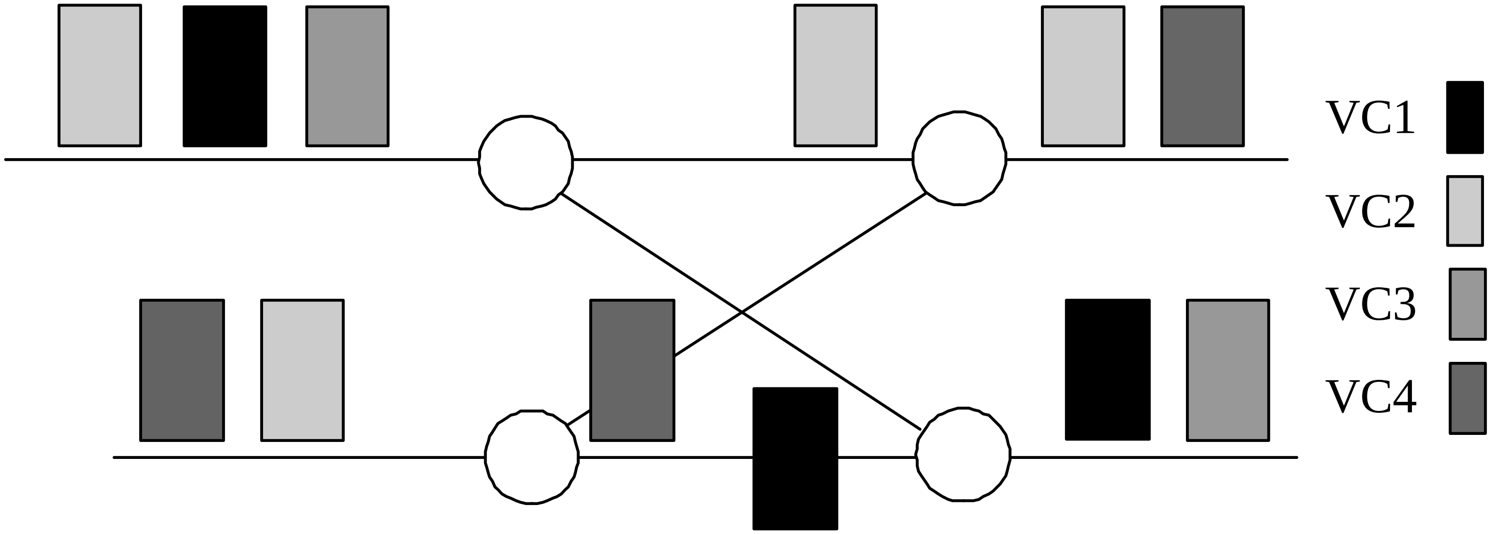
ATM (Asynchronous Transfer Mode)

- Main Features
- Network Layer
- ATM Header Structure
- ATM Adaptation Layer (AAL)
- Management and Control
- ATM over SONET
- Internetworking with ATM

ATM: Main Features

- Connection-Oriented Service
- 53-byte Cells
- Different QoS for different VCCs

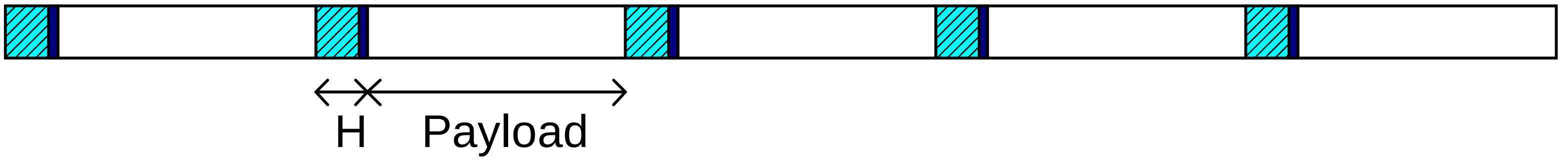
ATM - Main Features: Connections



Cells from the same VC follow the same path.
The cell header specifies the VC
Forwarding: Table (VCin, PORTin) -> (VCout, PORTout)
Note: Statistical multiplexing on every link (not TDM).

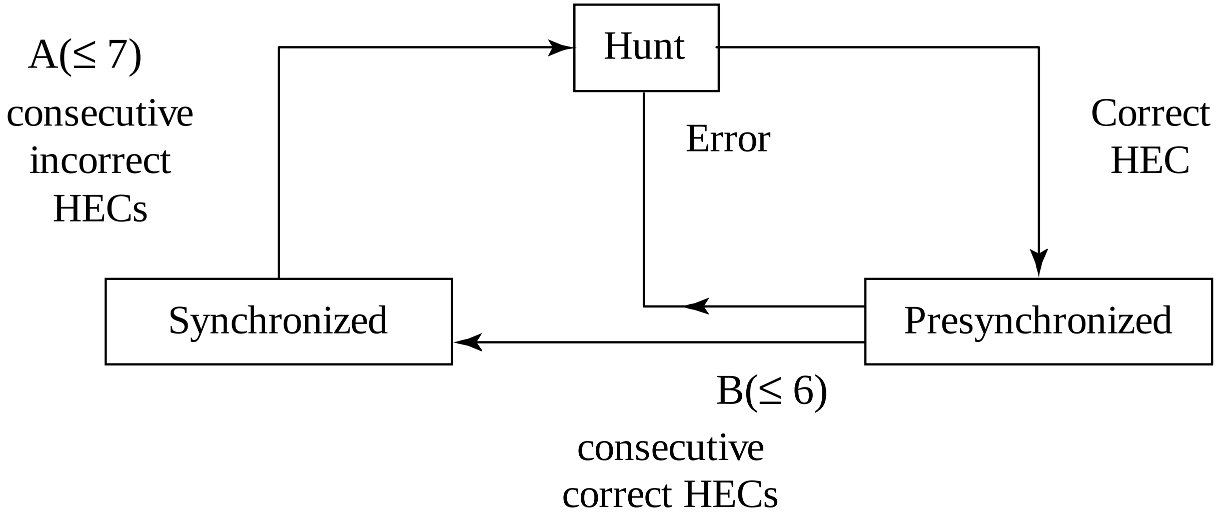
[6.1]

ATM - Main Features: Cells



H

Payload



Hunt

Error

Correct
HEC

A(≤ 7)
consecutive
incorrect
HECs

Synchronized

Presynchronized

B(≤ 6)
consecutive
correct HECs

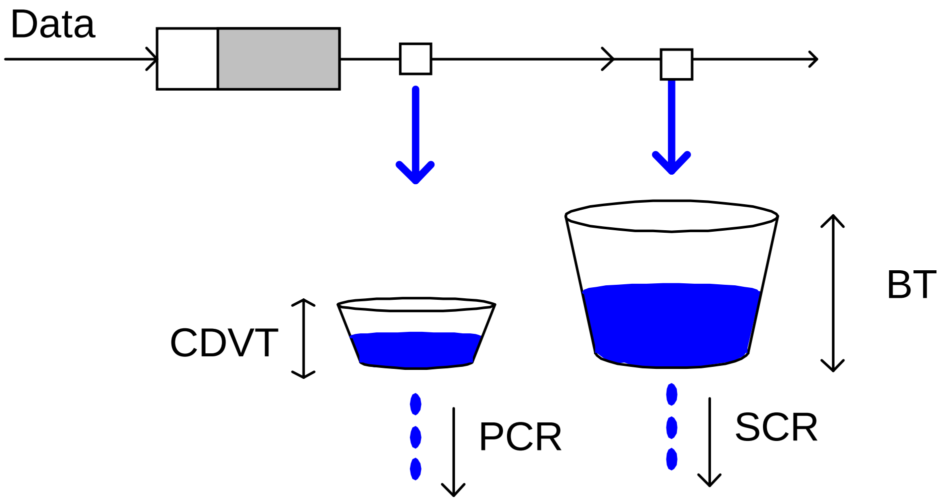
[6.2]

The fifth byte of the header is a checksum calculated from the previous four bytes. This feature enables a receiver to locate the cells in a byte stream.

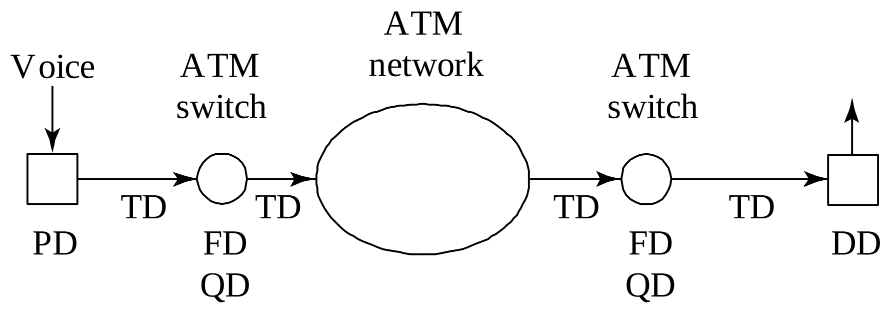
ATM - Main Features: QoS

← QoS → ← traffic →

Attributes	CBR	VBRrt	VBRnrt	ABR	UBR
CLR	Specified	Specified	Specified	Specified	Unspecified
CTD, CDV	CDV, max CTD	CDV, max CTD	Mean CTD	Unspecified	Unspecified
PCR, CDVT	Specified	Specified	Specified	Specified	Specified
SCR, BT	n/a	Specified	Specified	n/a	n/a
MCR	n/a	n/a	n/a	Specified	n/a
Congestion control	No	No	No	Yes	No



ATM - Main Features: QoS (cont.)

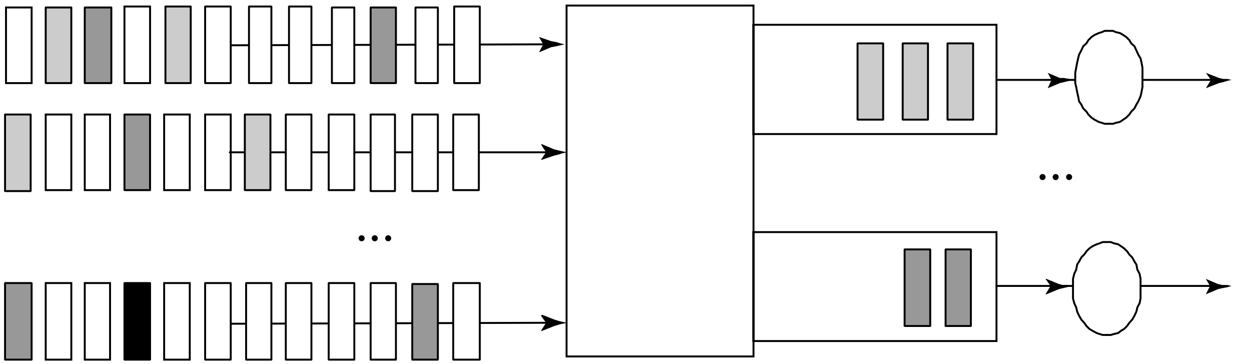


Assumptions
Voice transmission (64 Kbps)
Transmission rates = 155 Mbps
Length of path = 1,000 km
Path goes through 5 nodes

Delay	Value in μs
PD = Packetization delay	6,000
TD = Transmission delay (including propagation)	5,000
FD = Fixed processing delay	280
QD = Queuing delay	70
DD = Depacketization delay	70
Total delay	11,420
Delay jitter	70

Assumptions: Voice traffic over ten 155-Mbps links across USA.

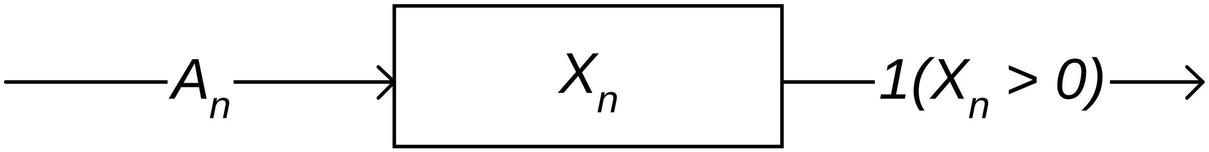
ATM - Main Features: QoS (cont.)



Analysis:

Flow of cells that arrive at an output queue is almost Poisson.
(Sampling of many independent flows. Each flow has a small sampling rate)

ATM - Main Features: QoS (cont.)



A_n

X_n

$1(X_n > 0)$

$$X_{n+1} = X_n + A_n - 1(X_n > 0)$$

Taking expectation gives $P(X_n > 0) = E(A_n) =: \rho$

Calculating $E(.)^2$ gives

$$E(X) = \frac{2\rho - \rho^2}{2(1 - \rho)}$$

ATM - Main Features: QoS (cont.)

Resource allocation:

- Signaling (Based on Q.2931):
  Setup, Ack, PNNI, Setup, {Connect, Release}, …, {Connect, Release},
- Call Admission Control
- Shaping at the source UNI, policing at the network UNI
- Buffer and capacity reserved for VC (=> Scheduling)

ATM: Network Layer

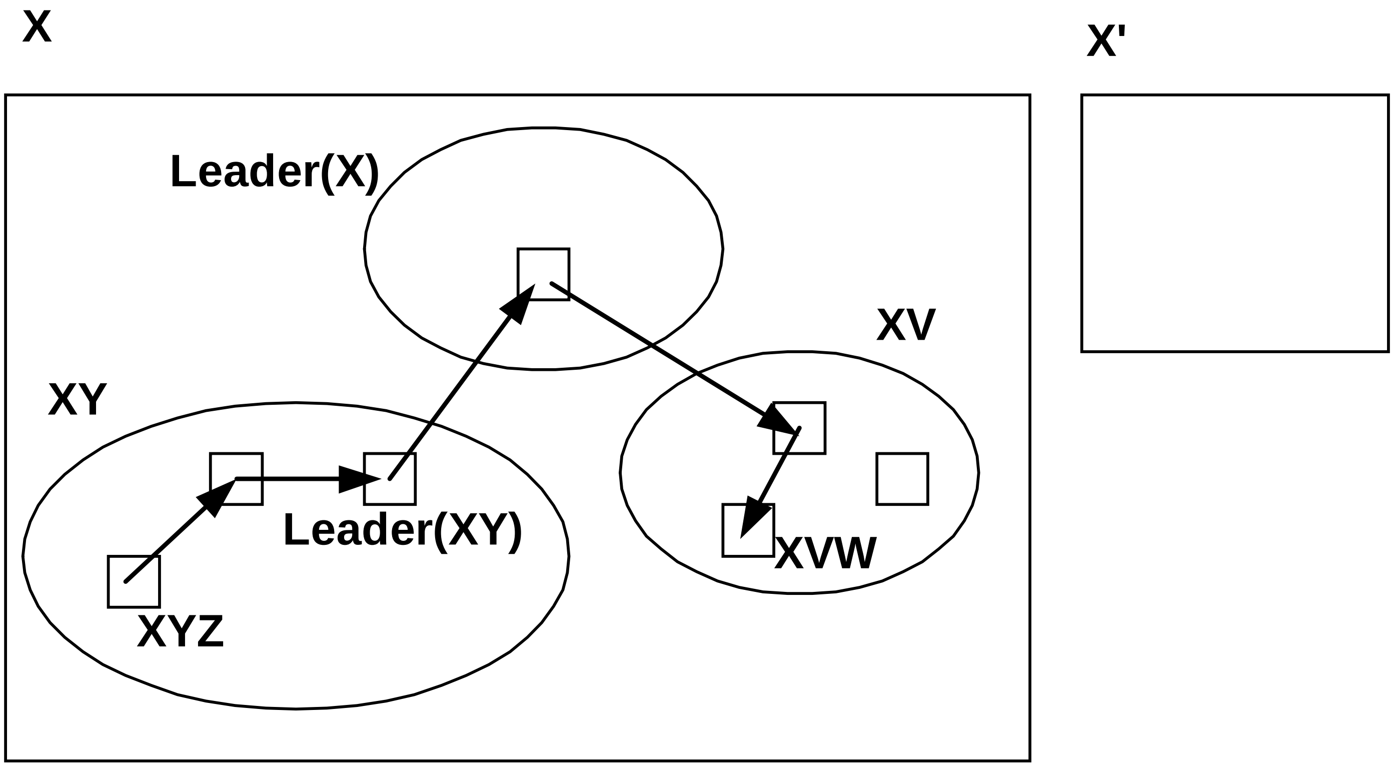
- Addressing
- Routing

ATM - Network Layer: Addressing

Hierarchical:

- nodes belong to a peer group with a leader
- leaders are nodes for the next level
- nodes compute intra-group routing tables
- leader maintains aggregate descriptions of group

ATM - Network Layer: Addressing (cont.)

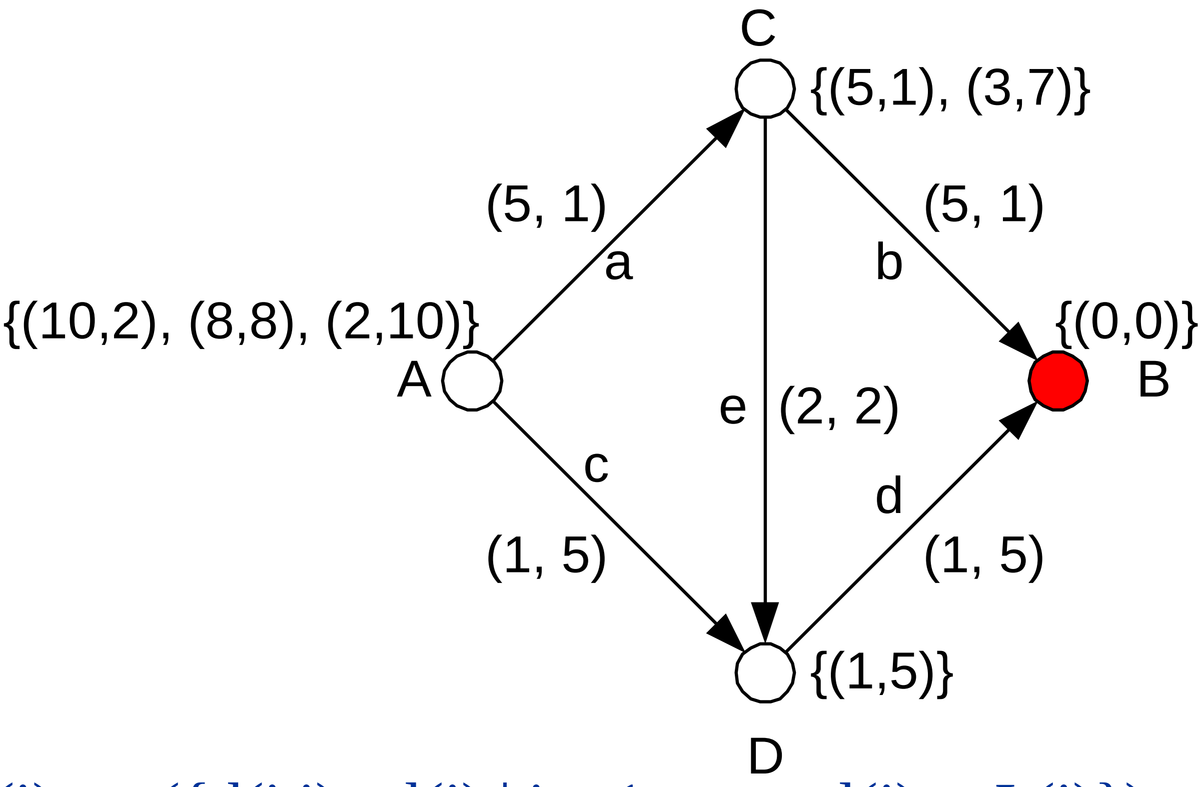


Hierarchical routing via group leaders.

ATM - Network Layer: Routing PNNI

- Algorithm: Dijkstra with vector of parameters
  - Nodal parameters (security, reliability)
  - Link parameters (delay, capacity)

ATM - Network Layer: Routing PNNI (cont.)



$L(i) = \mu(\{d(i,j) + l(j) \mid j = 1, \ldots, n; l(j) \in L(j)\})$

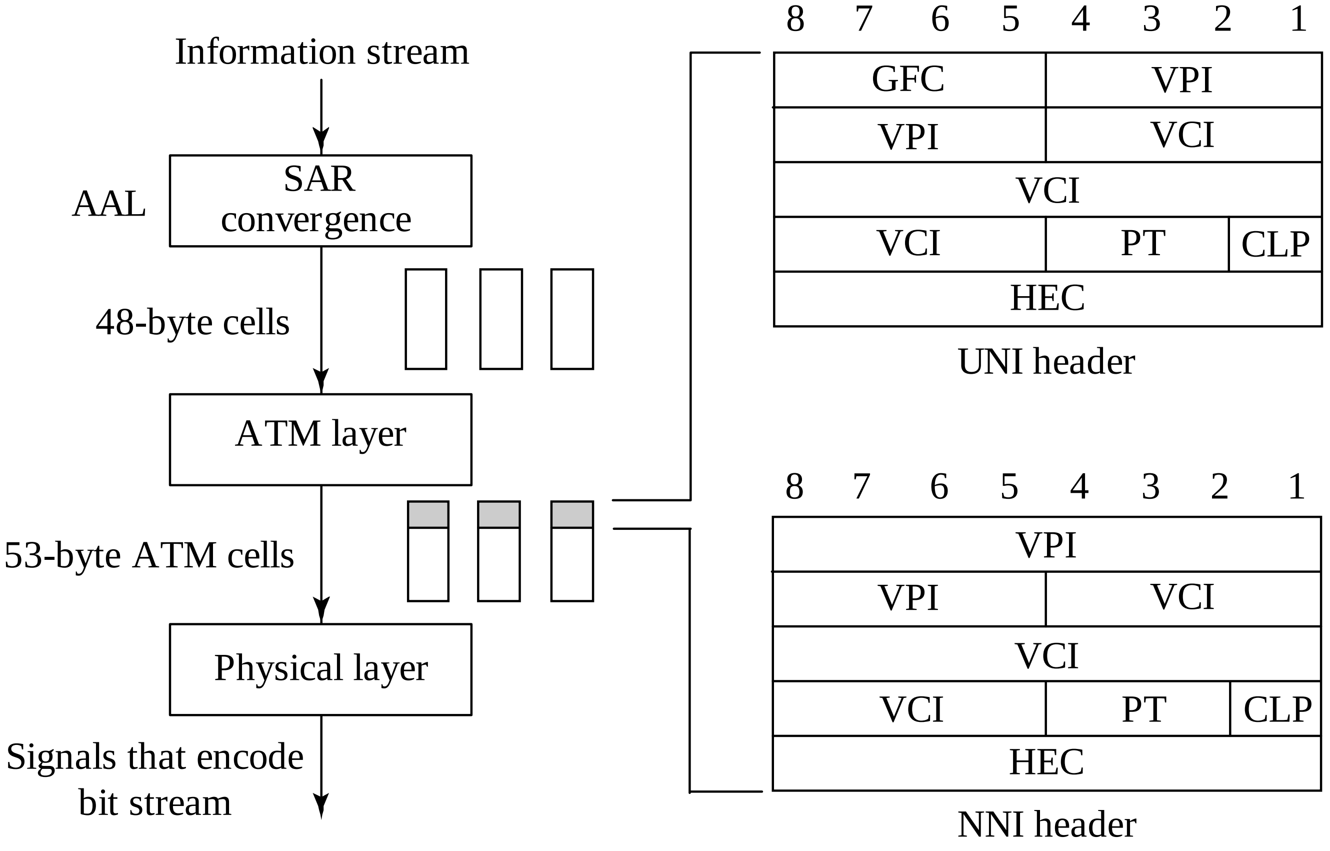
where $\mu(L)$ = subset of noninferior metric vectors.

Initial values: $L(B) = \{(0,0)\}$, $L(i) = \{(\infty, \infty)\}$ for $i \neq B$.

ATM: Header Structure

- Header
- VCI & VPI
- Other fields

ATM - Header Structure: Header



8	7	6	5	4	3	2	1
GFC				VPI			
VPI				VCI			
VCI							
VCI				PT			CLP
HEC							

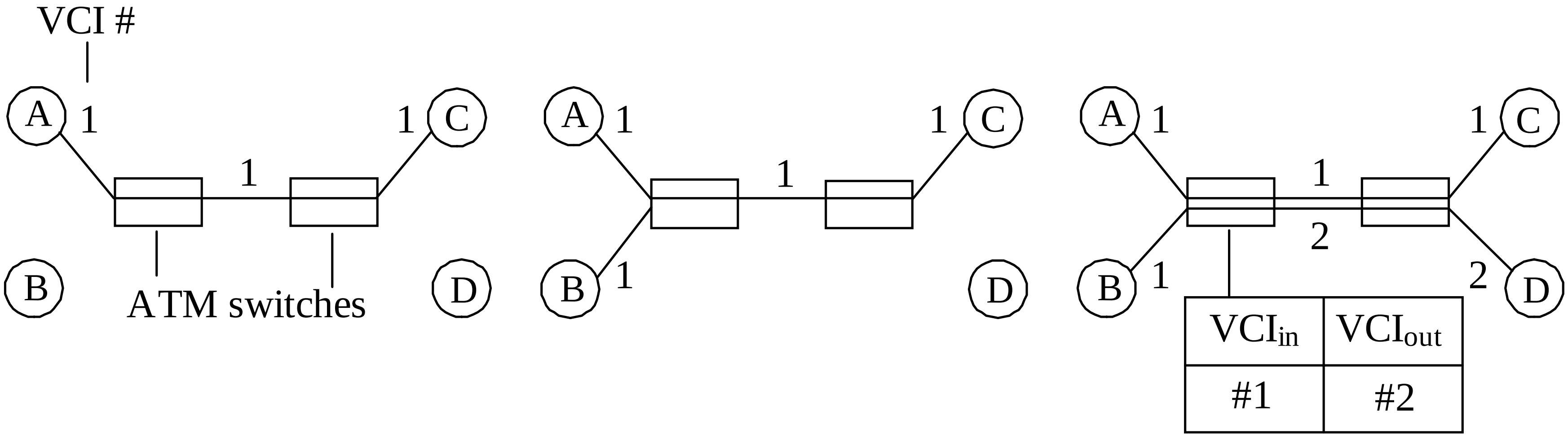
UNI header

8	7	6	5	4	3	2	1
VPI							
VPI				VCI			
VCI							
VCI				PT			CLP
HEC							

NNI header

GFC: generic flow control; VPI: virtual path identifier; PY: payload type; VCI: virtual circuit identifier; CLP: cell loss priority; HEC: header error control.

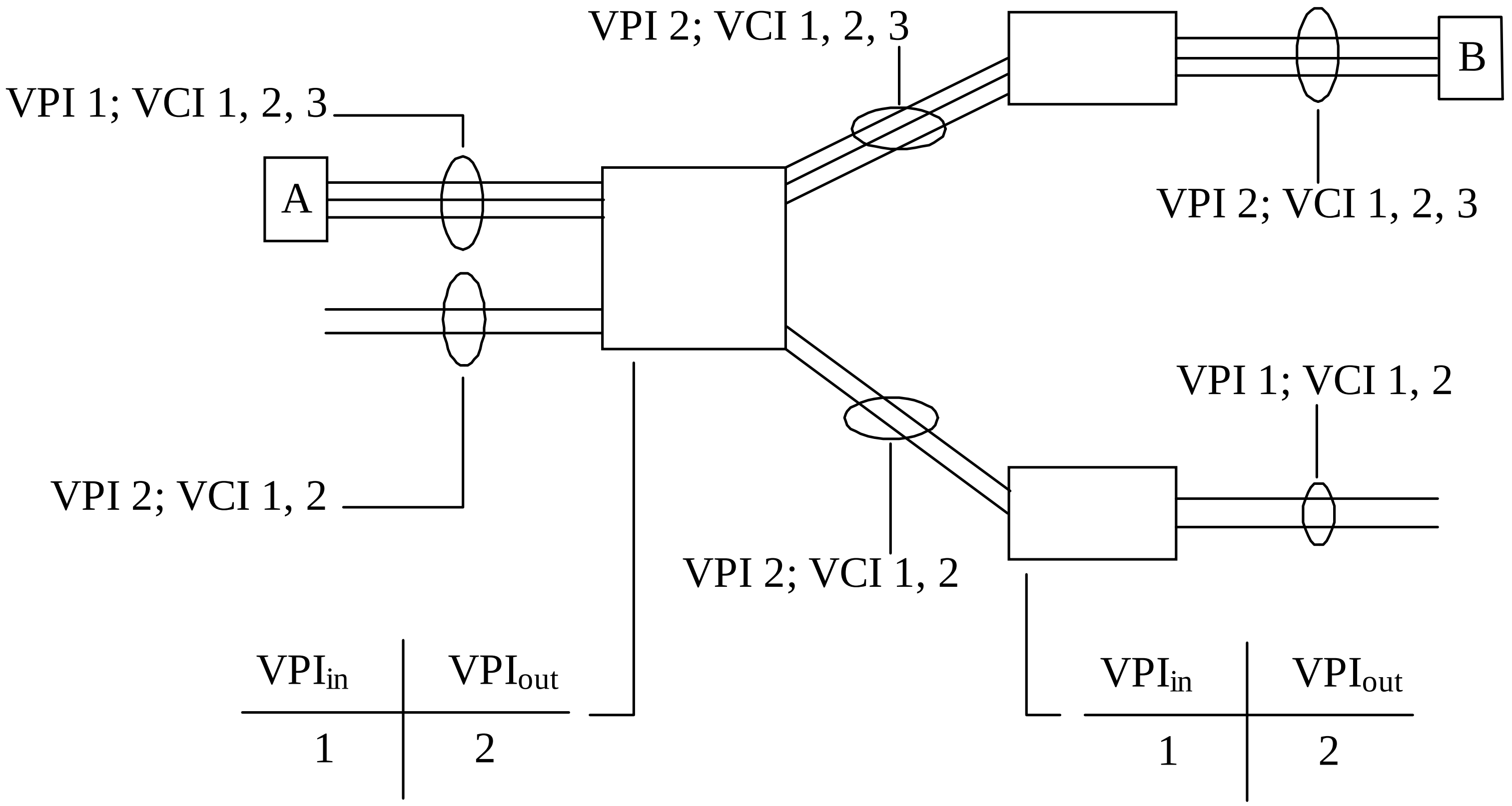
ATM - Header Structure: VCI



VCI$_{in}$	VCI$_{out}$
#1	#2

(1) (2) (3)

ATM - Header Structure: VPI



VPI_{in}	VPI_{out}
1	2

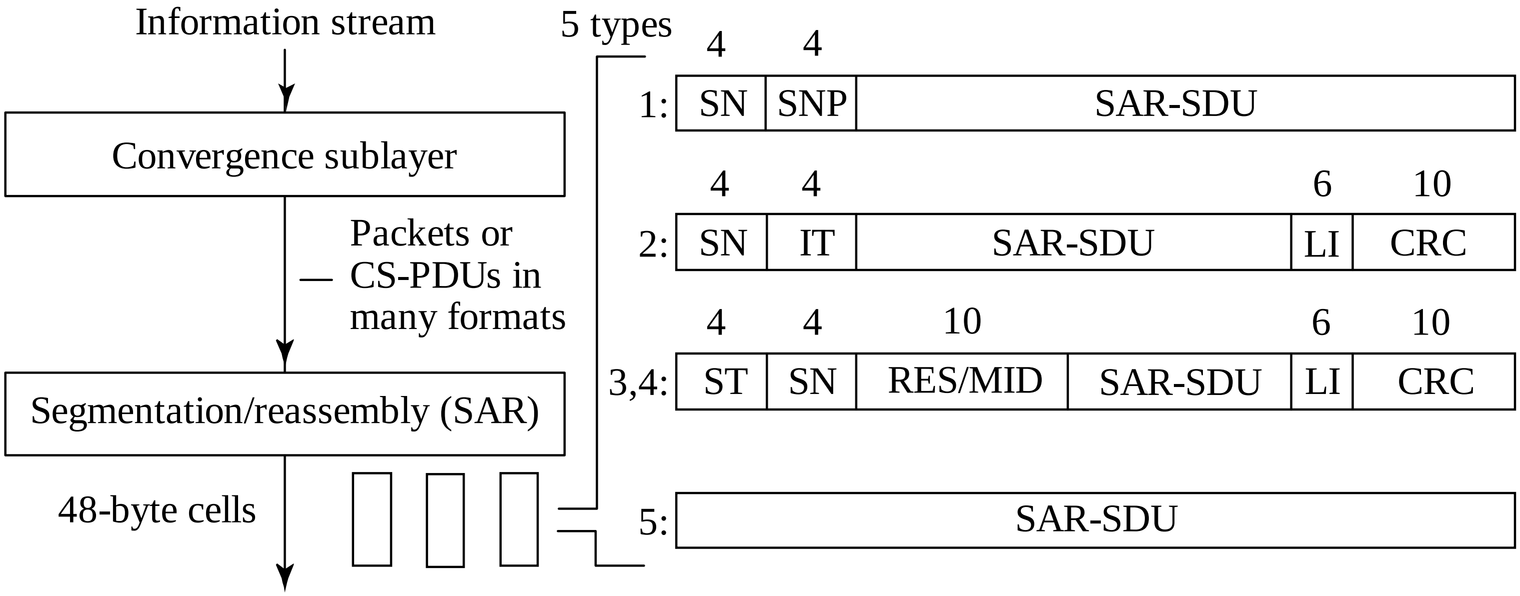
VPI_{in}	VPI_{out}
1	2

Note: Reserved VPI/VCI (Signaling, ILMI, OAM, idle cells)

ATM - Header Structure: Others

- GFC: At UNI, to change the rate ...
- PT: Payload Type
  - 0xx = user information
    (00x = no congestion, 01x = congestion)
  - 100, 101: control information
  - 110 = resource management cells
- CLP = cell loss priority
  - CLP = 0: do not discard
  - CLP = 1: may discard (non-conformant at policing)
- HEC = 01010101 + CRC on four previous bytes
  (Generator = 100000111)

ATM: Adaptation Layer



Information stream

Convergence sublayer

Packets or CS-PDUs in many formats

Segmentation/reassembly (SAR)

48-byte cells

5 types

1:

4	4	
SN	SNP	SAR-SDU

2:

4	4		6	10
SN	IT	SAR-SDU	LI	CRC

3,4:

4	4	10		6	10
ST	SN	RES/MID	SAR-SDU	LI	CRC

5:

SAR-SDU

1: CBR, either read at fixed rate or with explicit time stamp; SN(P) = seq. No. (protection)

2: Low bit rate, possible multiplexing and padding

SAR-PDU: start field (offset), header (channel ID for MUX), data, padding; LI = length

3/4: for interconnecting SMDS and MANs; error-free or with errors

ST = 10 for BOM, 00 for COM, 01 for EOM, 11 for single cell; MID = multiplex identifier

5: for IP packets: IP -> [length | data | pad | CRC] = nx48 bytes -> AAL5 cells;

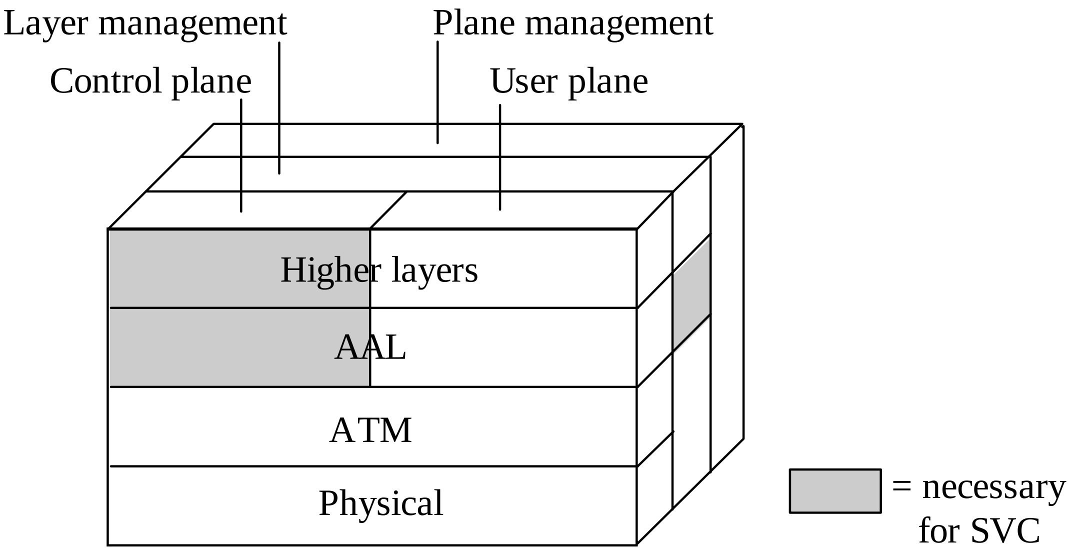
PT in cell header marks last cell of packet

ATM: Management and Control

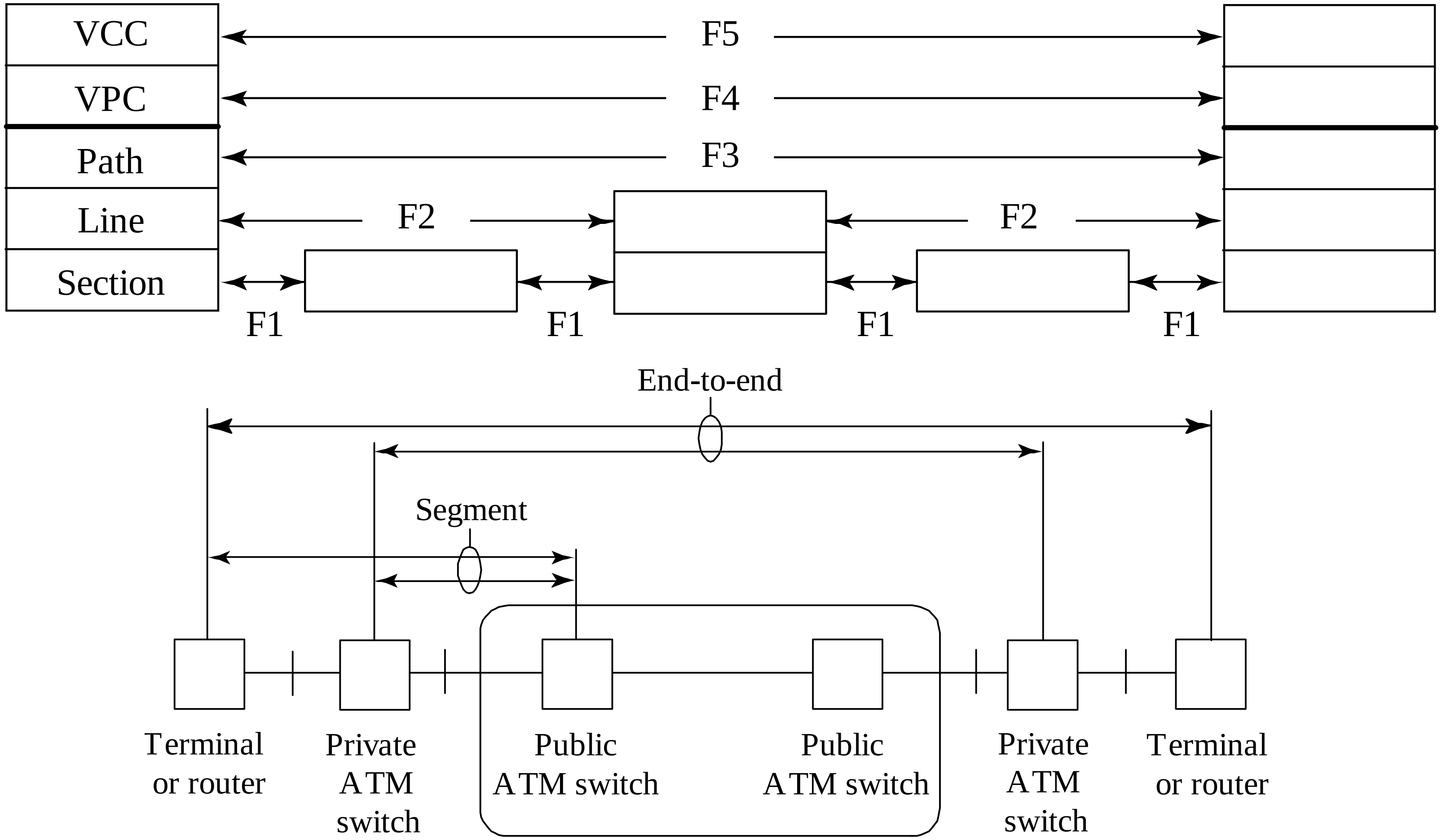
- Functions
- Fault
- Monitoring and Configuration
- User-Network Signaling

ATM - Mngt: Functions

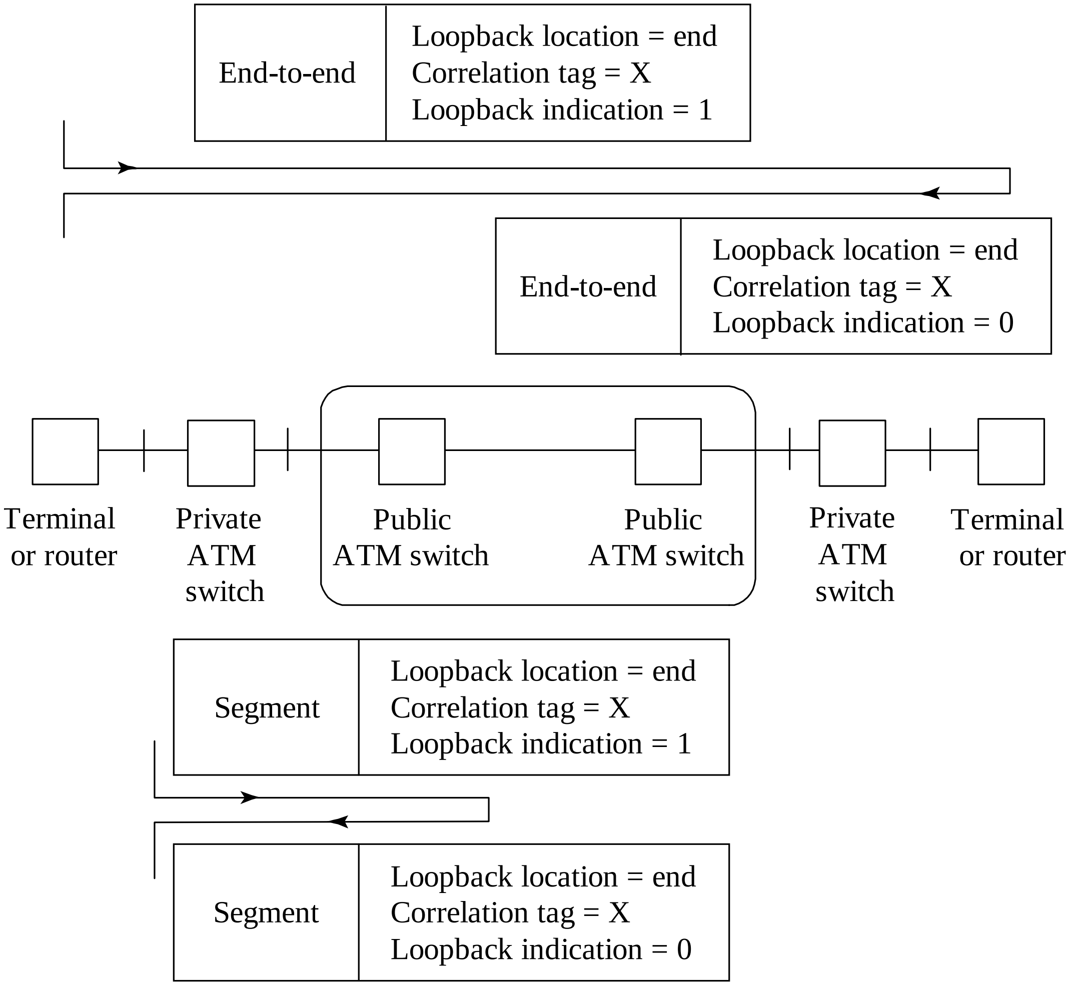
- Fault Management
- Traffic and Congestion Control
- Network Monitoring and Configuration
- User/Network Signaling



ATM - Mngt: Faults



ATM - Mngt: Faults (cont.)



End-to-end

Loopback location = end
Correlation tag = X
Loopback indication = 1

End-to-end

Loopback location = end
Correlation tag = X
Loopback indication = 0

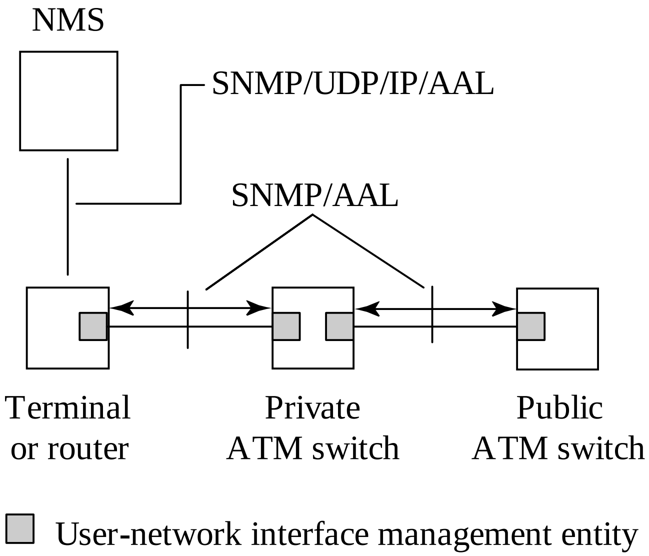
Segment

Loopback location = end
Correlation tag = X
Loopback indication = 1

Segment

Loopback location = end
Correlation tag = X
Loopback indication = 0

ATM - Mngt: Monitoring & Config



NMS

SNMP/UDP/IP/AAL

SNMP/AAL

Terminal
or router

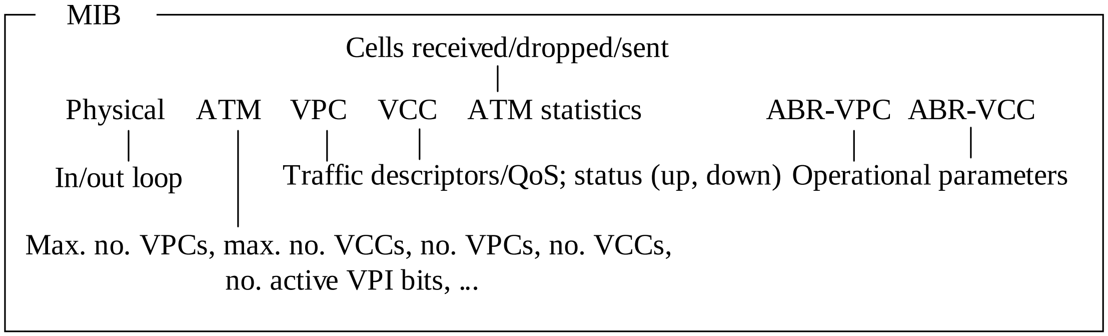
Private
ATM switch

Public
ATM switch

User-network interface management entity

ILMI

Status, configuration, control information
The ILMI attributes are organized as one MIB per IME.



MIB

Cells received/dropped/sent

Physical ATM VPC VCC ATM statistics ABR-VPC ABR-VCC

In/out loop

Traffic descriptors/QoS; status (up, down) Operational parameters

Max. no. VPCs, max. no. VCCs, no. VPCs, no. VCCs,
no. active VPI bits, ...

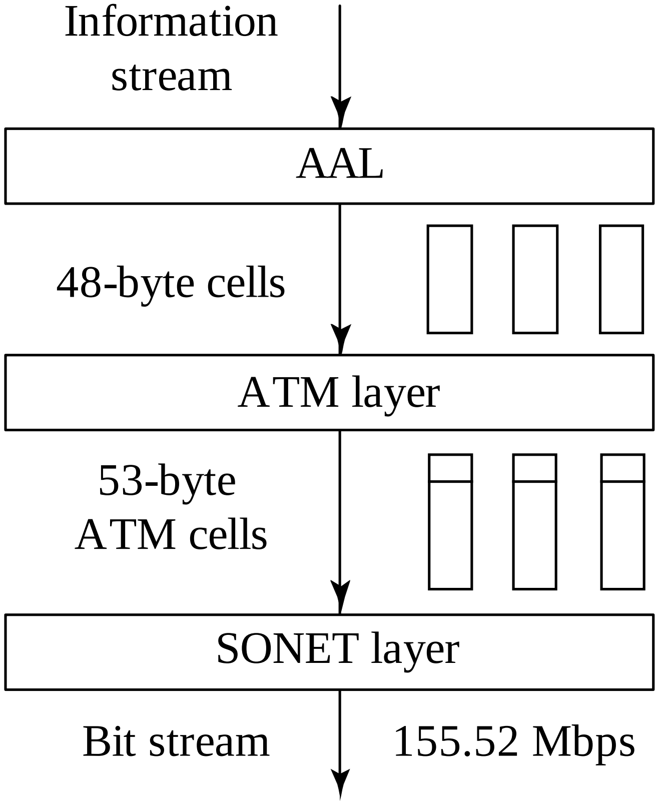
SNMP

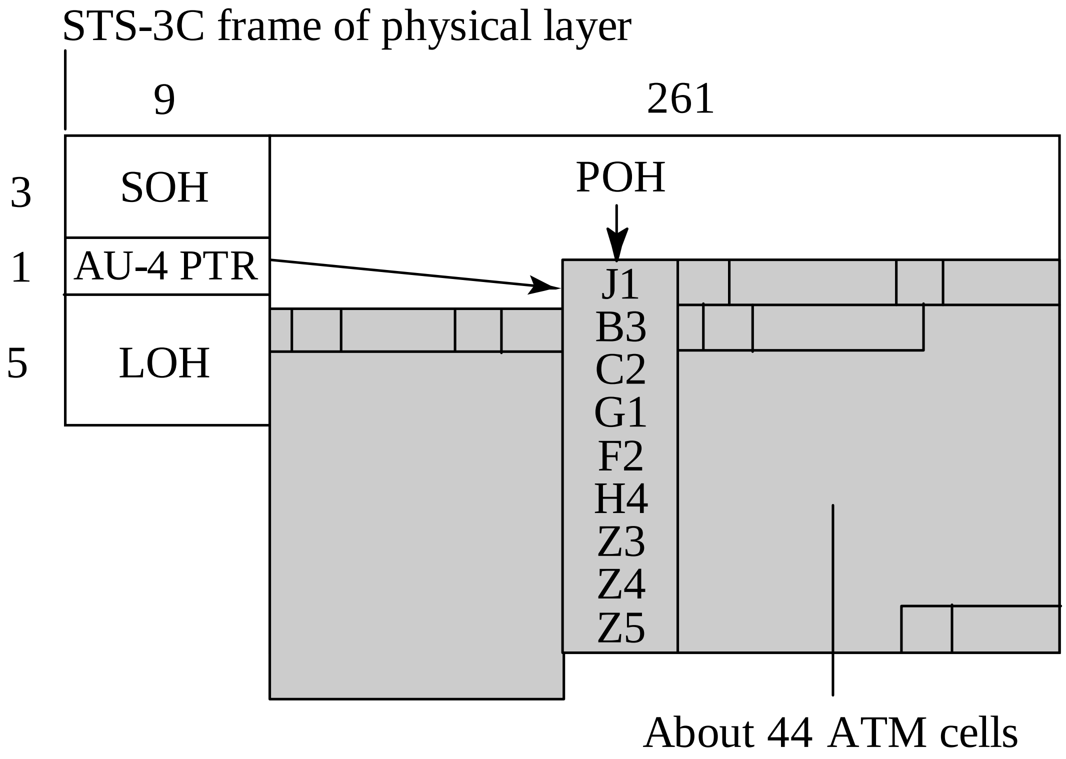
Get, Get-Next, Set, Trap

ATM - Mngt: U-N signaling

- User requrests an SVC
- Network accepts or rejects
- Network indicates error conditions with a connection

ATM: ATM/SONET

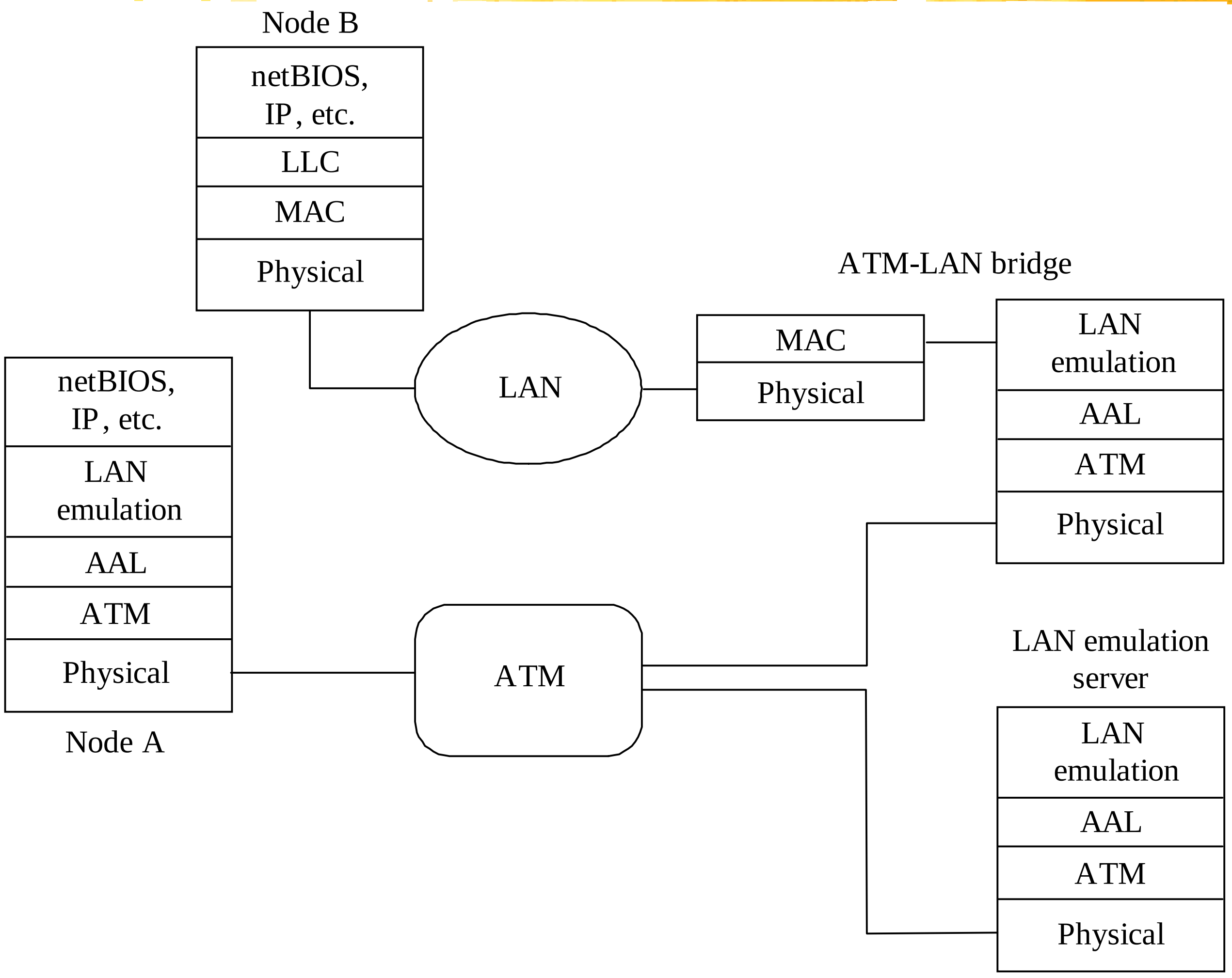




ATM: Internetworking

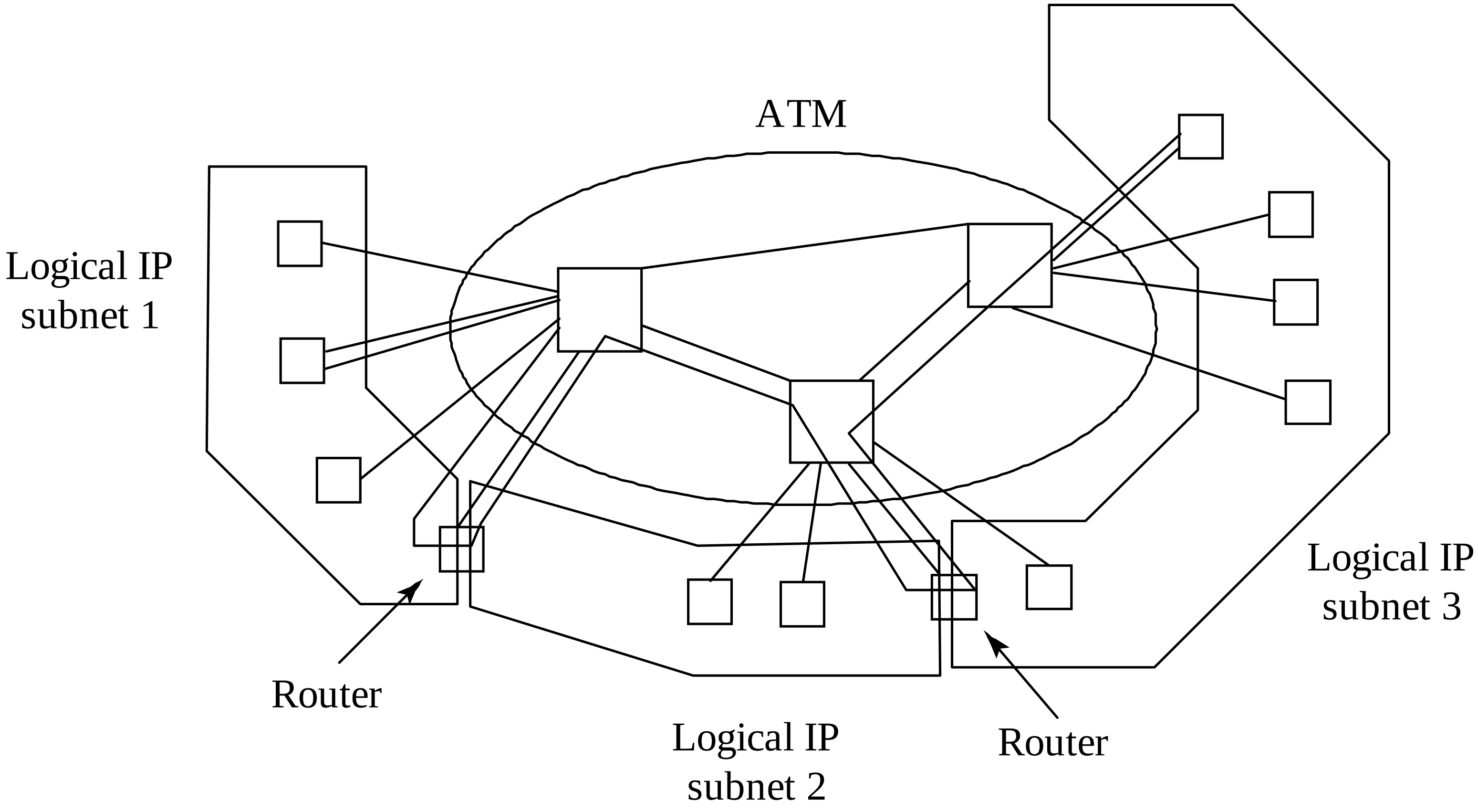
- LANE
- IP/ATM
- MPOA
- FR & SMDS/ATM

ATM - Internetworking: LANE



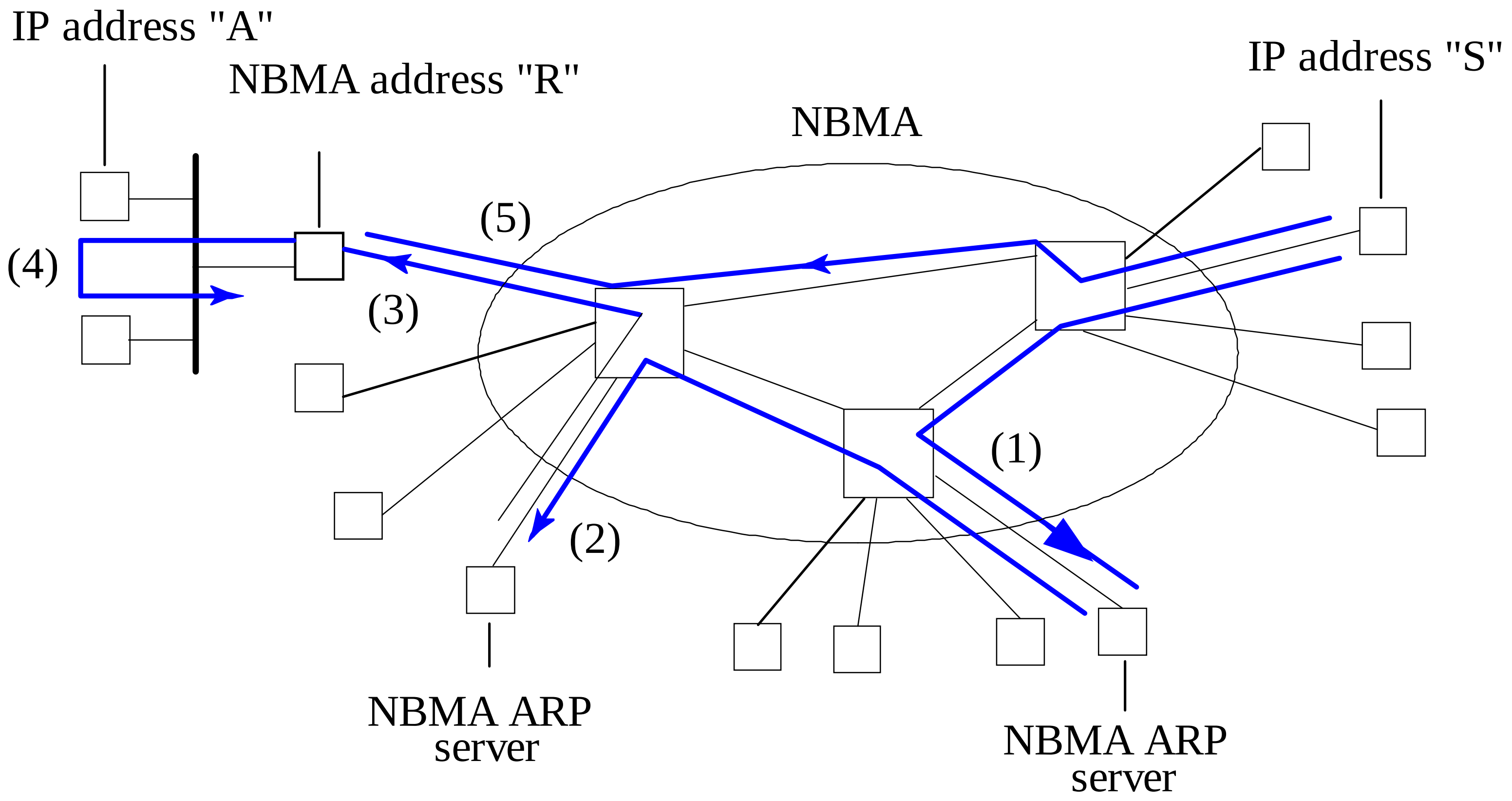
ATM - Internetworking: IP/ATM (1)

"Classic IP over ATM"



ATM - Internetworking: IP/ATM (2)

"Shortcut model: Next Hop Resolution Protocol NHRP"



ATM - Internetworking: IP/ATM (3)

"Multicast IP/ATM"

Multicast Address Resolution Server:

- Version 1: IP Multicast Address -> List or IP addresses

  In this version, the sender sets up one VC for each member of the M-group

  the join and leave requests go to the MARS

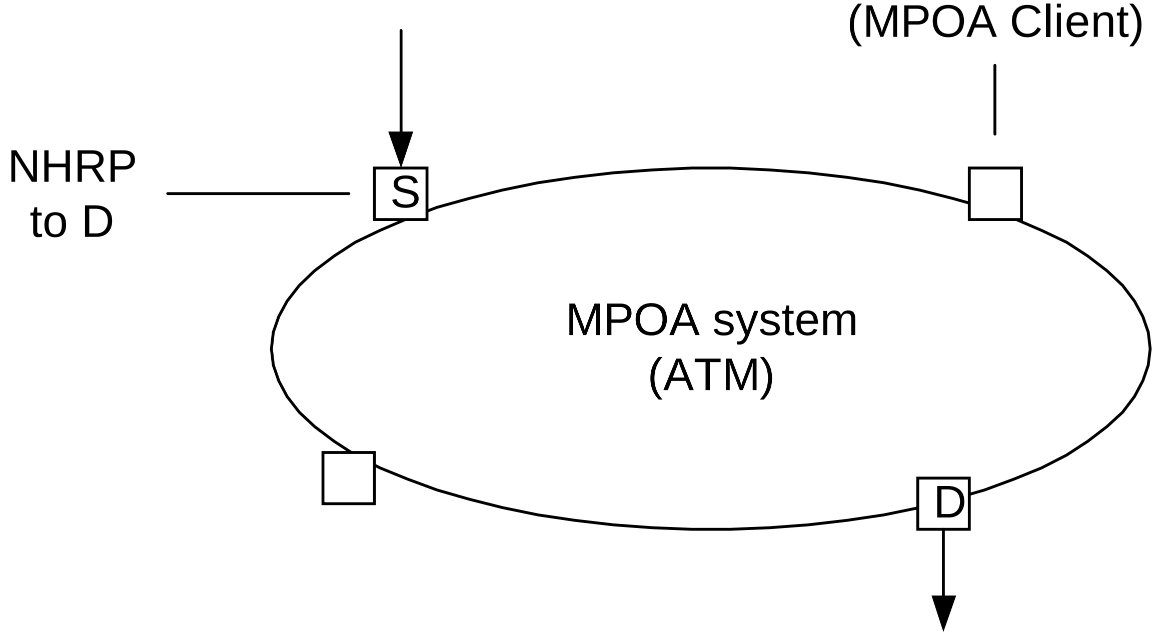
- Version 2: IP Multicast Address -> IP address of Multicast Server

  In this version, the Multicast server maintains the group and the VCs

  the join and leave requests go to the Multicast Server

ATM - Internetworking: MPOA

Integrates LANE and NHRP



ATM - Internetworking: FR & SMDS/ATM

- Frame Relay:
  - FR connection identifier -> ATM VPI/VCI
  - FR payload -> ATM cells
  - FR congestion/discard eligibility -> EFCI, CLP
  - FR CIR -> VBR parameters
- SMDS:
  - VPI/VCI
  - Packets -> AAL3/4

<

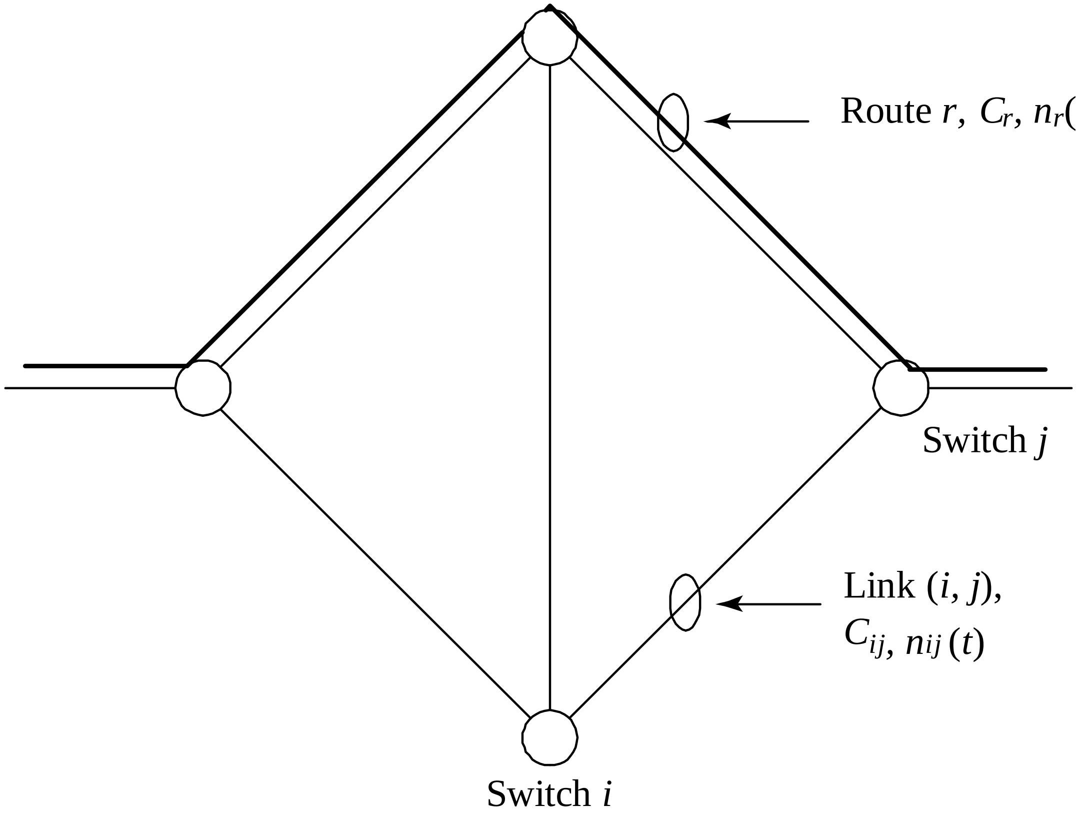
Circuit-Switched Networks

- Overview
- Performance
- SONET
- DWDM
- Fiber to the Home
- DSL
- Intelligent Networks
- CATV

CS: Overview

- Designed to carry constant bit rate streams
- Provides bitways for data networks
- Examples: Telephone, CATV
- Physical: Wireless, Satellite, Copper, Fiber
- Multiplexing:
  - TDM (isochronous or synchronous)
  - Fiber: WDM, SCM
  - Wireless: TDM/FM; CDMA; Cellular, ...
- Switching: Circuit switching

CS: Performance



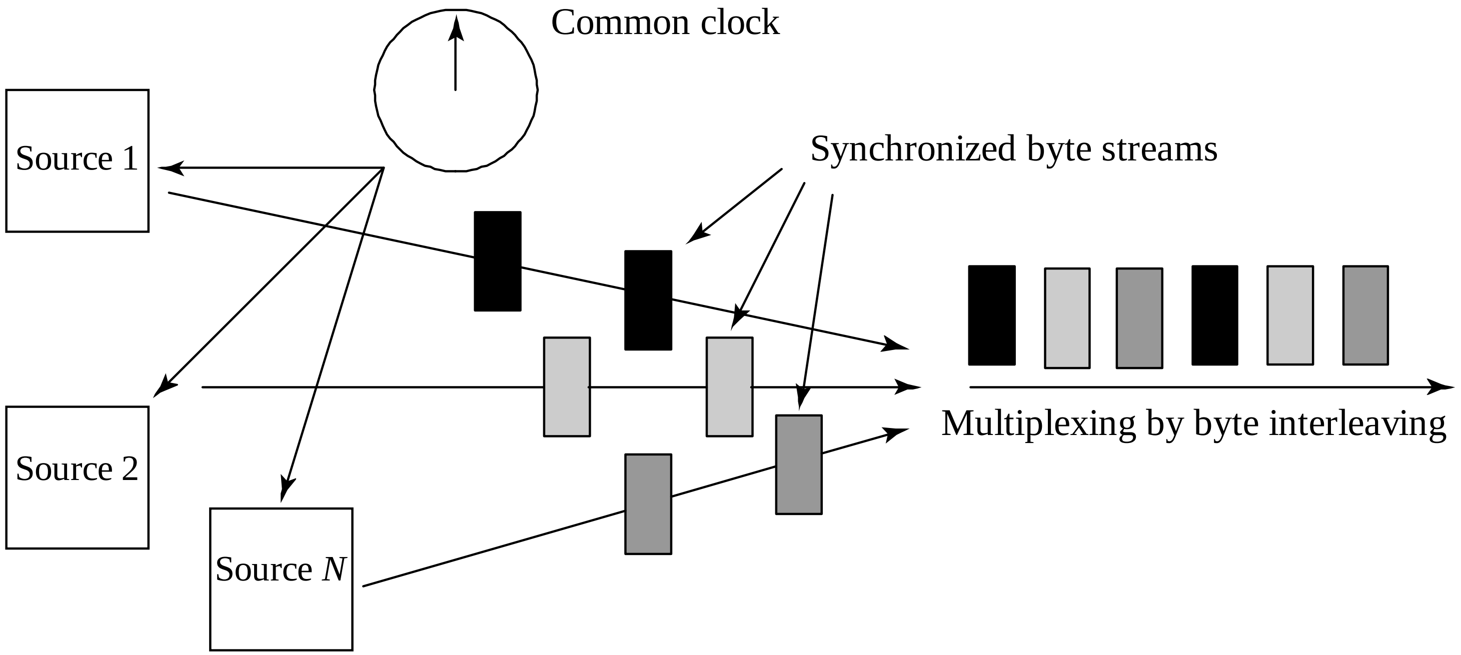
[5.1]

For a given load and routing -> Blocking probability

CS: SONET

- Operating Principles
- Layers
- Frame Structure
- Future of SONET

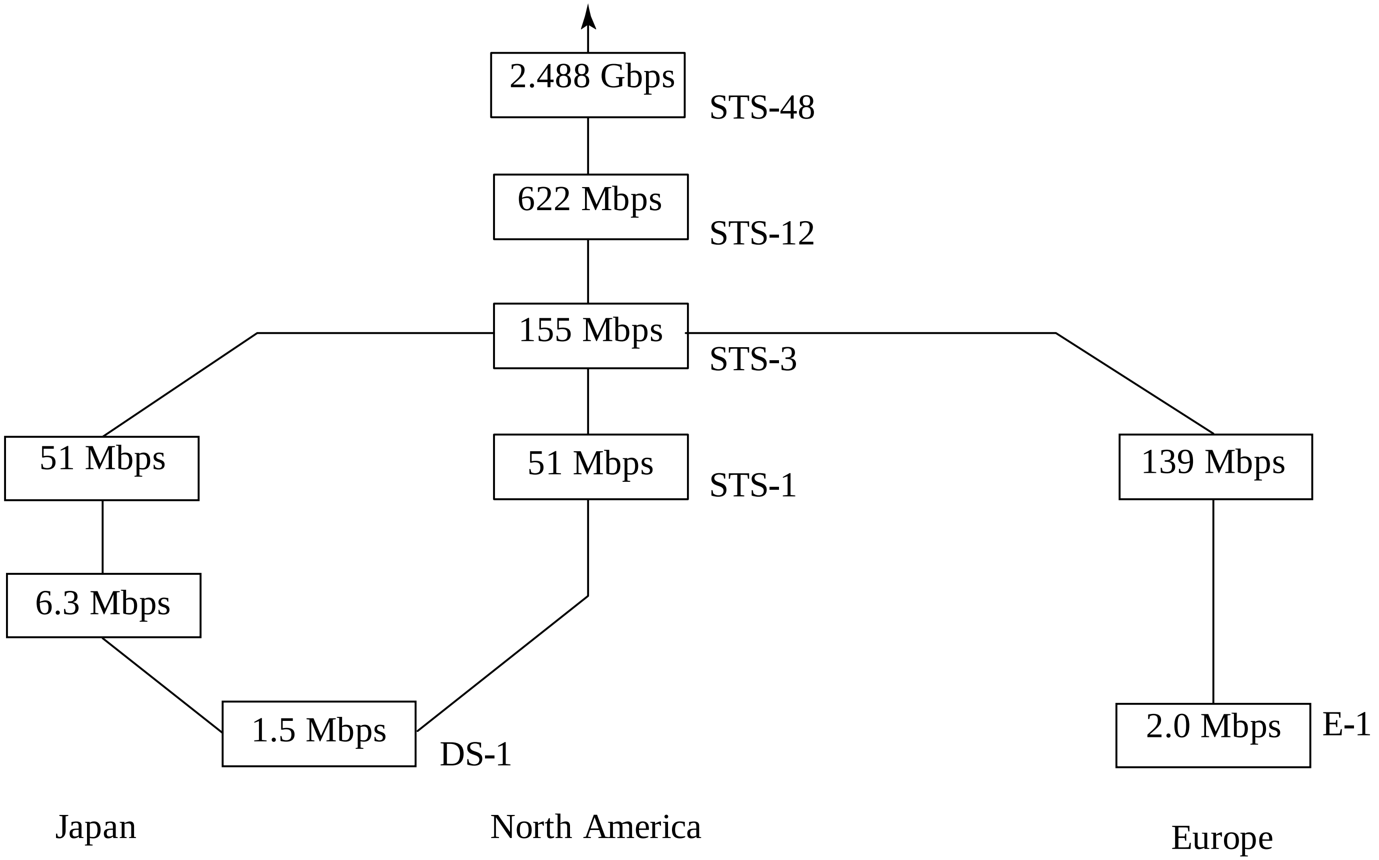
CS - SONET: Principles



[5.2]

TDM with synchronization of the sources

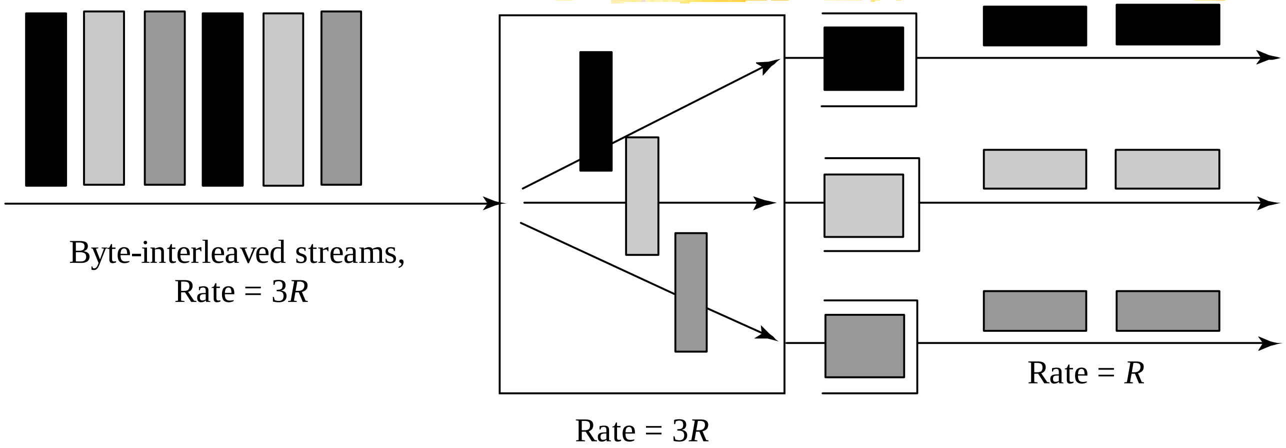
CS - SONET: Principles (cont.)



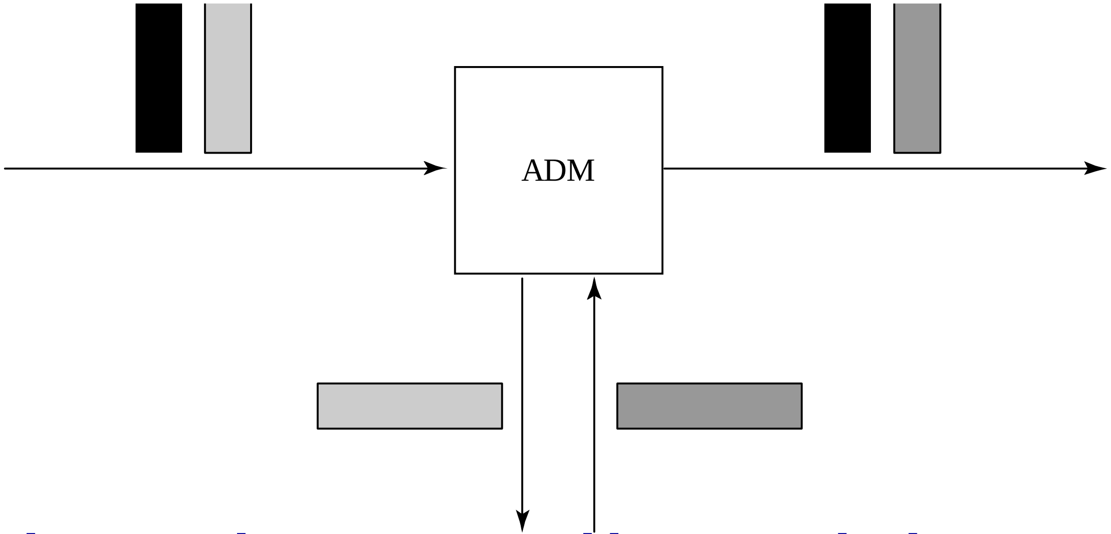
SONET/SDH Signal Hierarchy

[5.3]

CS - SONET: Principles (cont.)



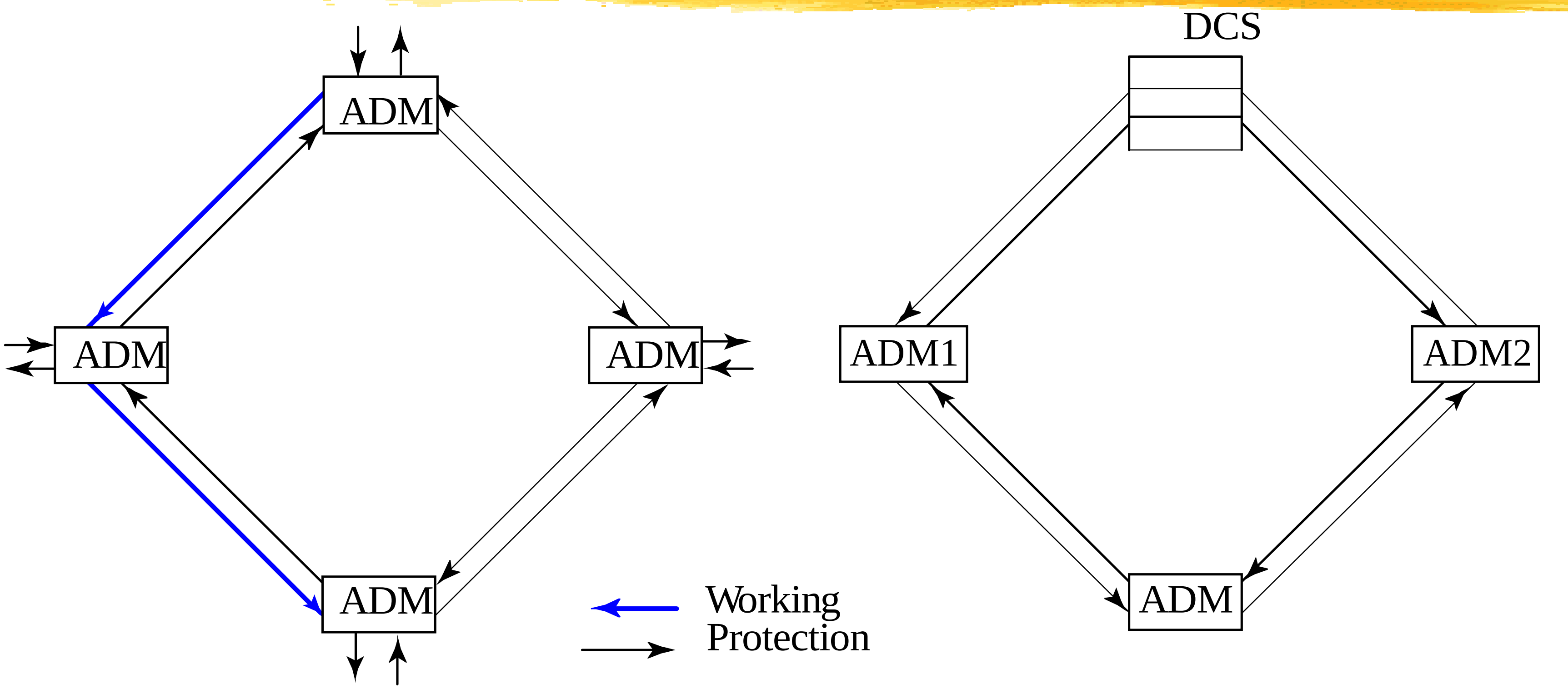
[5.4]



[5.5]

Simple demultiplexing and inexpensive Add-Drop Multiplexers

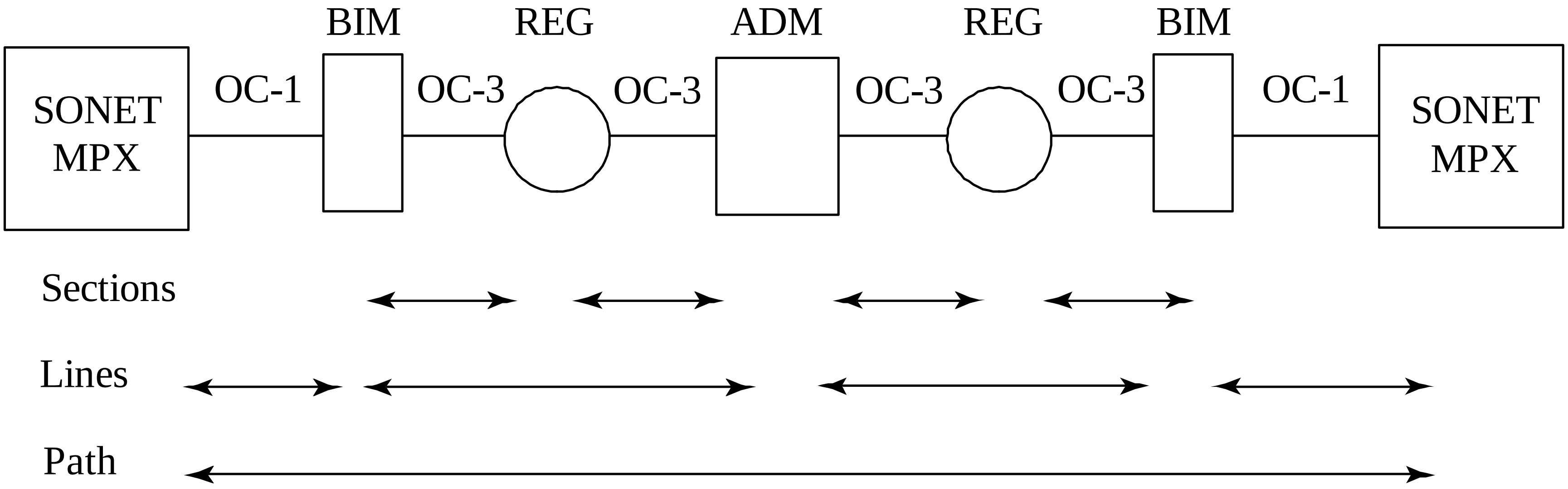
CS - SONET: Principles (cont.)



[5.6]

SONET rings for reliability

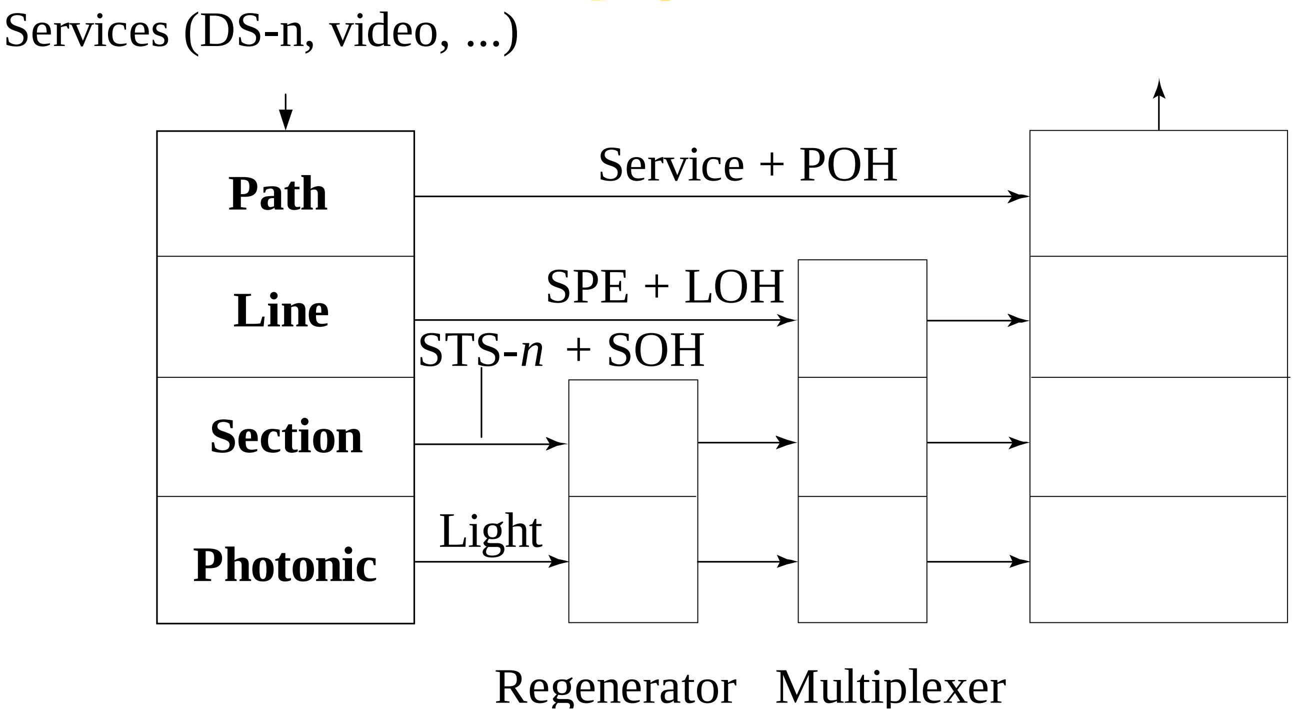
CS - SONET: Layers



BIM = Byte-interleaved multiplexer REG = Regenerator

[5.7]

CS - SONET: Layers (cont.)



[5.8]

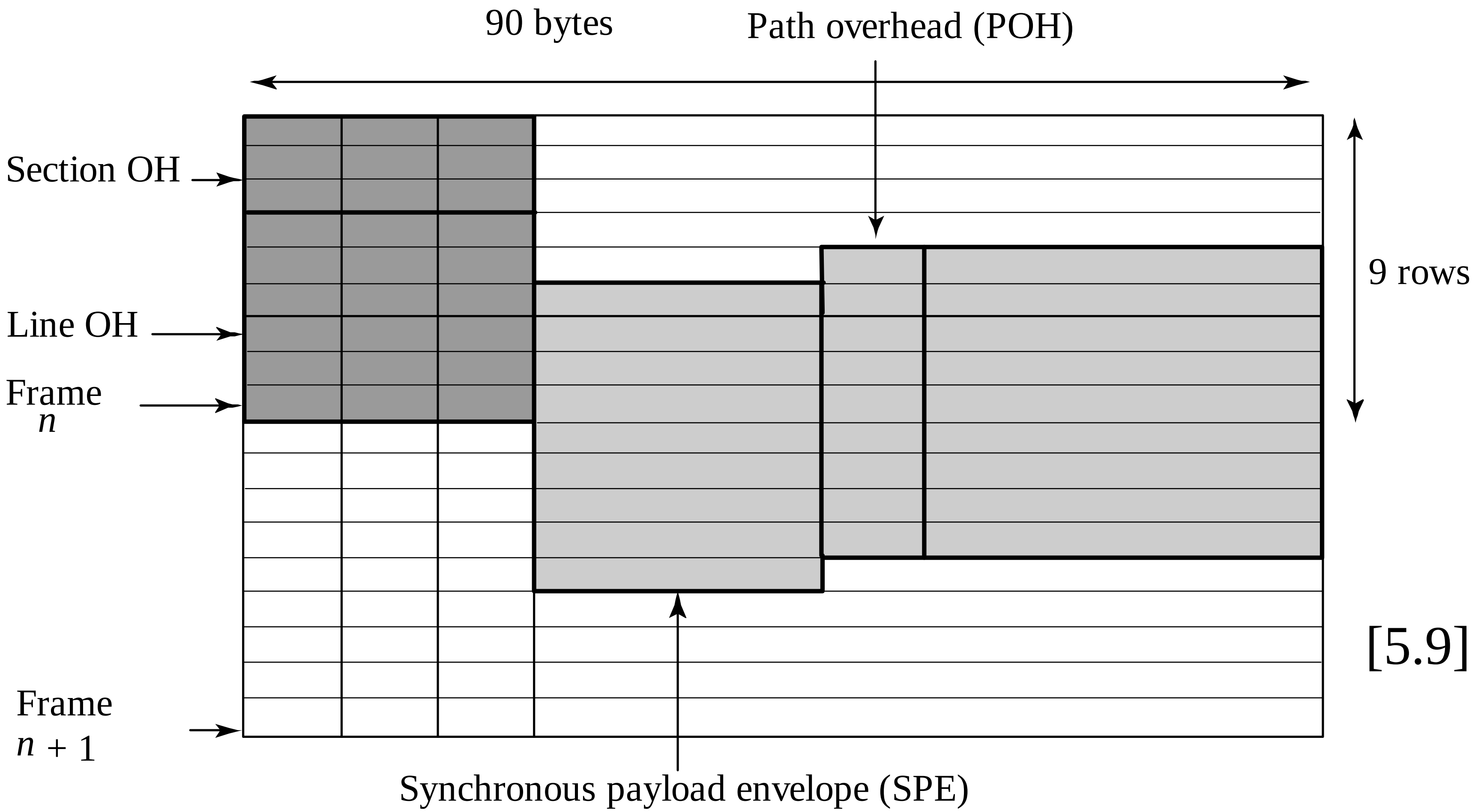
Path: Services; end-to-end error detection

Line: MPX (frame and frequency alignment); protection switching; data links

Section: Framing, scrambling, data links

Photonics: E/O and O/E conversions

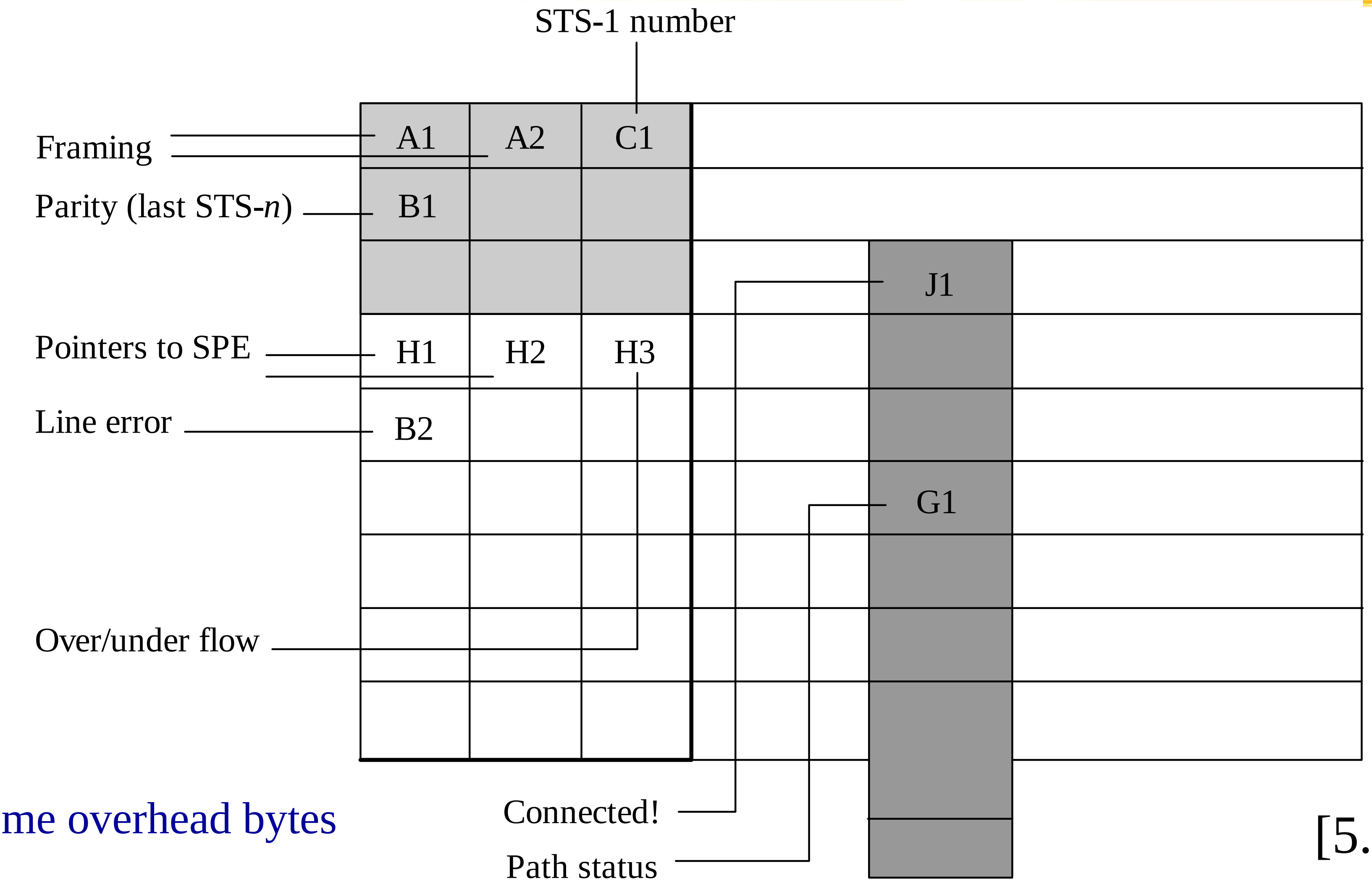
CS - SONET: Frames



[5.9]

SONET Frame

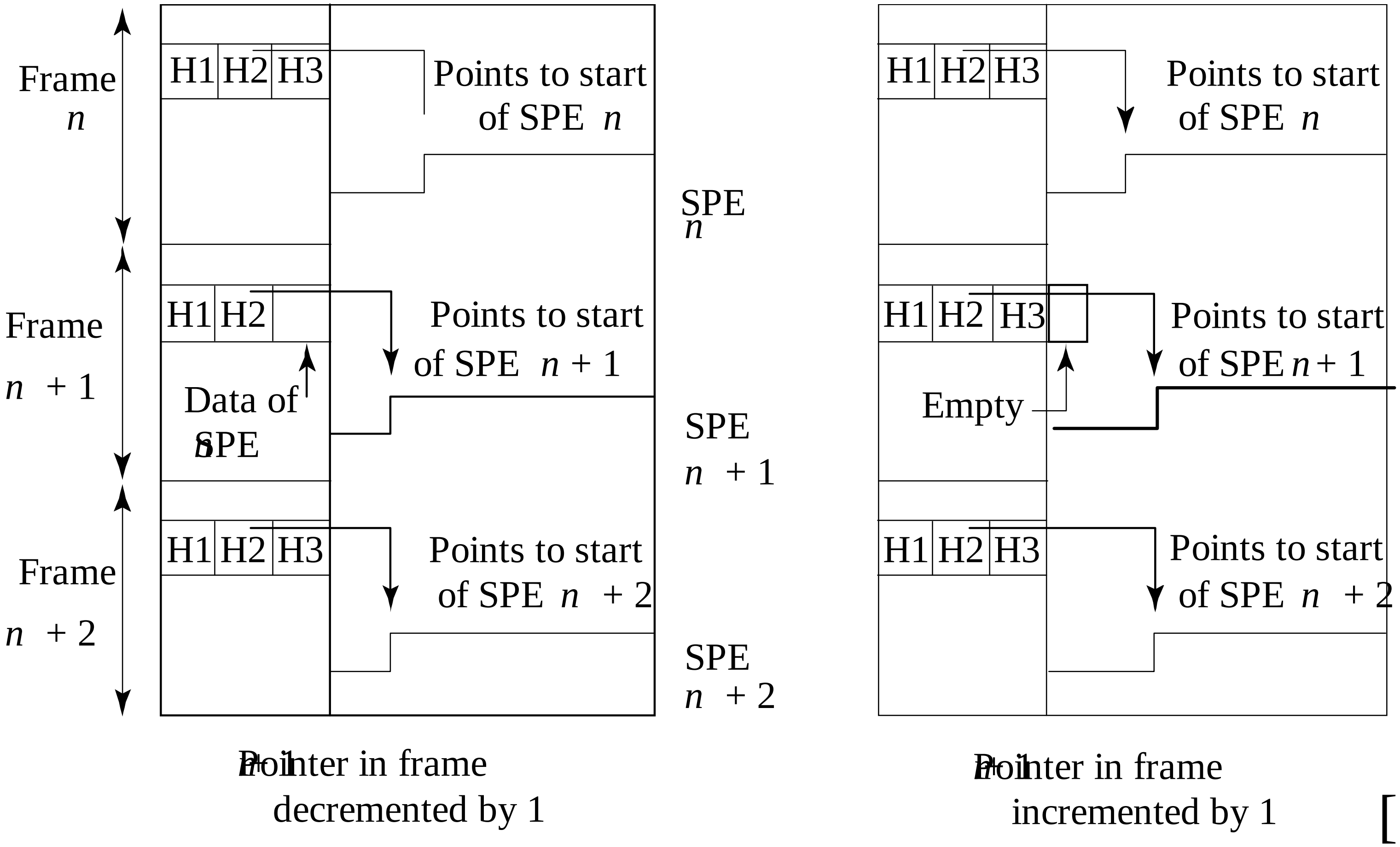
CS - SONET: Frames (cont.)



Some overhead bytes

[5.10]

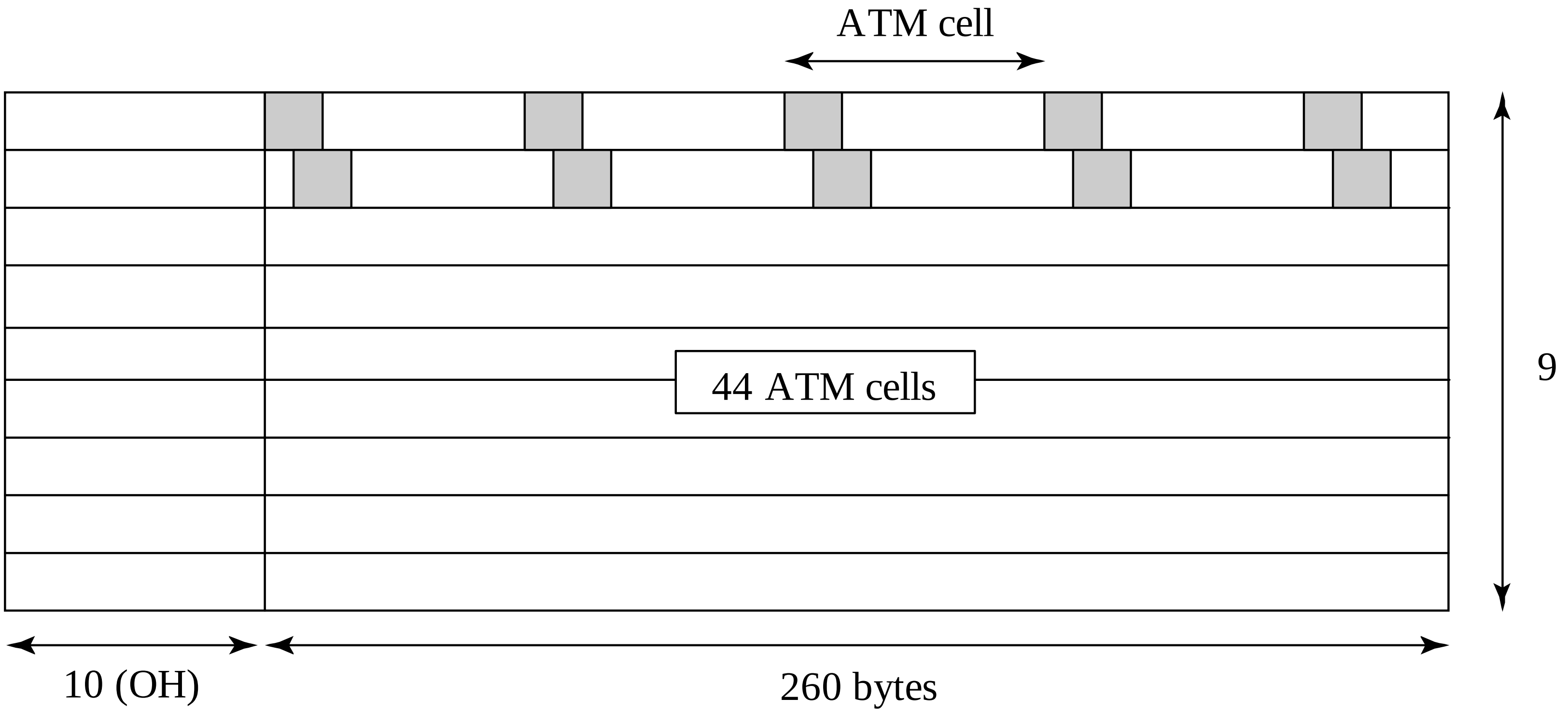
CS - SONET: Frames (cont.)



[5.11]

Frequency justification: add a byte (left) or stuff a byte (right)

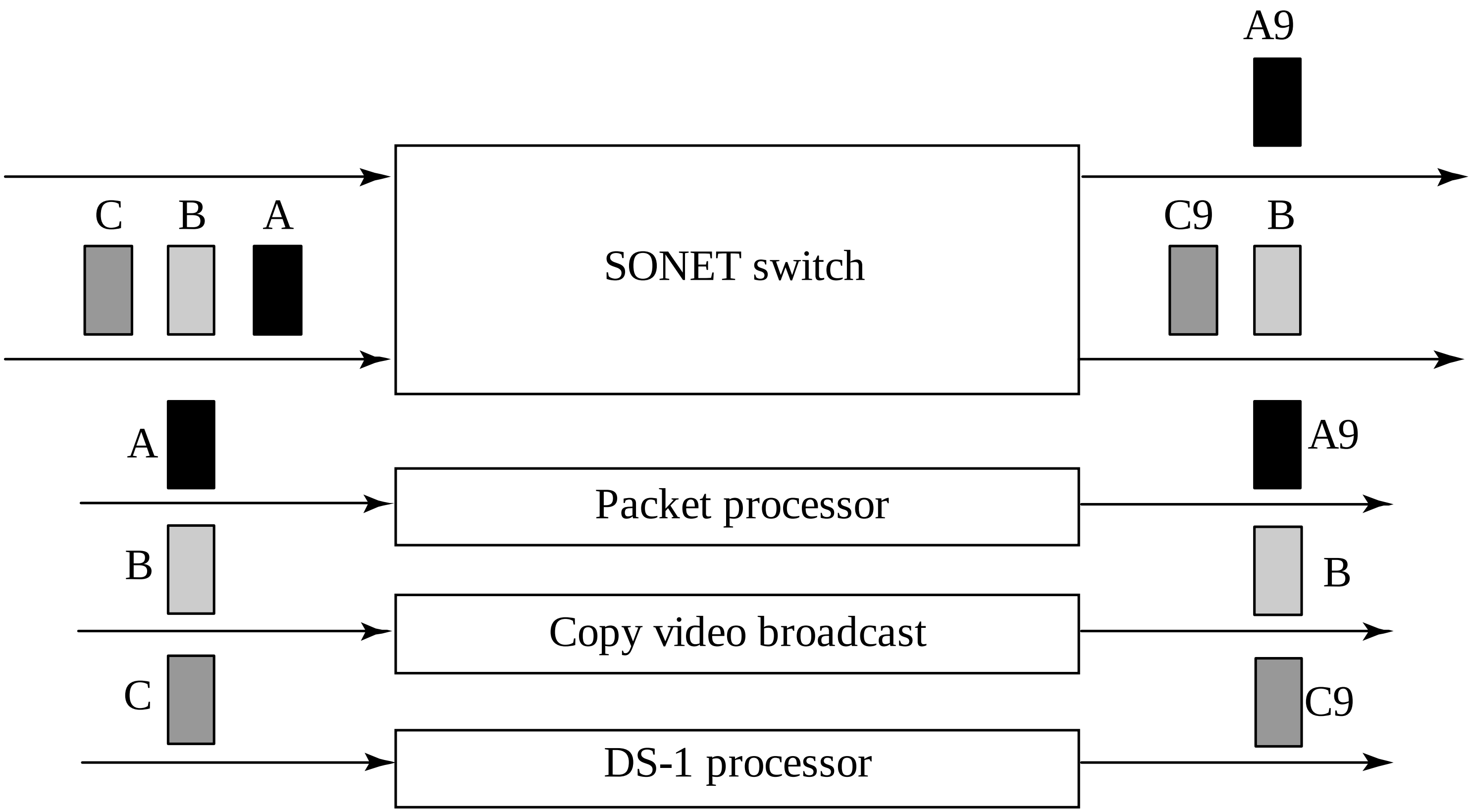
CS - SONET: Frames (cont.)



[5.12]

STS-3 frame accommodates 44 ATM cells. (Note: no framing bits.)

CS - SONET: Frames (cont.)



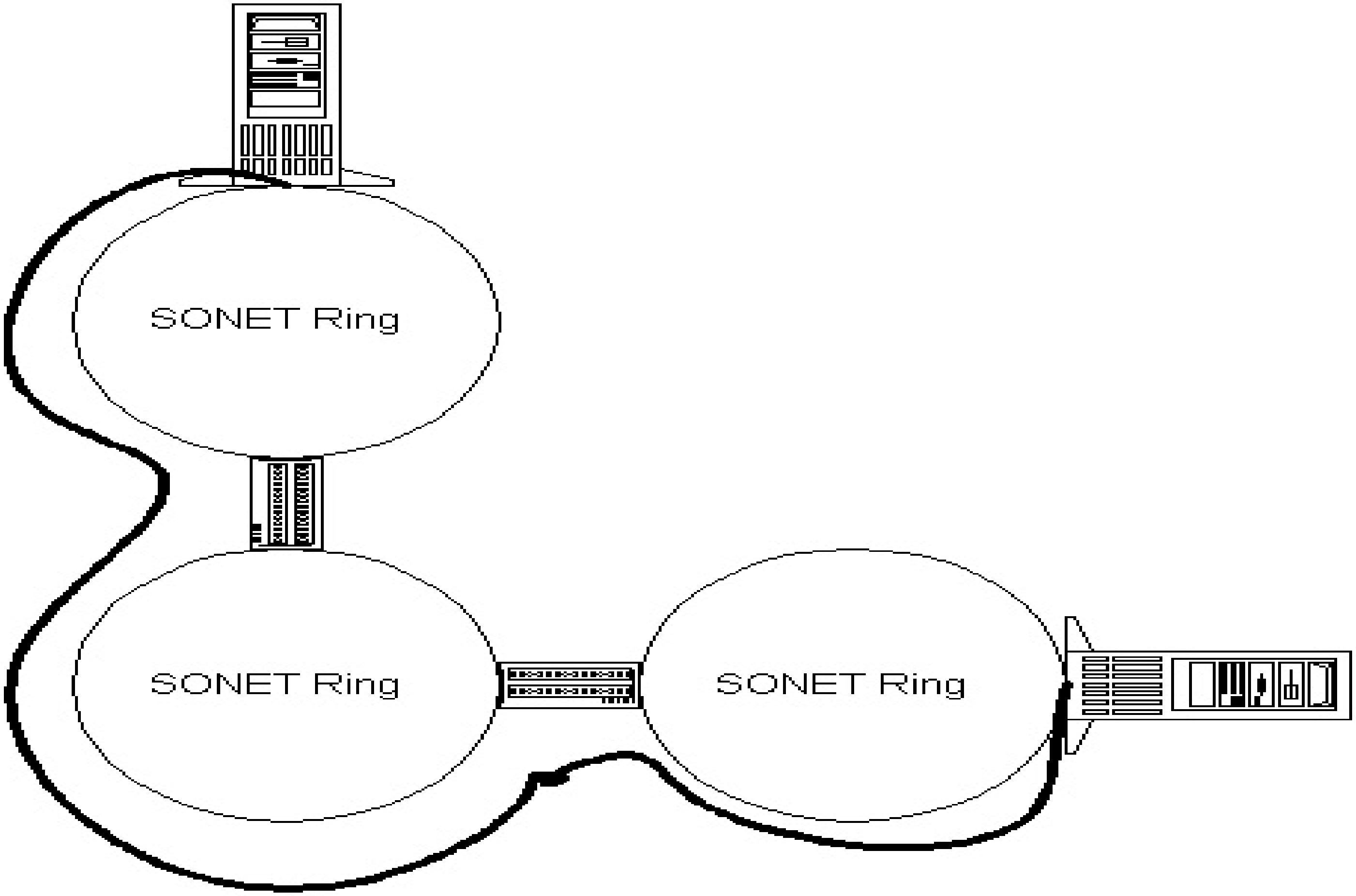
[5.13]

SONET switch: Demux streams, process, remux.

CS - SONET: Future

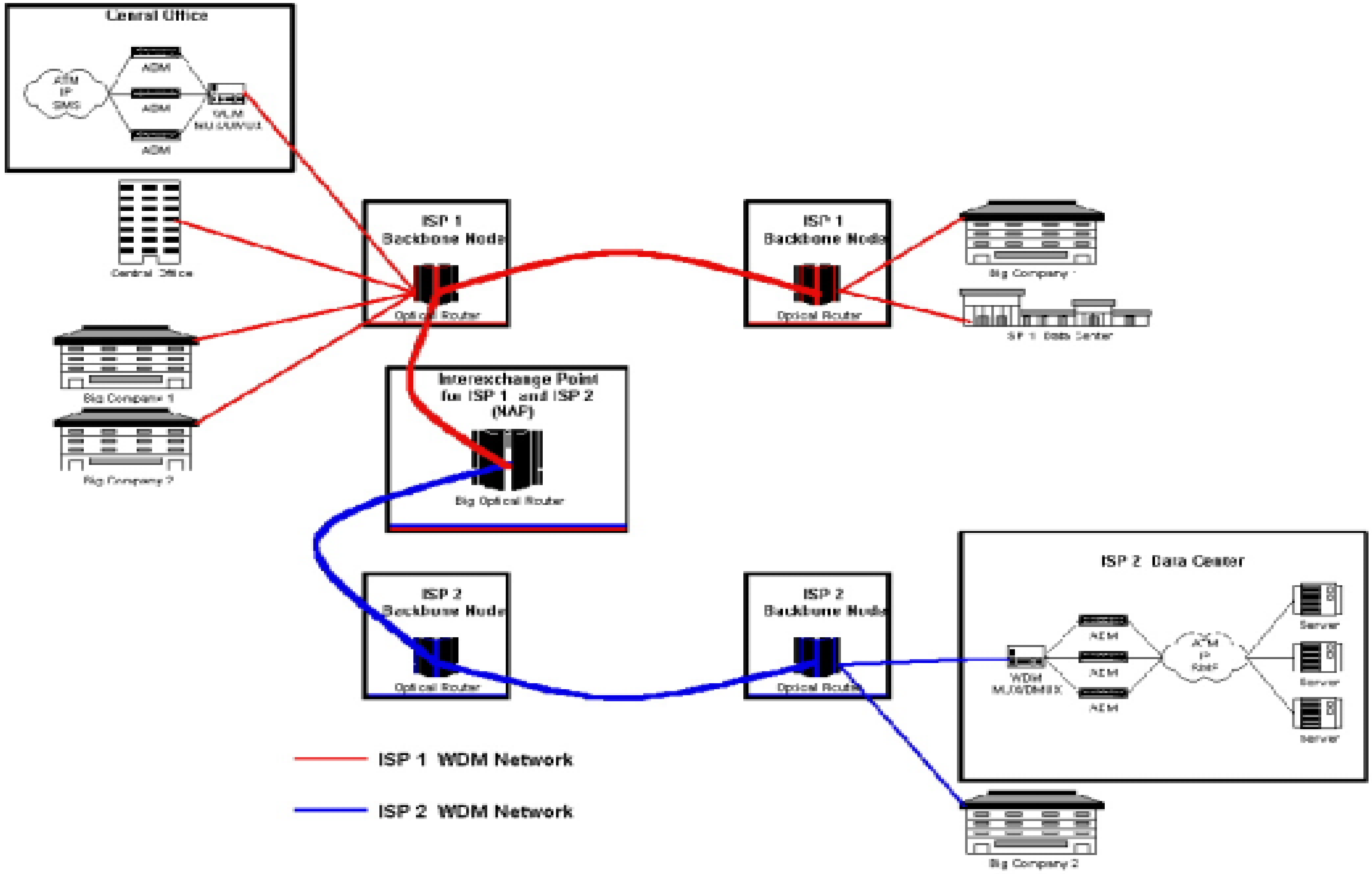
- SONET efficient for CBR
- SONET not necessary for data
- Lightweight SONET:
  - Eliminate synchronization
  - Keep link technology
  - Error protection: More efficient at higher layer
  - -> Packet over Sonet
- Gigabit Ethernet and 10G-E are competition

CS - SONET: Future (cont.)
From SONET to lightpaths



CS - SONET: Future (cont.)

Lightpath from C/O to backbone



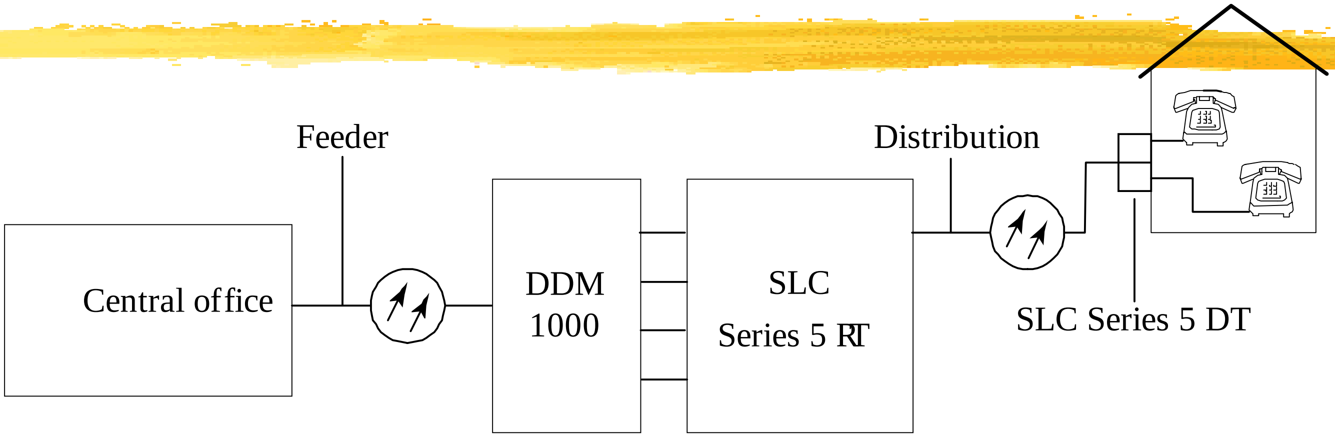
CS: DWDM

- Objective: Exploit the 25×10^{12} Hz bandwidth of fiber (A laser cannot be modulated that fast.)
- Method: Use lasers with narrow disjoint spectra
  - DFB lasers
  - Wideband optical amplifiers
  - Wavelength-selective switches
- 1996: 16 channels at 2.5Gbps
- 1998: 40 channels at 2.5Gbps
- 1999: 160 channels at 10Gbps

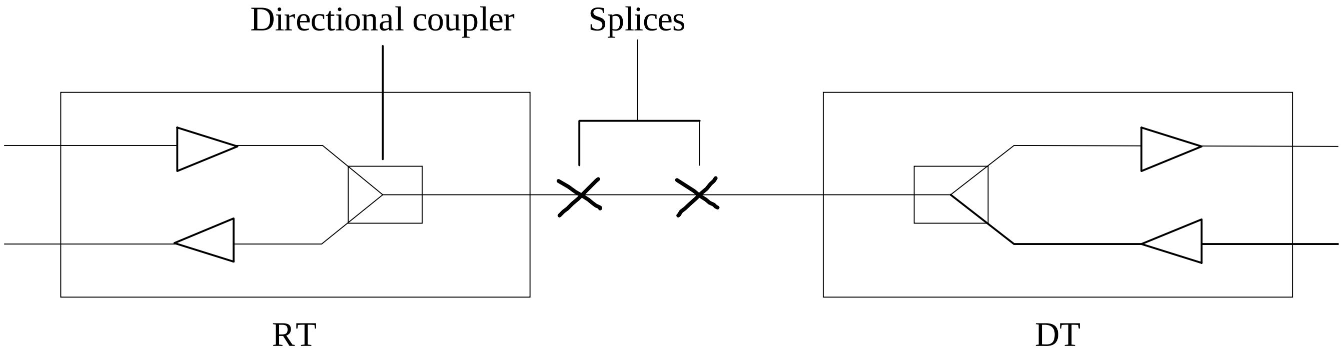
CS: Fiber to the Home

- AT&T’s Subscriber Loop System
- Passive Optical Networks
- Passive Photonic Loop
- Hybrid Scheme

CS - FTH: AT&T's SLC



[5.14]



[5.15]

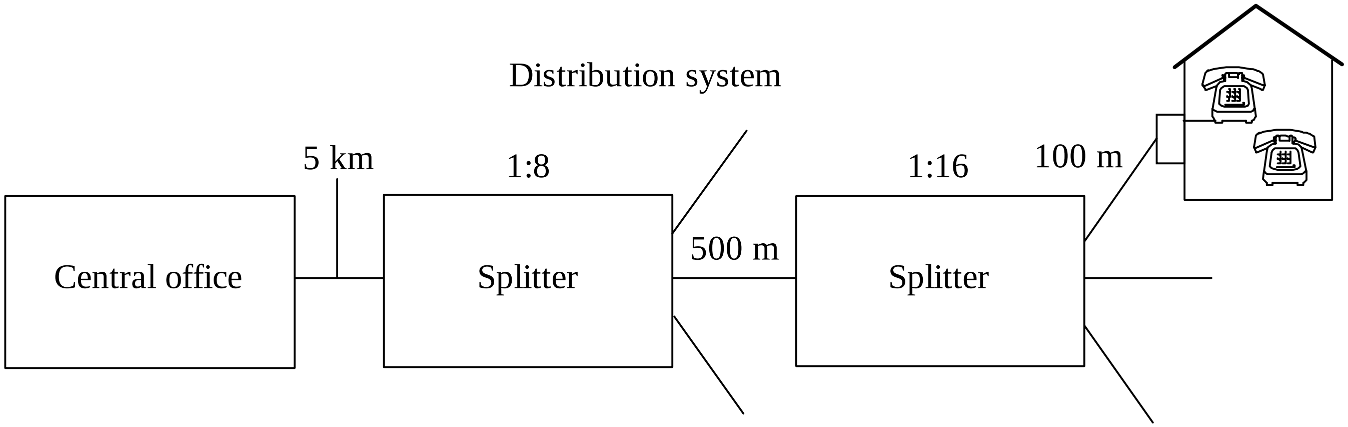
RT = remote terminal, DT = distant terminal, DDM = digital multiplexer

Single Mode Fiber, 3-dB directional couplers

InGaAs laser diode, 1.3µm, P_T = - 20dBm

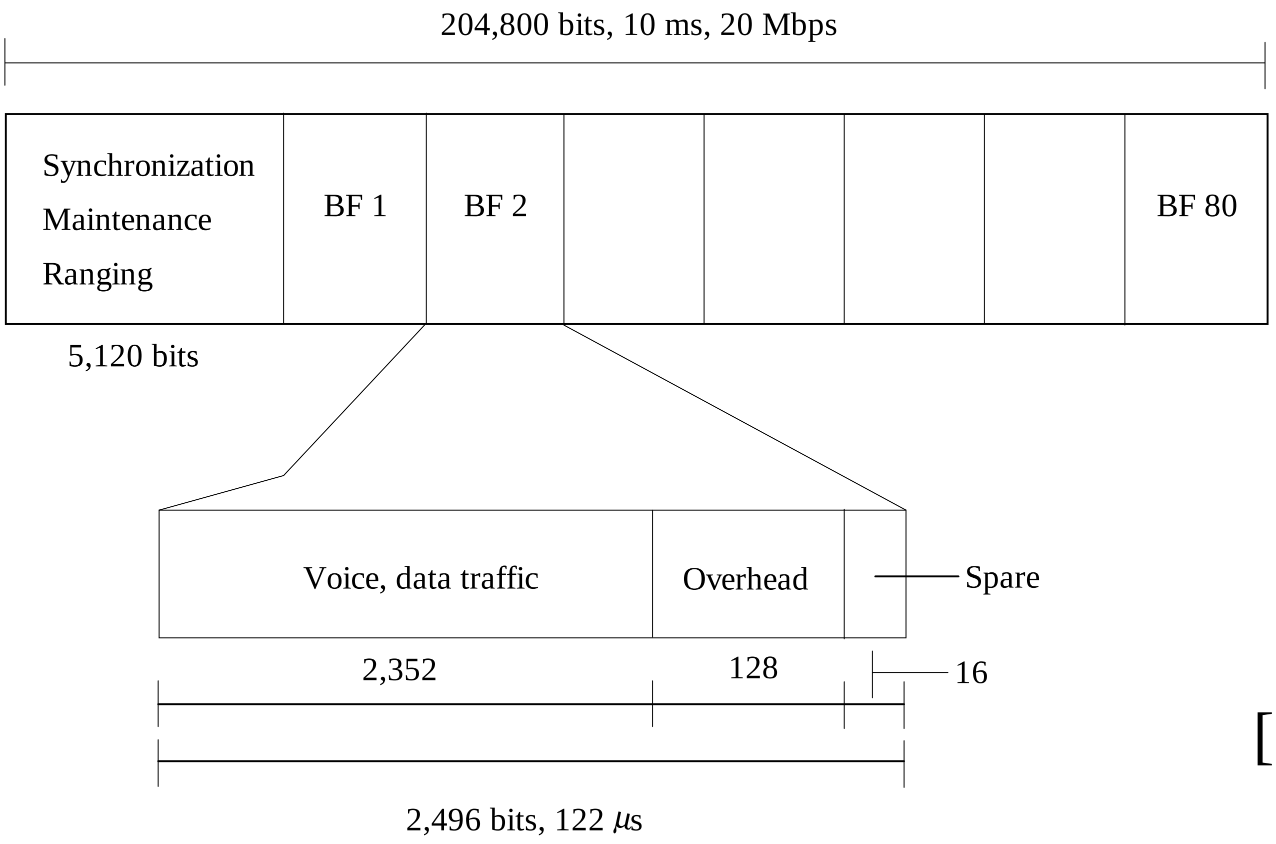
InGaAs PIN diode, - 46 dBm sensitivity at 1.5Mbps

CS - FTH: PON



[5.16]

CS - FTH: PON (cont.)



[5.17]

Capacity is 2,352× 8-Kbps channels divided into packages of

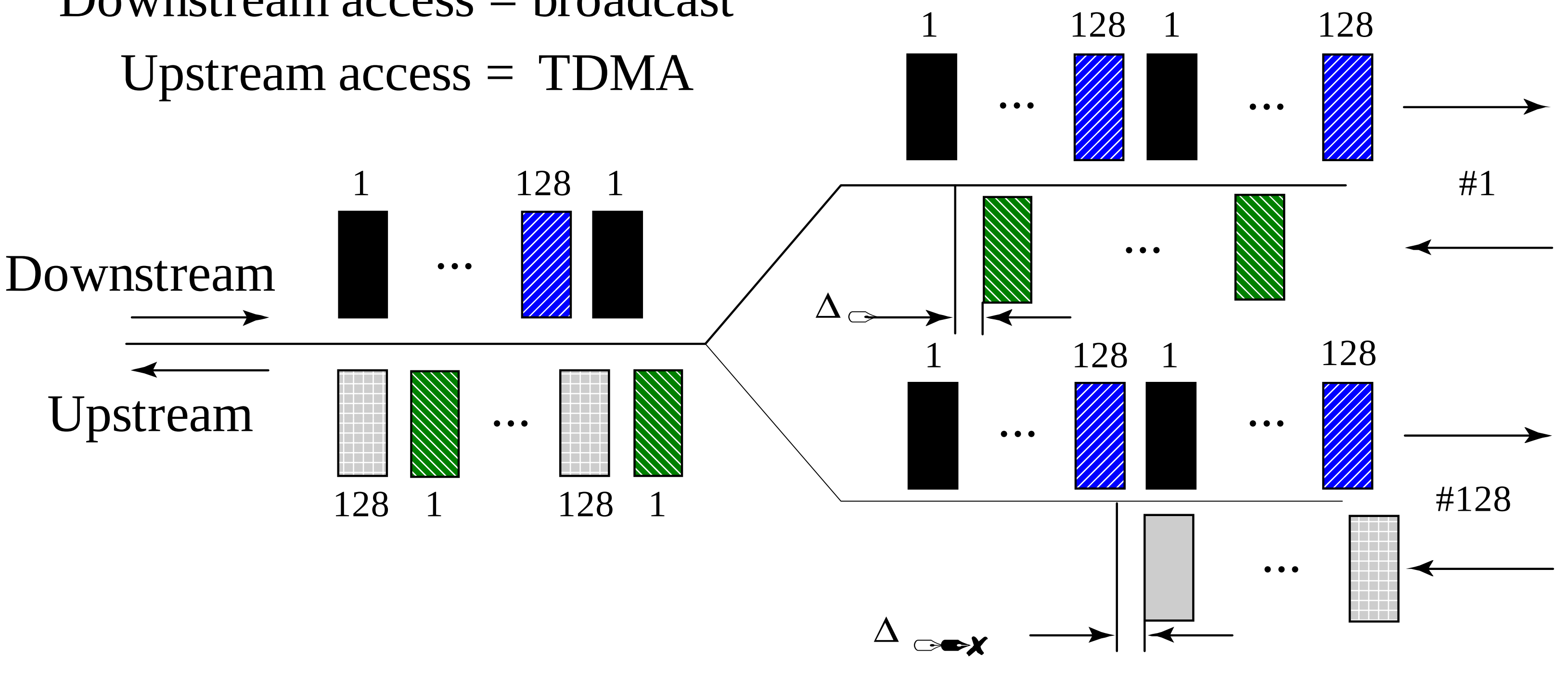
18 ×8 = 144-Kbps ISDN channels

8 ×8 = 64-Kbps voice channels

CS - FTH: PON (cont.)

Downstream access = broadcast

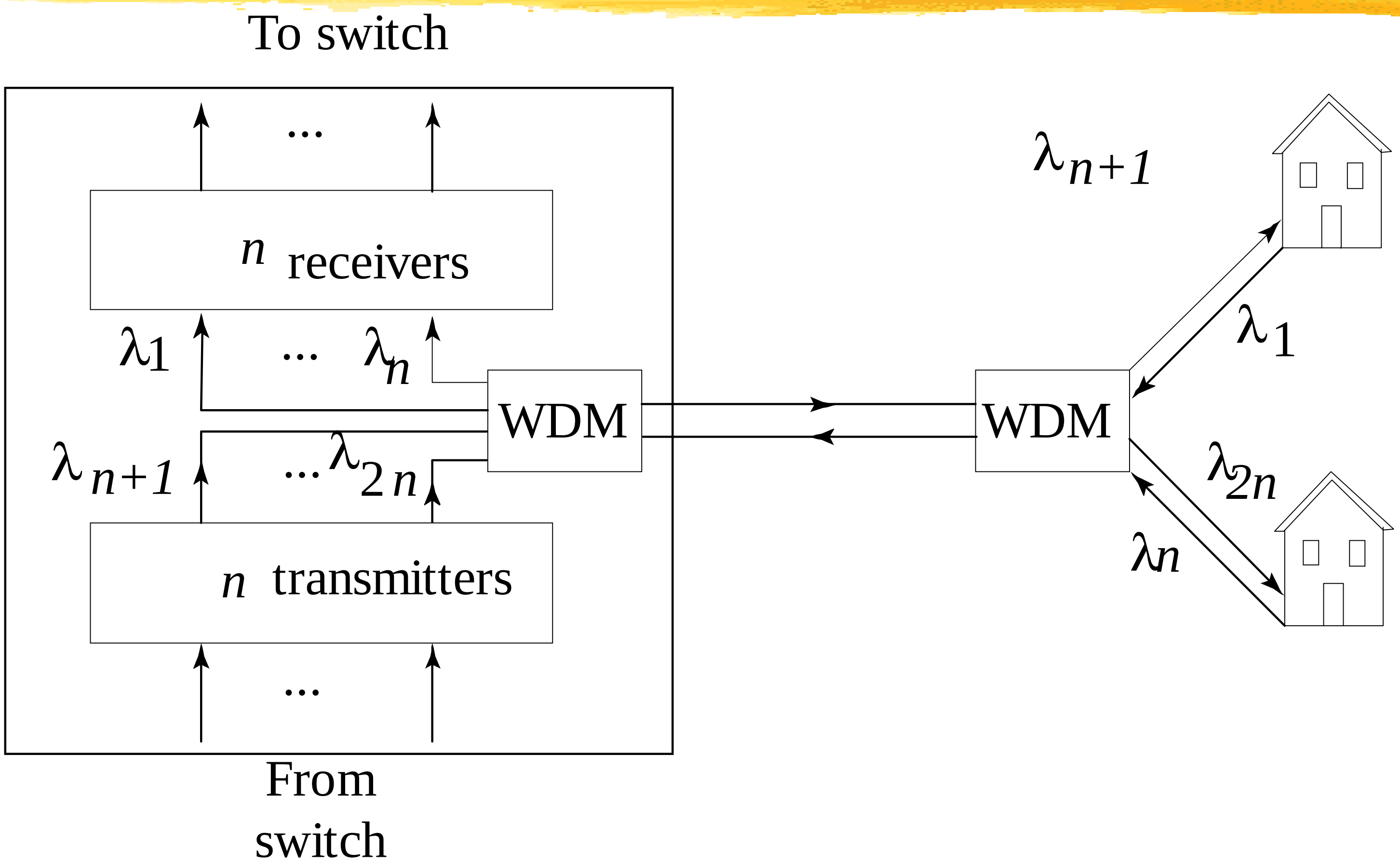
Upstream access = TDMA



[5.18]

Δs determined by ranging measurements

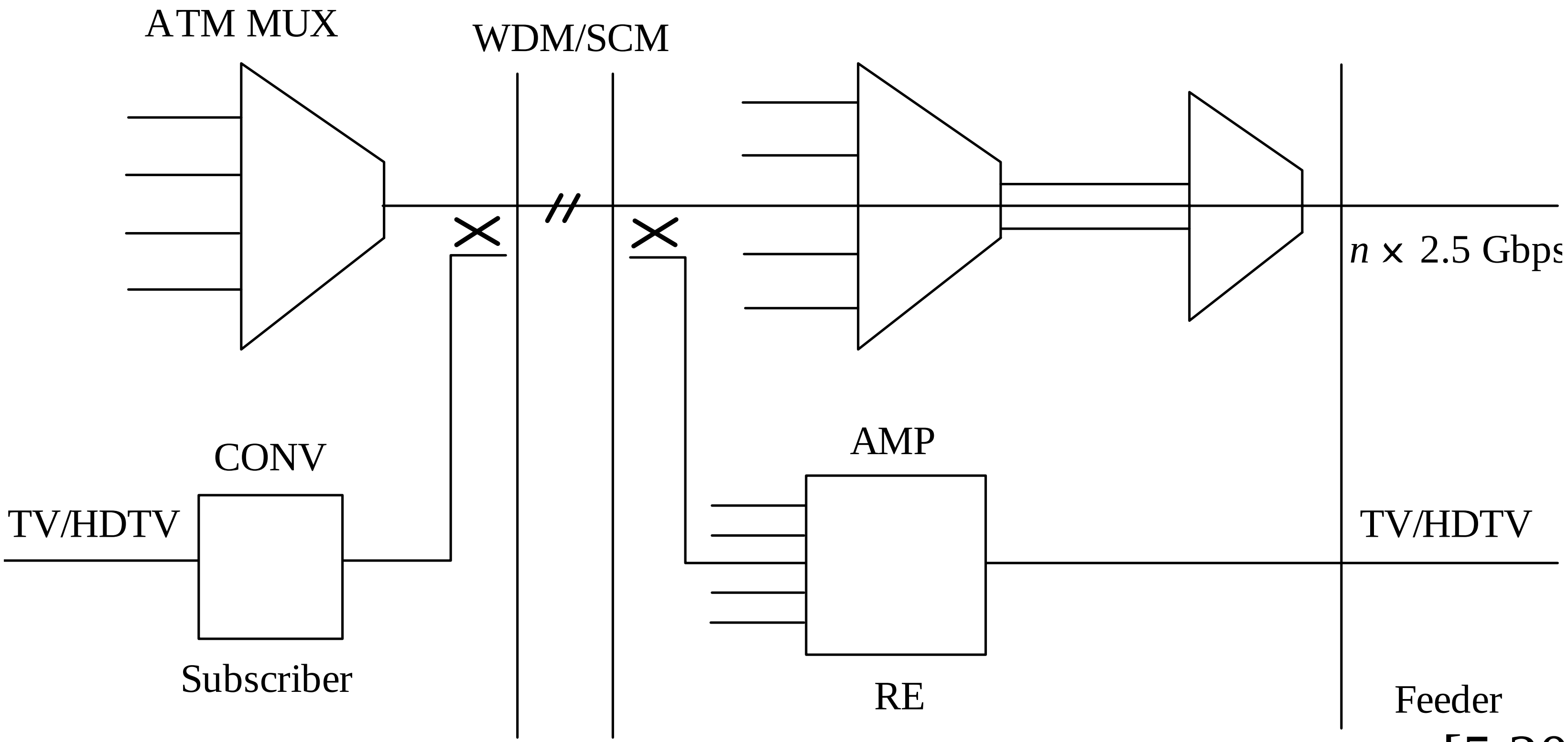
CS - FTH: PPL



[5.19]

One pair of wavelengths per subscriber.

CS - FTH: Hybrid



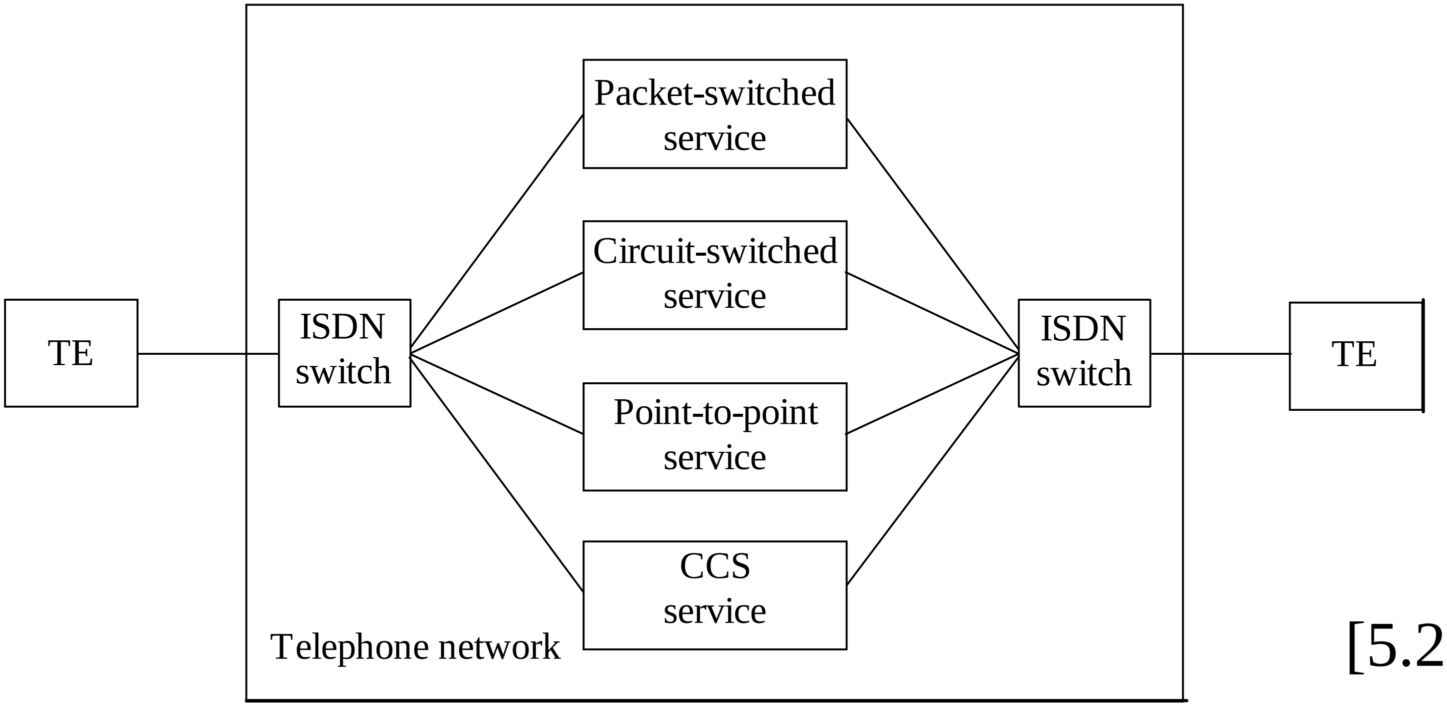
[5.20]

SCM divides one wavelength into channels (analog, digital).
Different wavelengths combined with WDM.

CS: Digital Subscriber Loop

- ISDN
- ADSL
- Modems

CS - DSL: ISDN



[5.21]

Architecture of ISDN

CS - DSL: ISDN (cont.)

Channel Types:

B: 64Kbps, CS, X.25
D: 16Kbps or 64Kbps
H: 384Kbps, 1,536Kbps, or 1,920Kbps (as B)

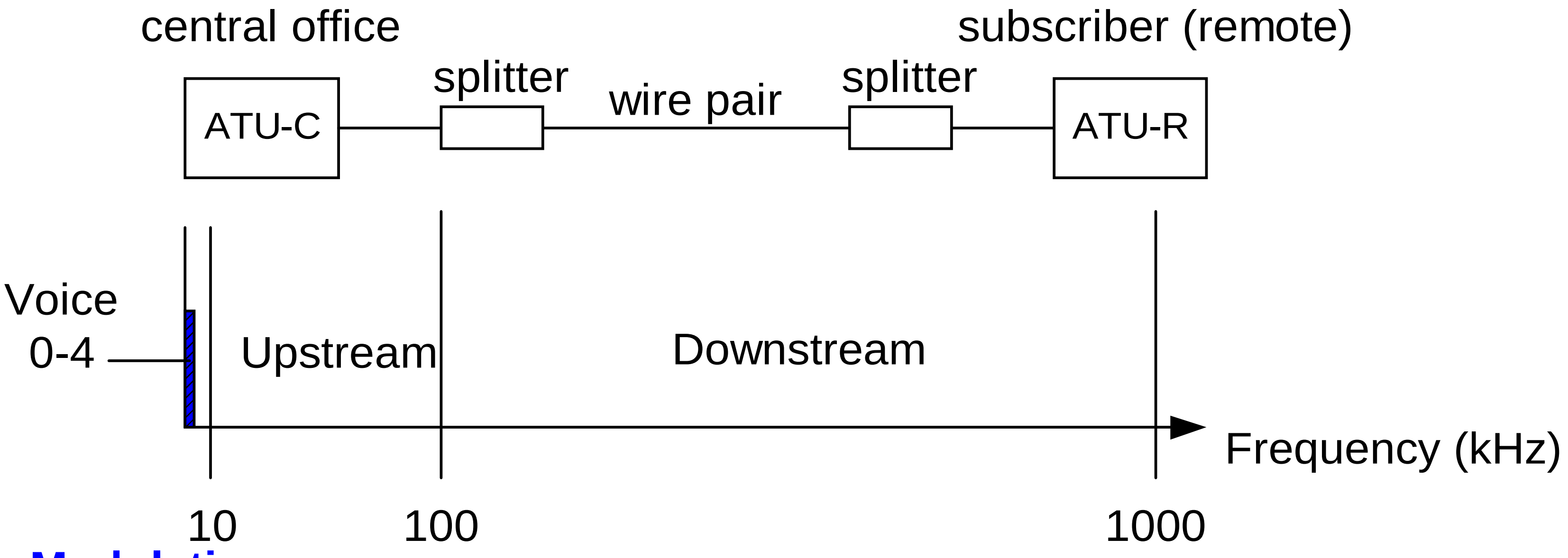
Basic Access: 2B + D

Primary Access: 30B + D(64): Europe
23B + D(64): US, Japan, Canada

Basic Interface:

Pseudo-ternary (1 = 0V, 0 = +/- 0.75V alt.)
Frame: 144Kbps (192Kbps with sync. & DC-balancing)
Link: B(PS): GBN, ACK+NACK; B(CS): user choice

CS - DSL: ADSL



Modulation:

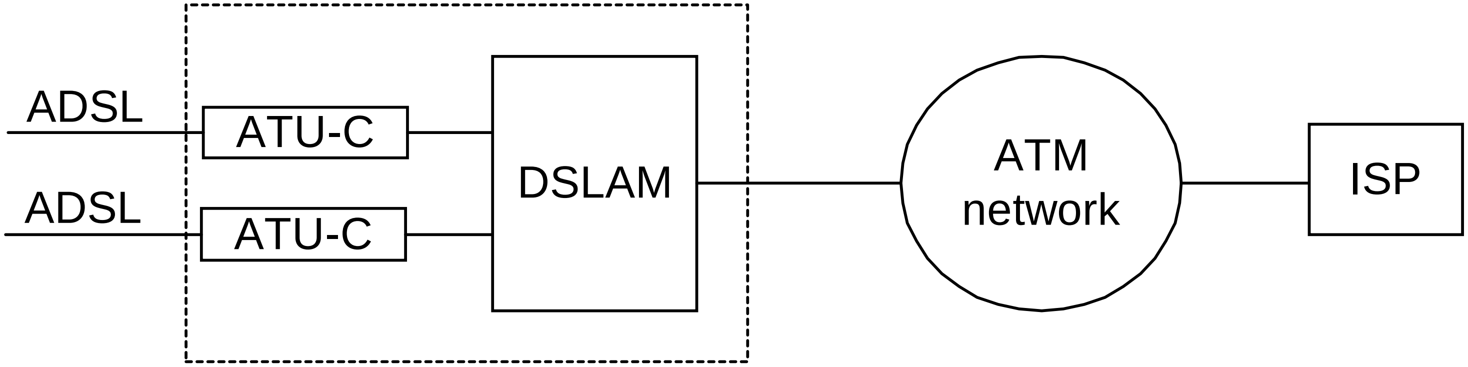
DMT: divide into 256 4-kHz channels, use QAM in each
CAP: carrierless amplitude modulation phase (single channel)

Data:

PPP (variable size frames with HDLC)
ATM, also with HDLC

Note: Splitter-less GLite

CS - DSL: ADSL (cont.)



Central Office

ADSL Modem

DSLAM

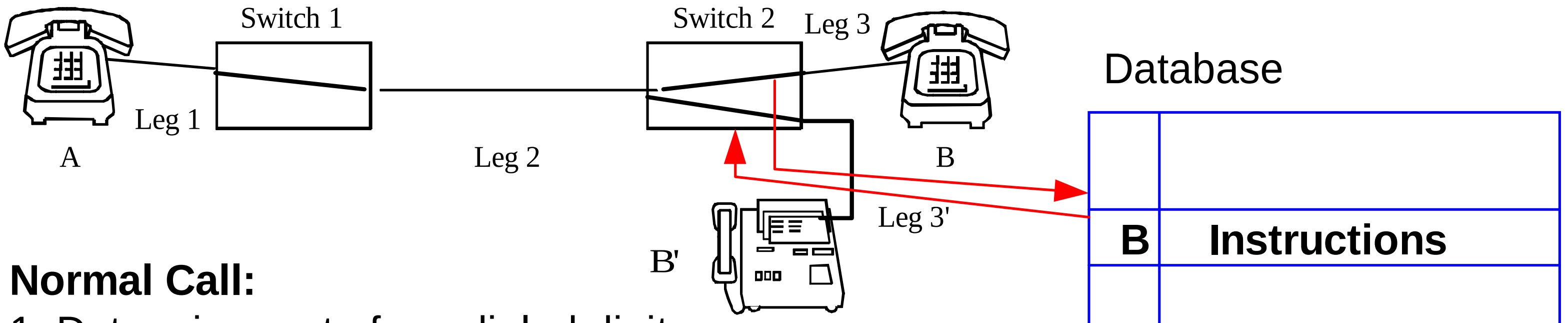
CS - DSL: Modems

Communications Protocols		
Protocol	**Maximum Transmission Rate**	**Duplex Mode**
Bell 103	300 bps	Full
CCITT V.21	300 bps	Full
Bell 212A	1,200 bps	Full
ITU V.22	1,200 bps	Half
ITU V.22bis	2,400 bps	Full
ITU V.29	9,600 bps	Half
ITU V.32	9,600 bps	Full
ITU V.32bis	14,400 bps	Full
ITU V.34	36,600 bps	Full
ITU V.90	56,000 bps	Full

CS: Intelligent Networks

- Service Examples
- Architecture
- Functional Components
- Summary

CS - IN: Service Examples



Normal Call:

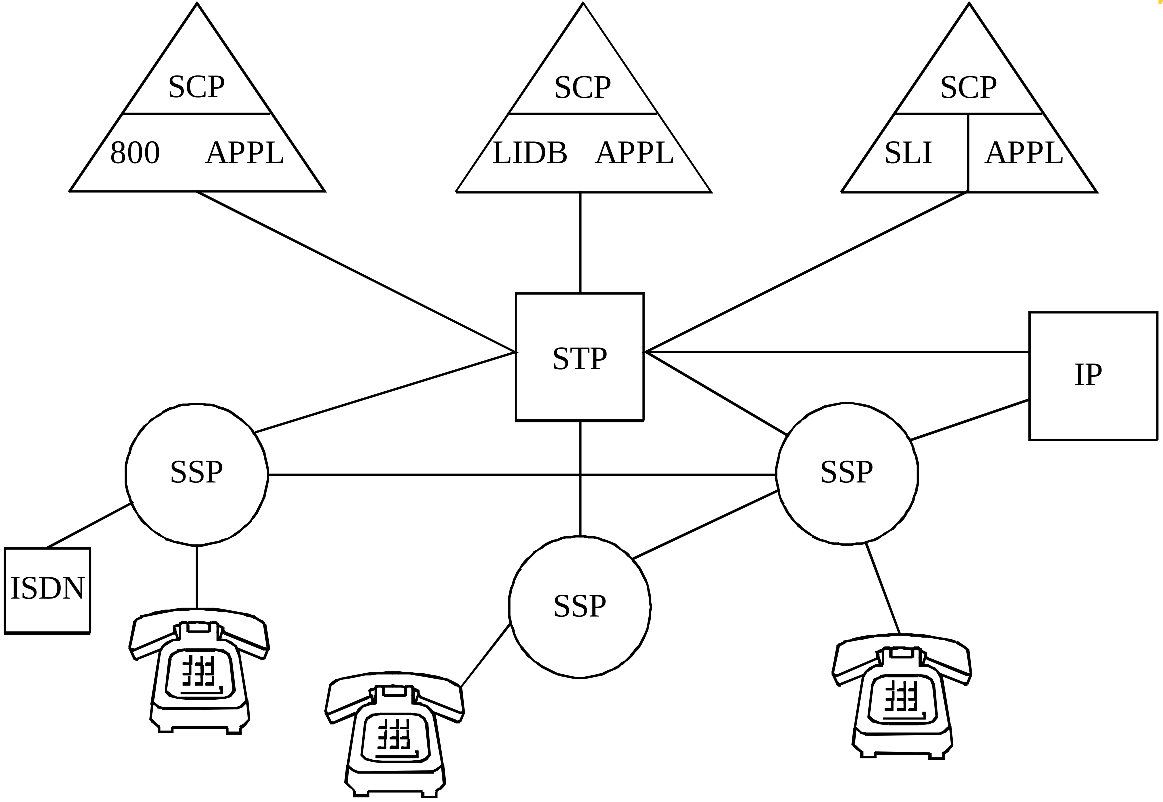
1. Determine route from dialed digits
2. Create and join legs 1, 2, 3
3. Verified called party is available
4. Conversation between A and B
5. Detect termination by participant set "on-hook"
6. Free legs 1, 2, 3.

Call Forwarding:

1. A dials B
2. Switch 1 creates and joins legs 1, 2
3. Address B sets off trigger at switch 2
4. Database responds to B: If no "off-hook after three rings, forward to B'
5. Eventually, switch 2 frees leg 3 and creates and joins leg 3'.

[5.26-28]

CS - IN: Architecture



[5.29]

SSP = service switching point, SCP = service control point,
STP = signal transfer point (packet switch for SS7), IP = intelligent peripheral.

CS - IN: Functional Components

- Control of processing
  - SCP provides instructions to SSP
  - SCP releases control to SSP at completion
- Connection request
  - Create leg between SSP and other component
  - Join leg to an ongoing call
  - split leg from an ongoing call
  - free a leg to release the resource
- User Interaction Request
  - Sending information (announcement, ringing)
  - Receiving information (dialed digits)

CS - IN: Functional Components (cont.)

- Network Resource Status Request
  - Used by SCP to process some call control
  - Instructs SSP to notify events (e.g., on-hook, flash)
- Network Inforamtion Revision Request
  - Enables SCP to modify information stored in SSP

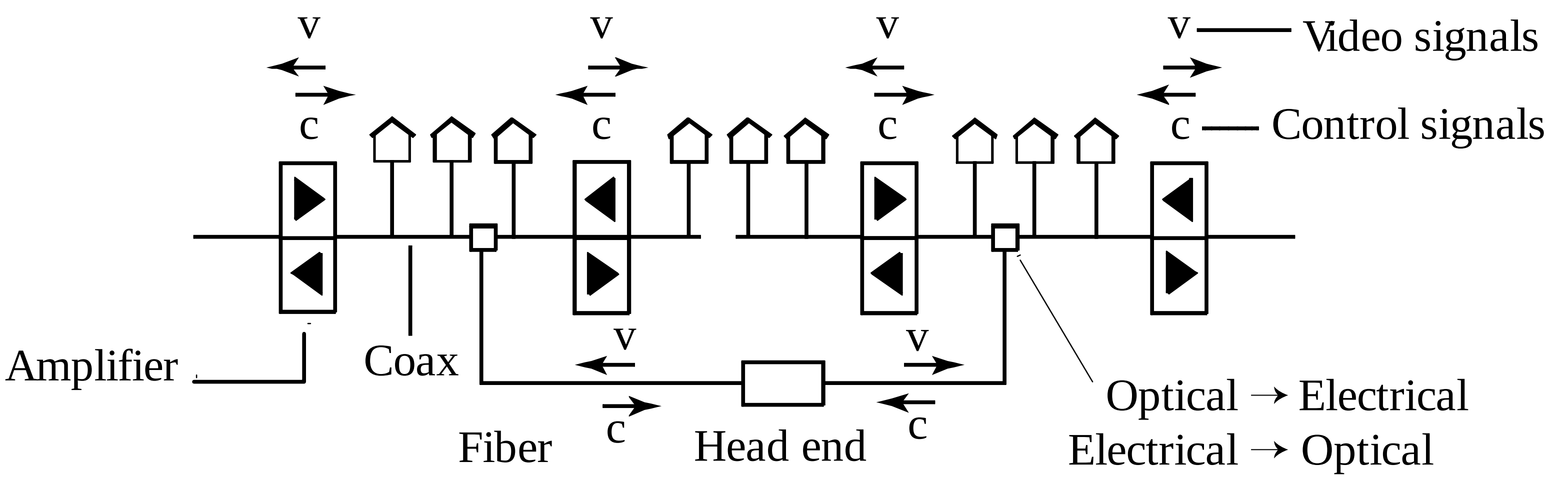
CS - IN: Summary

- Separation of operations and control
- Enables creation of new services as programmable sequences of functional components
- Sophisticated customers can program these services (e.g., 800-number services)

CS: CATV

- Layout
- Layers
- Services
- MPEG

CS - CATV: Layout



v — Video signals

c — Control signals

Amplifier

Coax

Fiber

v

c

Head end

v

c

Optical → Electrical

Electrical → Optical

[5.32]

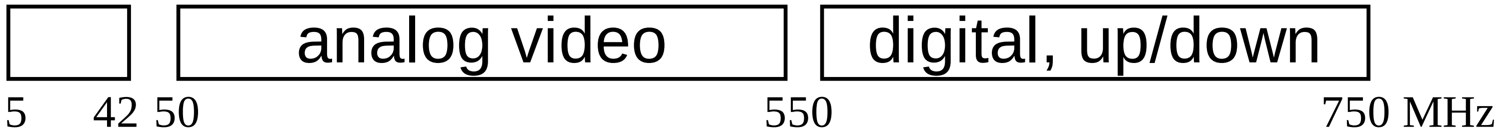
6-MHz digital channels
QPSK upstream: 10.8 Mbps
64 QAM downstream 32.4 Mbps

16 VSB downstream 42.8 Mbps

Reverse
channel

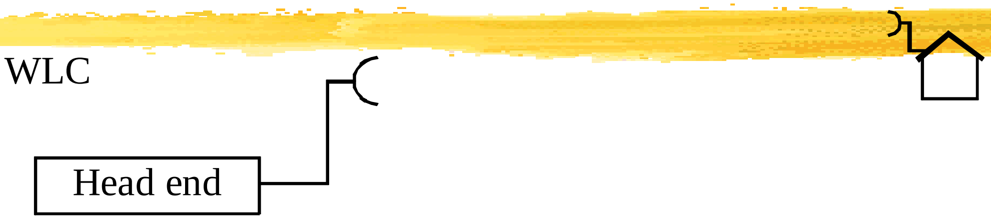
6 (United States) or 8 (Europe) MHz
analog channels (VSB)

[5/31]



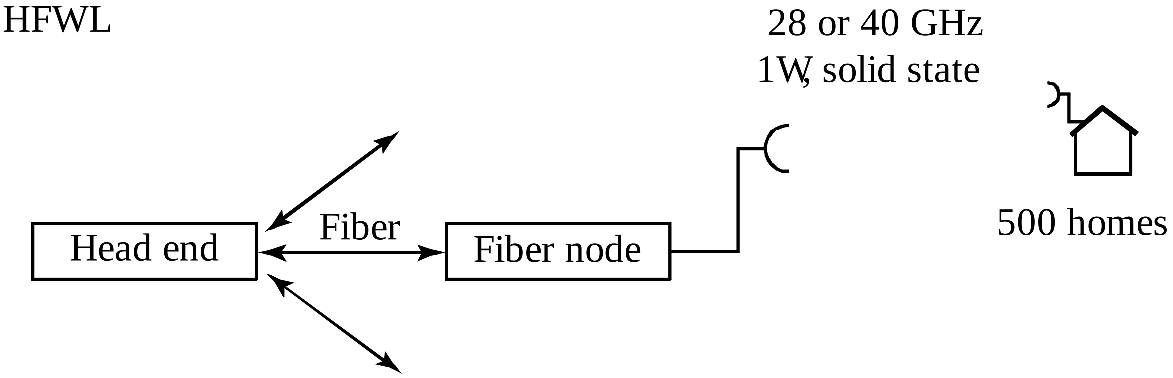
CS - CATV: Layout (cont.)

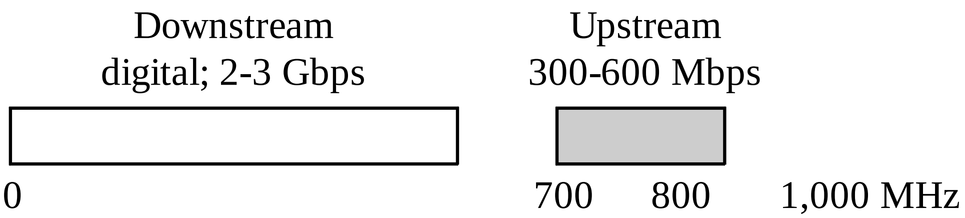
WLC



2-3 GHz, < 40 km, 33 6-MHz analog channels, or
28-30 GHz, < 25 km, 100W TWT , 49 20-MHz FM analog

HFWL





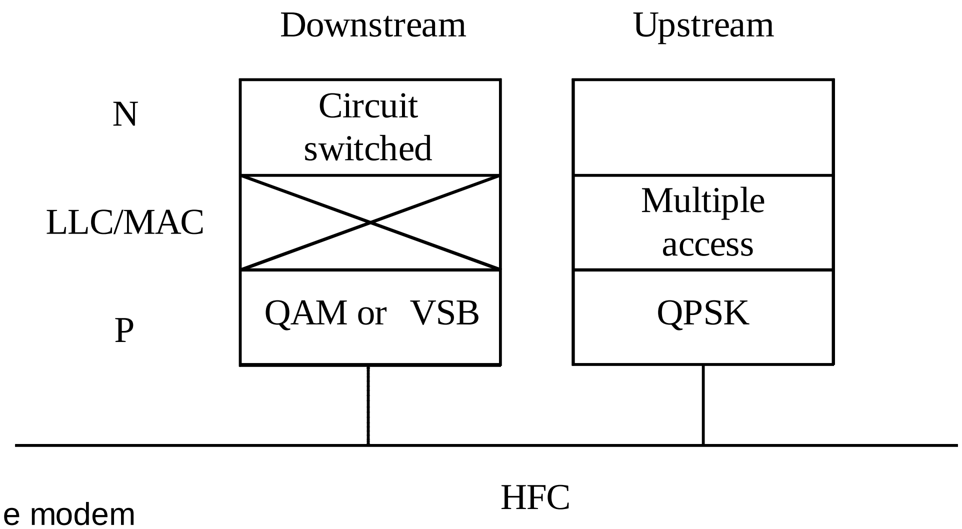
CS - CATV: Layers

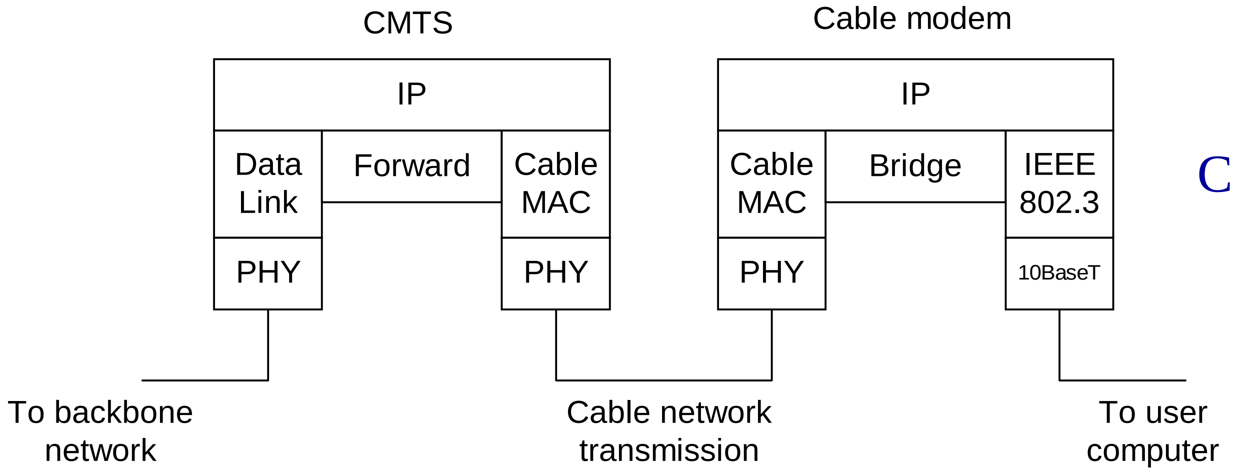
DOCSIS (data over cable service interface specification):

Synchronization as in TPON

Upstream:reservation minislots

Backoff after collision



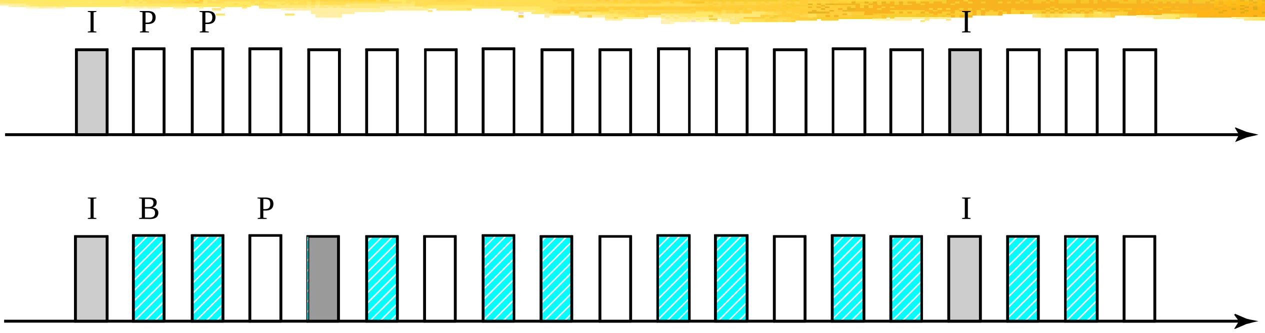


CMTS = Cable Modem Termination System

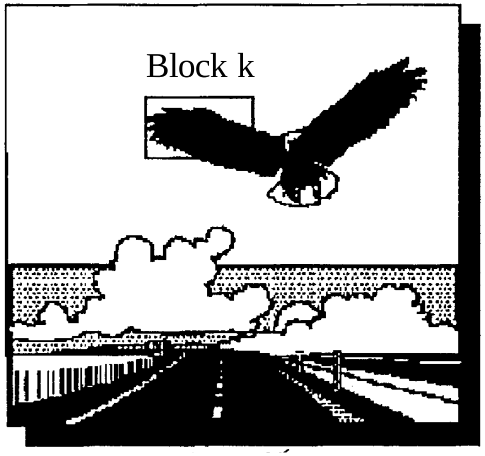
CS - CATV: Services

- Video on demand (MPEG2 over an available channel)
- Internet Access (shared 3Mbps up and 38Mbps down)
- Telephone service

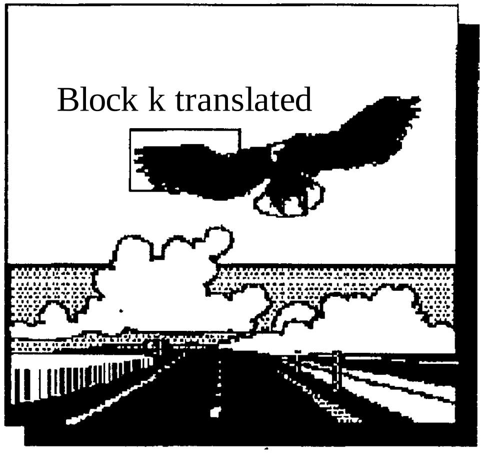
CS - CATV: MPEG



I = Intraframe compression: DCT + run-length encoding of DCT coefficients
P = Predictive compression: motion compensation
B = Forward/backward compression: interpolation between previous and future frames

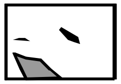


Frame *n*



Frame *n* + 1

Translation vector for
block k to achieve best
match in frame *n* + 1



Difference between
block k in frame *n* and
its best match after
translation in frame *n* + 1

<

QoS: Network Performance

- Overview
- Routing
- TCP
- ATM
- Measurements

QoS - Overview

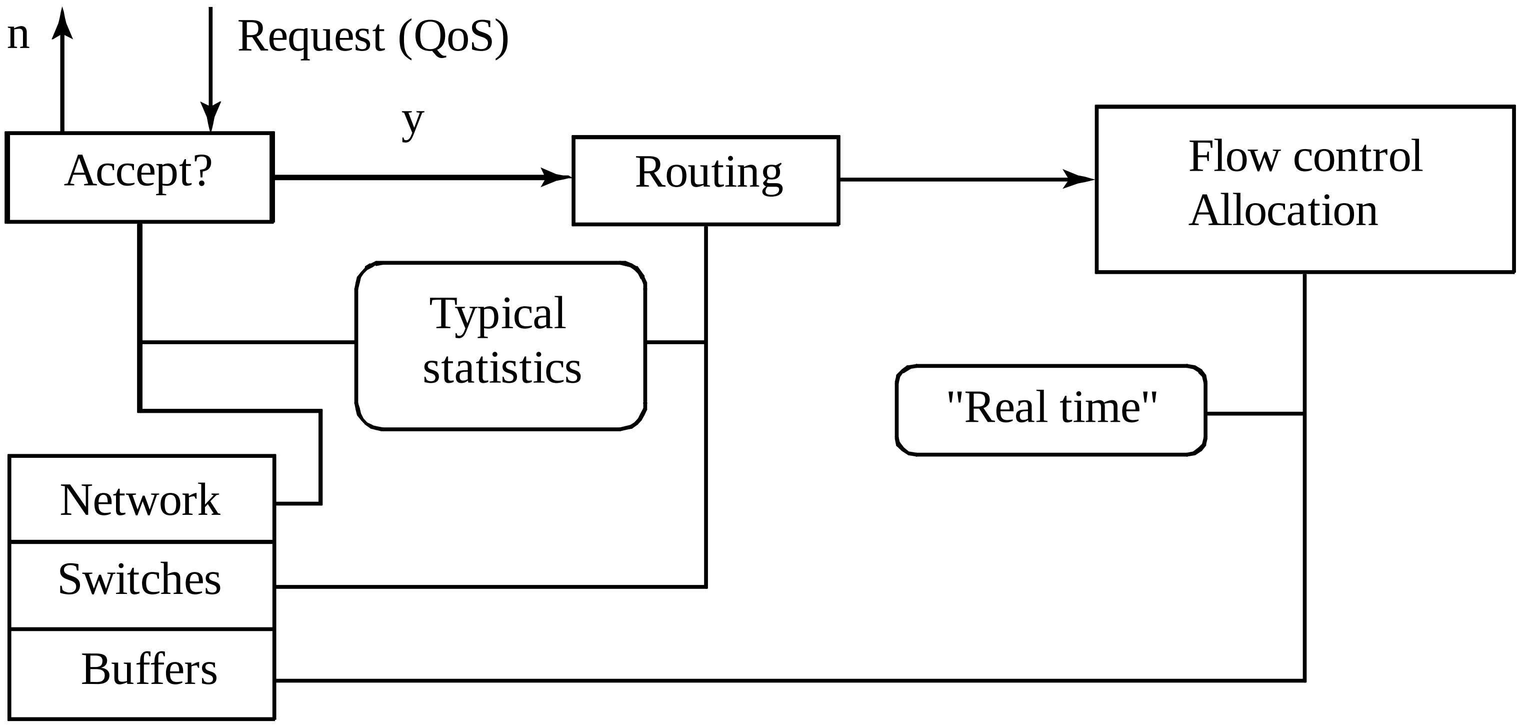
- Performance Measures:
  - Delays (mean, jitter)
  - Throughput
  - Loss Rate
- Different types of applications:
  - Closed-loop (TCP): throughput for transfer time

    Throughput = F(network, number of connections)
  - Open-loop (UDP): delay, jitter, loss rate

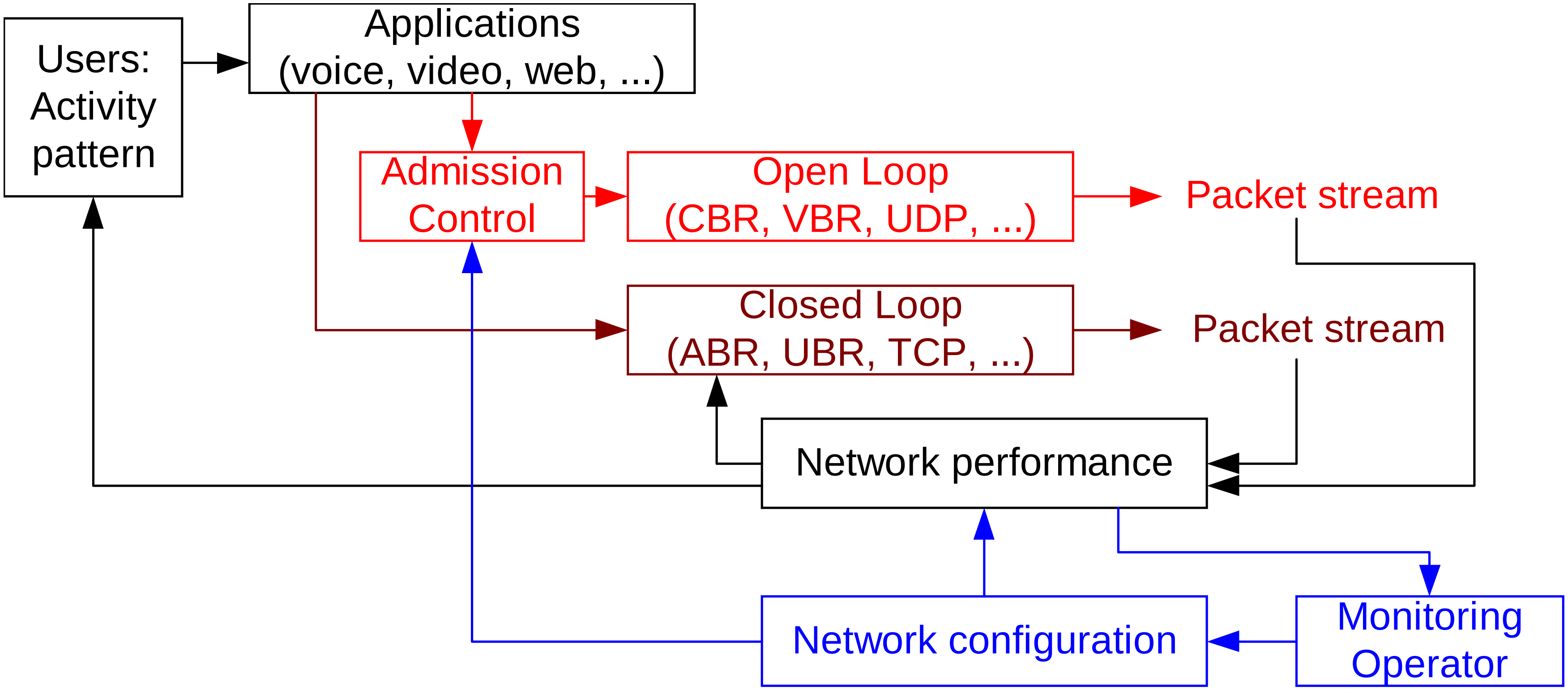
    Source generate specific traffic

QoS - Overview (cont.)

- Different time scales and actions



QoS - Overview (cont.)



Users: Activity pattern

Applications (voice, video, web, ...)

Admission Control

Open Loop (CBR, VBR, UDP, ...)

Packet stream

Closed Loop (ABR, UBR, TCP, ...)

Packet stream

Network performance

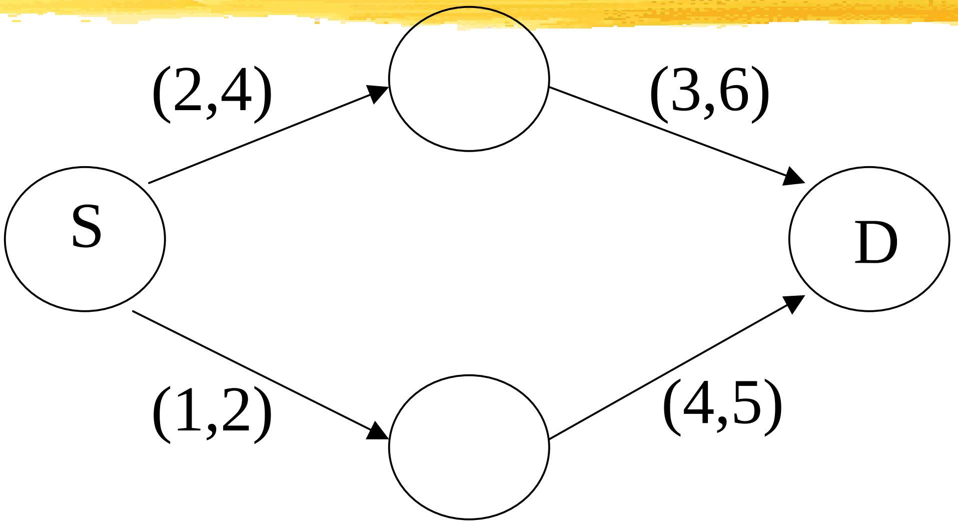
Network configuration

Monitoring Operator

QoS: Routing

- Problem
- Complexity
- Geographical Routing
- Algorithm

QoS - routing: problem



- each edge e labeled with (cost(e), delay(e), …)
- find a path p = $e_0\, e_1\, e_2 \ldots e_k$ from S to D so that

$$\sum \text{cost}(e_i) \leq C$$
$$\sum \text{delay}(e_i) \leq D$$

QoS - routing: complexity

- The problem is NP complete
- Theorem. Using a distributed Bellman-Ford algorithm, the problem can be solved in pseudo-polynomial time

  $$O(|V|.|E|.\min\{C,D\})$$

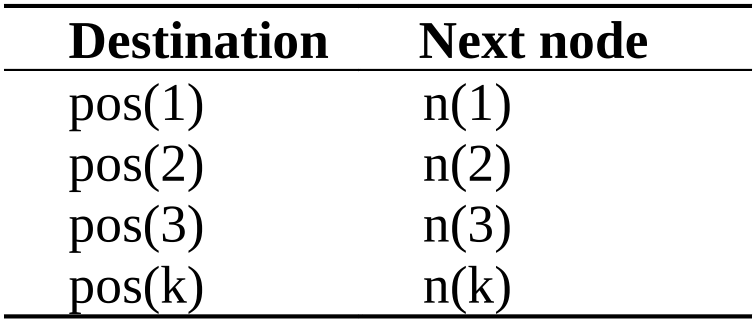
  where |V| (|E|) is number of nodes (edges)

QoS - routing: Geographical

- Nodes (1, 2, ..., n). Address of i is pos(i). Find route from S to D
- Routing decision.
  - j gets packet for i
  - j forwards to neighbor k such that pos(k) is closest to pos(i) than pos(j)
  - problem: there may be no such k known to j
  - solution: start route discovery

QoS - routing: Algorithm

- Routing table looks like this:



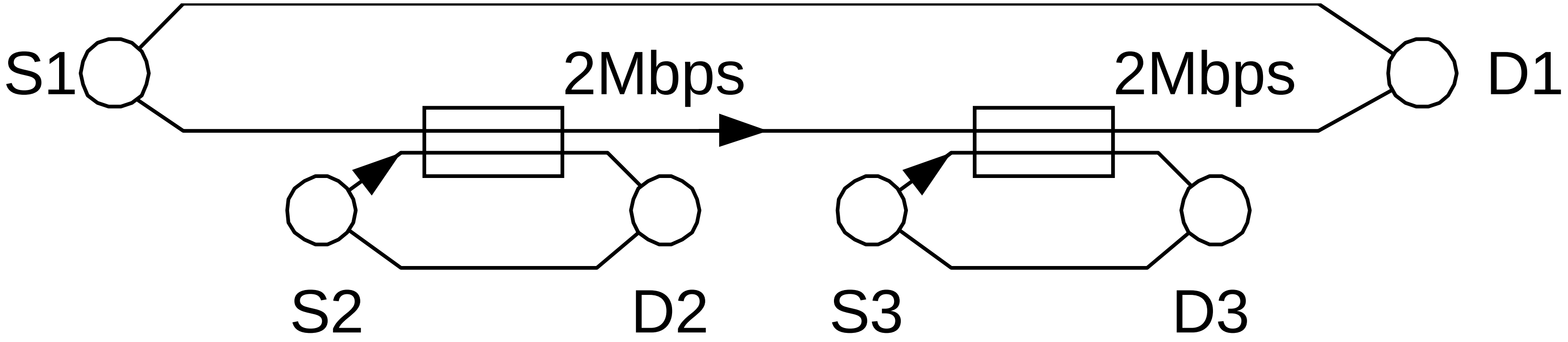
Destination	Next node
pos(1)	n(1)
pos(2)	n(2)
pos(3)	n(3)
pos(k)	n(k)

- When packet for i is stuck, find path to pos(i) and insert pos(i) in routing tables along the path.
- Theorem. There are no cycles in routing table.

QoS: TCP

- Difficulty of predicting QoS
- RIO for control
- Other form of control
- Model of TCP: R(p)
- Limiting the TCP bias

QoS - TCP: Throughput



Assume three long-lasting active TCP connections (S1->D1, …).
The rates R1, R2, R3 of the connections are determined by
the network topology and the link rates, not by the sources!
If TCP were “fair”, one would expect

R1 = 1, R2 = 1, R3 = 1 (max-min equilibrium)

or

R1 = 2/3, R2 = 4/3, R3 = 4/3 (proportionally fair).

With the bias of TCP in favor of short RTTs (proportional to $1/RTT^{1.5}$, say),

R1 = 0.03, R2 = 0.97, R3 = 0.97 (RTT1 = 10×RTT2 = 10 × RTT3).

QoS - TCP: Throughput (cont.)

The example shows the difficulty of predicting the throughput of TCP connections, even in a small network.
The network engineering problem, choosing the capacities so that the throughputs of most connections are satisfactory, is correspondingly difficult.

Other complication: The user behavior depends on the network performance.
That is, the fraction of “active” users may decrease if the network gets slow.

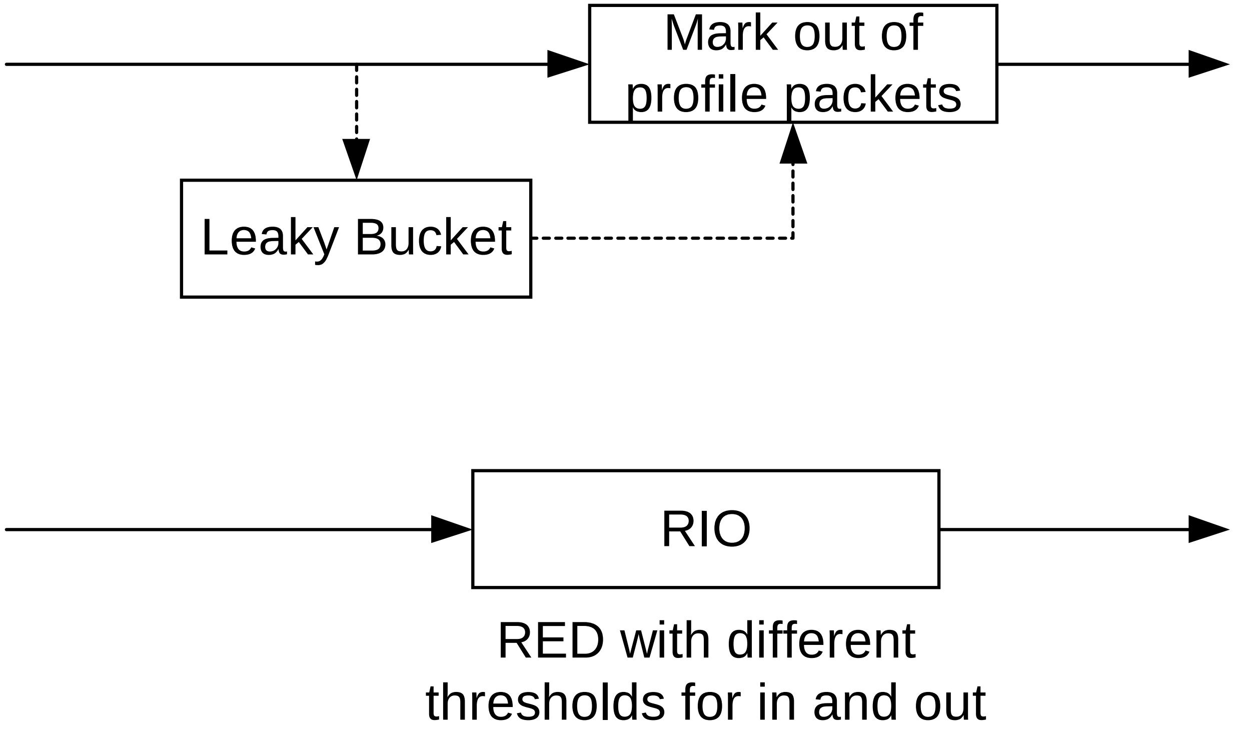
Guidelines:
Assume some rate per user and a number of users to engineer the network

Limit the number of users by some “class of service” mechanism (as in Metro)

Allocate bandwidth to different classes (DiffServ, MPLS)

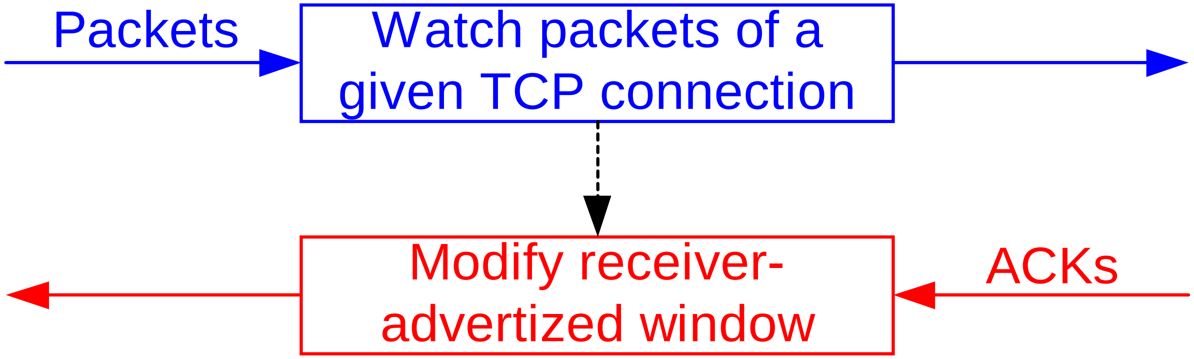
QoS - TCP: RIO

- RIO:


Mark out of profile packets
Leaky Bucket
RIO
RED with different thresholds for in and out

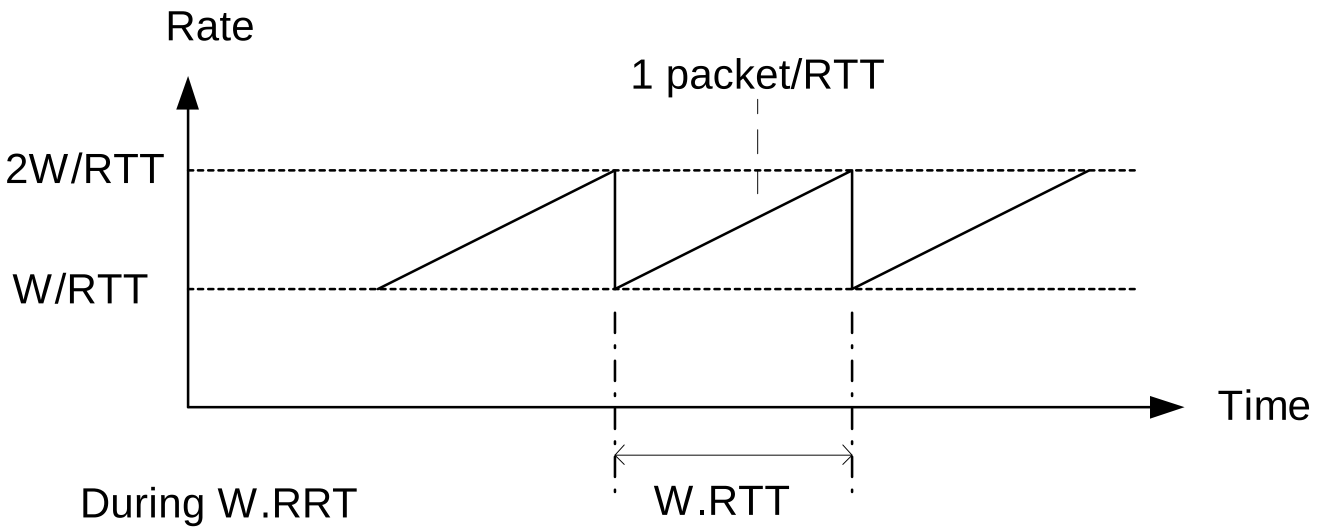
QoS - TCP: Other control

- Bandwidth allocation; Prevent cheating:


Packets
Watch packets of a given TCP connection
Modify receiver-advertized window
ACKs

QoS - TCP: R(p)

- Model of TCP (1)



During W.RRT
1 packet is lost
(3W/2).W packets are sent

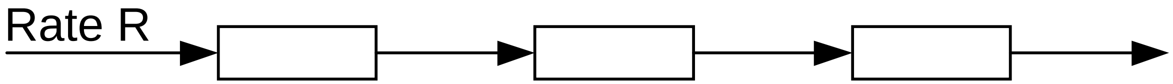
Hence, Throughput = R = 3W/2RTT and Loss rate = p = 1/[(3W/2).W].

Thus, $R = \frac{1.25}{RTT.p^{0.5}}$

Same p => R inversely proportional to RTT: TCP bias.

QoS - TCP: Limiting bias

- To limit bias: RED, ECN, DiffServ, MPLS (see Internet Chapter)
- Other approach: Fix increase rate of window (problem: difficult to find universally good rate)
- Other method: Vegas-like: target backlog:



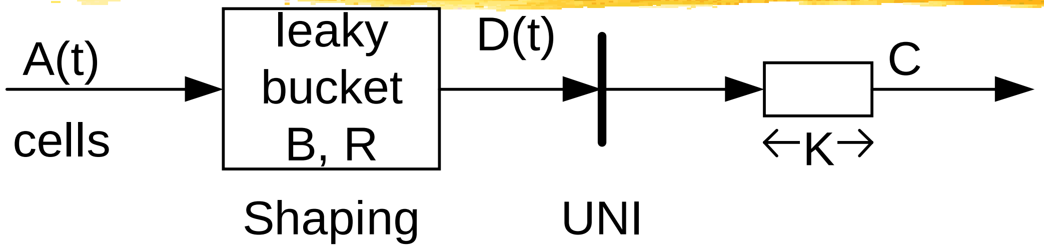
Propagation time = d
Measured RTT = T
Backlog in routers = Q

Window size W = R.d + Q => Q = W - R.d
Algorithm: Adjust W to maintain a target backlog Q
Motivation: comparable backlog => comparable rates

QoS: ATM

- Deterministic upper bounds
  Leaky buckets limit the burst size and the rate
  => Limiting the number of LB connections,
  one can avoid all losses
- Statistical approaches
  Deterministics bounds are unnecessarily conservative
  => Derive “99.9%”-guarantees

QoS - ATM: Deterministic



A(t)

cells

leaky bucket B, R

Shaping

D(t)

UNI

C

←K→

Definition:

D(.) is LB(B, R) if

D(t+s) - D(s) < Rt + B for all t

Fact 1:

D(.) is LB(B, R)
K > B, C > R

→ No loss

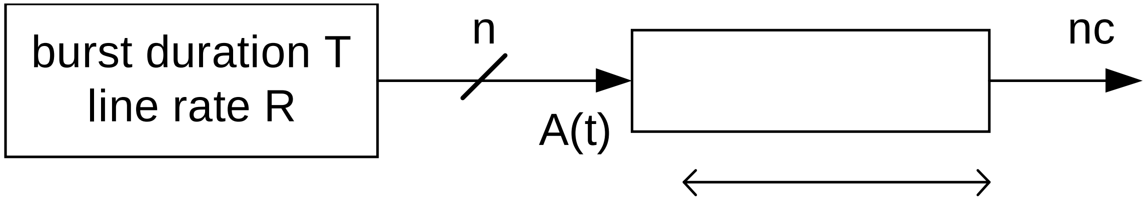
Fact 2:

B and R are additive for LB(B, R) streams

Fact 3:

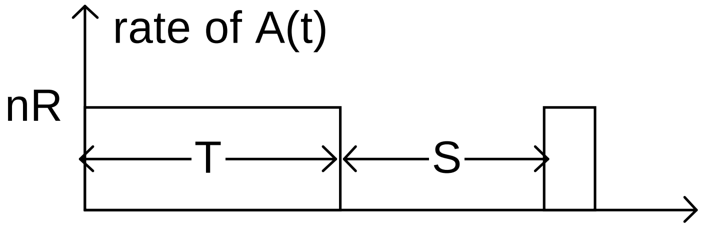
Leaky bucket is the best way to shape traffic into LB(B, R) stream.

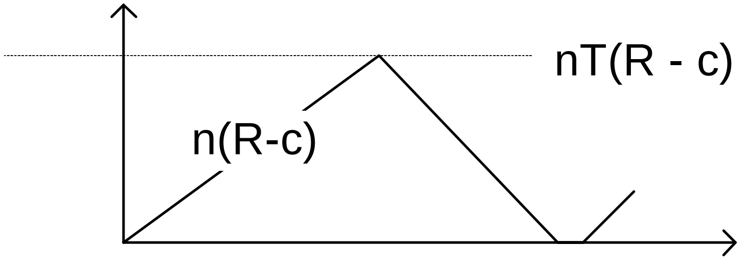
QoS - ATM: Deterministic (cont.)

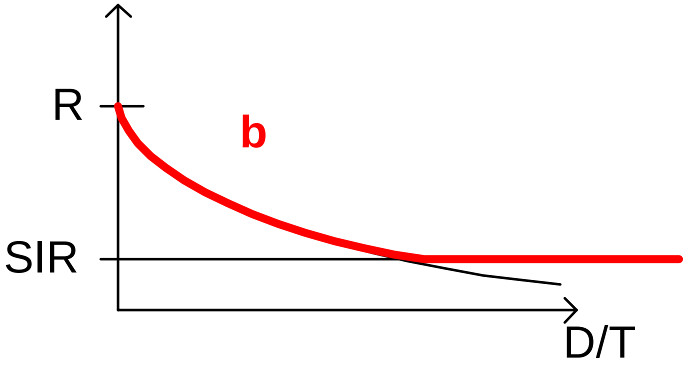


K < ncD to limit delay to D

Worst case:







Need nT(R - c) < ncD => c > max{R.T/(T + D), SIR} =: b

b is a measure of the bandwidth required per connection.
b is an "**effective bandwidth**"

QoS - ATM: Statistical

Motivation:

Consider a video source with T = 7s, D = 1s => b = 7R/8

Assume source is ON (10Mbps) w.p. 1/6, OFF (1Mbps) otherwise.

For 100 sources, mean number ON = 16.6, variance = 20(1/6)(4/6) = 2.2, sigma == 1.5

Hence, P(more than 16.6 + 3x1.5 = 21 sources ON) = 0.1%

Thus, the rate 21x10Mbps + 79x1Mbps = 289Mbps is sufficient for 100 sources

==> Deterministic allocation gives 8.75Mbps per source

Statistical allocation gives 2.9Mbps per source

Gain = factor 3 in required capacity for MPEG2

QoS - ATM: Statistical (cont.)

More general theory: Large Deviations

Version 1: Zero-buffer

$P(X1 + ... + Xn > nc) = (A/n^{0.5})\exp\{- nI(c)\}$

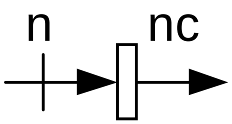
Version 2: Large buffer

$P(\text{queue occupancy} > x) = \exp\{ - xd\} \Leftrightarrow c > b(d)$

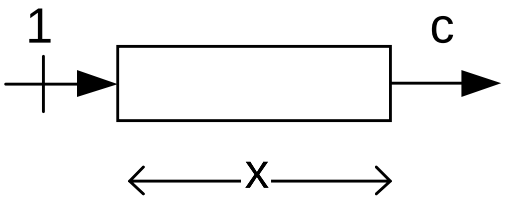
$b(d)$ = effective bandwidth

Version 3: Many connections

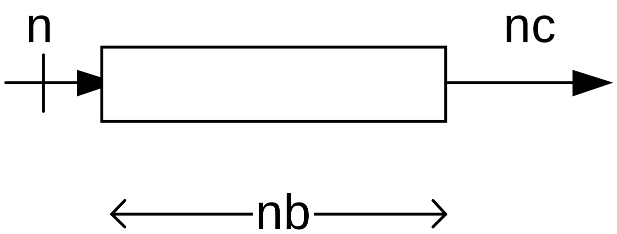
$P(A1(t) + \quad + An(t) > b + nct) = \exp\{ - nI(c,b)\}$



version 1



version 2

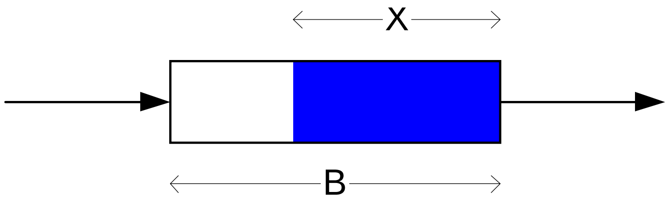


version 3

QoS: Measurements

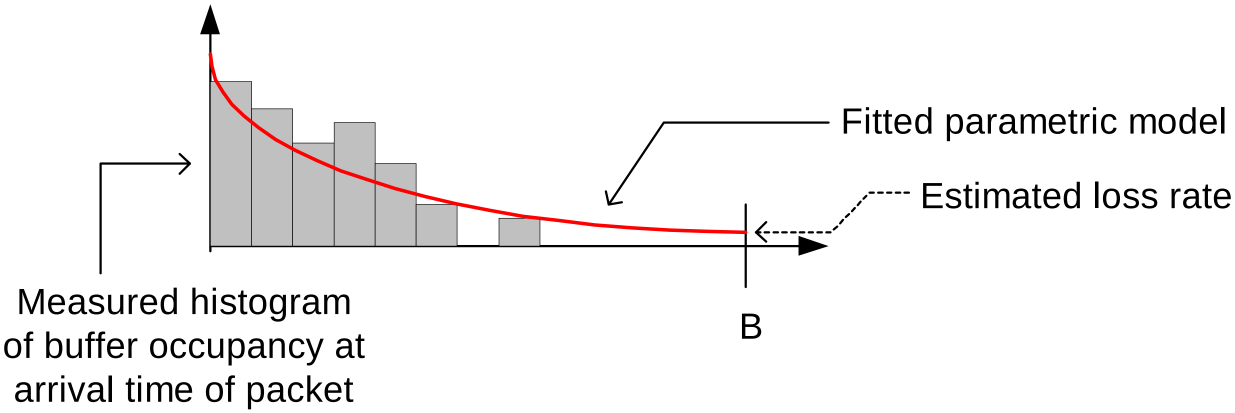
- Loss Rate
- Rate-Latency

QoS - Measurements: Loss



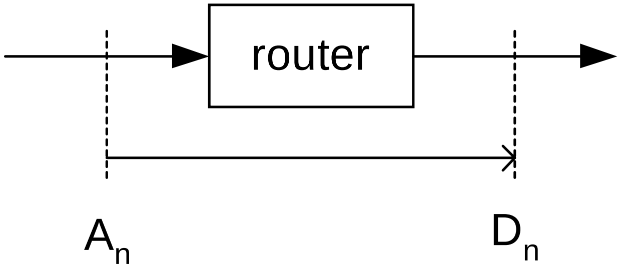
Direct measurements of losses are slow. Must observe a few hundred losses to estimate the loss rate. If loss rate = 1 cell/10^5 cells, and if the cell rate is 60% of 155Mbps, it takes 90 seconds to see 200 losses. This estimation may be too slow for control or for tracking changes.

Speed-up:



Suitable model: trade-off between
potential accuracy (complexity) and ease of fitting.

QoS - Measurements: rate-latency



Model of QoS: rate-latency (R, T)

Almost all the packets (say 99%) leave faster than they would if they were served by parallel servers with rate R each, followed by a delay of T.

Estimation of (R, T):

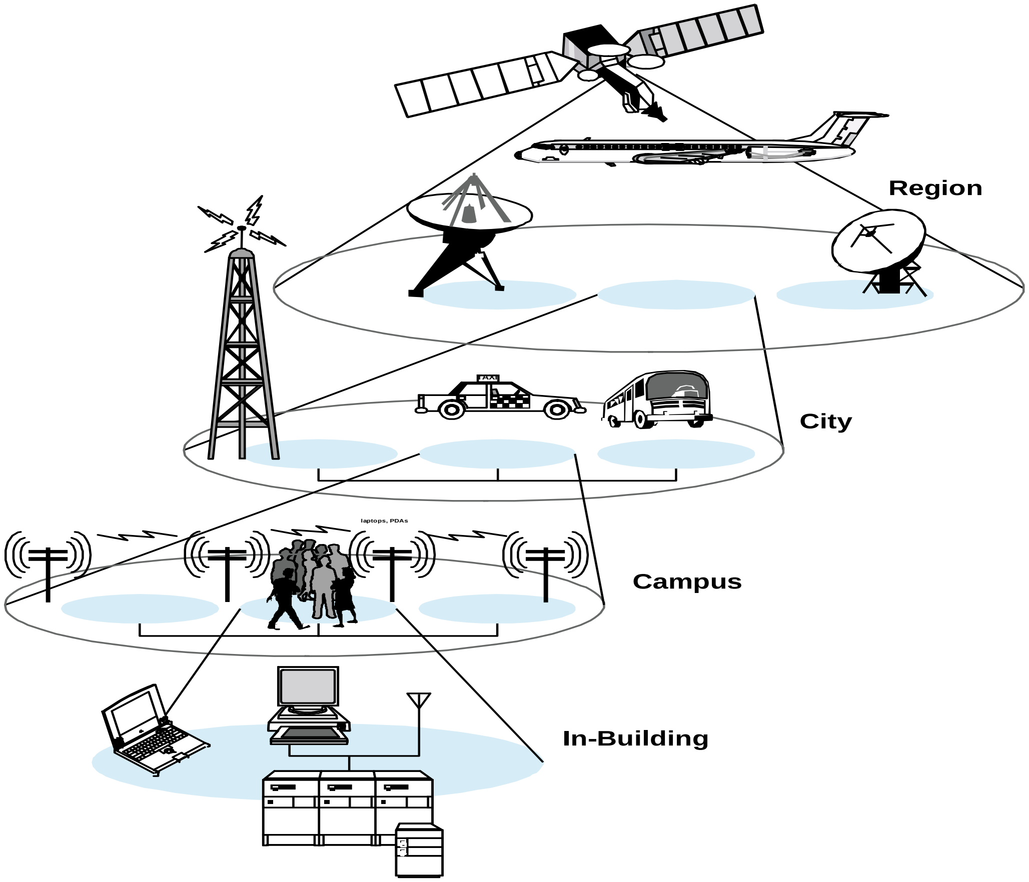
P(R, t) = fraction of packets that meet the (R, t) QoS
Parametric fit of P(R, .) and extrapolate to find the value of T so that P(R, T) = 99%.

<

Wireless networks

- Vision: Network Hierarchy
- Physical layer
- Network architectures
- Cellular telephony
- Wireless LANs and Ad Hoc
- WAP

WL - Vision: network hierarchy

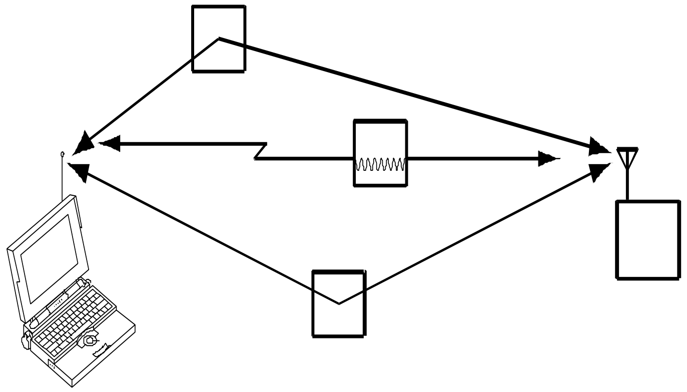


WL - Physical Layer

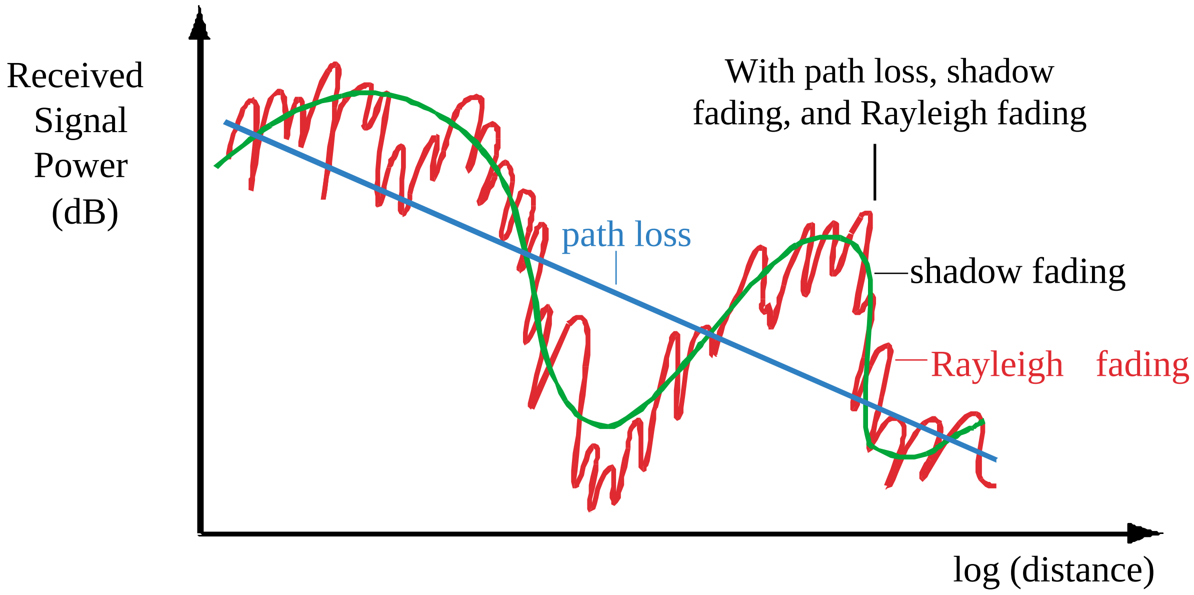
- Channel Characteristics
- Channel Summary
- Link Design Considerations

Physical - Channel characteristics

- path loss
  - $P_R \propto (G \times P_T)/(f^2 \times d^\alpha)$
- shadow fading (4-12db)
  - Random Gaussian fluctuations from hills, ..
- multipath (20-30db)
  - Narrow-band Rayleigh or Ricean
  - ISI for large multipath spread



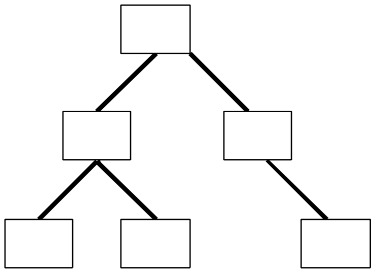
Physical - Channel summary



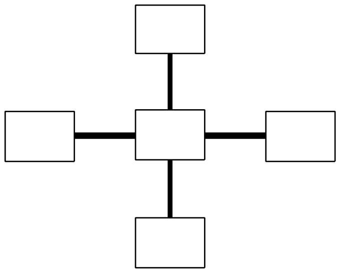
Physical - Link design considerations

- frequency band
- modulation (linear, nonlinear)
- channel coding (FEC, convolution codes)
- flat-fading countermeasures
  - diversity (spatial, temporal-interleaving)
  - ISI compensation (equalizers, OFDM)
  - spread spectrum

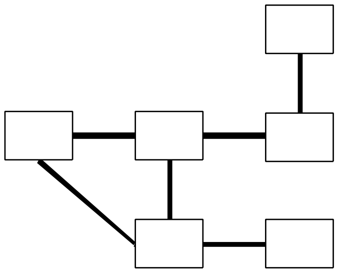
WL - Network arrangements



Hierarchical
(cellular)



Star
(bluetooth)



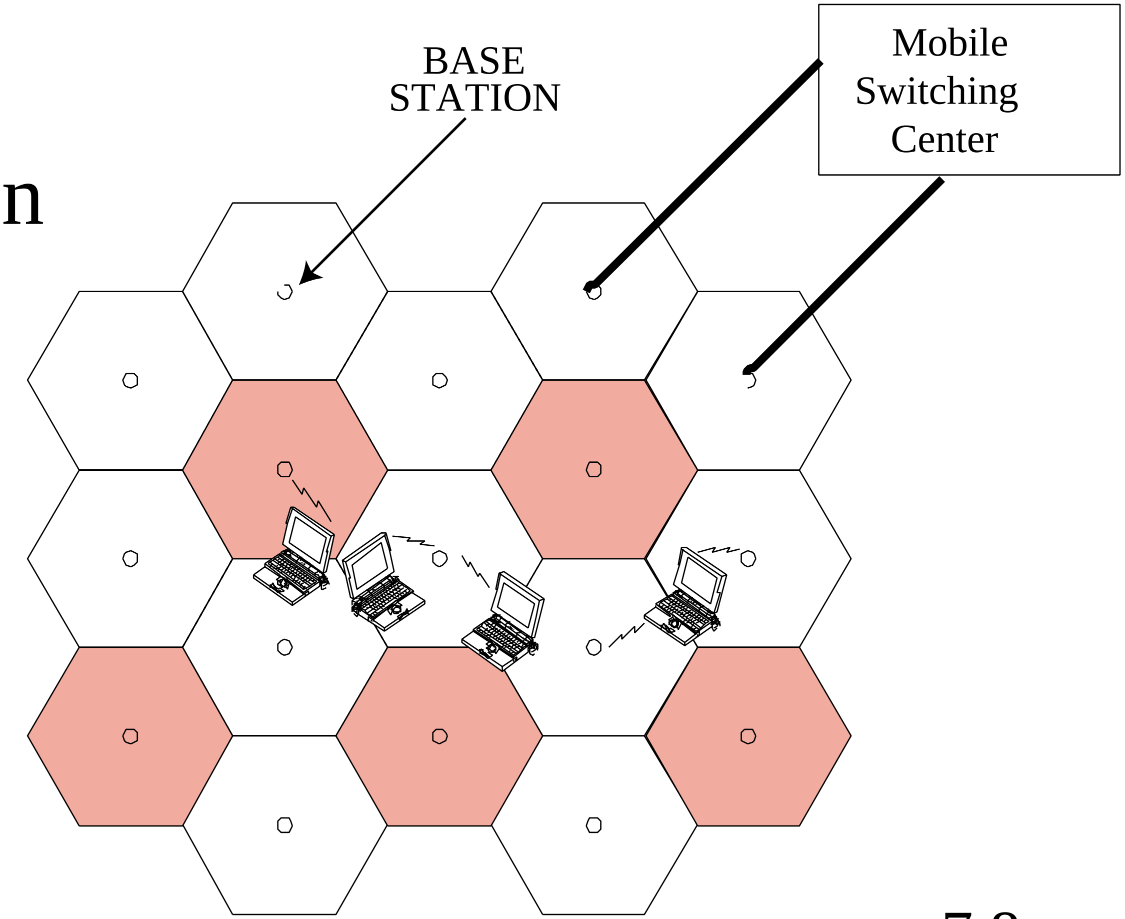
Peer-to-Peer
(wavelan)

WL - Cellular telephony

- Principles
- 1G
- 2G
- 3G

Cellular: Principles

- frequency reuse in cells;
- base to mobile communication



Cellular - 1G cellular

- Analog transmission, AMPS
  - 800 MHz, 30 kHz FDMA channels
  - frequency reuse factor 1/7 gives one call / 210 kHz /cell or 60 concurrent calls per 12.5 MHz of operator spectrum

Cellular - 2G cellular: digital

- IS-54/136 (3 TDMA channels/30 kHz, 800/1900 MHz), 10 Kbps voice, increases capacity over AMPS by 3
- IS-95 (1.23 MHz, DSS channels, 800/1900 MHz), variable bit rate voice, adaptive power control, diversity reception, increases capacity over AMPS by 10-20
- GSM (8 TDMA channels/200 kHz, 900 MHz), increases capacity through smaller cell size

Cellular - 3G cellular (1/2)

- IMT-2000 goals
  - common global frequency band (1.8-2.2 GHz)
  - common air interfaces for vehicular (144 Kbps), pedestrian (384K), indoor (2.0 M), satellite (9.6K)
  - circuit-switch, packet-switch, multimedia
  - compatibility with fixed networks
  - global roaming

Cellular - 3G cellular (2/2)

Proposals

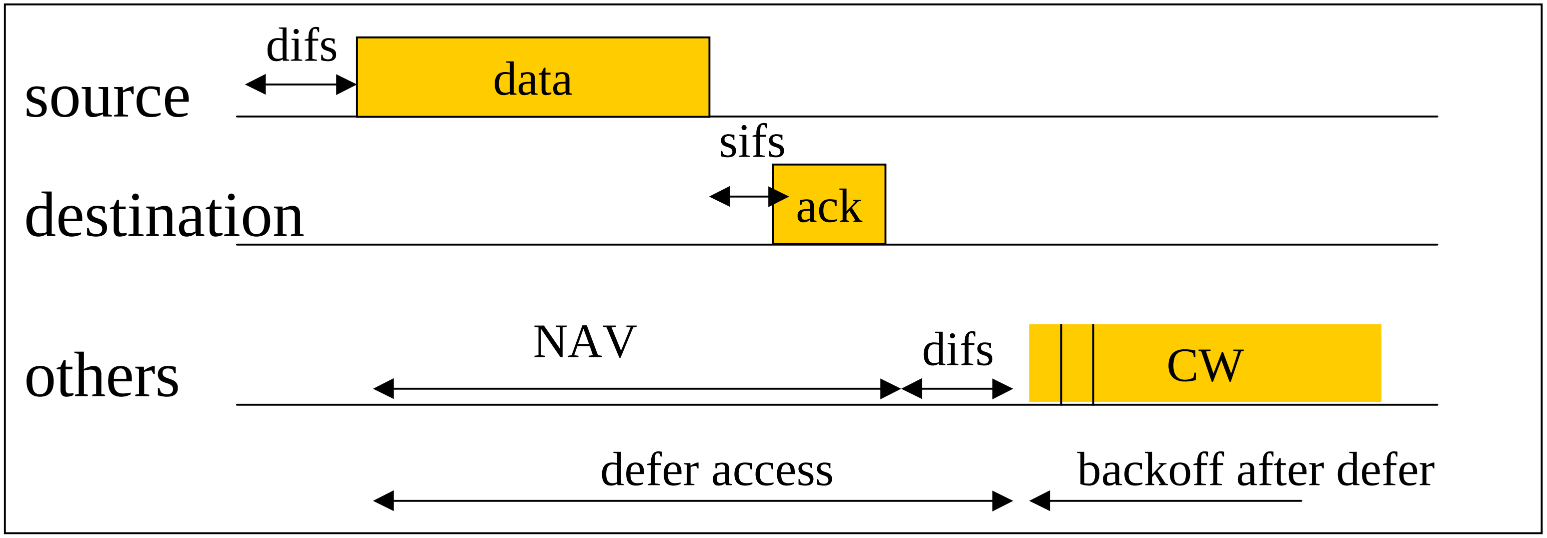
- UWC-136/HS (30,200,1600 KHz channels), support of GPRS-based packet data, EDGE
- GMS-SMS(short message service), GPRS(general packet radio service, X.25 and IP compatible), EDGE(enhanced data rates for global evolution, 200 KHz/8TDMA slots)
- IS-95B up to 14.4×8 Kbps, IS-95C (mutlicarrier CDMA)

WL - LANs and ad hoc

- 802.11
- Bluetooth
- Ad hoc

WL Lans ...: 802.11 (1/2)

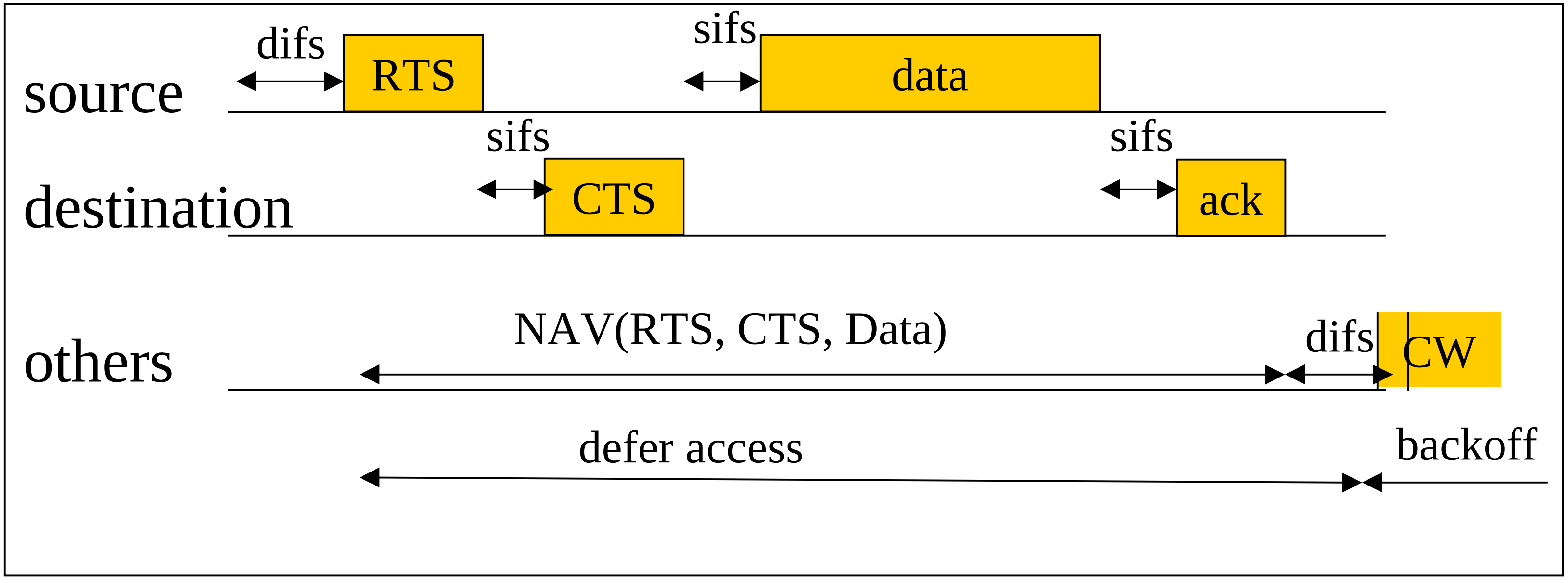
- Physical: FHSS 79, 1MHz channels, 2.4 GHz. (2.5 hops/s)



Contention mode MAC

WL Lans ...: 802.11 (2/2)

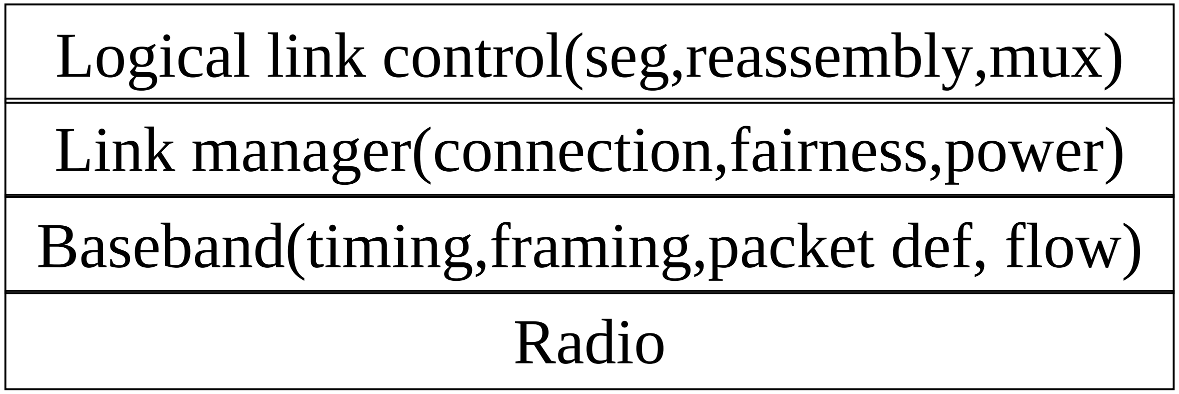
- RTS/CTS



Using RTS/CTS

WL Lans ...: Bluetooth

- slotted channel (SCO, ACL), 10m,1Mbps, 2.4 GHz; 1600 hops/sec through 79, 1-MHz band
- up to 8 nodes in a 1 Mbps piconet, access controlled by master through polling



Logical link control(seg,reassembly,mux)

Link manager(connection,fairness,power)

Baseband(timing,framing,packet def, flow)

Radio

WL Lans ...: Ad hoc network

- Types of addressing
  - Unique, global IP ?
  - Geographical addressing (GPS-based?)
  - Group based (e.g. vehicle.xxx)?
- Each addressing scheme has its own uses and can be implemented by special “domain name service.” Similar to PBX, VPN and some alternatives to ICAAN

WL: Wireless Application Protocol

- What is Wireless Application Protocol?
- Architecture Overview (layers)
- Accessing Web from Cell Phones
- Bottom up description of the stack

WAP: What is WAP?

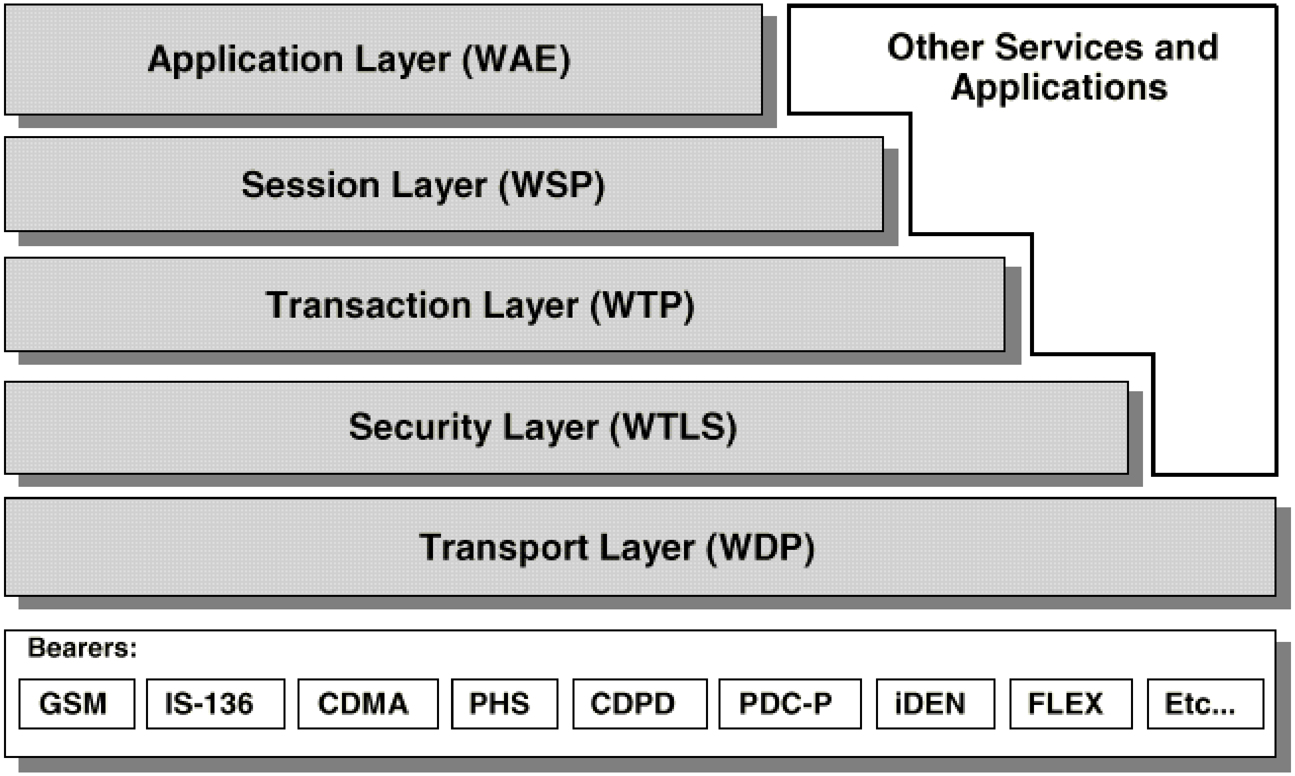
- "WAP specifies an application framework and network protocols for wireless devices such as mobile telephones, pagers, and personal digital assistants (PDAs)."
- WAP is the mobile phone industry’s answer to interactive web applications.
- WAP defines its own set of protocols but models after existing web protocols.

WAP Protocols (1/2): vs. IP world

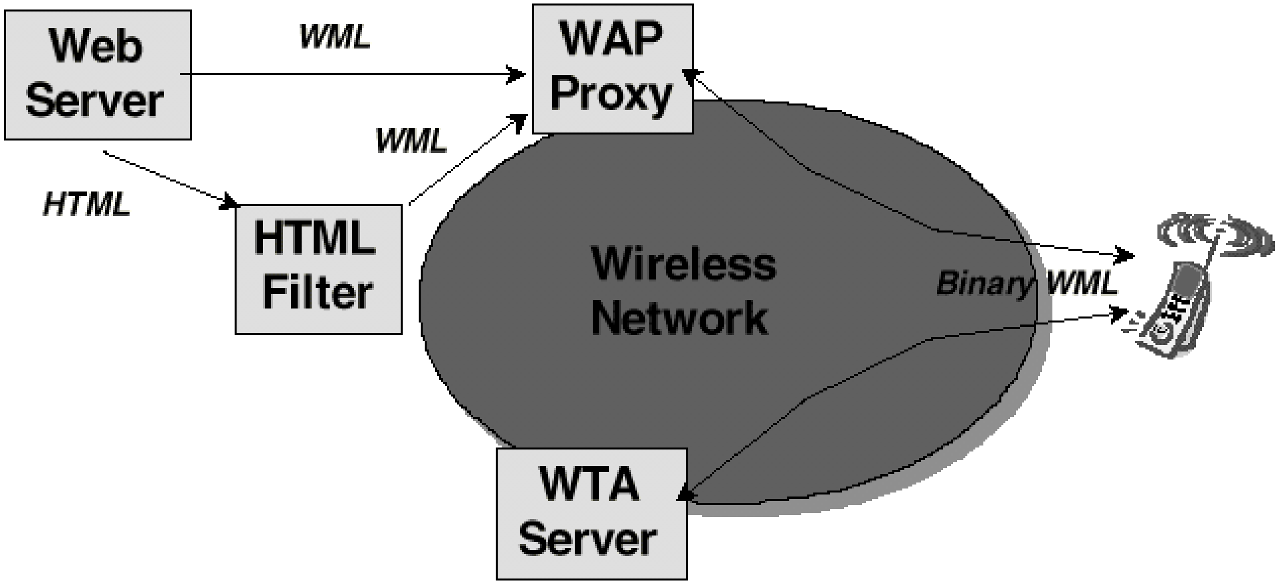
GSM, CDMA, IS-136	IP
WDP (datagram)	UDP
WTLS (security)	TLS (from SSL)
WTP (transaction)	no counterpart
WSP (session)	HTTP
WML (markup)	HTML
WML Script (scripting)	JavaScript

Rationale: reuse as much as possible from IP world, but optimize for the wireless world (i.e. compression, adapt to high-loss rate.)

WAP Protocols (2/2)



WAP: Accessing Web from Cell Phones



WAP: Datagram Protocol (WDP) (1/3)

- **Goal**: To allow transport, security, and session protocols to operate independent of the underlying bearer (e.g. GSM, CDMA, etc.)
- To provide a UDP-like interface

WAP: Datagram Protocol (WDP) (2/3)

Basic Features:

- Port Numbers: Demux lower layer packets to different higher layers
- Segmentation/Reassembly: only present if underlying bearer does not support it already
- Details: More than you ever want to know about the differences between some 30 kinds of cellular data network.

WAP: Datagram Protocol (WDP) (3/3)

Summary:

- If bearer (e.g. GSM USSD) speaks IP, WDP equals UDP.
- Otherwise, WDP adapts to the underlying network and provides Demux & Segmentation normally provided by UDP/IP.
- Reliable transport builds on top of WDP, not the underlying bearer even if bearer speaks IP

WAP: Transaction Protocol (WTP) (1/5)

- Goal: A light-weight transaction protocol necesaary for interactive "browsing" applications.
- T stands for Transaction, not transport.
- WTP is message-oriented protocol, not stream-based.
- Each new transaction has a new Transaction ID (~ seq no.)
- Provides 3 classes of service

WAP: Transaction Protocol (WTP) (2/5)

Class 0 (Unreliable 1-way)

- Sender sends a message but does not wait for an ack
- Receiver delives message to application w/o checking for duplicates
- No acks are sent
- Sample app: unreliable weather update (push)

WAP: Transaction Protocol (WTP) (3/5)

Class 1 (Reliable 1-way)

- Sender sends a message and waits for an ack (retx if necessary)
- Receiver checks for duplicates before delivering message to application
- Acks are sent and retransmitted if client retransmits request erroneously

WAP: Transaction Protocol (WTP) (4/5)

Class 2 (Reliable 2-way)

- Sender sends a message and waits for an ack (retx if necessary)
- Receiver checks for duplicates before delivering message to application
- Ack for request is piggy-backed on top of the reply from the server application
- Client acks the receipt of the result

WAP: Transaction Protocol (WTP) (5/5)

WTP: Other interesting features

- User-level acks (more precisely, application-level acks)
- May do Segmentation and Reassembly
- Allow multiple messages (PDUs) to be concatenated into one SDU (link-layer frames)
- Re-transmit bit to distinguish fresh vs. retransmitted packets

WAP: Session Protocol (WSP) (1/2)

- Currently defined WSP services include only browsing services WSP/B
- WSP/B is a binary equivalent of HTTP/1.1

WAP: Session Protocol (WSP) (2/2)

WSP Features Highlight

- Supports both connection-oriented and connectionless modes
- Pull: Support all HTTP/1.1 request methods (GET, PUT, POST, etc.) (confirmed and non-confirmed)
- Push: Server push (confirmed, non-confirmed)
- Suspend/Resume sessions independent of transport sessions

WAP: Application Layer (WAE) (1/3)

- Specifies an application framework for wireless devices such as mobile telephones, pagers, and PDAs.
- Defines a model suitable for building interactive applications that function well in ***narrow-band*** environment with ***medium to high latencies***.

WAP: Application Layer (WAE) (2/3)

Model

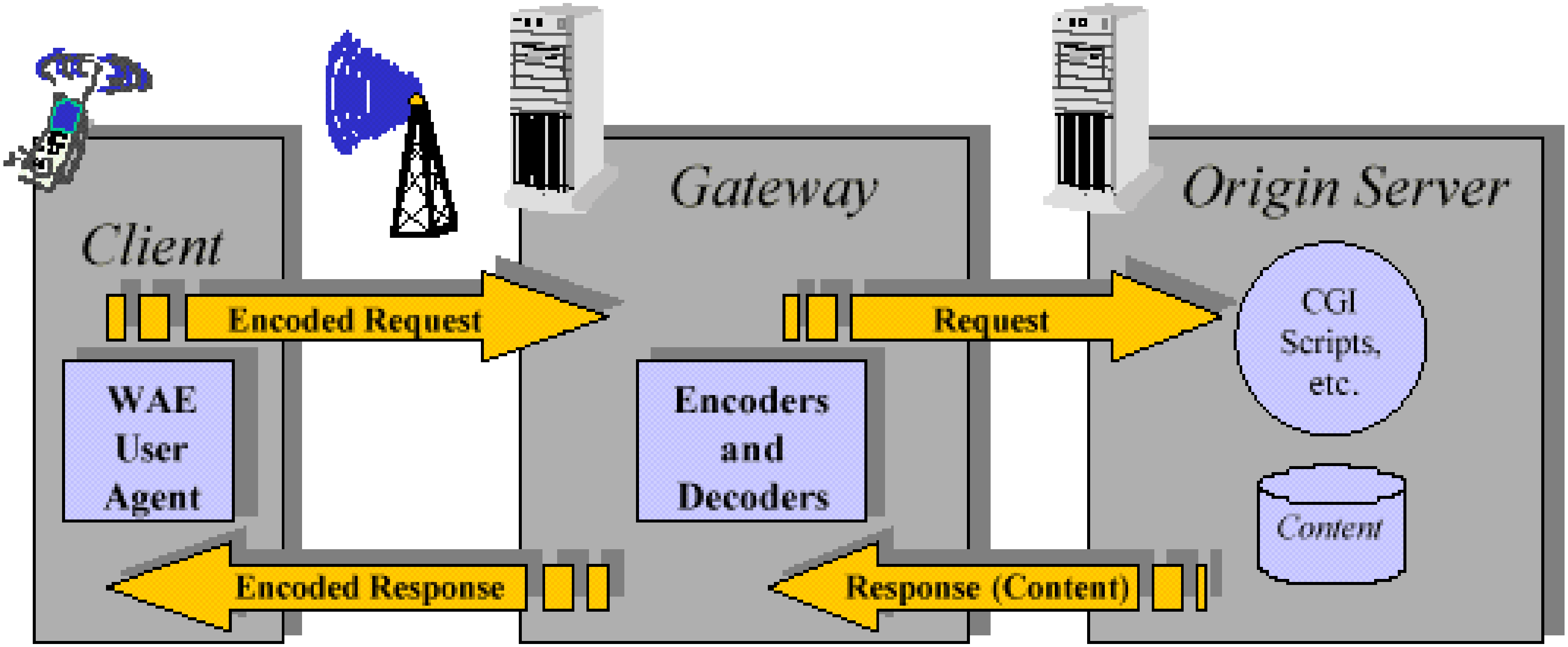


Figure 2: WAE Logical Model

WAP: Application Layer (WAE) (3/3)

Components:

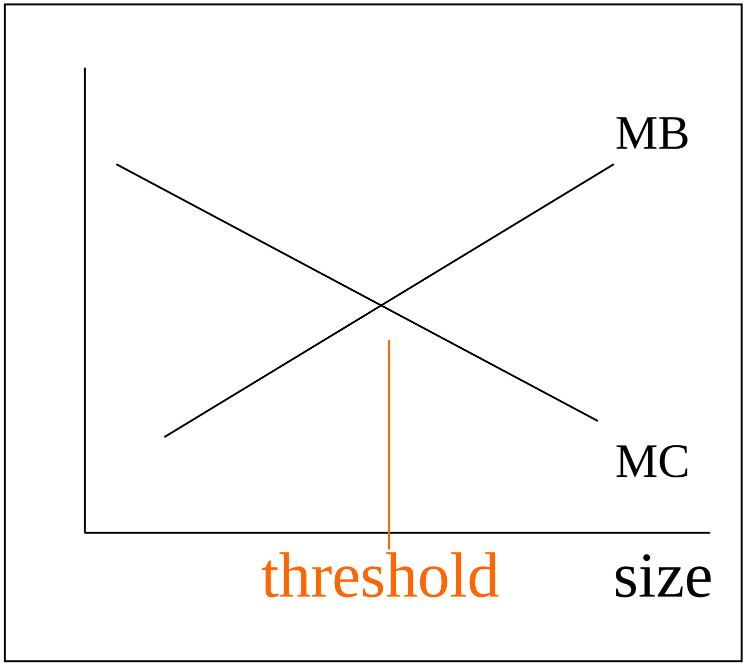
- User Agents: in-device software that retrieves and displays content to the user. Understands WML and WML-Script
- Gateways: Conversion between HTTP & WSP and/or HTML & WML
- Various WAE specific formats: WML, WML-Script, WBMP(bitmap)

Economics

- Principles
- eCommerce
- Resource Allocation

Economics: Principles (1/3)

- Network externalities
  - marginal cost decreases, marginal benefit increases
  - eg. telephone, email
  - threshold effect
  - first mover advantage

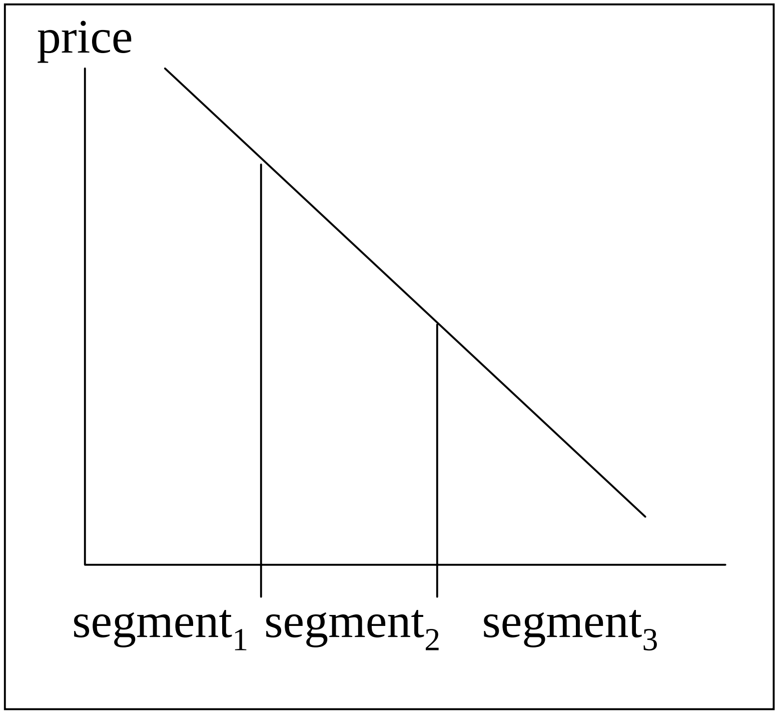


Economics: Principles (2/3)

- information goods
  - high fixed cost of production
  - zero marginal cost
  - zero distribution cost
  - eg. software

Economics: Principles (3/3)

- market segmentation
  - means to extract surplus
  - requires exclusion, product differentiation, stickiness (high switching cost)
  - eg. DSL speed



Economics: eCommerce

- transaction cost reduction
  - search (Yahoo)
  - intermediaries (Chemdex)
  - price discovery (Priceline)
  - market making (eBay)
- portals
  - collect site rents in form of ads, commissions

Economics: Resource allocation

- use of market mechanism for efficient allocation of resources
- equate marginal cost = marginal benefit = price (fixed cost problem)