Digital Communications I

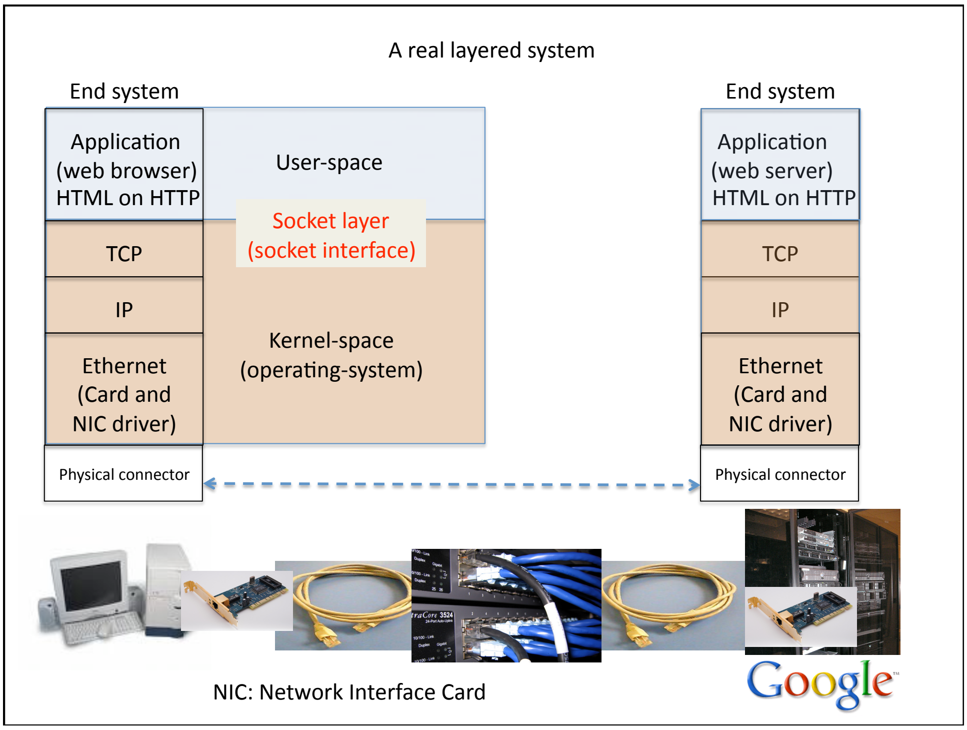
(Introduction to Digital Communications)

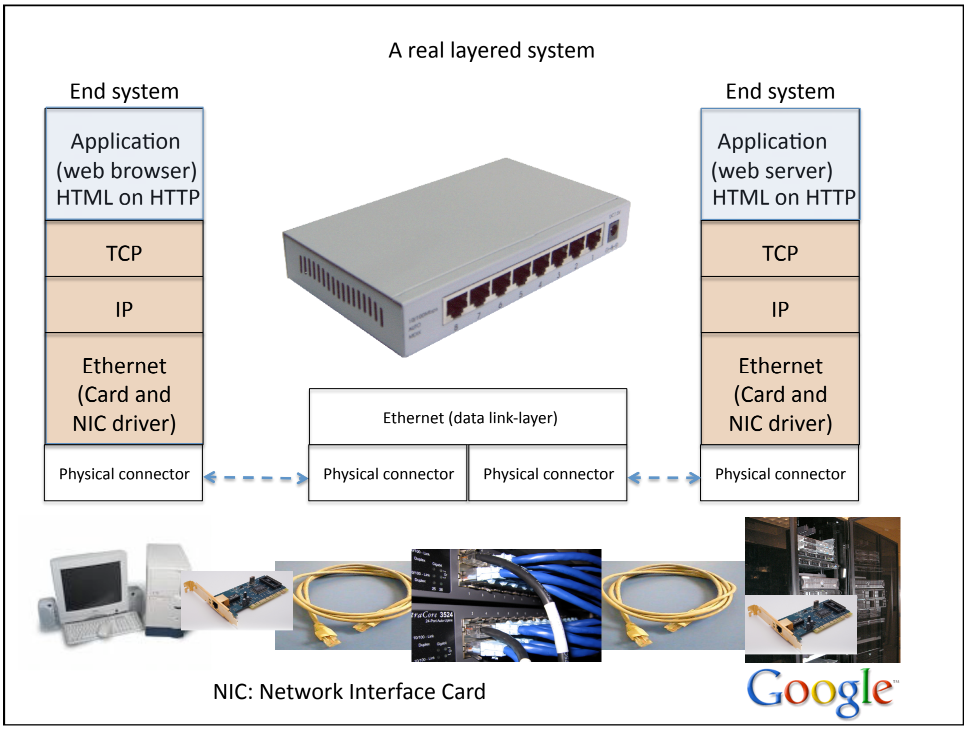
Lent Term — 2009

Handout 2

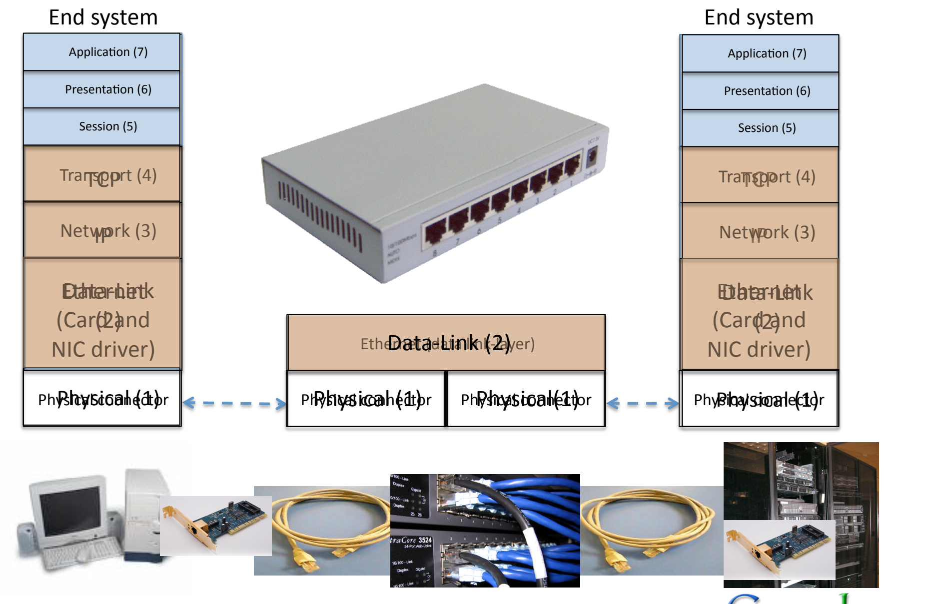
Andrew W Moore

andrew.moore@cl.cam.ac.uk


A real layered system
End system
Application
(web browser)
HTML on HTTP
TCP
IP
Ethernet
(Card and
NIC driver)
Physical connector
User-space
Socket layer
(socket interface)
Kernel-space
(operating-system)
End system
Application
(web server)
HTML on HTTP
TCP
IP
Ethernet
(Card and
NIC driver)
Physical connector
NIC: Network Interface Card
Google

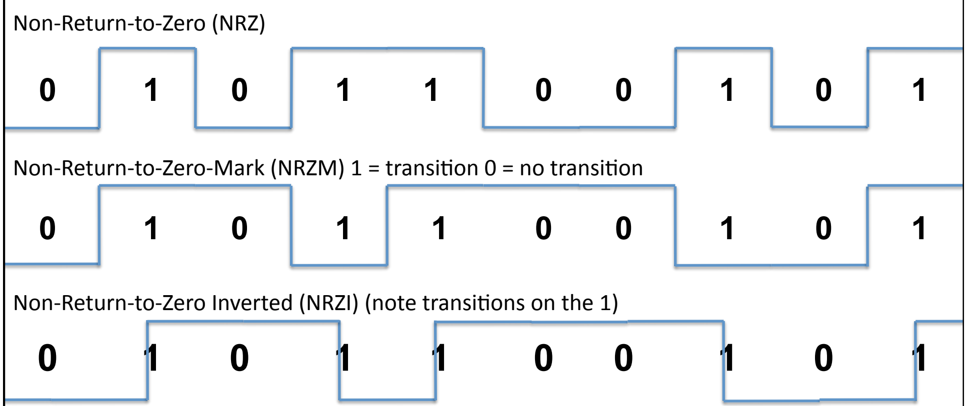

A real layered system
End system
Application
(web browser)
HTML on HTTP
TCP
IP
Ethernet
(Card and
NIC driver)
Physical connector
Ethernet (data link-layer)
Physical connector
Physical connector
End system
Application
(web server)
HTML on HTTP
TCP
IP
Ethernet
(Card and
NIC driver)
Physical connector
NIC: Network Interface Card
Google

A real layered system

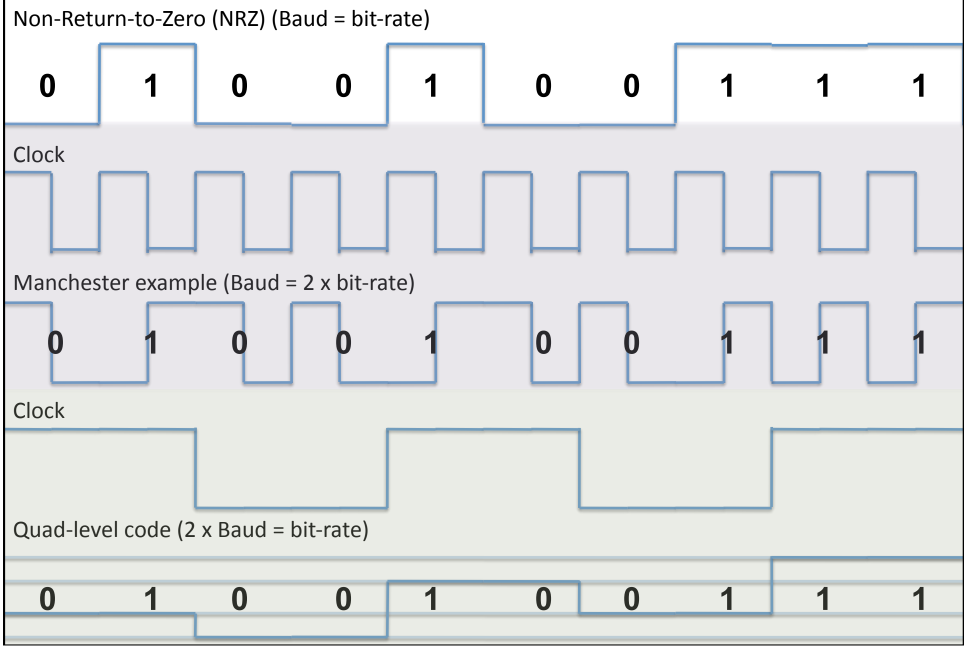


NIC: Network Interface Card

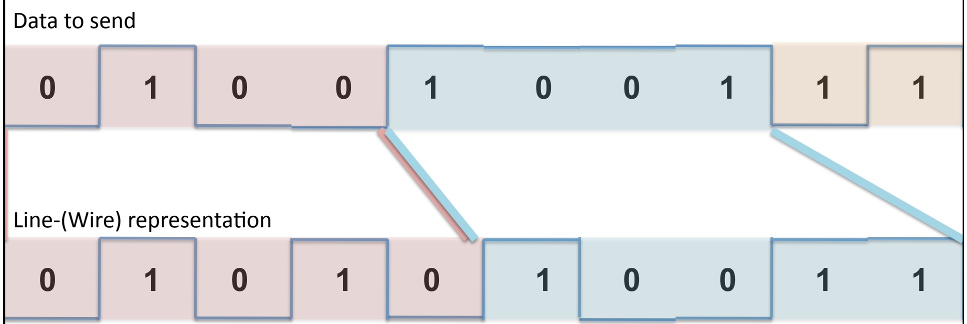
Line Coding Examples where Baud=bit-rate



Line Coding Examples - II



Line Coding Examples - III



Name	4b	5b	Description
0	0000	11110	hex data 0
1	0001	01001	hex data 1
2	0010	10100	hex data 2
3	0011	10101	hex data 3
4	0100	01010	hex data 4
5	0101	01011	hex data 5
6	0110	01110	hex data 6
7	0111	01111	hex data 7
8	1000	10010	hex data 8
9	1001	10011	hex data 9
A	1010	10110	hex data A
B	1011	10111	hex data B
C	1100	11010	hex data C
D	1101	11011	hex data D
E	1110	11100	hex data E
F	1111	11101	hex data F

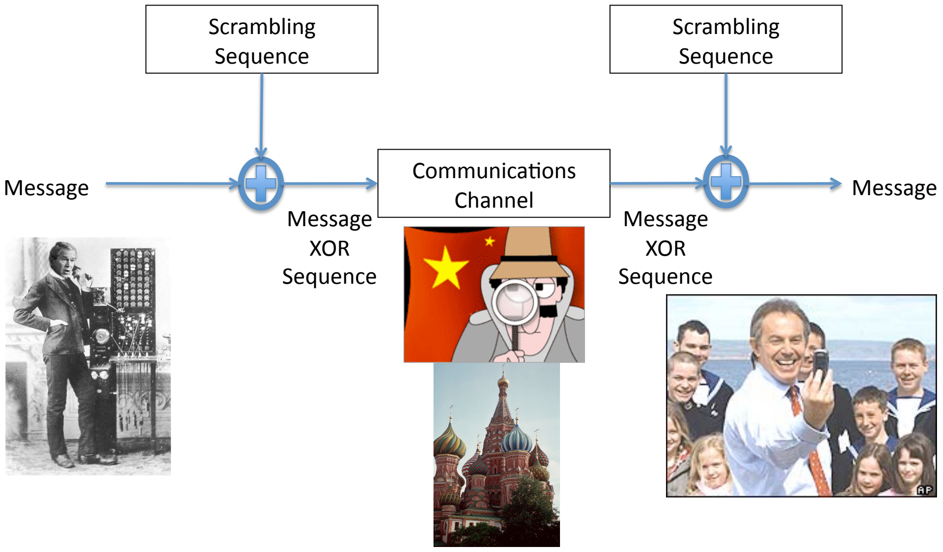
Name	4b	5b	Description
Q	-NONE-	00000	Quiet
I	-NONE-	11111	Idle
J	-NONE-	11000	SSD #1
K	-NONE-	10001	SSD #2
T	-NONE-	01101	ESD #1
R	-NONE-	00111	ESD #2
H	-NONE-	00100	Halt

Block coding transfers data with a fixed overhead: 20% less information per Baud in the case of 4B/5B

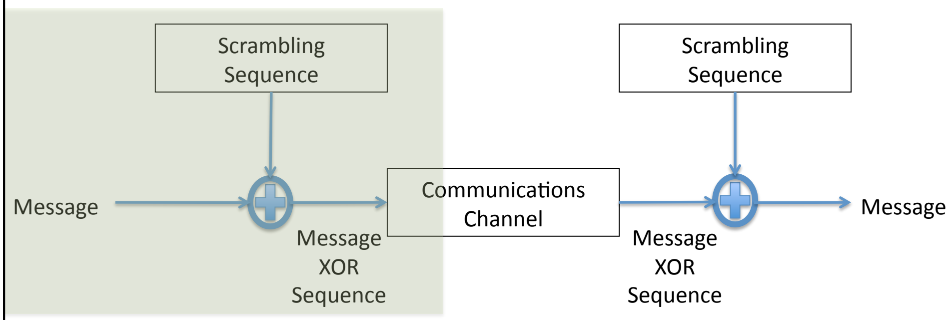
So to send data at 100Mbps; the line rate (the Baud rate) must be 125Mbps.

1Gbps uses an 8b/10b codec; encoding entire bytes at a time but with 25% overhead

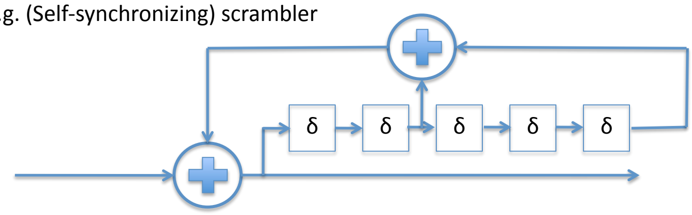
Line Coding Examples - IV



Line Coding Examples - IV



e.g. (Self-synchronizing) scrambler



Line Coding Examples – V (Hybrid)

…100111101101010001000101100111010001010010110101001001110101110100…

…10011110110101000101000101100111010001010010110101001001110101110100…

Inserted bits marking "start of frame/block/sequence"

Scramble / Transmit / Unscramble

…0100010110011101000101001011010100100111010111010010010111011101111000…

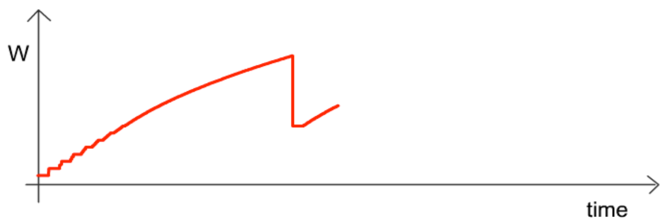
Identify (and remove) "start of frame/block/sequence"
This gives you the Byte-delineations for *free*

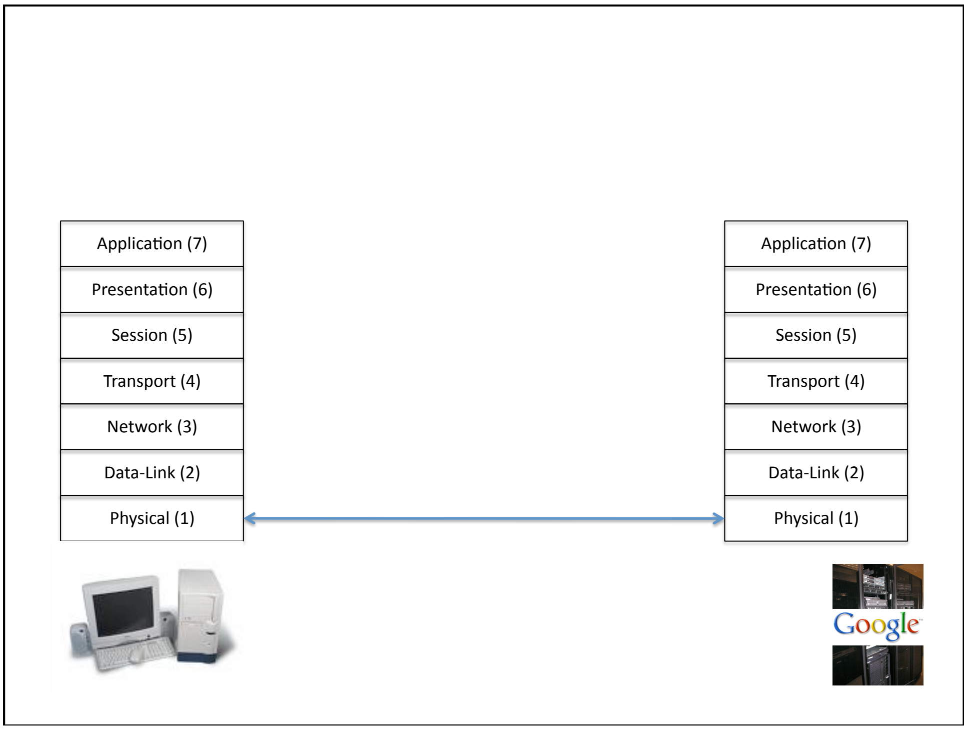
64b/66b combines a scrambler and a framer. The start of frame is a pair of bits 01 or 10: 01 means "this frame is data" 10 means "this frame contains data and control" – control could be configuration information, length of encoded data or simply "this line is idle" (no data at all)

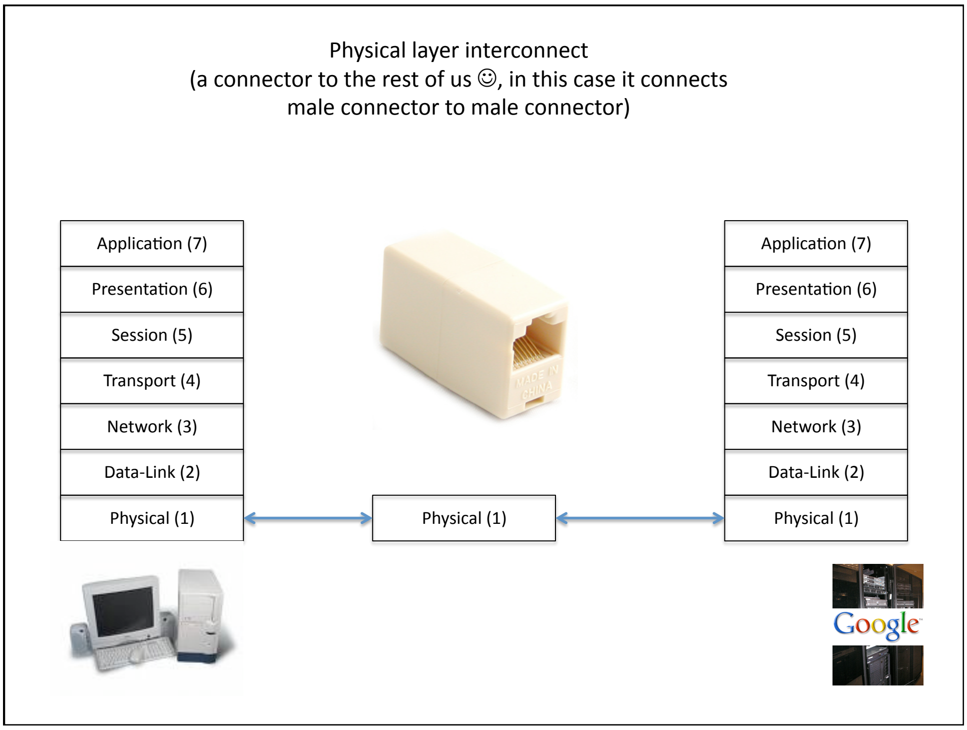
Single Flow ARQ in operation

Only W packets may be outstanding

Rule for adjusting W
- If an ACK is received: W ← W+1/W
- If a packet is lost: W ← W/2

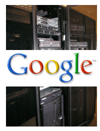



Application (7)
Presentation (6)
Session (5)
Transport (4)
Network (3)
Data-Link (2)
Physical (1)
Application (7)
Presentation (6)
Session (5)
Transport (4)
Network (3)
Data-Link (2)
Physical (1)
Google


Physical layer interconnect
(a connector to the rest of us ☺, in this case it connects
male connector to male connector)
Application (7)
Presentation (6)
Session (5)
Transport (4)
Network (3)
Data-Link (2)
Physical (1)
Physical (1)
Application (7)
Presentation (6)
Session (5)
Transport (4)
Network (3)
Data-Link (2)
Physical (1)
Google

Physical device – media convertor
speaks fluent Physical-layer
But only a bit of the Data-Link (Ethernet)

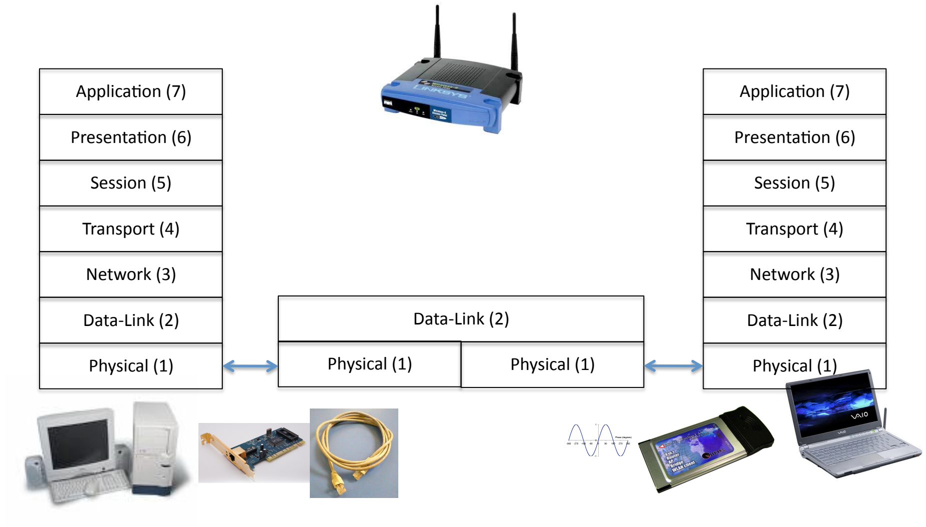
Application (7)			Application (7)
Presentation (6)			Presentation (6)
Session (5)			Session (5)
Transport (4)			Transport (4)
Network (3)			Network (3)
Data-Link (2)	Data-Link (2)		Data-Link (2)
Physical (1) ↔	Physical (1)	Physical (1)	↔ Physical (1)



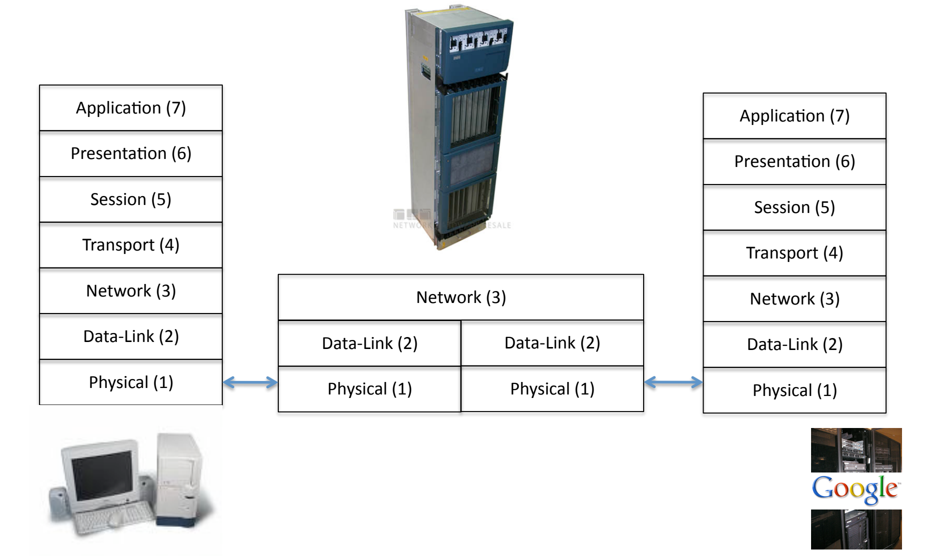
Data-Link device aka an Ethernet switch
speaks fluent Ethernet and often the extra/new
Ethernet-related standards (802.1q for VLAN and
802.1x for Network Access Control)

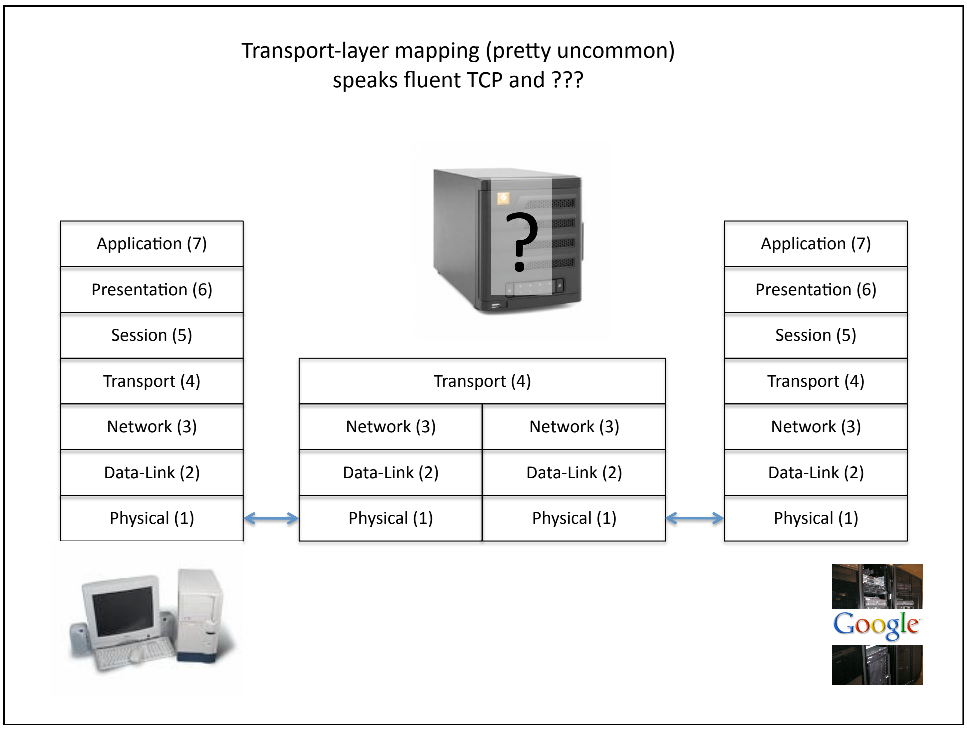
Application (7)			Application (7)
Presentation (6)			Presentation (6)
Session (5)			Session (5)
Transport (4)			Transport (4)
Network (3)			Network (3)
Data-Link (2)	Data-Link (2)		Data-Link (2)
Physical (1) ↔	Physical (1)	Physical (1)	↔ Physical (1)

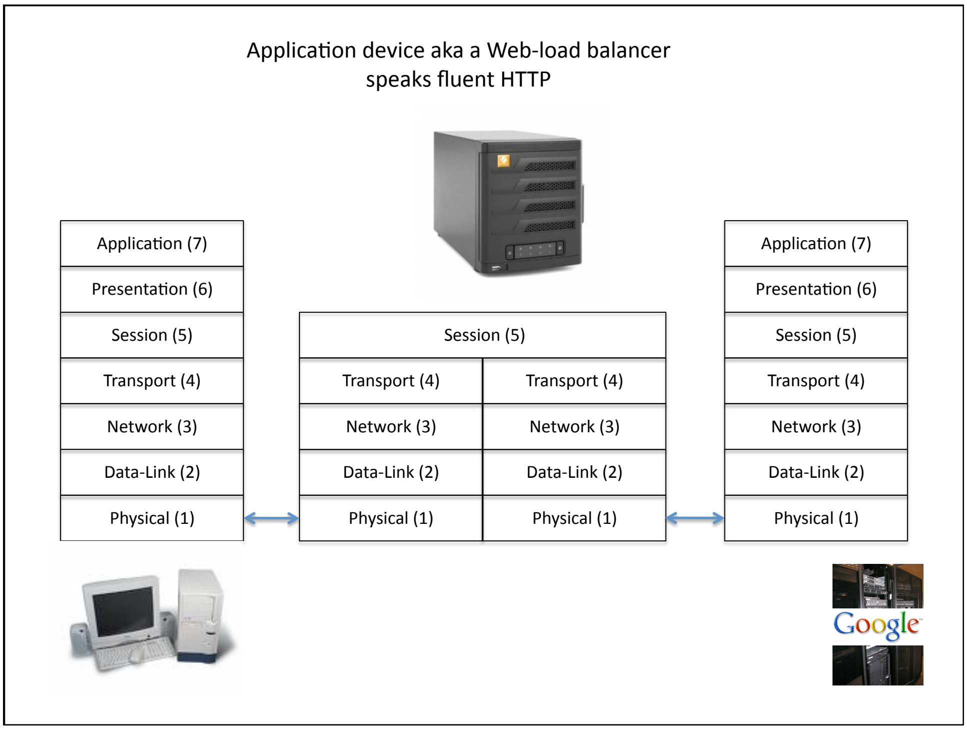
Data-Link device aka an Ethernet switch
speaks fluent Ethernet (two standards: wire and wireless!)
Another example with media conversion



Network device aka an IP Router
speaks fluent IP and can convert between any data-link
and any physical standard to any other.




Transport-layer mapping (pretty uncommon)
speaks fluent TCP and ???
?
Application (7)
Presentation (6)
Session (5)
Transport (4)
Network (3)
Data-Link (2)
Physical (1)
Transport (4)
Network (3)
Network (3)
Data-Link (2)
Data-Link (2)
Physical (1)
Physical (1)
Application (7)
Presentation (6)
Session (5)
Transport (4)
Network (3)
Data-Link (2)
Physical (1)
Google


Application device aka a Web-load balancer
speaks fluent HTTP
Application (7)
Presentation (6)
Session (5)
Transport (4)
Network (3)
Data-Link (2)
Physical (1)
Session (5)
Transport (4)
Transport (4)
Network (3)
Network (3)
Data-Link (2)
Data-Link (2)
Physical (1)
Physical (1)
Application (7)
Presentation (6)
Session (5)
Transport (4)
Network (3)
Data-Link (2)
Physical (1)
Google

ARP

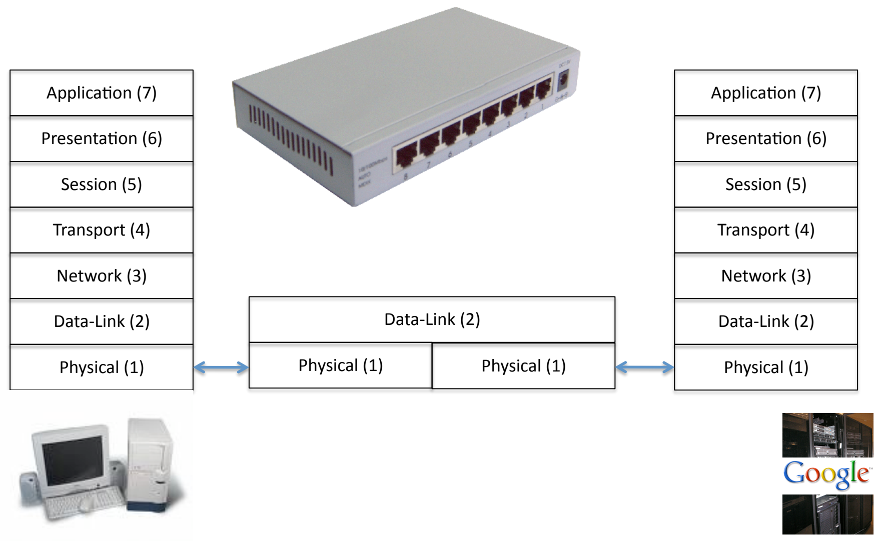
Recall from the primary slide-stack (slide 8-22) we need a mechanism for higher-layer entities to know what address to utilize in the lower entities.

ARP (address resolution protocol) provides a mechanism for establishing a host's link-layer (e.g. Ethernet) address using only a network-address (e.g. IP address)

ARP is not limited to either Ethernet or IP; although that is the most common current use.

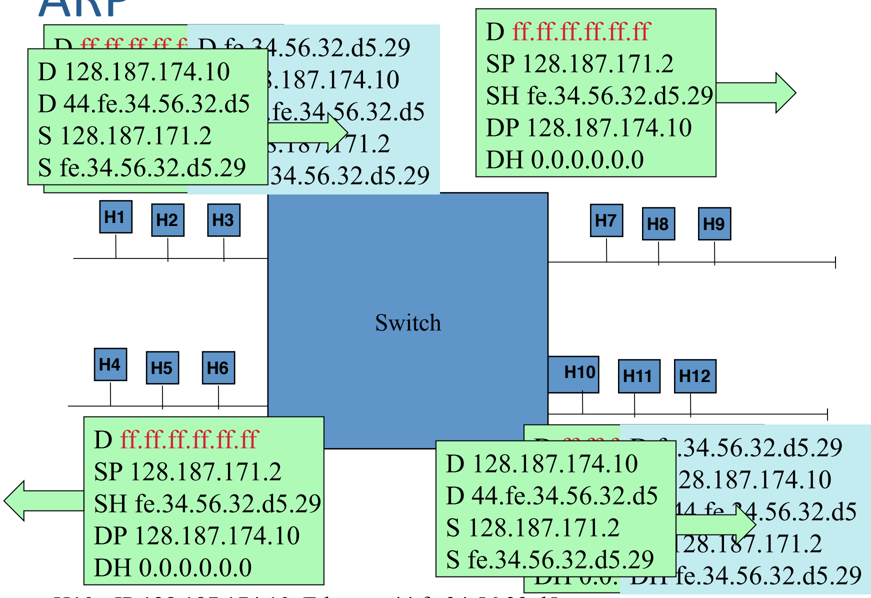
A linux machine's ARP table can be shown using the command `arp -a`

Consider the following situation:



An Ethernet (switch) example

ARP



H10= IP 128.187.174.10, Ethernet 44.fe.34.56.32.d5

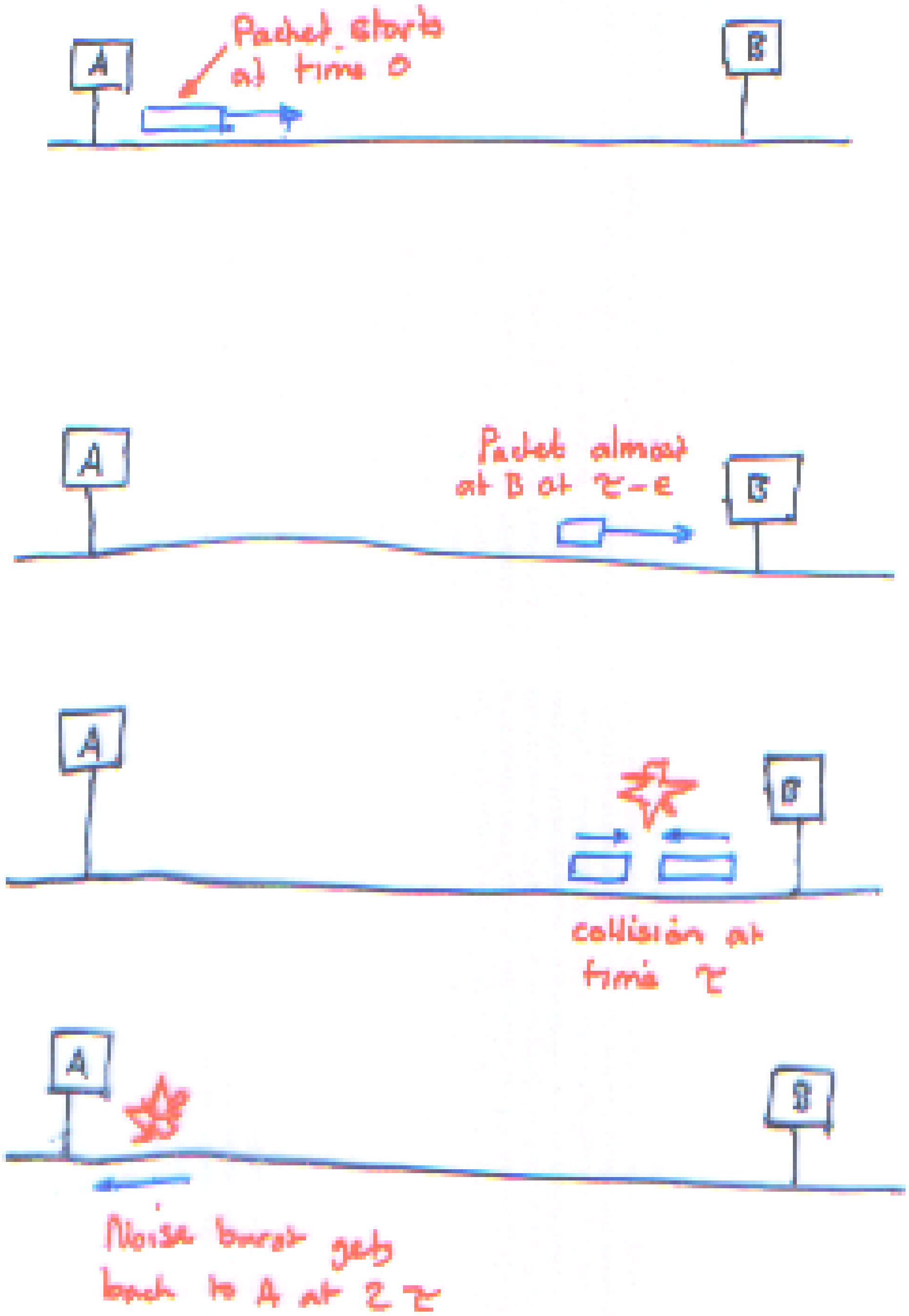
Notes

- ARP table entries timeout in about 10 minutes

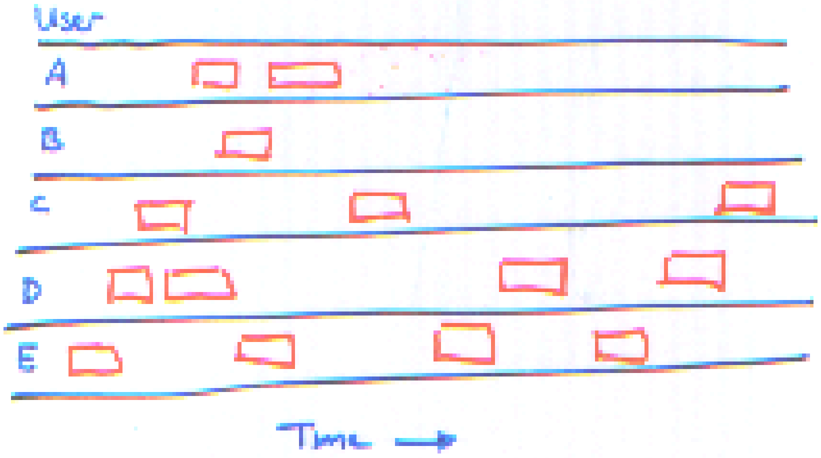
ARP table rules:

- update table with source when you are the target
- update table if you already have an entry
- do not refresh table entries upon reference

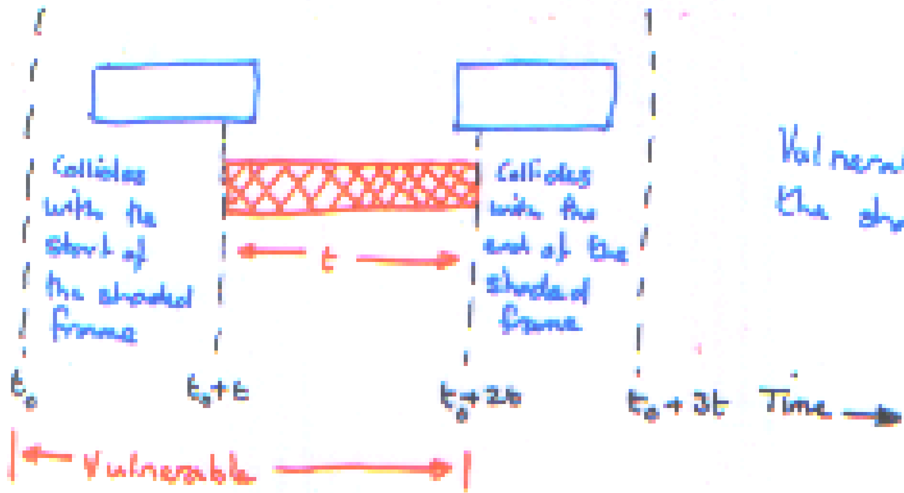
Collision detection can take as long as 2 τ



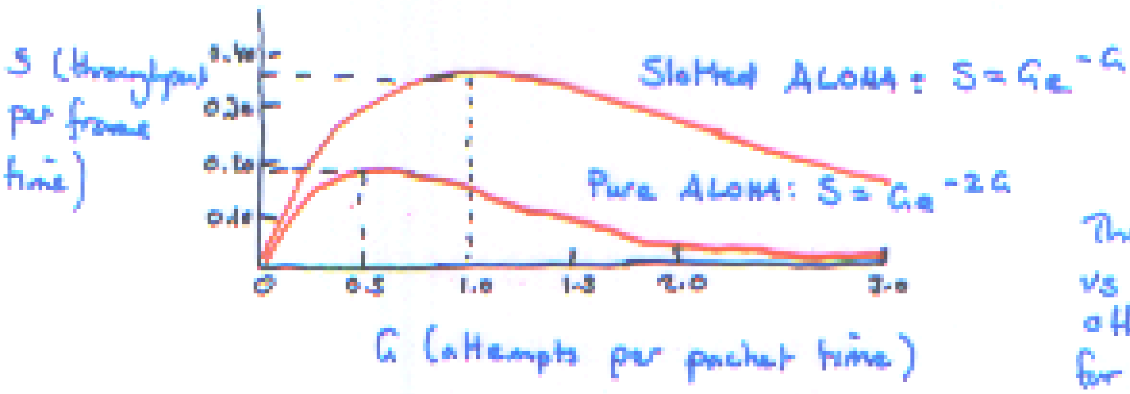
Aloha



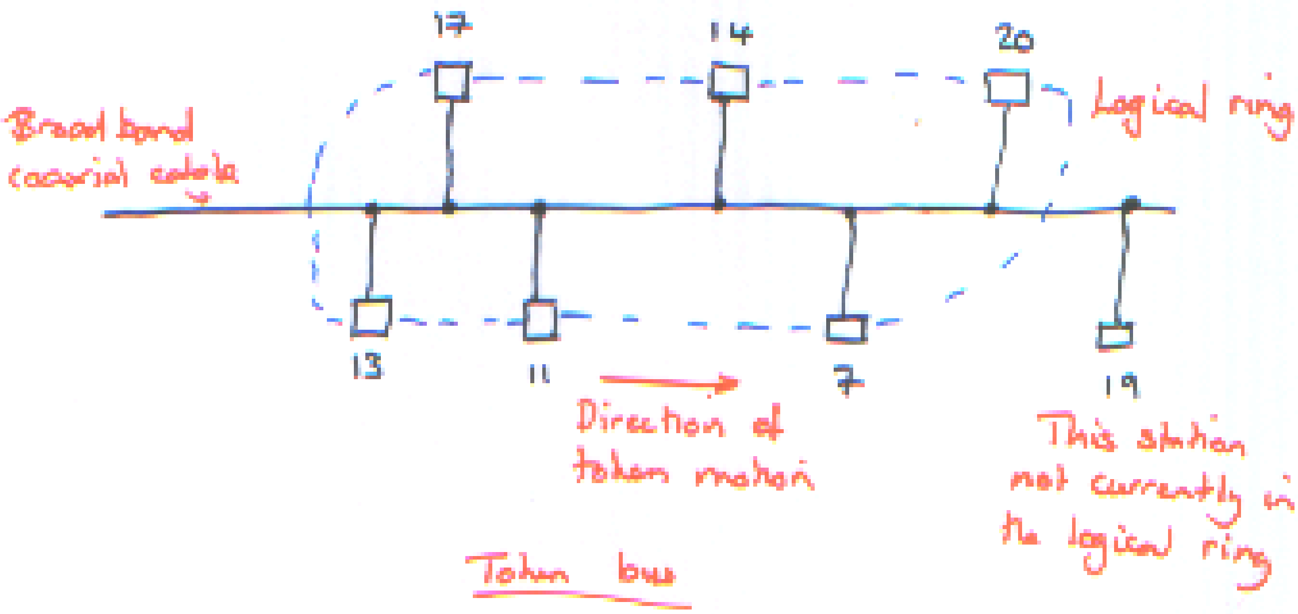
In pure ALOHA, frames are transmitted at completely arbitrary times.



Vulnerable period for the shaded frame.



Throughput vs offered load for ALOHA systems.



17
14
20
Logical ring
Broad band
(coaxial) cable
13
11
7
19
Direction of
token motion
This station
not currently in
the logical ring

Token bus

PCM codec.

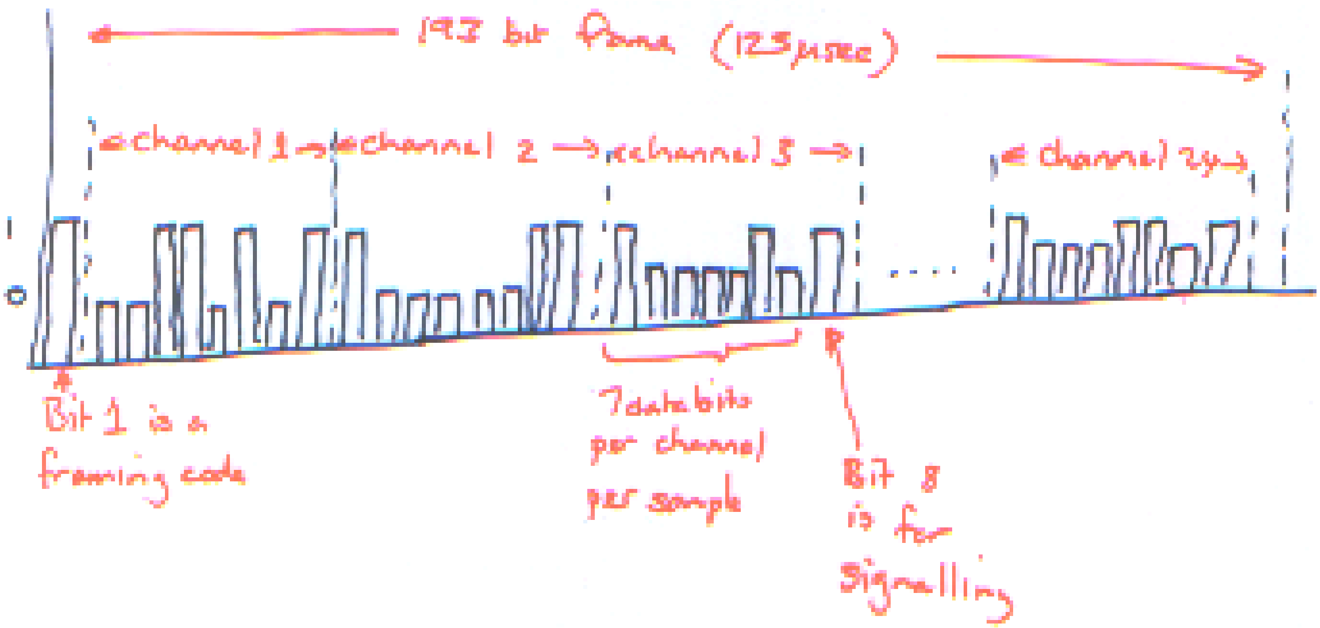
produces 7 or 8 bit number

8000 samples per second → 125 μsec / sample

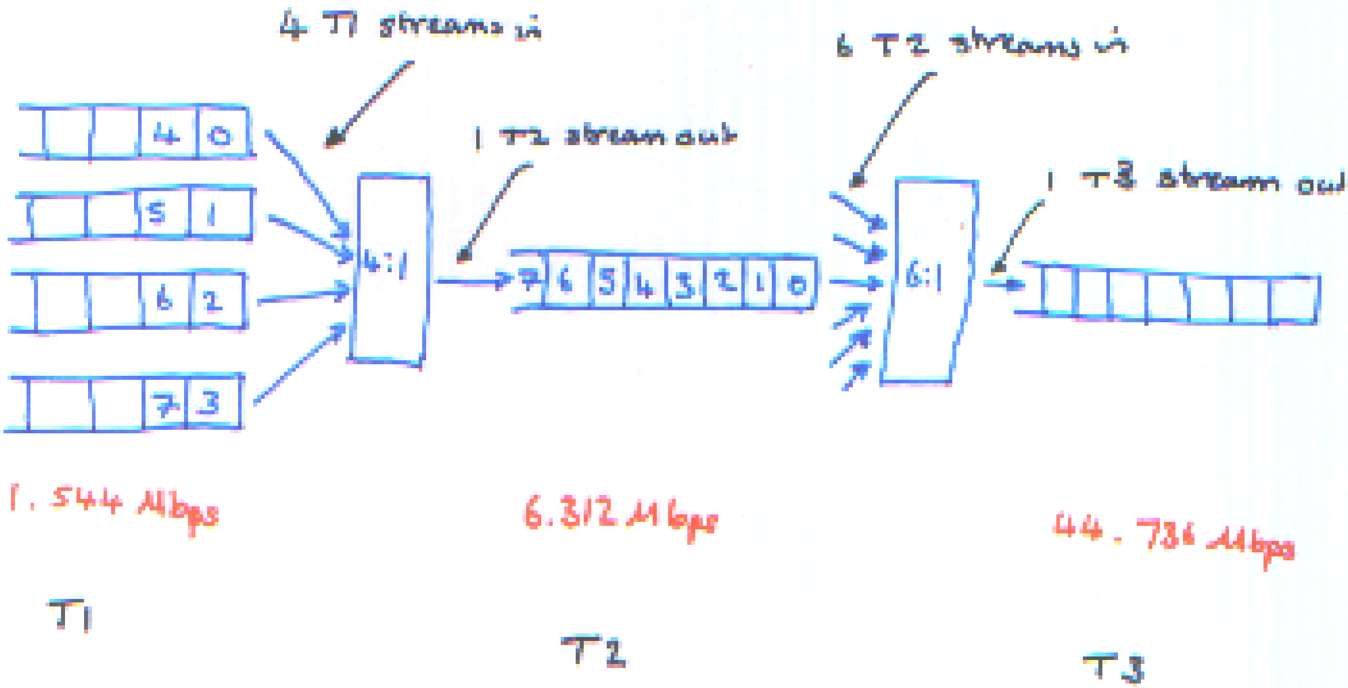
∵ Nyquist sampling rate for 4-kHz channel (telephone channel is 4 kHz)

→ virtually all channels in telephone system are multiples of 125 μsec.

e.g. T1 channel carries 24 voice channels.

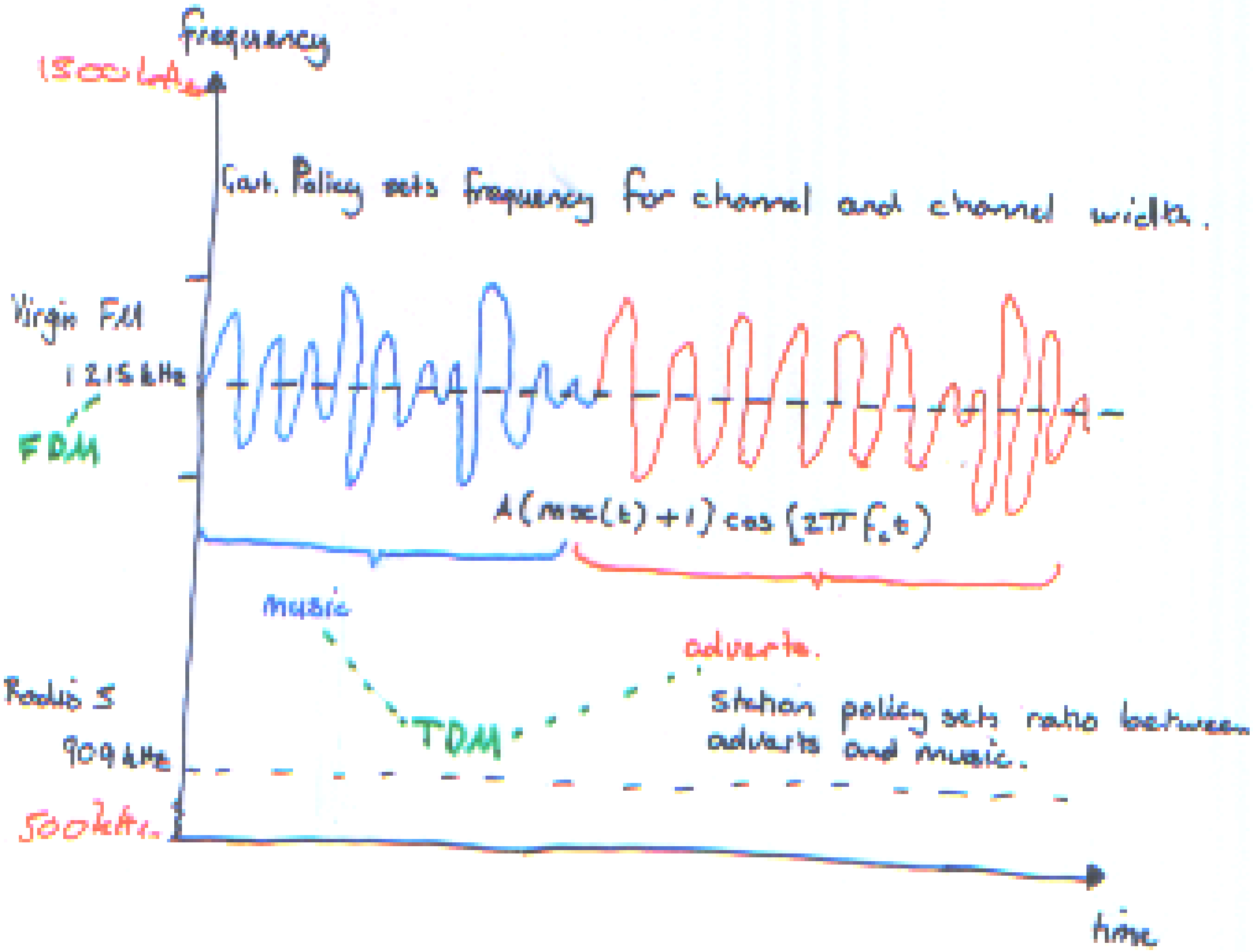


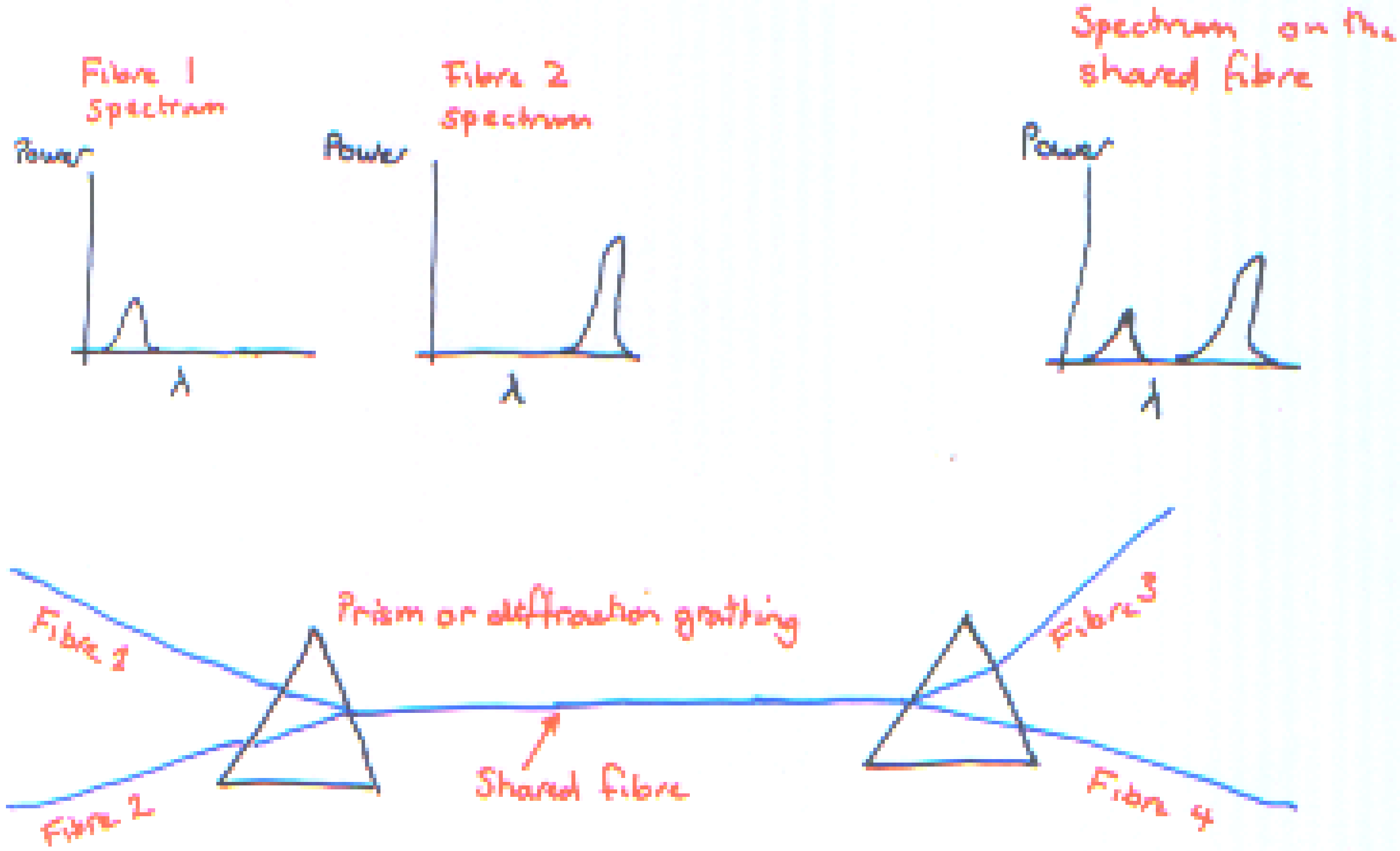
T1 carrier (1.544 Mbps)



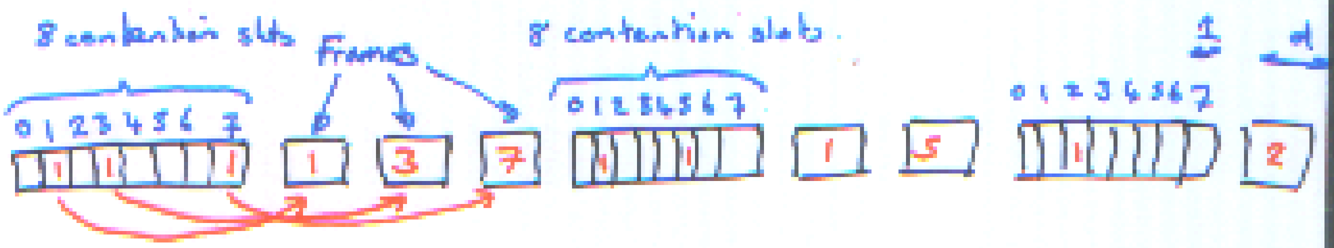
Multiplexing T1 streams onto higher carriers.

AM radio example.





Wavelength division multiplexing



The basic bit map protocol

Forwarding Algorithm.

```
D = destination IP address
for each forwarding table entry (subnet number, subnet mask, NextHop)
    D1 = SubnetMask & D
    if D1 = SubnetNumber
        if NextHop is an interface
            deliver datagram directly to destination
        else
            deliver datagram to NextHop (a router)
```

Link State Routing.

OSPF: Open Shortest Path First Protocol is a version of link state Routing.

In Link state Routing.

- □ Each node knows the state of the links to its neighbours and the cost of the link.
- □ Every node knows how to reach its neighbours. (those directly connected)
- □ Relies on two mechanisms
  - Reliable flooding
  - calculation of routes based on the sum of all accumulated link-state knowledge.

Reliable Flooding.

- □ Each node forwards its link state information to all neighbours, in the form of a "Link-state Packet" (LSP)
- □ Upon receiving an LSP packet a node forwards it on all its links.
- □ Process continues until info. reaches all nodes in the network.

LSP packet contains.

- ID of node that created the LSP.
- A list of directly connected neighbours of that node, with the link cost to each one.
- a sequence number
- a time to live (ttl)

Route Calculation

- Based on Dijkstra's algorithm.

Information Stored at Node	Distance to Reach Node						
	A	B	C	D	E	F	G
A	0	1	1	2	1	1	2
B	1	0	1	2	2	2	3
C	1	1	0	1	2	2	2
D	2	2	1	0	3	2	1
E	1	2	2	3	0	2	3
F	1	2	2	2	2	0	1
G	2	3	2	1	3	1	0

Final destinations stored at each node (global view)

Destination	Cost	NextHop
B	1	B
C	1	C
D	2	C
E	1	E
F	1	F
G	2	F

Final Routing Table at node A.

Information Stored at Node	Distance to Reach Node						
	A	B	C	D	E	F	G
A	0	1	1	∞	1	1	∞
B	1	0	1	∞	∞	∞	∞
C	1	1	0	1	∞	∞	∞
D	∞	∞	1	0	∞	∞	1
E	1	∞	∞	∞	0	∞	∞
F	1	∞	∞	∞	∞	0	1
G	∞	∞	∞	1	∞	1	0

Initial distances stored at each node (global view)

Destination	Cost	NextHop
B	1	B
C	1	C
D	∞	—
E	1	E
F	1	F
G	∞	—

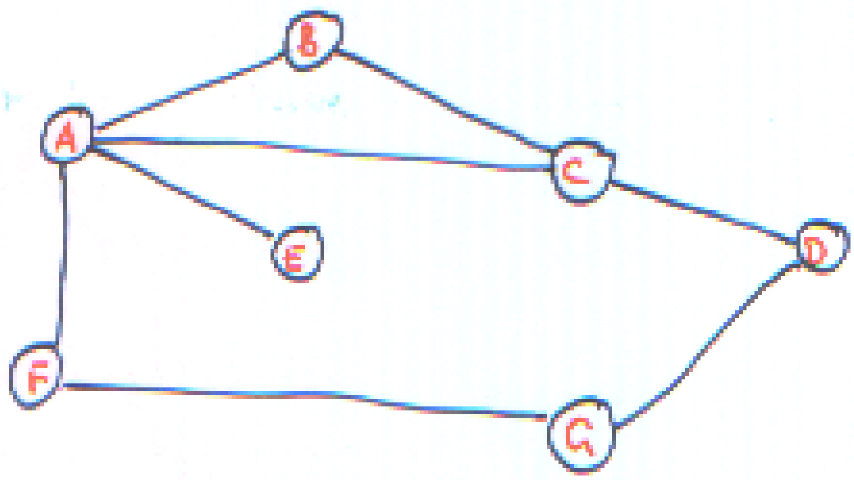
Init routing table at node A.

Distance Vector Routing. (also called Bellman-Ford)

RIP: Routing Information Protocol form of this.

- Each node constructs a 1-D vector contain the "distances" (costs) to all other nodes and it distributes the vector to all neighbours.
- Each node knows the cost of the links to all its directly connected neighbours at time t_0
- A link that is down has infinite cost.
- At each time step a node sends its vector to its neighbours

e.g.



Classless Routing [CIDR] pronounced "cider"

Solves two issues of scaling in the internet.

- ☐ Growth of the backbone routing table.
- ☐ Address space exhaustion due to address space assignment inefficiencies

e.g. A class C with 2 hosts has an assignment efficiency of

$2/255 = 0.78\%$

A class B with 256 hosts has an assignment efficiency of

$256/65535 = 0.39\%$

The rigid structure of the class based addressing model forced these issues on us.

Subnetting continued.

- Exactly one subnet mask per subnet
- subnet number is :-

host address AND subnet mask

e.g H1: 128.96.34.15 AND 255.255.255.128 →
128.96.34.0

- when a host wants to send a packet to an IP address it does a bitwise AND between its subnet mask and the destination IP. If the result is the same it knows it is on the local subnet then it can deliver locally otherwise it is sent to a router.

e.g. H1 → H2

@ H1 : 255.255.255.128 AND 128.96.34.139 →
(subnet mask for H1) (H2's address)

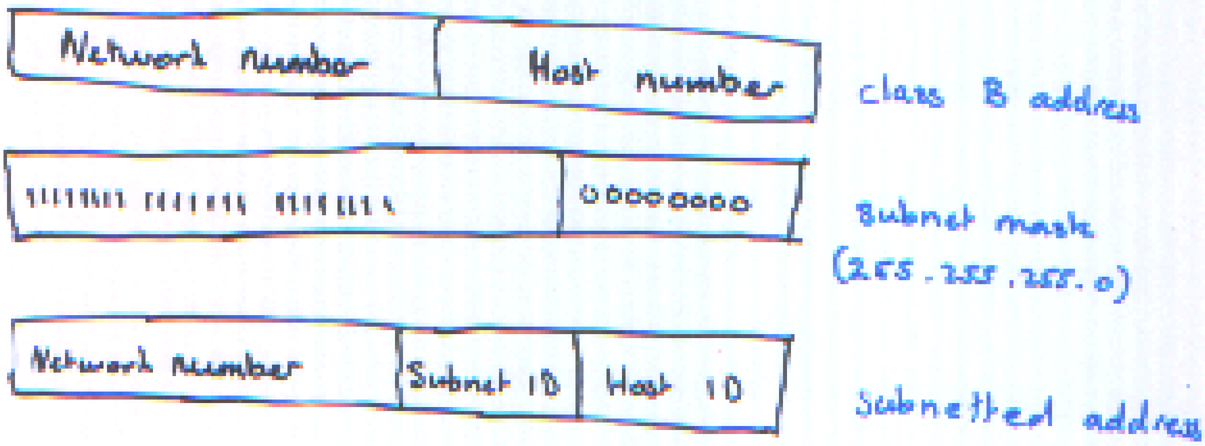
128.96.34.128
not H1's subnet number which is 128.96.34.0

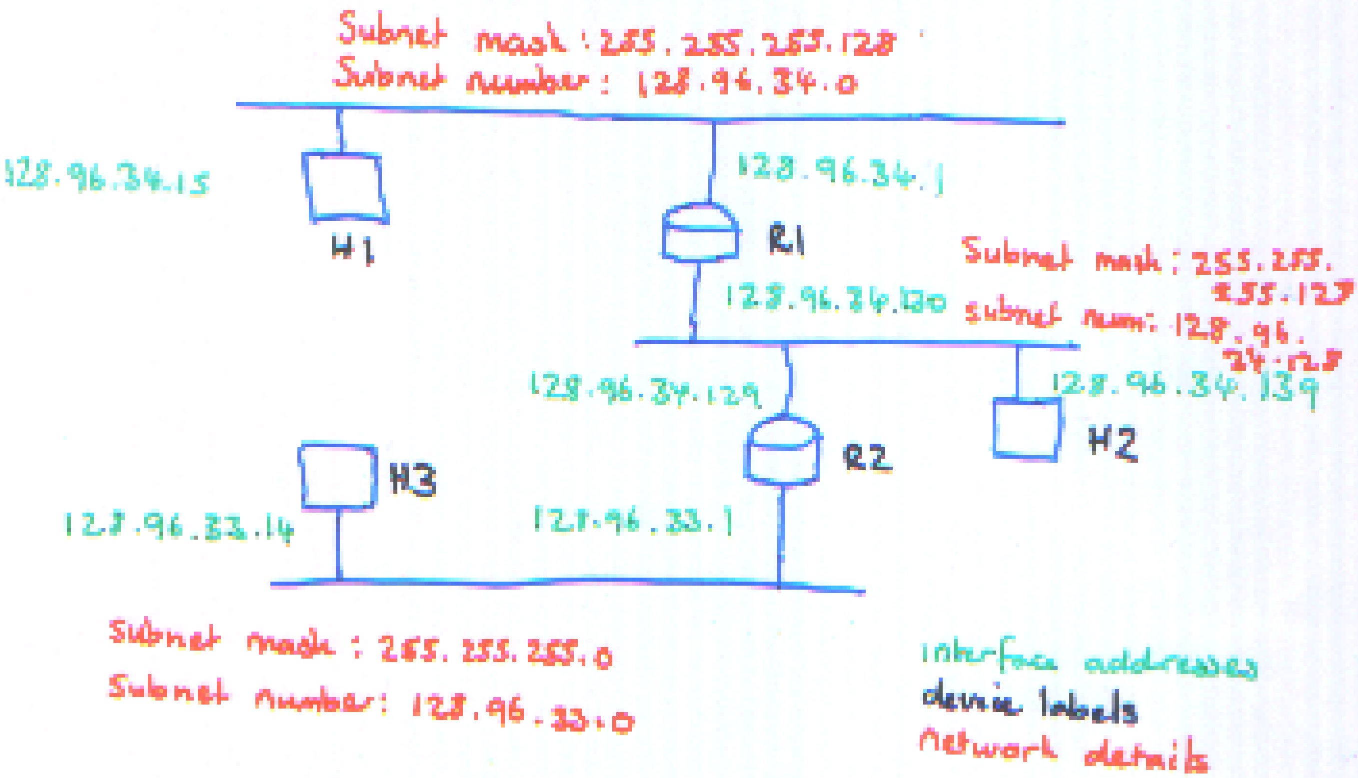
H1 can't deliver locally so sends its packet to its default router, R1. Packet arrives at R1.

@ R1 has the following Routing Table.

	Subnet number	Subnet mask	Nexthop
	128.96.34.0	255.255.255.128	Interface 0
	128.96.34.128	255.255.255.128	Interface 1
	128.96.33.0	255.255.255.0	R2
(default route)	0.0.0.0	0.0.0.0	R0 (not shown)

subnetting example. pages 301 to 305 of Computer Networks, Peterson + Davie





Subnetting cont.

R1 compares the bitwise AND of the packets destination AND each rows subnet mask with the subnet number for that row, if they match the router has found the destination subnet for the packet.
If no subnet is found the packet is dropped or sent on to the default router

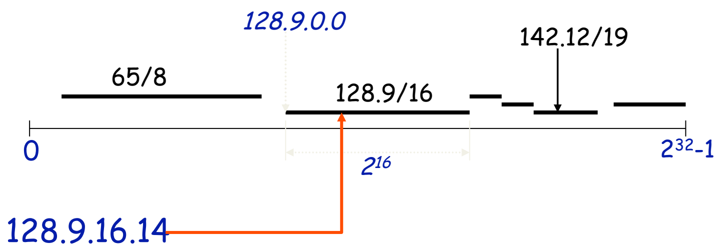
e.g.

dest	subnet mask	dest + subnet mask	subnet num	match
128.96.34.139	255.255.255.128	128.96.34.128	128.96.34.0	✗
128.96.34.139	255.255.255.128	128.96.34.128	128.96.34.128	✓

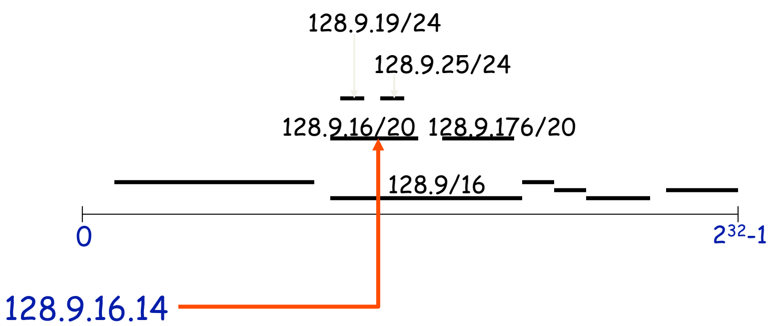
Here R1 would match its second table entry with the destination and so send the packet on interface 1.

CIDR and Longest Prefix Matches

- The IP address space is broken into line segments.
- Each line segment is described by a *prefix*.
- A prefix is of the form x/y where x indicates the prefix of all addresses in the line segment, and y indicates the length of the segment.
- e.g. The prefix 128.9/16 represents the line segment containing addresses in the range: 128.9.0.0 … 128.9.255.255.

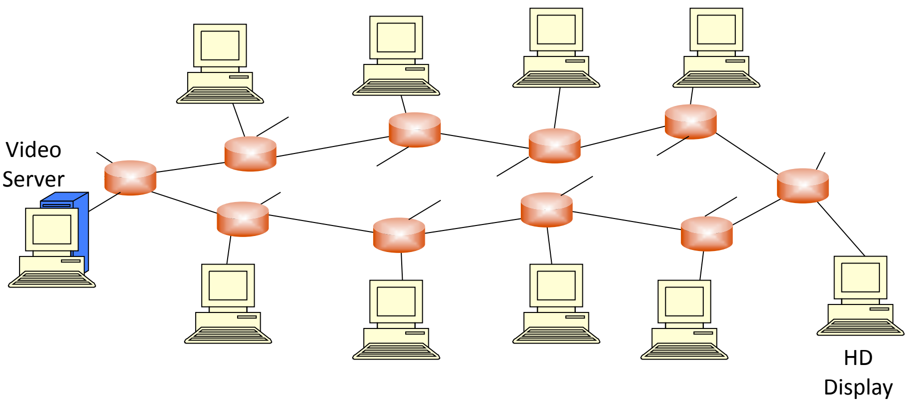


Classless Interdomain Routing (CIDR)

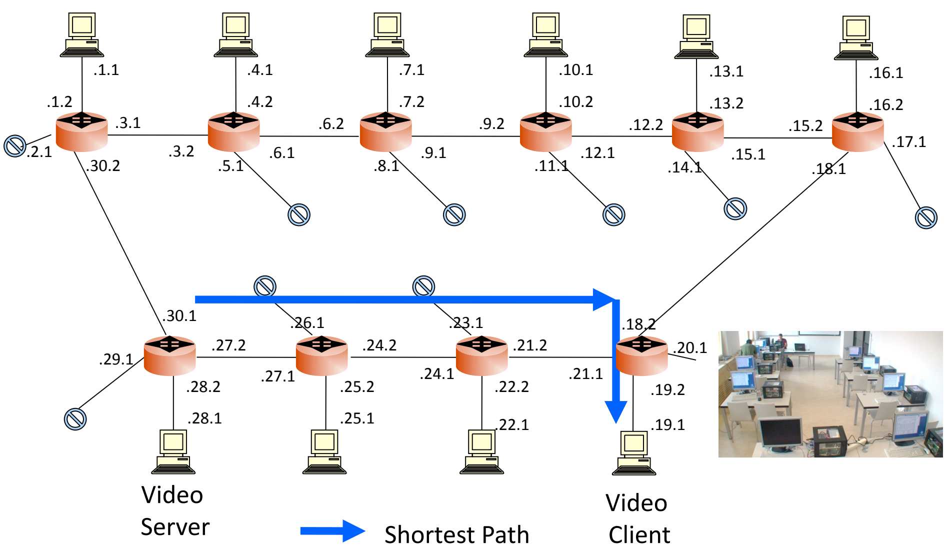


Most specific route = "longest matching prefix"

Topology of Routers



Subnet Configuration

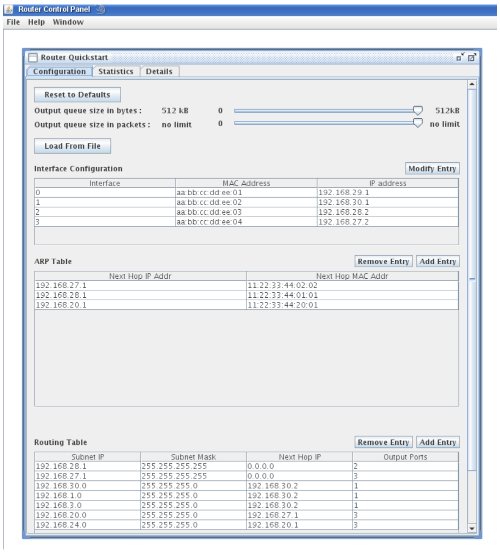


Step 1 – Observe the Routing Tables

The router is already configured and running on your machines

The routing table has converged to the routing decisions with minimum number of hops

Next, break a link …



Interface	MAC Address	IP address
0	aa:bb:cc:dd:ee:01	192.168.29.1
1	aa:bb:cc:dd:ee:02	192.168.30.1
2	aa:bb:cc:dd:ee:03	192.168.28.2
3	aa:bb:cc:dd:ee:04	192.168.27.2

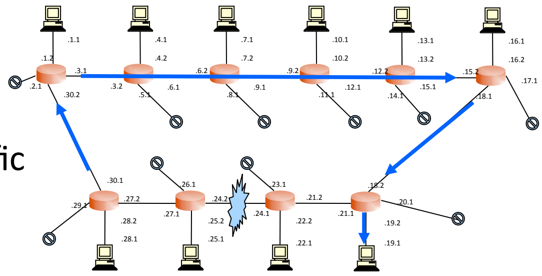
Next Hop IP Addr	Next Hop MAC Addr
192.168.27.1	11:22:33:44:02:02
192.168.28.1	11:22:33:44:01:01
192.168.20.1	11:22:33:44:20:01

Subnet IP	Subnet Mask	Next Hop IP	Output Ports
192.168.28.1	255.255.255.255	0.0.0.0	2
192.168.27.1	255.255.255.255	0.0.0.0	3
192.168.30.0	255.255.255.0	192.168.30.2	1
192.168.1.0	255.255.255.0	192.168.30.2	1
192.168.3.0	255.255.255.0	192.168.30.2	1
192.168.20.0	255.255.255.0	192.168.27.1	3
192.168.24.0	255.255.255.0	192.168.20.1	3

Step 2 - Dynamic Re-routing

Break the link between video server and video client

Routers re-route traffic around the broken link and video continues playing



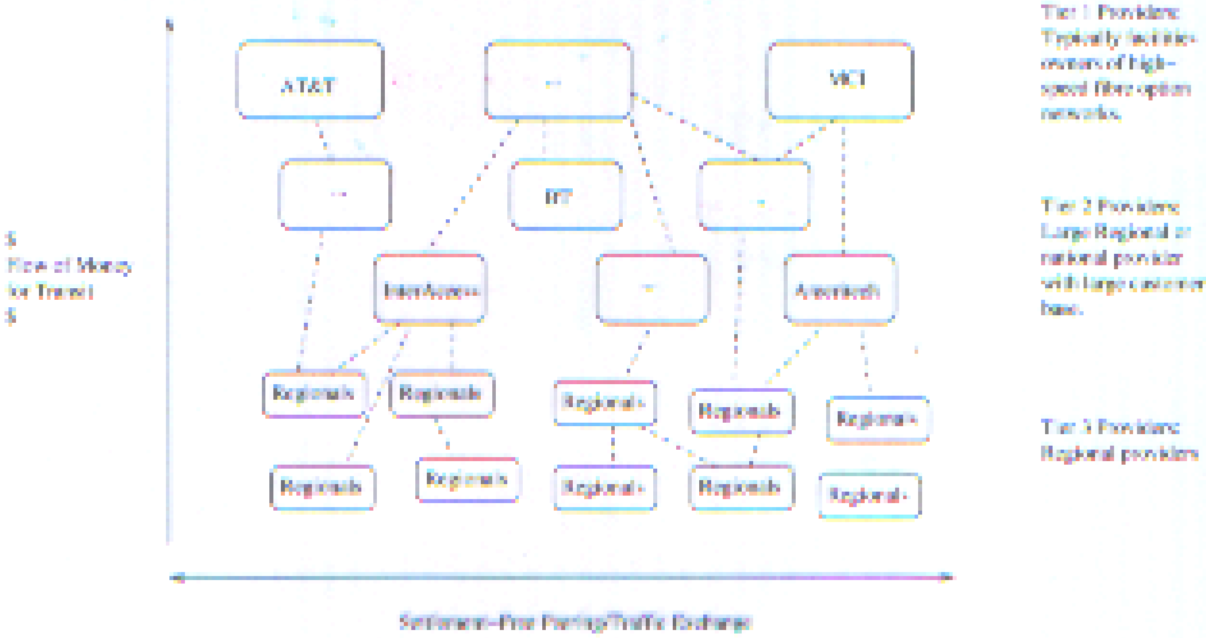
Routing Table Remove Entry Add Entry

Subnet IP	Subnet Mask	Next Hop IP	Output Ports
192.168.28.1	255.255.255.255	0.0.0.0	2
192.168.27.1	255.255.255.255	0.0.0.0	3
192.168.24.2	255.255.255.255	192.168.20.1	3
192.168.24.1	255.255.255.255	192.168.30.2	1
192.168.30.0	255.255.255.0	192.168.30.2	1
192.168.1.0	255.255.255.0	192.168.30.2	1
192.168.3.0	255.255.255.0	192.168.30.2	1
192.168.20.0	255.255.255.0	192.168.30.2	1
192.168.25.0	255.255.255.0	192.168.20.1	3

Working IP Router

- Observe PW-OSPF re-routing traffic around a failure
- Video is temporarily disrupted then resumes playing as TCP recovers from packet-loss

Structure of the internet



Not all ISPs are equal.

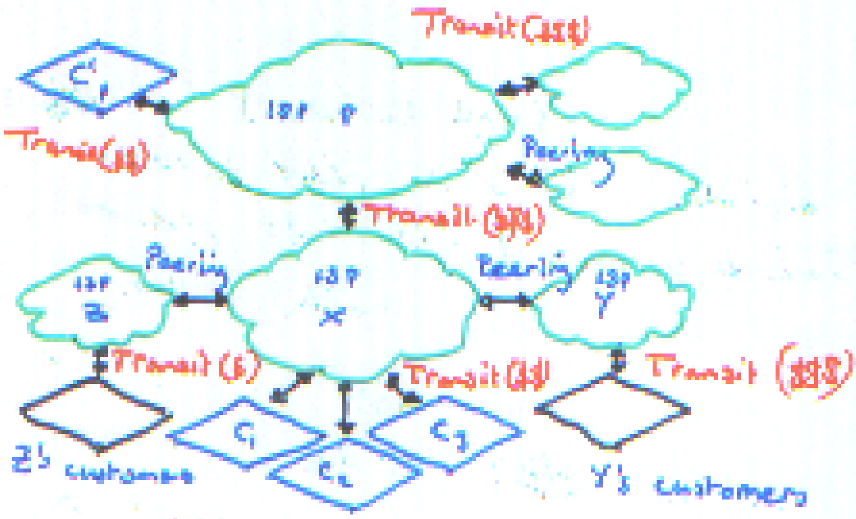
- Tier 1 ISP   AT&T, MCI
  - Really huge! Only a handful worldwide.
  - Global scope
    - Routing tables contain all currently reachable addresses.
    - Global PoPs (Points of presence.)
- Tier 2 ISP   BT, France Telecom, Pipex
  - Regional scope (e.g. a US state, or non US country)
- Tier 3 ISP   Plusnet, ~~[illegible]~~ Bulldog.
  - Local ISP

* Optimal Routing is often at odds with optimal economics. *

- Moving packets costs money!
- Why would I move packets for someone who isn't paying me to do so?
- But need to interconnect to provide connectivity.

Who should pay?

Policy based routing (example)



Inter-AS relationships; transit and peering.

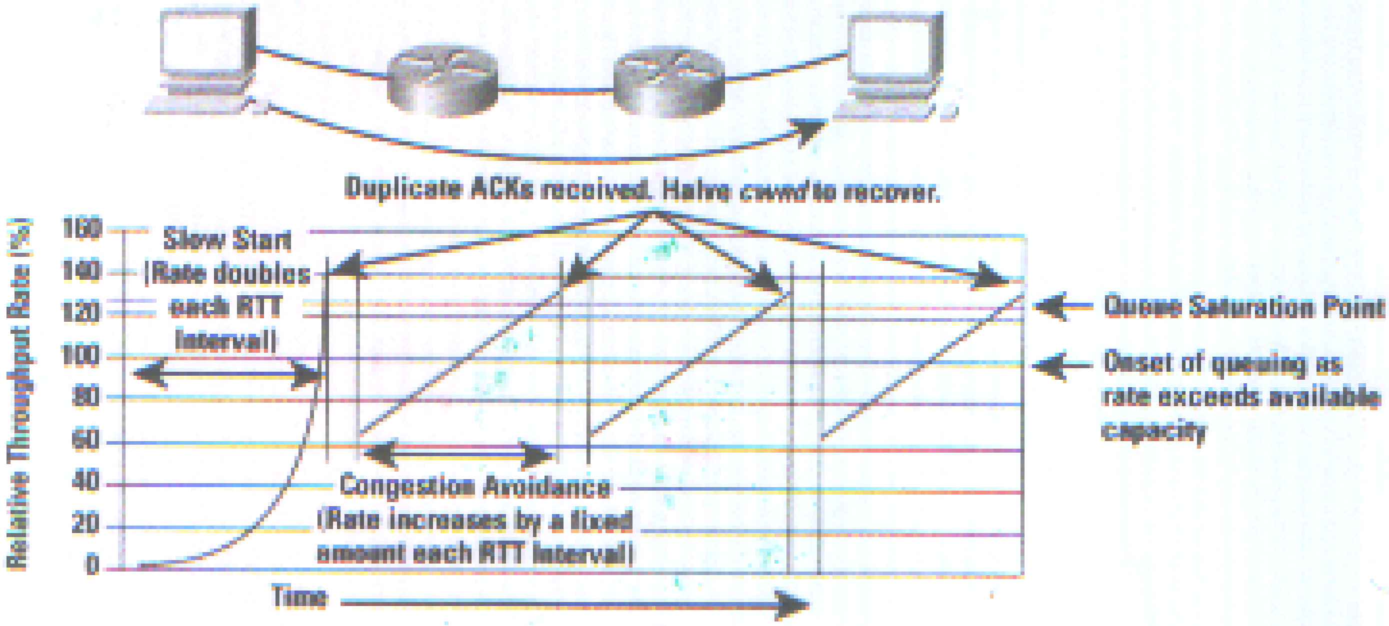
- C'p is a customer of P, who advertises C'p route to P's neighbours.
- X does not want to be a path to C'p unless they are being paid to provide connectivity.
  - X advertises C'p on to its customers i.e. those buying transit. C_1, C_2, C_3
  - X does not advertise C'p on its peering links i.e. to ISP Z or Y.

X does not make any money moving packets for Y's customers so it doesn't provide transit for them.

BGP: Border Gateway Protocol.

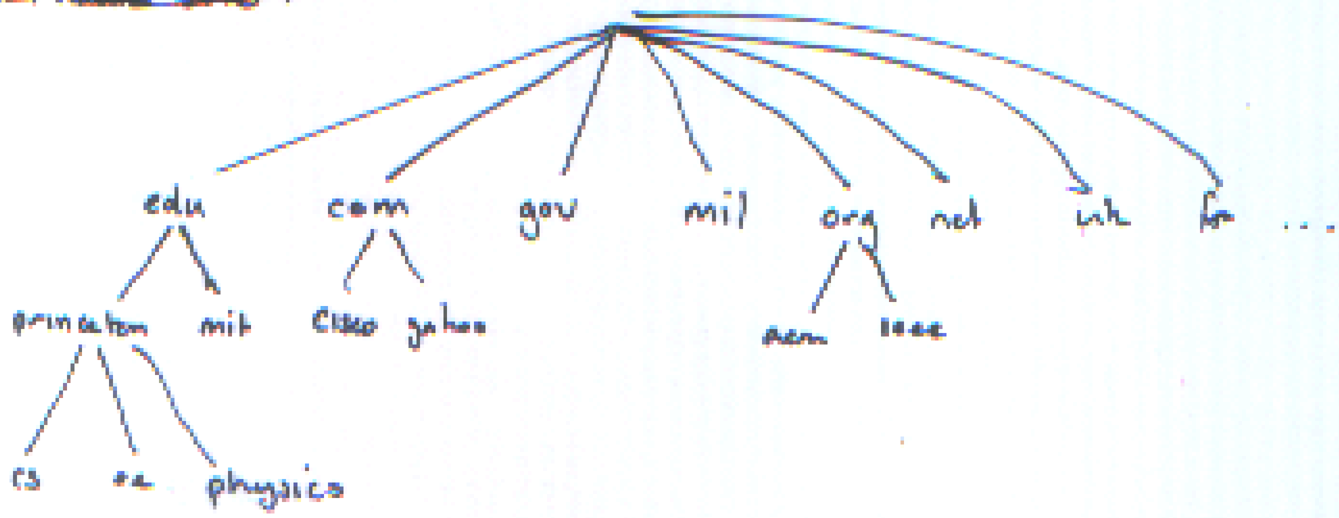
- It is the only inter-domain
- Two forms :-
  - eBGP (external bgp)
    external routing where bgp exchanges routing information between autonomous system (ASs)
  - iBGP (internal bgp)
    used to re-advertise paths internally to a domain.
- BGP is a complex protocol whose key features are
  - scaling
  - policy
  - wide spread deployment.

See text book for details.

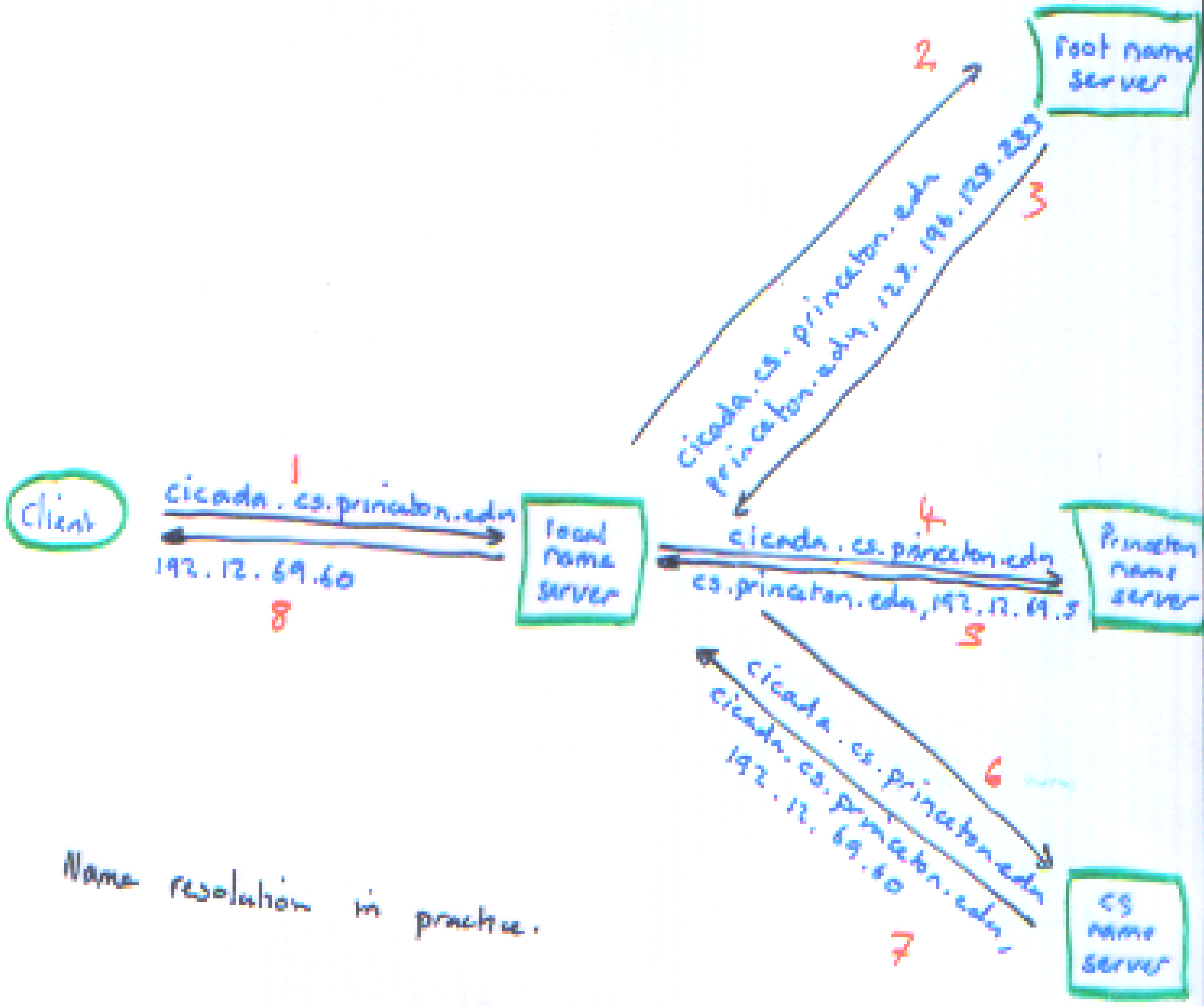


The Internet Protocol Journal

Vol 9, Number 2



Example name server hierarchy.



Name resolution in practice.

fig 9.5, p642, Comp. Networks, a systems approach
Peterson + Davie

34,677,374 : How do I represent it?

Big-Endian form: the most significant bit of a word is in the byte with the lowest address.

e.g. Motorola 680x0

(2)	(17)	(34)	(126)
00000010	00010001	00100010	01111110

Little-Endian form: the most significant bit is in the byte with the highest address.

e.g. Intel 80x86

(126)	(34)	(17)	(2)
01111110	00100010	00010001	00000010

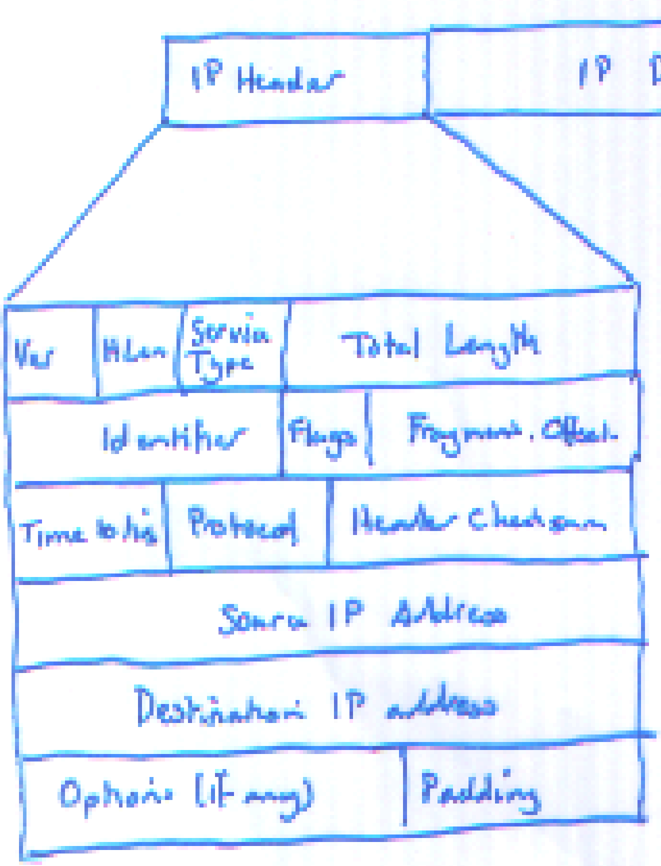
IP networks use Big-Endian encoding as defined in the IP protocol spec.

see http://en.wikipedia.org/Endianness

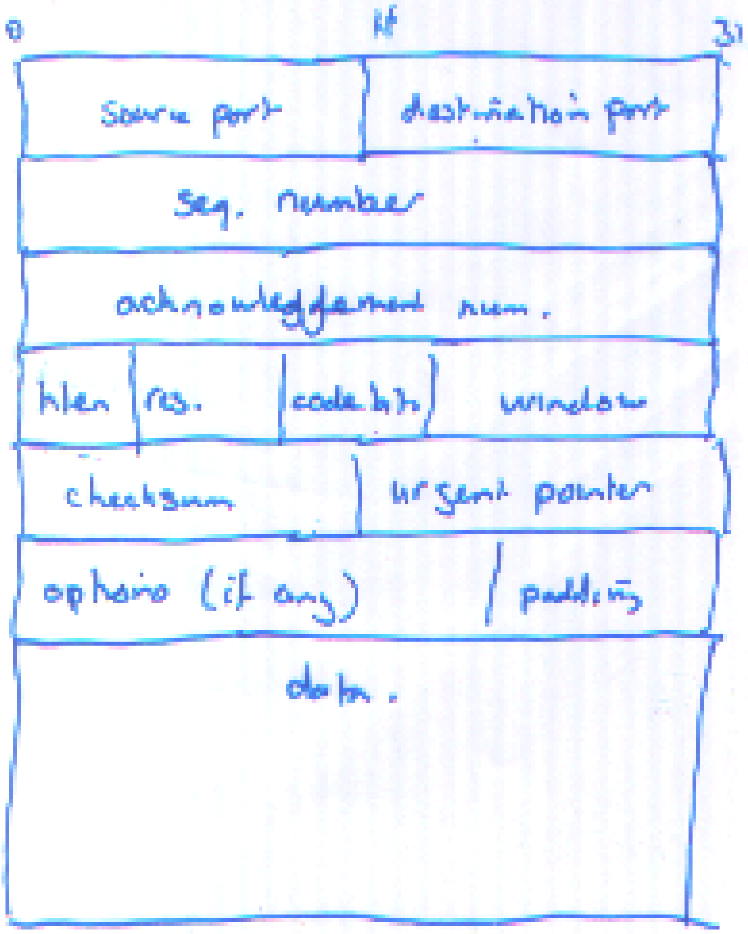
ping google.com ~ 20ms

example from Computer Networks, A Systems Approach, 3rd Ed. Peterson + Davie.

IP DATAGRAM

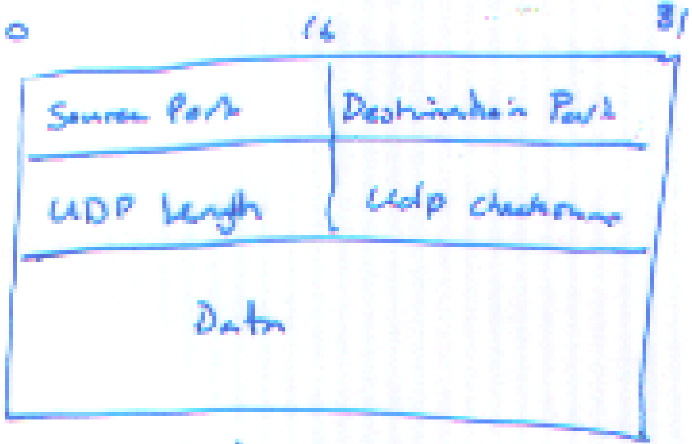


Slide 28-5



TCP header

slide 28-R



UDP header.

IP multicast : "a method of forwarding IP datagrams to a group of interested receivers"

"A multicast address is associated with a group of interested receivers. According to RFC 3171, addresses 224.0.0.0 to [redacted] 239.255.255.255 are designated as multicast addresses."

http://en.wikipedia.org/wiki/IP_multicast.

uses : media streaming (1:n)
file distribution
stock tickers.

Common IP port numbers from /etc/services.

ftp-data	20 /tcp
ftp	21 /tcp
telnet	23 /tcp
smtp	25 /tcp
whois	43 /tcp
domain	53 /tcp
domain	53 /udp
www	80 /tcp
www	80 /udp
pop3	110 /tcp
pop3	110 /udp