

MULTIACCESS COMMUNICATION

Network Layer

medieval

Comp

F

1

2

3

LANS

Two types of networks:

John
①

- **Switched:** interconnection by means of transmission

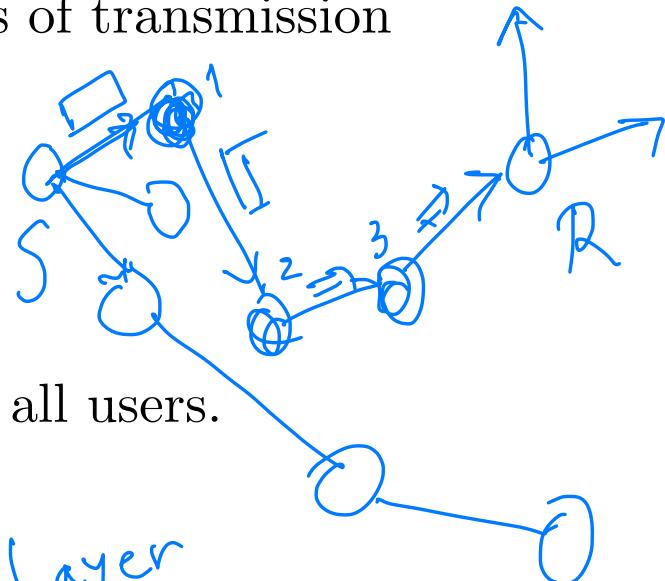
- lines, multiplexers, switches.
- Addressing scheme hierarchical.
- Routing tables are required.

②

- **Broadcast:** information received by all users.

- No routing is necessary.
- Addressing scheme is flat.
- Medium Access Control is required to orchestrate transmissions.

S ① ② ③ R



Because of its simplicity, broadcast networks are the preferred LAN technology.



john@scs.carleton.ca

ca → carleton



. scs



john

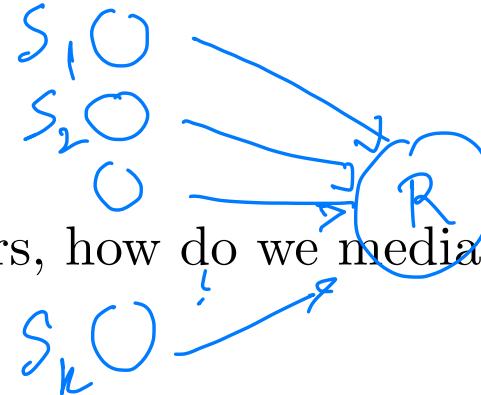
Mediating Access



- In point-to-point networks received signal is a function¹ of single transmitted signal



- In broadcast networks a single transmission medium is shared. Received signal is a function of possibly more than one transmitted signal



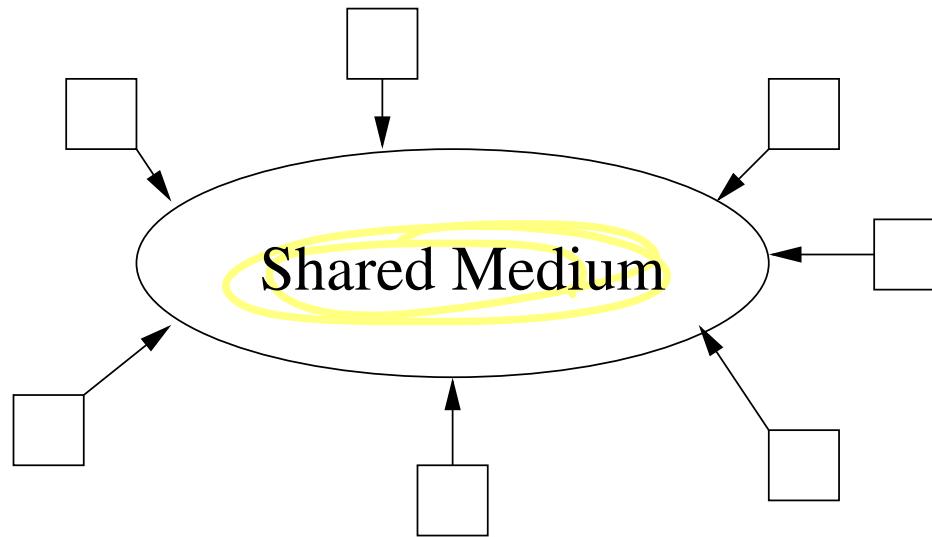
- **Problem:**

Given that there are multiple users, how do we mediate access to a shared channel?

- Medium Access Control (MAC) sublayer between Physical and DLC (Data Link Control) is used to solve this problem

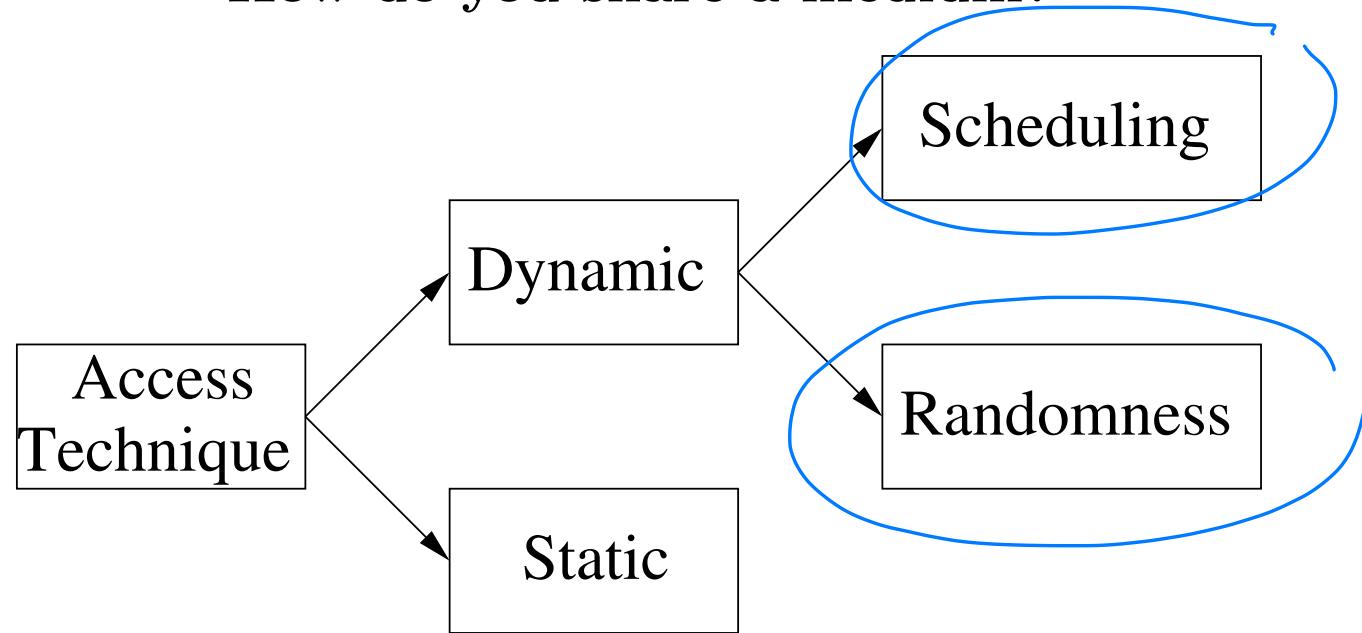
MAC Protocols

- **Centralized:** A distinguished node (master) makes access decisions for the remaining nodes (slaves).
- **Distributed:** All nodes are equivalent and the access decision is derived together in a distributed fashion.



Centralized schemes are too dependent on master failure and generally less efficient.

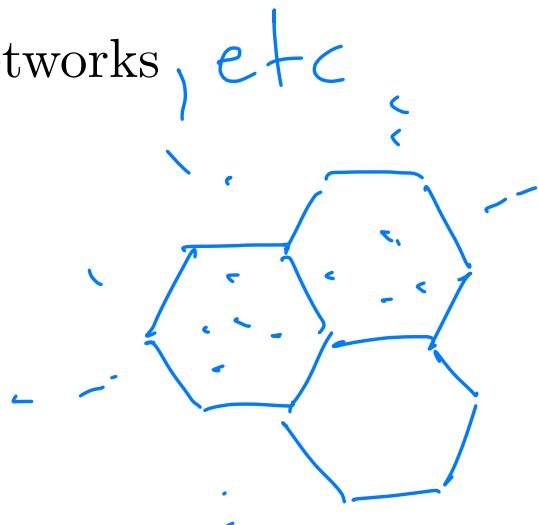
How do you share a medium?



1. **Static Partitioning Schemes:** Partition transmission medium into separate dedicated channels.
2. **MAC Schemes:** Dynamic and on-demand. However, must minimize collisions.

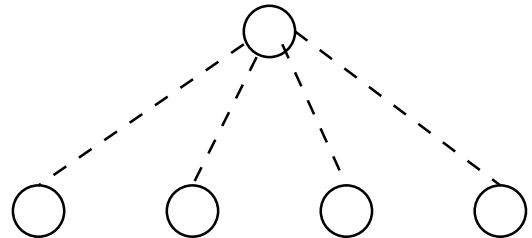
Some Examples: Types of Networks

- Satellite channels (wireless)
~~Iridium network~~ *Space X, Amazon*
- Multitapped bus (wired):
Ethernet
- Star topology with hub (wired):
Fast Ethernet
- Packet radio networks (wireless)
Ad Hoc, Bluetooth, Piconets, Wireless networks
- Cellular networks (wireless)
Cell phones, Wireless LANs, etc

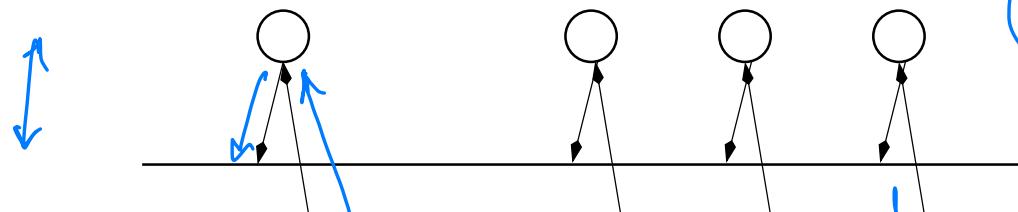


Some Examples: Network Topologies

Satellite



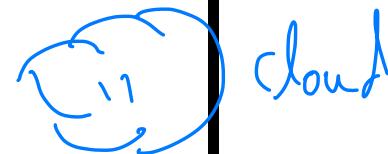
Ground Stations



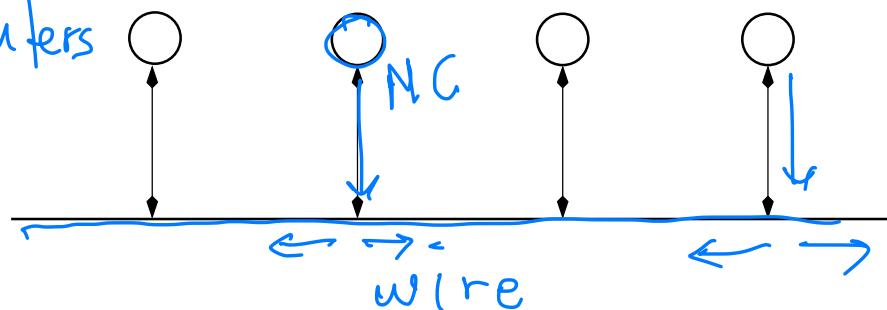
I do not remember!

Multihop

Telephone Line



Computers



Multiaccess

Channel

3 COM

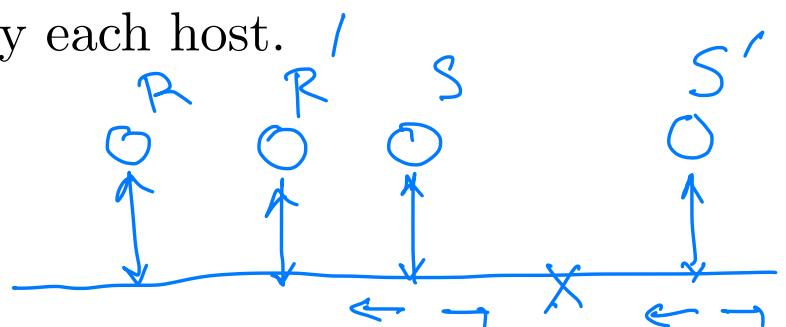
How Do You Mediate Access?

- Given that there are many users, several issues must be taken into account.
 - Give access to each user that wants to communicate.
 - Decide who talks first.
 - Be fair to all.
- How do you accomplish all these tasks?
- It is inevitable that we must employ some measurement on how long medium is used by each host.

quite complex
Some randomization is needed!

Computers must tap the line

$$S \sim R, S' \sim R'$$



X: noise
O: message

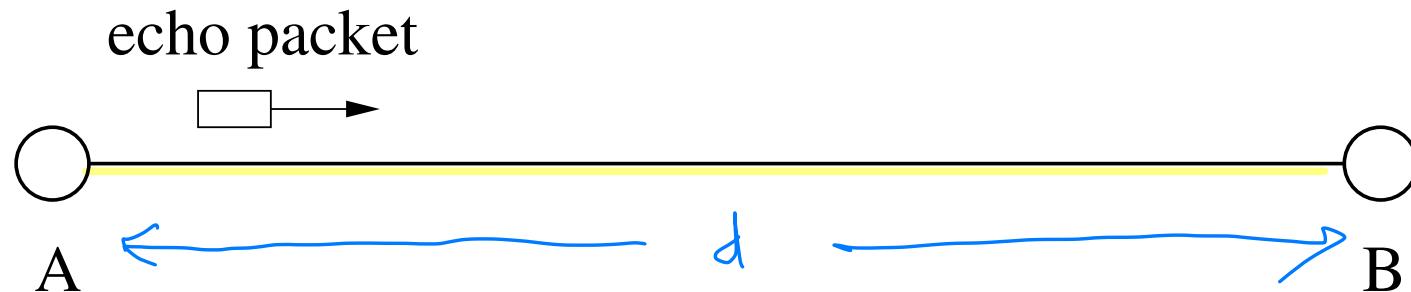
The Case of Two Hosts

- Access mediation is complex even for two users.
- A lot of subtle issues arise that must be taken into account.
- Lets try to understand the problem for two hosts first!
- To be specific, we will address the following problems.
 1. Measure the Propagation Time
 2. Coordinate access.
 3. Select a winner.
- We will address the access mediation problem for many users later.

Measuring the Propagation Time

- Let T_{prop} be the bit-propagation time of a channel.
- If d = “distance between the two stations” and v = “the speed of the medium” then

$$T_{prop} = \frac{d}{v}.$$



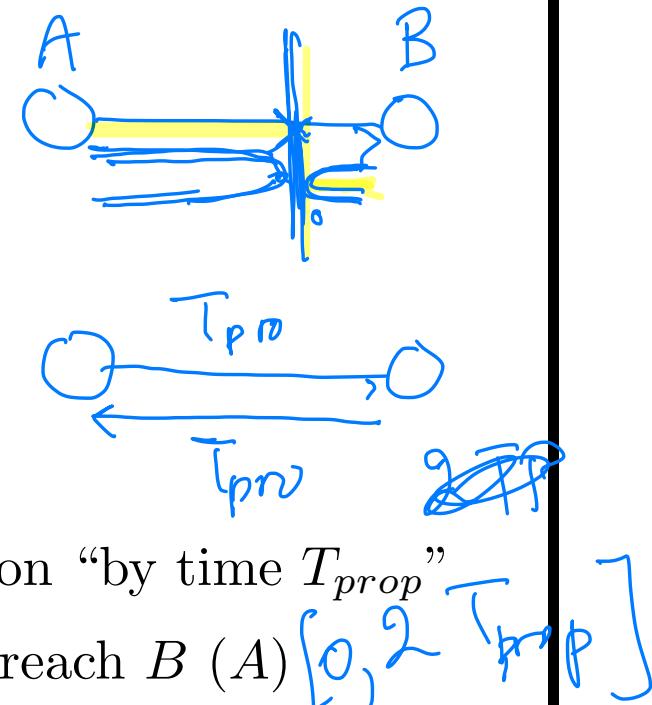
- Both stations can measure T_{prop} , e.g., can use ping.
- So we can assume they both have the same value for T_{prop} .

How do you coordinate access?

- **Access Coordination Algorithm:**

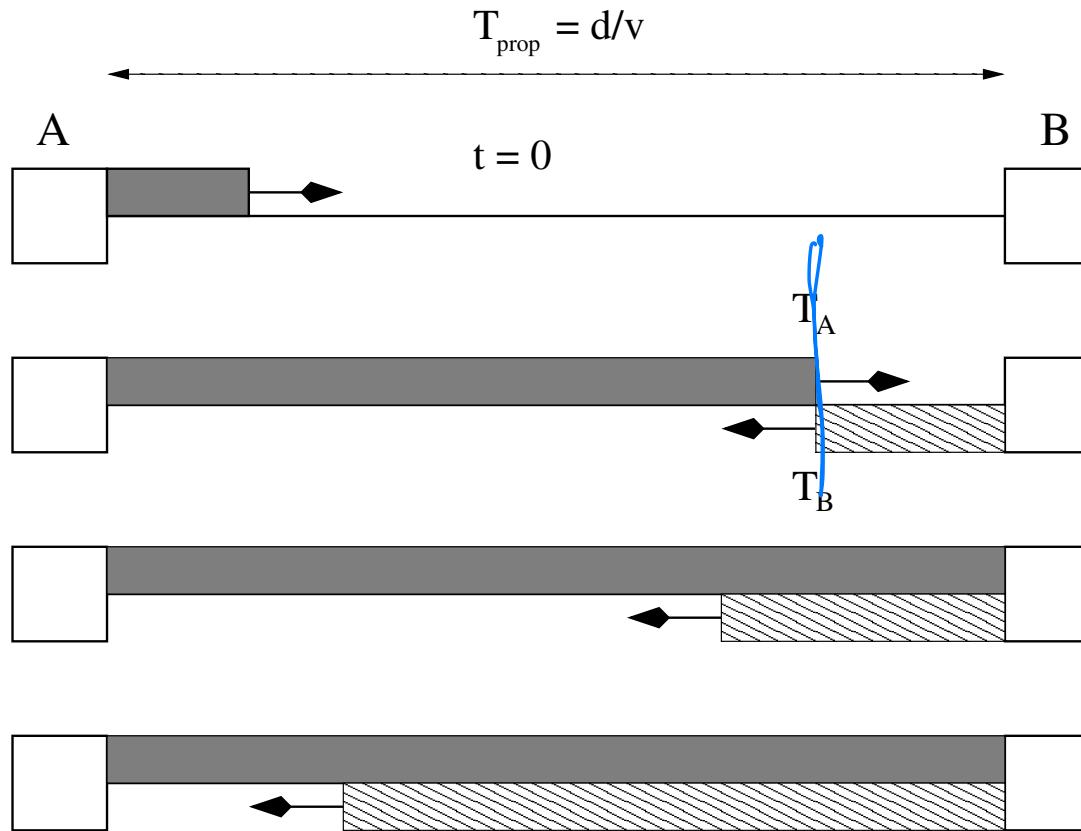
1. A (B) listens to channel
2. if channel not busy
3. then A (B) transmits packet
4. A (B) continues to listen to channel
5. if B (A) has not began transmission "by time T_{prop} "
6. then A (B) is certain packet will reach B (A) $[0, 2T_{prop}]$
7. else A (B) detects collision and retransmits.

- If user A is to be able to **detect a collision** it must occupy the channel for a time period of $2T_{prop}$ time units.
- **Note:** Since both stations can measure T_{prop} , at the latest, by time $2T_{prop}$, A will know if a collision occurred.



Measuring Time

Stations measure time T_A (T_B) from the beginning of (their own) packet transmission to the time a collision occurs.



User that sends a packet occupies
the channel

And the Winner Is!

- Stations A and B can compare T_A and T_B with T_{prop} .
 $T_A < T_B$ iff $T_A < T_{prop}$
 1. A wins iff $T_A < T_B$.
 2. Losing station remains quiet until winner completes transmission.
 3. For the sake of fairness, after completing transmission, the winner remains quiet for $2T_{prop}$ time units to allow the loser to capture channel.

A wins iff $2T_A < 2T_B$ | Process
iff $T_A < T_B$ | Very quick
 | us
 | microsec

Efficiency!

- So for each packet sent, $2T_{prop}$ time is required to coordinate access.
- If bit rate is R and packet length is L then channel efficiency is

$$\frac{L}{L + 2T_{prop}R} = \frac{1}{1 + \frac{2T_{prop}R}{L}} = \frac{1}{1 + 2a},$$

≈ 1
 $2a \approx 0$

where $a := \frac{T_{prop}R}{L}$, must be small

- The closer to 0 the number a is, the more efficient the channel.
- If $a = \frac{T_{prop}R}{L}$ (i.e., $a \sim 0$) is small then $1 + 2a \sim 1$ and therefore the efficiency is ~ 1 , i.e.,

$$\frac{L}{L + 2T_{prop}R} \sim 1.$$

Measurements and LANS

- Measurements made depend on the technical specifications of the networks being used.
- Recall that

$$\begin{aligned} T_{prop}R &= \frac{d}{v}R \\ a &= \frac{T_{prop}R}{L} = \frac{dR}{vL}, \end{aligned}$$

d, R, v, L

$R \uparrow \Rightarrow L \uparrow$
 $v \uparrow$

where d is distance, v speed of medium, L , is the packet length, and R is the bit transmission rate.

- Clearly, these parameters depend on the network technology.

Comparing Performance of Some Networks

Use transmission speed $v = 3 \cdot 10^8 \text{ m/s}$, and packet length $L = 1,500B = 12,000b$. Vary distance d and transmission rates R .

d	Rate $R =$	Rate $R =$	Rate $R =$	
Network	10 Mbps	100 Mbps	1 Gbps	
100 m	$3.33 \cdot 10^0$	$3.33 \cdot 10^1$	$3.33 \cdot 10^2$	$= T_{prop}R$
LAN	$2.77 \cdot 10^{-4}$	$2.77 \cdot 10^{-3}$	$2.77 \cdot 10^{-2}$	$= a$
10 km	$3.33 \cdot 10^2$	$3.33 \cdot 10^3$	$3.33 \cdot 10^4$	$= T_{prop}R$
MAN	$2.77 \cdot 10^{-2}$	$2.77 \cdot 10^{-1}$	$2.77 \cdot 10^0$	$= a$
1000 km	$3.33 \cdot 10^4$	$3.33 \cdot 10^5$	$3.33 \cdot 10^6$	$= T_{prop}R$
WAN	$2.77 \cdot 10^0$	$2.77 \cdot 10^1$	$2.77 \cdot 10^2$	$= a$

For each d and R we compute $T_{prop}R$ and $a = \frac{T_{prop}R}{L} = \frac{dR}{vL}$.

$$\frac{1}{1+2a}$$

What Does the Table Tell Us?

- For “large” distances a is computed to be very large and therefore the efficiency

$$\frac{1}{1 + 2a}$$

is very small and so unacceptable!

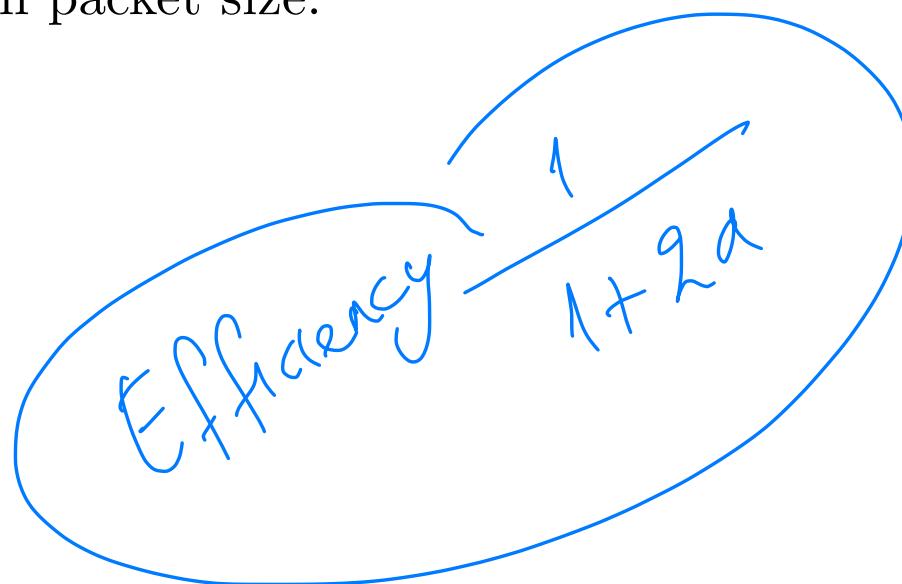
- The reason for this is that “one is forced to occupy the medium” so as to be sure nobody else is using it at the same time.

Scaling Ethernet

- In Ethernet, where there is broadcasting type of message passing, every node is always listening to the network and may initiate transmission only when the network is silent.
- The network is a broadcast media in which every node can hear every other node.
- In order for two nodes not to send data simultaneously in a quiet network, nodes must listen to their transmissions, and if the data a node reads from the Ethernet does not match the data it is placing on the Ethernet, it knows that a collision has occurred.
- Whenever a collision occurs, a node stops sending and waits a random time before attempting to retransmit.

Limitations of Ethernet: Distance Factor

- In a 10 Mb Ethernet, the minimum packet size is 64 bytes for a 5 km cable.
- In a 1 Gb Ethernet, the minimum packet size is about 6400 bytes.
- From an architectural perspective 6400 bytes is too large a number for the minimum packet size.



Other Issues

- Medium access protocol is very technology dependent!
- Can we be sure that measurements are accurate?
- Even “Echo” measurements may differ for two hosts!

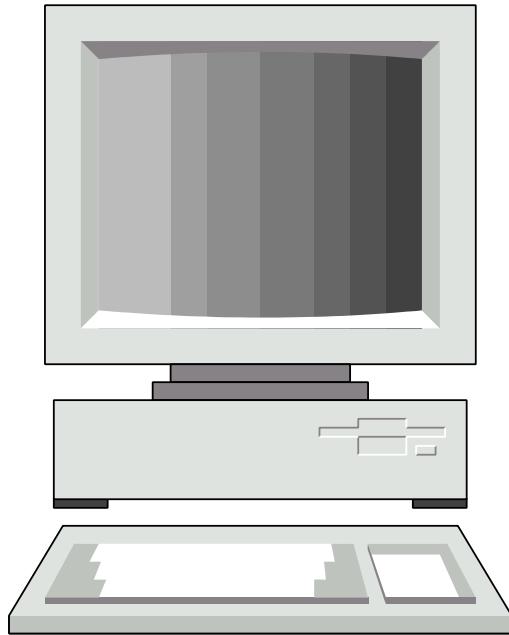
Nevertheless, resulting protocols are realistic and efficient because they are on-line.

Comparison of Peer-to-Peer and MAC Protocols

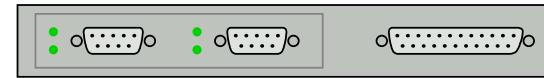
- Peer-to-Peer concern communication between two users as opposed to MAC protocols that concern many.
- A rough comparison of tradeoffs is given in the following table.

	Peer-to-Peer	MAC
# Nodes	Two	Many <i>Two</i>
Concern	Loss/Delay	Interference
Method	Sequencing	Randomization
Mechanism	ACK	Coordination
Performance	$\text{Delay} \times \text{Bandwidth}$	$\text{Delay} \times \text{Bandwidth}$
Node-Status	Independent	Coordinated

Some LAN Devices



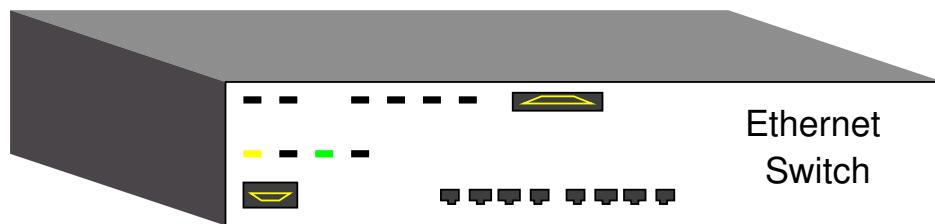
Host



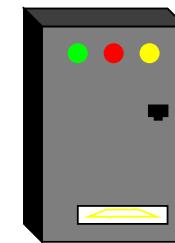
Network Bridge



Network Hub



Ethernet
Switch



Network Transceiver

Exercises^a

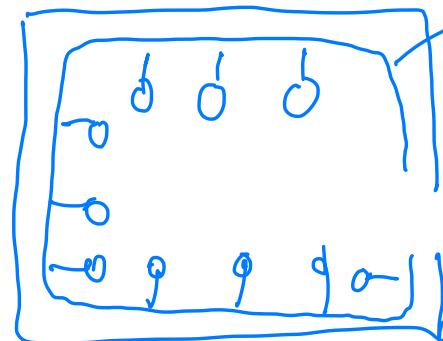
1. Discuss advantages and disadvantages of dynamic, static access control.
2. Why does a multiaccess algorithm depend on the underlying network?
3. List three differences between switched and broadcast networks.
4. Explain why in medium access control we must obey the following design principle: the longer the distance between hosts the larger the length of the packet that must be used.
5. The distance between two hosts is 1 km and the speed of the medium is $4 \cdot 10^8 \text{ m/s}$. What is the propagation time? What is the RTT?
6. Why does ethernet require that a packet must have a minimum

^aDo not submit

length?

7. Show that if d is the distance between two hosts, v speed of medium, L , is the packet length, and R is the bit transmission rate then $\frac{T_{prop}R}{L} = \frac{dR}{vL}$.
8. The bit rate of a channel between two hosts A and B in one direction is R and in the other directions is $3R$. Assume the packet length is L . What is the channel efficiency?
9. In the previous exercise, determine the channel efficiency if in addition to propagation delays we have transmission delays, i.e., the transmission delays at hosts A and B are t_A and t_B , respectively.
10. What alternative methods could you use to decide the winner between two hosts in medium access control?

LANs: Ethernet



channel
is the
wire.

Outline

1. Contention
 - (a) Unlimited
 - (b) Limited
2. MAC Protocols
 - (a) CSMA-CD,
 - (b) Aloha
 - (c) Persistence
3. Limited Contention
4. Reservation
 - (a) Bit-Map
 - (b) Binary Countdown
 - (c) Splitting Algorithms

In Wireless we use
a diff paradigm: CA

Contention

MAC and Contention

- MAC (Multiple Access Control) requires contention resolution,
 - i.e., prioritize the use of the medium so as to ensure good performance.
- There are two types of contention.
 - Uncoordinated
 - Coordinated
- In the former, no scheduling is required, while in the latter a coordination algorithm must be performed by all users.

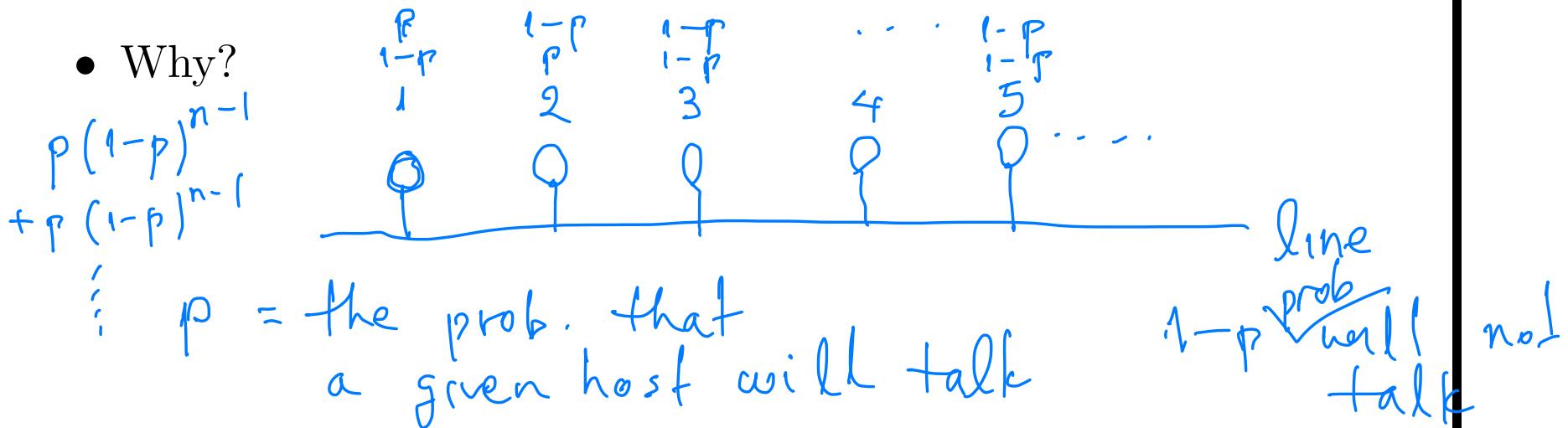
Uncoordinated Contention (1/3)

- Suppose n nodes are contending for a channel.
- A node transmits during a contention with probability p .
- Probability of a successful transmission is equal to the probability that exactly one node transmits:

$$\Pr[\text{Success}] = np(1-p)^{n-1}.$$

- $\Pr[\text{Success}]$ is maximized (as a function of p) when $p = 1/n$.

- Why?



Uncoordinated Contention (2/3)

- Take the derivative of the function

$$f(p) := np(1 - p)^{n-1}$$

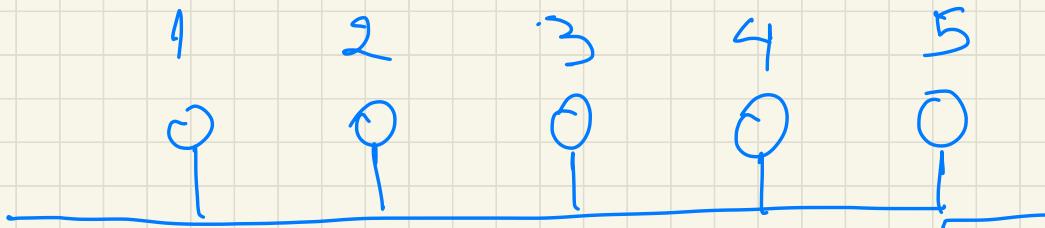
with respect to p .

$$\begin{aligned}\frac{d}{dp} f(p) &= \cancel{n(1-p)^{n-1}} + np(n-1)(-1)\cancel{(1-p)^{n-2}} \\ &= \cancel{n(1-p)^{n-2}} \cancel{((1-p) - (n-1)p)}\end{aligned}$$

- If you set $\frac{d}{dp} f(p) = 0$ we see that

$$1 - p - (n-1)p = 0. \quad \leftarrow$$

- Hence, $p = 1/n$.



$$\Pr[\text{Success}] = 5 \frac{1}{5} \left(1 - p\right)^4$$

$$= 5 \frac{1}{5} \left(1 - \frac{1}{5}\right)^4$$

167 $\Pr[\text{Success}] = 167 \frac{1}{167} \left(1 - \frac{1}{167}\right)$

Uncoordinated Contention (3/3)

- Hence, $f(p)$ obtains its maximum value when $p = 1/n$.
- In which case

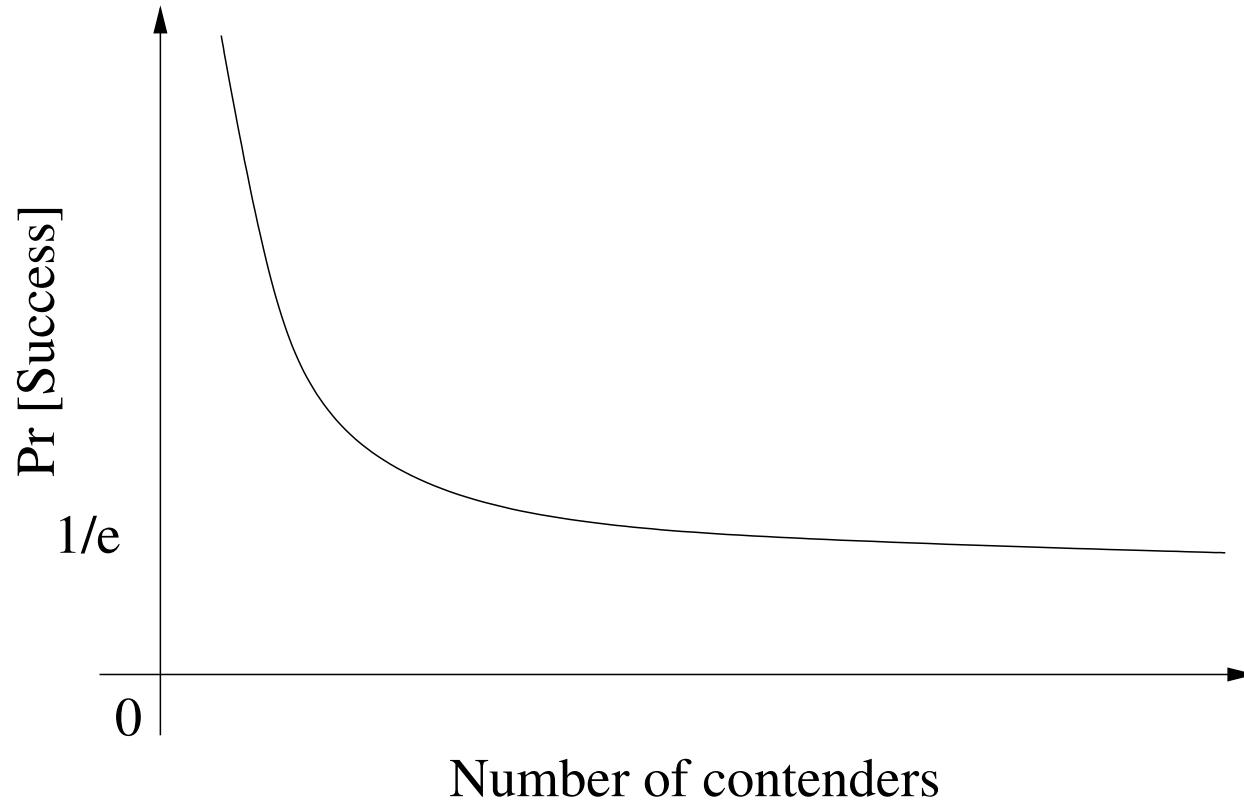
$$\begin{aligned} P_{\max} &:= \Pr[\text{Success}] \\ &= n \frac{1}{n} \left(1 - \frac{1}{n}\right)^{n-1} \\ &= \left(1 - \frac{1}{n}\right)^{n-1} \quad e=2.781 \\ &\approx \boxed{1/e} \text{ as } n \rightarrow \infty. \end{aligned}$$

- The average number of contentions until success occurs is measured as the mean of a geometric distribution and is equal to

$$\frac{1}{1/e} = e.$$

Limits of Contention

The only way to improve the performance of contention is by limiting the number of users.



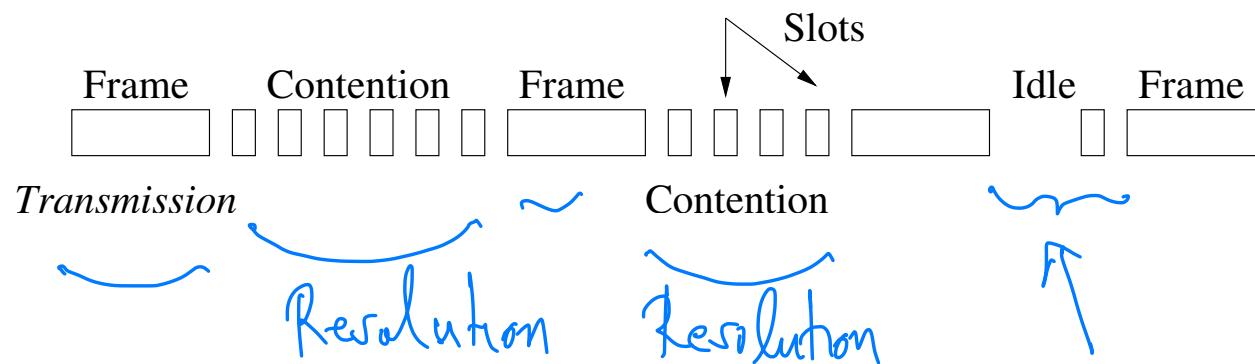
Limited Contention

Access Method: Limited Contention Algorithms

- The problem with the previous approach is that it assumes that each of the contenders wants to talk all the time.
- In practice this is not the case.

Access Method: Limited Contention Algorithms

- After a collision if some way could be found of resolving the collision quickly it may be possible to increase the throughput
- Algorithms for resolving collisions after they occur are sometimes called limited contention algorithms
- Periods of normal operation (with no collisions) are punctuated by periods where a different algorithm is used to resolve a collision



Limited Contention Algorithms

- **Limited Contention** protocols combine best properties of
 1. contention (contention at low load provides low delay) and
 2. collision-free (good channel efficiency at high loads).

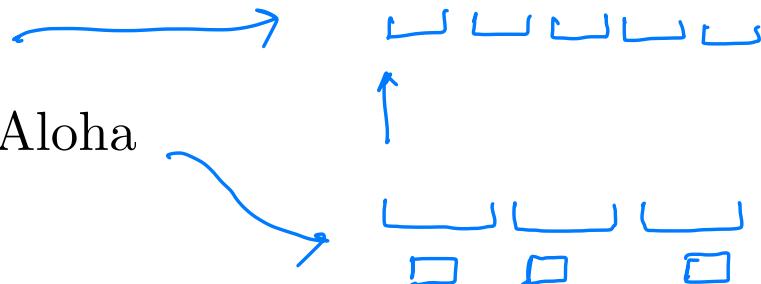
Limited Contention are algorithms that limit the number of users contending in a given round.

Types of MAC Protocols

Two basic classes of MAC (Multiple Access Control) protocols.

1. Random Access

- (a) Slotted Aloha
- (b) Unslotted (or Pure) Aloha
- (c) CSMA
- (d) CSMA-CD



2. Scheduling

- (a) Reservation
- (b) Polling
- (c) Token-Passing

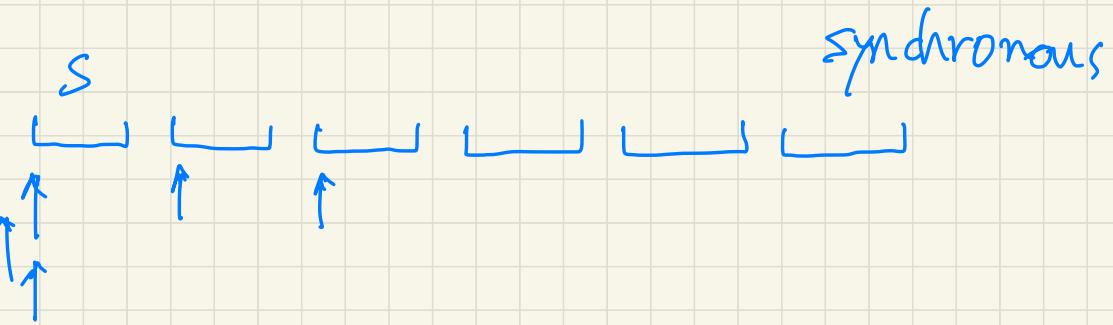
Aloha Ethernet: In two Varieties

- Devised in 60's by Abramson at U of Hawaii to interconnect terminals located at campuses of different islands. *Created for a wireless net.*
- A radio transmitter was attached to the terminals; a message is transmitted as soon as it becomes available.
- If packets collide they are retransmitted.
- Adapted by Boggs and Metcalf at Xerox in 1973 to get Ethernet
- Aloha Ethernet come in two varieties: Slotted and Uslotted.
 - **Slotted:** time divided into slots that handle fixed length packets with transmissions only at slot boundary
 - **Uslotted:** no restrictions on packet size or time of transmission

Slotted.

multiple
users

Unslotted



2s

↑
↑
↑

synchronous

asynchronous

Backoff Protocols

- Based on a suitable probability distribution and a queue.
- There is a queue of stations (nodes) waiting to transmit.
- Each station (node) keeps track of the number of attempts to transmit
- A function $p(x)$ (known to all hosts) is defined in advance:
 - $p(x)$ is the probability you transmit in the x -th attempt
 - same function is used by all nodes
 - $p(x)$ decreasing in x ($x = \# \text{ of attempts}$)

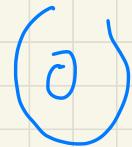
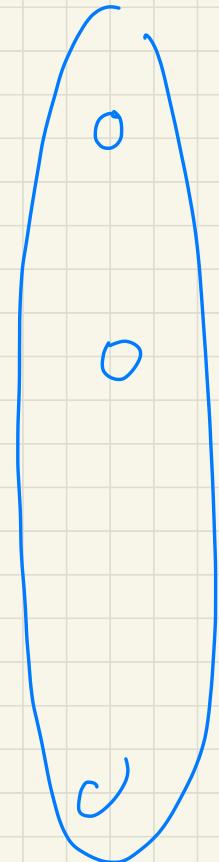
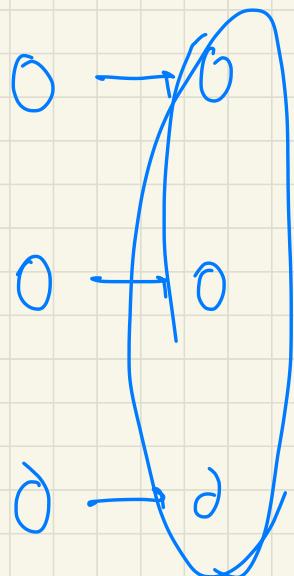
$$p(1), p(2), p(3), \dots \dots \dots$$

Backoff Algorithms

- The main backoff algorithm is as follows.
 - Station i has a variable bck_i :
 1. Initialize $bck_i \leftarrow 0$;
 2. Station attempts transmission if queue is not empty with probability $p(bck_i)$;
 - (a) if it fails to transmit due to collision it increments the variable bck_i (e.g., $bck_i \leftarrow bck_i + 1$);
 - (b) Else it assigns 0 to the variable bck_i ;
 3. If queue is empty and no transmission is made the variable bck_i remains unchanged.
 - Major concern:
 - What functions $p(x)$ should we use and are what criteria?
 - How stable is traffic under various functions $p(x)$?

Collision $\frac{1}{2}$

$\frac{1}{4}$



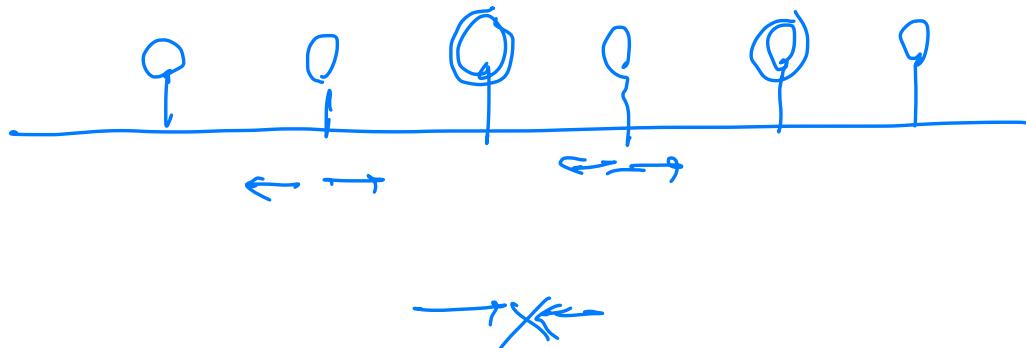
Aggressiveness: Binary Exponential Backoff

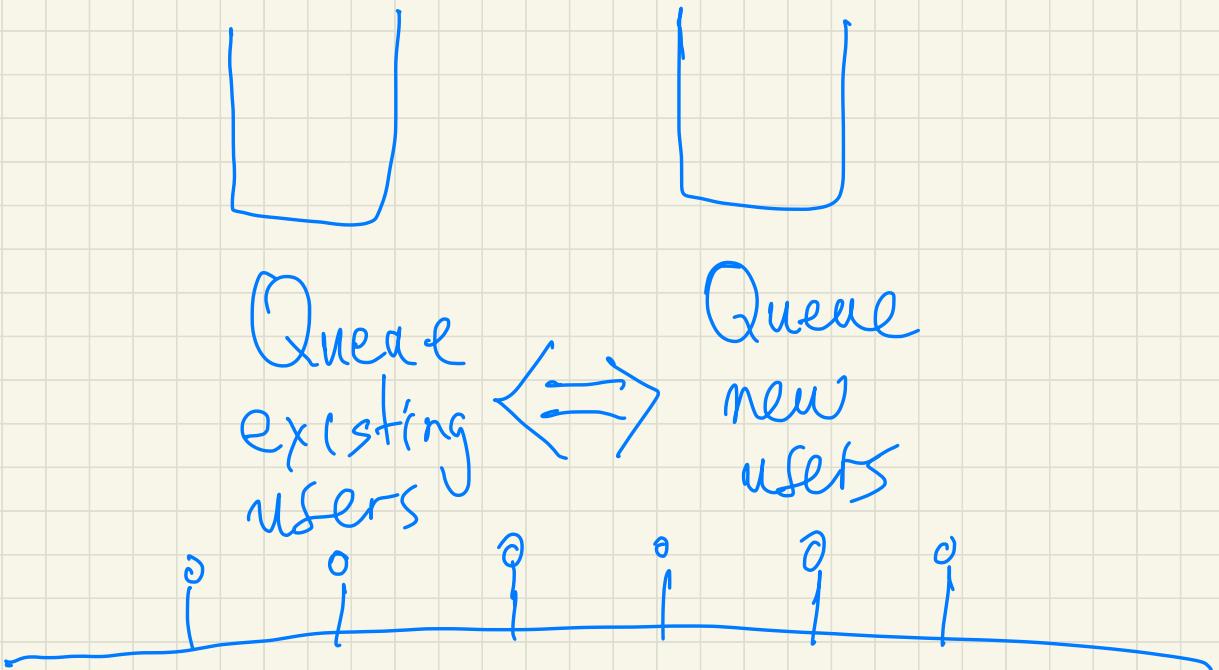
- The function $p(x)$ is important: $x = \# \text{ of round you're in}$
 - it tells you how aggressive you should be in your x -th attempt to transmit.
- Backlogged packet attempts with probability $p(x)$ where x is the # of unsuccessful attempts.
 1. Exponential backoff: $p(x) = 2^{-x}$. IEEE
 2. Polynomial backoff (k is constant): $p(x) = \frac{1}{(x+1)^k}$.
 - $k = 0$: constant backoff
 - $k = 1$: linear backoff
 - $k = 2$: quadratic backoff
- Polynomial backoff has been shown to be stable for $k > 1$.
- Proving these claims requires sophisticated research!

Carrier Sense Multiple Access/Collision Detection

There are two important aspects to Ethernet:

- **CSMA** *you need a NIC (network card)*
 - whereby the carrier is sensed for the presence of signals, and
- **CD**
 - whereby the hosts attempt detection of collisions.
 - This is used to cut down on unnecessary collisions!





Round:

1. choose value x
2. compute 2^{-x}
3. w.r. $\frac{1}{2^x}$ you talk
4. listen

Round : $\left[\begin{matrix} \text{Receive} \\ \text{Inform} \end{matrix} \right] \left(\begin{matrix} \text{Computes} \\ \text{Some} \end{matrix} \right) \left(\begin{matrix} \text{Send} \\ \text{Infor} \end{matrix} \right)$

Ethernet CSMA/CD

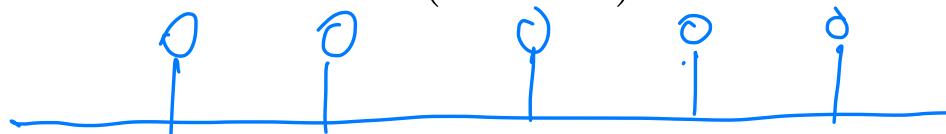
- **CSMA (Carrier Sense Multiple Access)**

1. A station wishing to transmit must first listen for existing traffic on the line. *tap line with NIC card.*
2. If no voltage detected, the line is considered to be idle and a transmission is initiated.

 noise  packet

- **CD (Collision Detection)**

1. If no voltage detected, a transmission is initiated. During transmission, station checks line for extremely high voltages that indicate collisions.
2. If collision detected station waits a predetermined amount of time for the line to clear (backoff) and sends data again.



What Do I Do after a Collision? Exponential Backoff!



1. After 1st collision, each station “waits at random” either 0 or 1 time slots and tries again.
2. If they pick same random value they collide again. After 2nd collision, each station “waits at random” either 0 or 1 or 2 or 3 time slots and tries again.
3. If they pick same random value they collide again. After 3rd collision, each station “waits at random” either 0 or 1 or 2 or 3 or 4 or 5 or 6 or 7 time slots and tries again.
4. And so on...

.

$$x=3$$



$$x=t$$

How Long Does it Take before a Successful Transmission?

- Remember, **collisions** are random events!
- It makes no sense to talk about deterministic time!
- It is more precise to talk about expected time!
- The expected time depends on the backoff protocol being used!

Expected Time

- Suppose that two synchronous ethernet stations A, B are at contention period t :
- This means, they have failed in previous contention periods $0, 1, \dots, t - 1$, and are now in contention period t .
- They each pick a random number in the interval $0..2^t - 1$.
 - A picks random value a in $0..2^t - 1$. 0 ... 7
 - B picks random value b in $0..2^t - 1$ 0 ... 7
- The expected number of steps to reach contention period t is cumulative,
 - i.e., you must add the “waiting times”, for all the trials before t .

How Persistent Should I Be?

- Successful transmission (i.e., no collision) can occur only when the two stations select different values at random, i.e., $a_t \neq b_t$.
- Hence, if that happens then

$$\Pr[\text{no collision}] = \frac{2^t(2^t - 1)}{(2^t)^2} = 1 - \frac{1}{2^t}$$

$\overset{2^t = m}{\curvearrowleft \curvearrowleft \curvearrowleft \dots \curvearrowright}$
 $\frac{m(m-1)}{2}$

- The waiting time for a success in t -th attempt will be

$$\min\{a_t, b_t\},$$

where a_t, b_t are the values chosen above!

- NB:** The bigger the t is the higher (closer to 1) is the probability that you will succeed next time ($t + 1$).
- NB:** Don't forget that there are many stations (not just two) contending for the medium!

100 people try to talk

→ 50 remain

→ 25 remain

→ 12 n

:

:

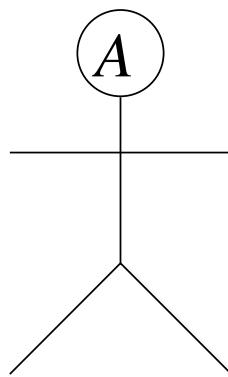
:

:

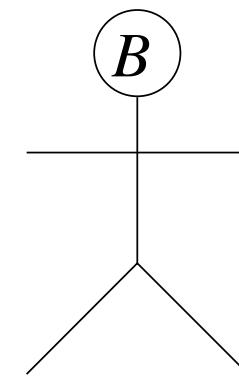
$$2^t \geq 100$$

$$t = 7$$

How Persistent Should I Be?

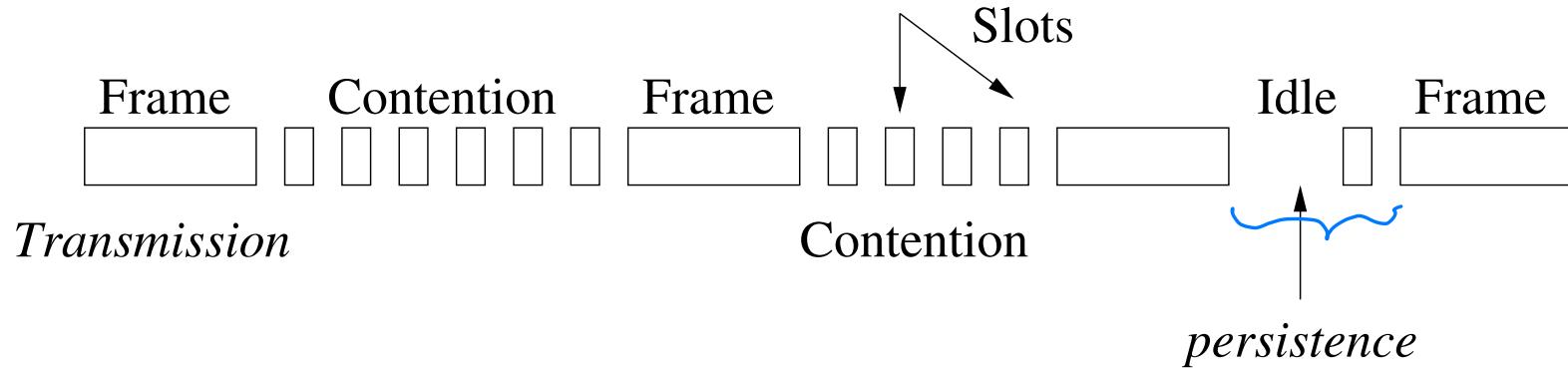


What do I do now?



How persistent should I be?

Variants of CSMA: Smart Transmitters



- When a station has data to send and line is busy it waits for the line to go idle.
- How persistent is a transmitter in capturing an idle channel?

Proactive Mechanism
to take advantage of the
idle line.

Avoid transmissions that are certain to cause collisions!

- In most LAN systems, a node can sense if another node is using the channel after some delay for the signal to reach it
- CSMA senses the medium for the presence of carrier signal.
- In this case, a node can wait until the channel becomes idle before transmitting
- Schemes that take advantage of this are called Carrier Sense Multiple Access (CSMA)

- 1. 1-Persistent CSMA
- 2. Non-persistent CSMA
- 3. p -persistent CSMA

$$0 \leq p \leq 1$$

with carefully chosen prob. p will have the best performance

Slotted CSMA: 1-Persistence

- Remember, there are many stations competing!
- When node has data to send.
 1. First listens to channel to see if anybody else is transmitting.
 2. If channel busy the node waits until channel becomes idle.
 3. When node detects idle channel it transmits frame.
 4. If collision occurs node waits a random amount of time and starts over again.
- 1-persistent means that a transmitter with a packet to send transmits with probability 1 whenever a busy line goes idle.



Slotted CSMA: Non-Persistence

- **Nonpersistent:** Do not persist! Act like a collision and wait a random number of slots before retrying.
- When node has data to send.
 1. First listens to channel to see if anybody else is transmitting.
 2. When node detects idle channel it transmits frame.
 3. If channel is busy the node does not continually listen in order to seize it for transmission when idle. Instead it waits a random period of time and then repeats the algorithm.
- Non-Persistent behavior is better than 1-persistence.

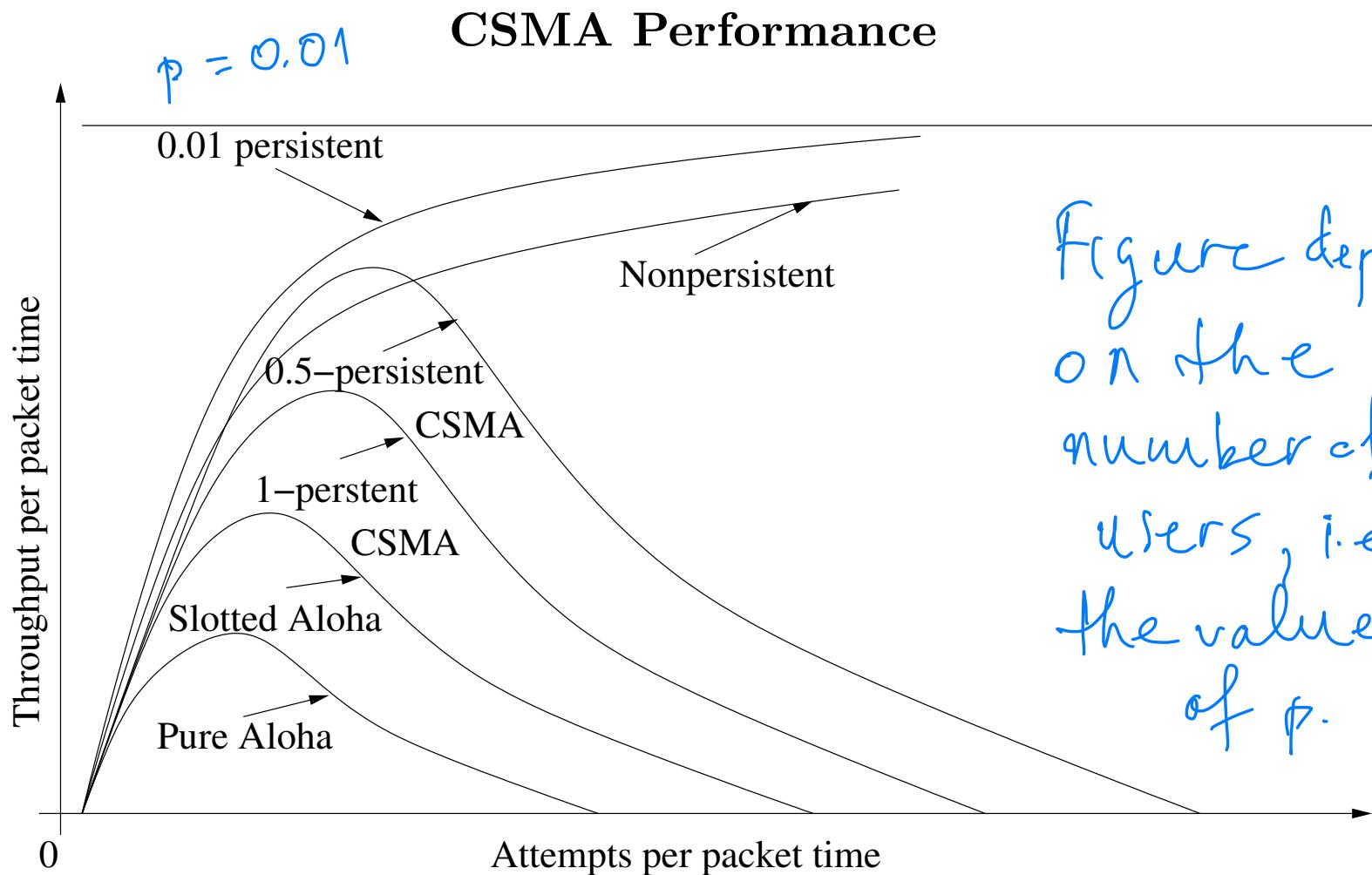
Slotted CSMA: p -Persistence

- **p -Persistent:** send with probability p in next available slot
- When node has data to send.
 1. First listens to channel to see if anybody else is transmitting.
 2. If channel busy wait until next slot and try again.
 3. If channel idle then **Repeat** until either sucess or another node begins transmitting.
 - (a) it transmits with probability p .
 - (b) it defers until next slot with probability $1 - p$.
 4. if another node begins transmitting node acts as if there is collision: waits a random amount of time and starts again.

$$P = \frac{1}{3} \quad . \quad 3 \text{ users}$$

(Example) Slotted CSMA: p -Persistence

- A p -persistent protocol transmits with probability p , where $0 \leq p \leq 1$, after a line becomes idle.
- E.g., Assume $p = 1/4$ and 100 stations are waiting for the busy line to become idle.
- Each of these stations transmits with probability $1/4$.
- Hence only ≈ 25 stations will attempt to transmit and the remaining ≈ 75 stations defer until next time slot.



Ethernet CSMA-CD

- CSMA improves over Aloha because it senses the carrier.
- Still, however, collisions involve entire packets!
- An obvious improvement is to abort transmission when collision is detected!
 1. Station senses channel
 2. **if** channel idle **then** transmit
 3. **else** use a CSMA strategy (p -, non-persistent)
 4. **if** collision detected during transmision
 5. **then** transmit jamming signal **and** abort transmission
 6. schedule future transmission with backoff.

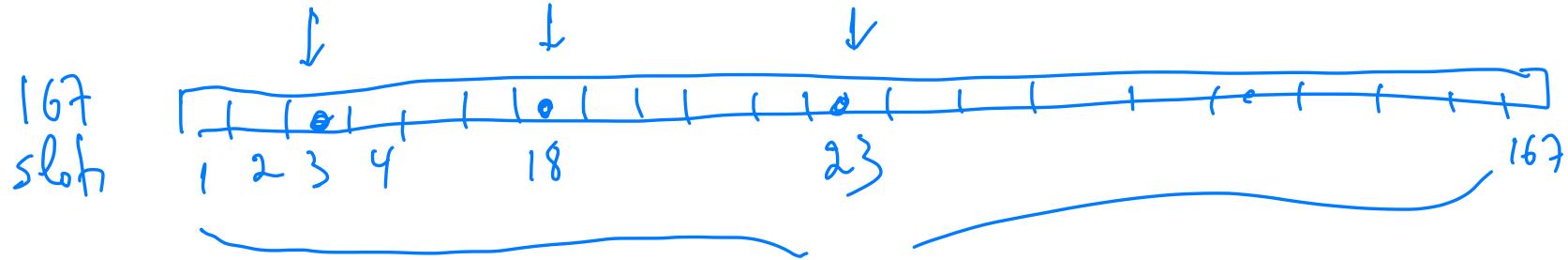


Reservation

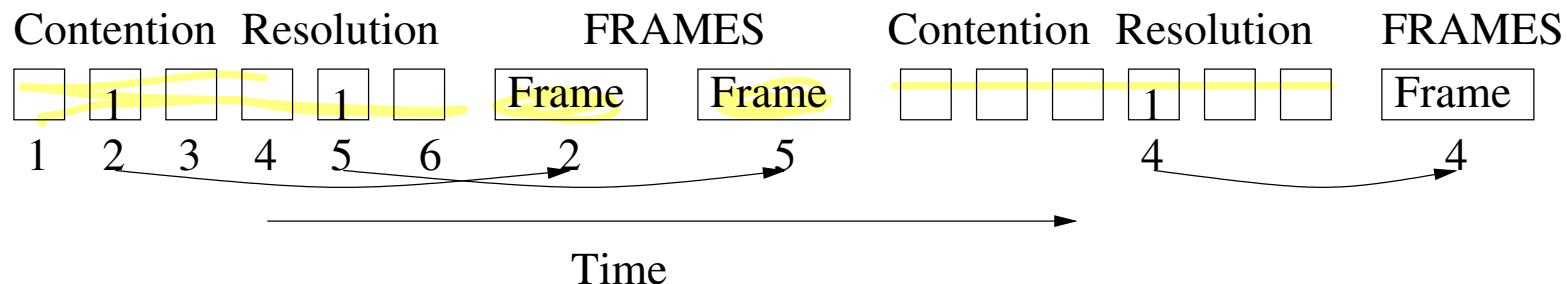
You reserve the channel,
and you talk when your
turn comes up.

Collision-Free Reservation Protocols: Basic Bit-Map

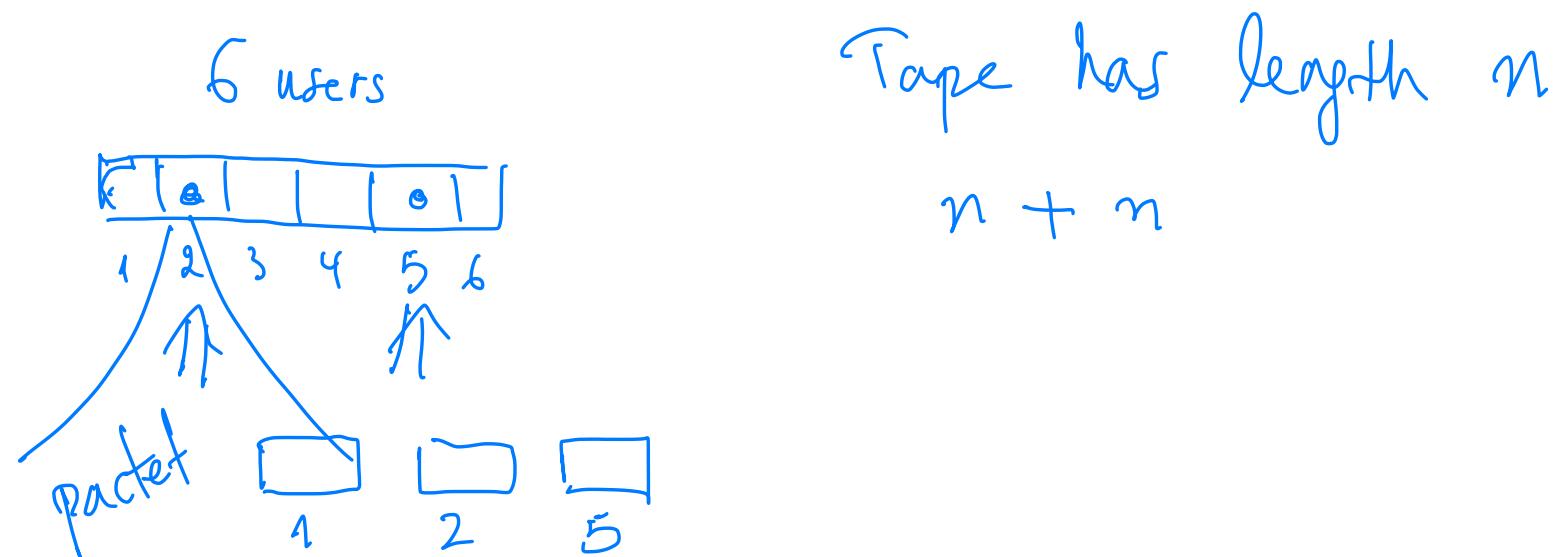
- Assume there are exactly N stations labeled 0 to $N - 1$. Let the propagation delay be negligible.
- **Basic Bit-Map Method:**
 1. Each contention period is exactly N slots.
 2. If station i has a packet to send then it inserts a 1 bit in the i -th slot. No other station is allowed to transmit during this slot.
 3. After N slots have passed by, each station has complete knowledge of which stations wish to transmit.
 4. Stations now begin transmitting in numerical order.



Analysis of Basic Bit-Map

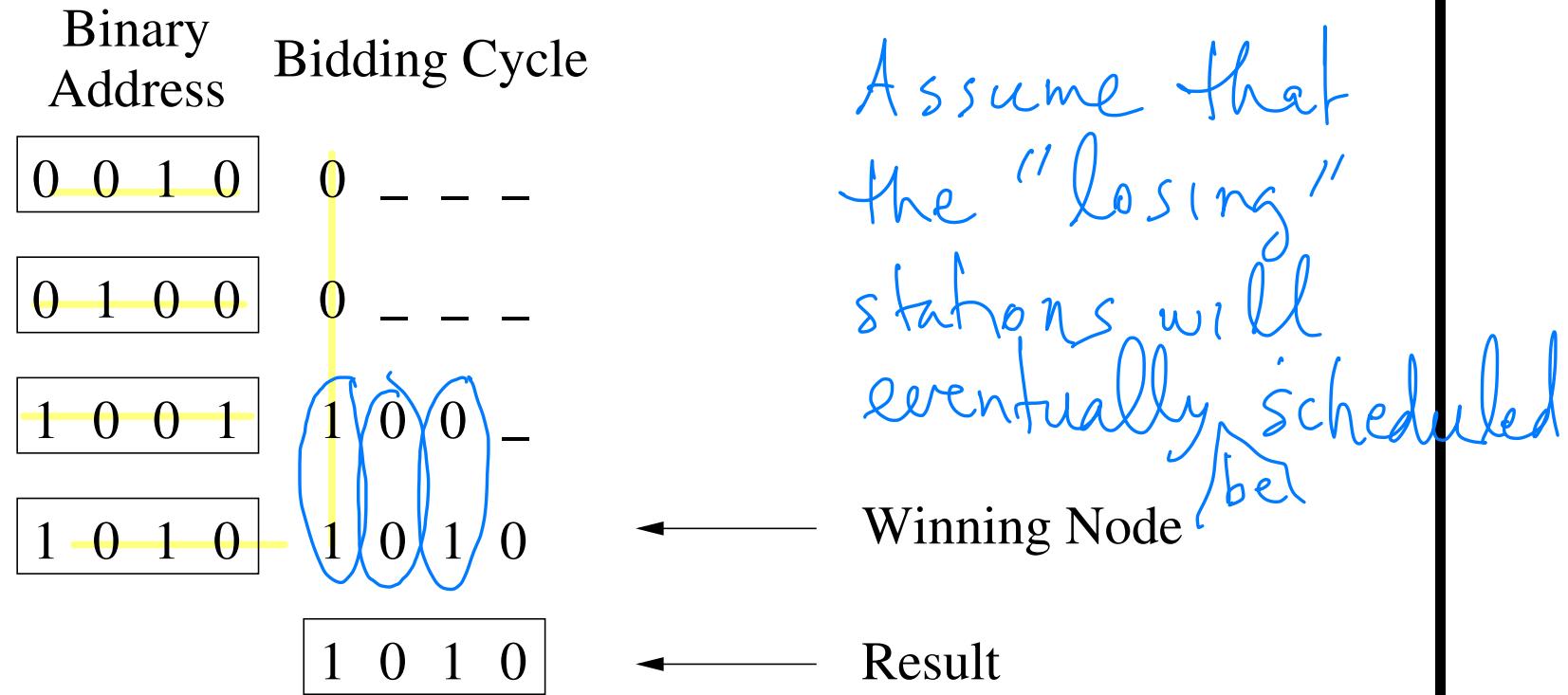


A station is out of luck if it becomes ready just after its bit slot passes by.



Collision-Free Reservation Protocols: Binary Countdown

Here we assume all addresses are in binary and of the same length.



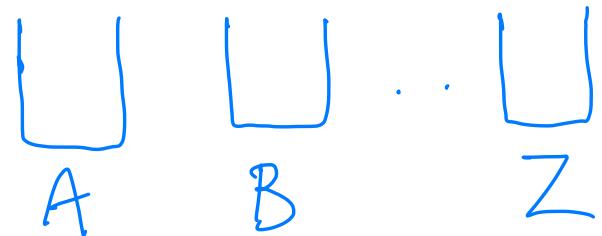
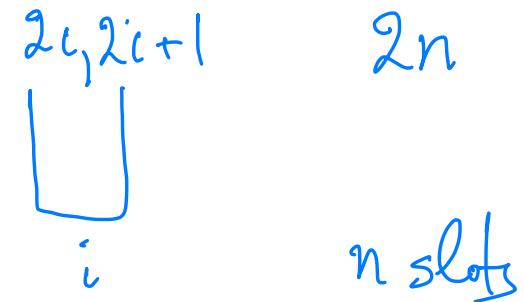
Nodes write their bit starting from highest to lowest order to a register. They drop out as long as another node has written a 1 in current position, but continue otherwise.

Splitting Algorithms

- Idea is to partition set of nodes into “groups”.
- Each group is assigned to a slot.
- Thus a collision can only occur within a group.
- **Example 1:** Assume the nodes have unique (fixed length) identifiers (e.g., $1, 2, \dots, m$)
 - Divide nodes into pairs.
 - Node $2i$ or $2i + 1$ goes in slot i
 - If collision occurs they go in order

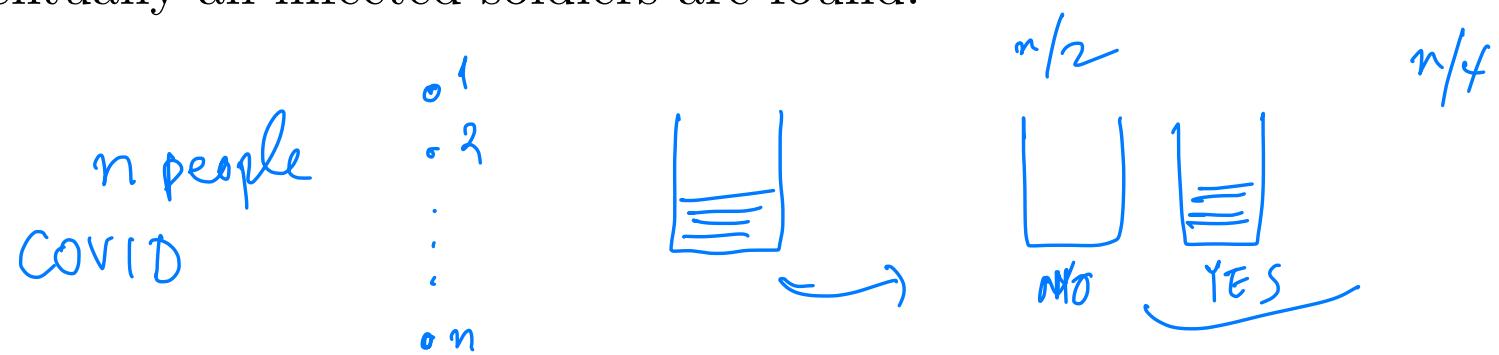
167 students

A B C ... Z



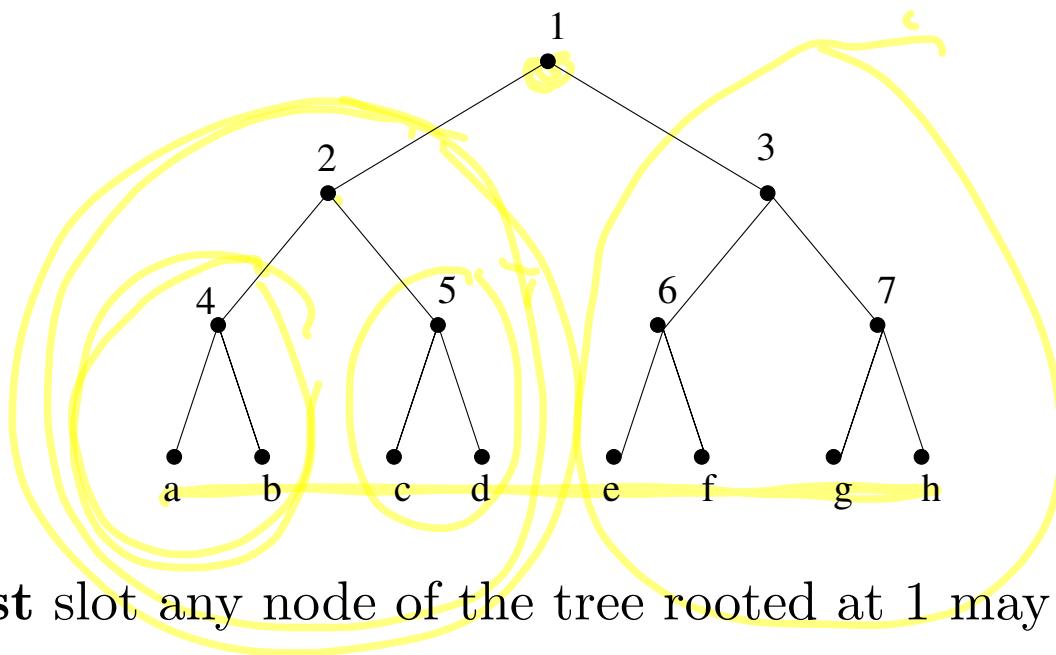
Paradigm

- Similar to US Army testing for syphilis. Take blood test from n soldiers.
- A portion of each sample is poured into a single test tube. This is tested for antibodies.
- If none found all soldiers in group declared healthy.
- If antibodies were found then split soldiers into two groups: $1..n/2$ and $n/2 + 1..n$ and iterate.
- Eventually all infected soldiers are found!



Tree Splitting Algorithm

Nodes are leaves of a tree.



During **first** slot any node of the tree rooted at 1 may contend.

If collision occurs, in **next** slot only nodes in left subtree (rooted at 2) may contend.

Again, if collision occurs, in **next** slot only nodes in left subtree (rooted at 4) may contend.

Tree Algorithm: Example

Eight stations a, b, \dots, h contend for a shared channel. Stations a, c, d, f, g suddenly become ready at once!

Slot 1: a, c, d, f, g ; (collision)

Slot 2: a, c, d ; (collision)

Slot 3: a ; (success)

Slot 4: c, d ; (collision)

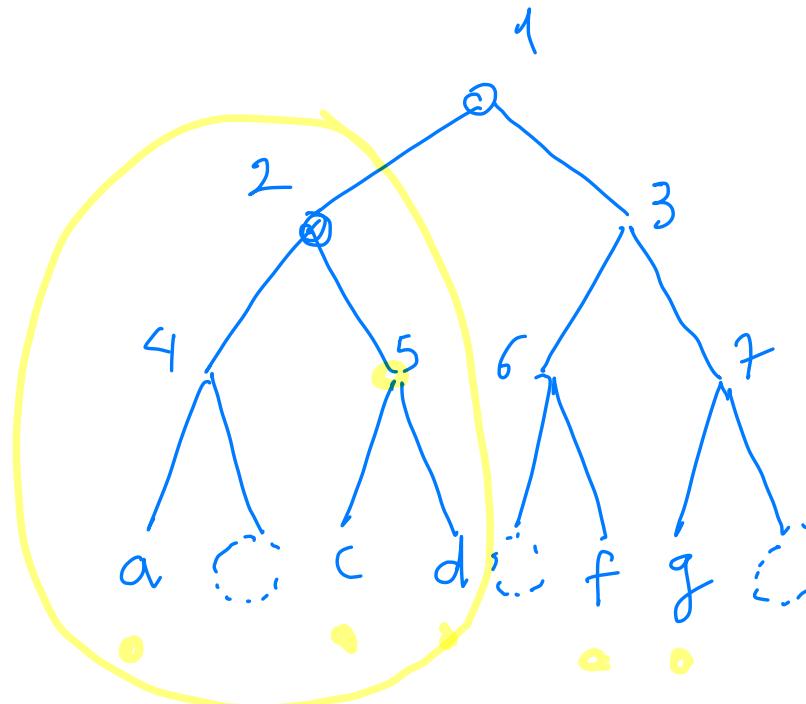
Slot 5: c ; (success)

Slot 6: d ; (success)

Slot 7: f, g ; (collision)

Slot 8: f ; (success)

Slot 9: g ; (success)



Exercises^a

1. What is the difference between coordinated and uncoordinated access?
2. Suppose that n nodes in uncoordinated contention select $p = \frac{1}{kn}$, where $k \geq 1$ is a given integer. What is the $\text{Pr}[\text{Success}]$ as a function of k ? Derive the formula.
3. Suppose that in uncoordinated contention at any given time at most \sqrt{n} of the hosts contend for the channel. What is the $\text{Pr}[\text{Success}]$?
4. What makes unslotted Aloha less efficient than slotted Aloha? What is the tradeoff?
5. In what ways do exponential and polynomial backoff algorithms differ?

^aDo not submit.

6. Eight stations a, b, \dots, h contend for a shared channel. Stations c, d, f, g suddenly become ready at once!. Describe how the splitting algorithm allocates slots.
7. Eight stations a, b, \dots, h contend for a shared channel. Stations a, d, f, g suddenly become ready at once!. Describe how the splitting algorithm allocates slots.
8. Use binary countdown to resolve contention between hosts with the following addresses:

01001011, 01001010, 00000011
9. (*) A medical lab tests the blood of x individuals for a blood-borne pathogen that affects a random individual w.p. p . The x individual blood samples are sent to the lab. If each sample is individually tested, x tests are required. Another option would be to pool the x blood samples forming one composite. If the composite tests negative then all x individuals

are assumed to be pathogen free and no additional testing is required. On the other hand, if the pathogen is detected and the composite tests positive, then each individual sample would have to be tested to determine exactly which of the samples are effected. In such a case a total of $x + 1$ tests would be performed, the original composite test plus x individual tests.^b

- (a) What is the expected number of tests performed?
- (b) What is the expected number of tests performed per individual?
- (c) Composite testing would be preferred to individual testing if the expected number of tests per individual is < 1 . What is the maximum value of p for which this can happen?
(Hint Set the “expected number of tests per individual to 1” and optimize.)

^bA more sophisticated but similar technique (tree splitting) is used in Ethernet to find collisions.

IEEE Standards

Network Interface Card

A number of computers and devices interconnected by a shared transmission medium (wired or wireless) may be arranged in

- a bus,
- ring,
- star topology.

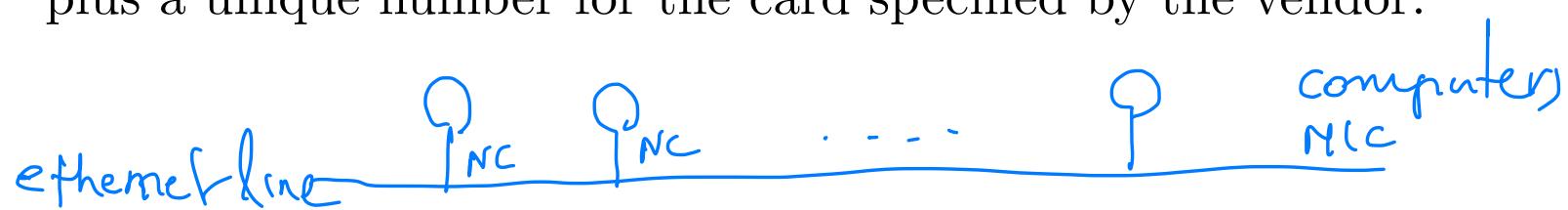
Delay Tolerant Network
Line ethernet

Regardless of the connection topology the stations need to “communicate to the network”.

The device achieving this goal is called “Network Interface Card”.

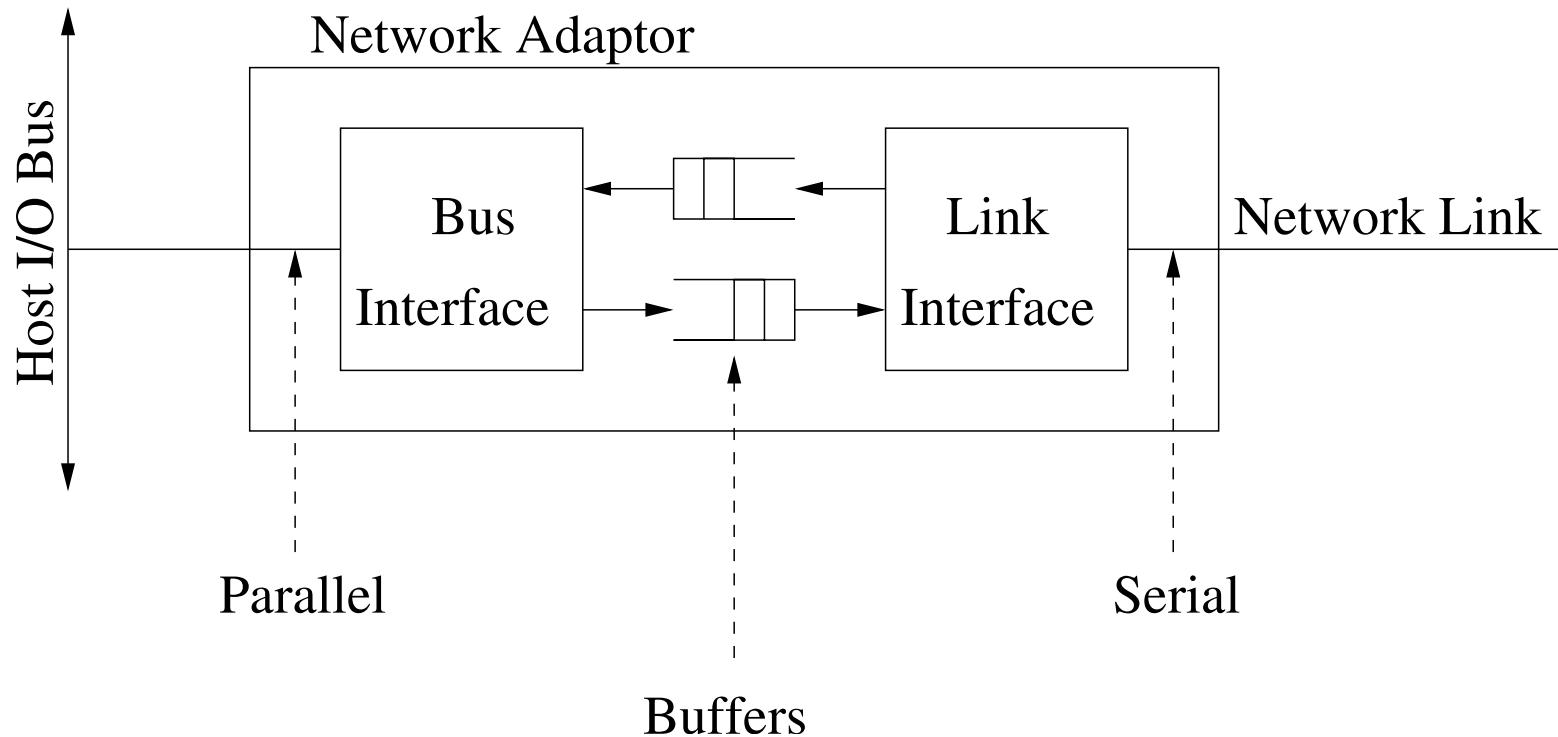
IEEE 802: The Network Interface Card

- Computers and devices connected to the system via a NIC (Network Interface Card) or LAN adaptor card which coordinates information transfer between computer and network.
- The NIC communicates in parallel with computer RAM and serially with network. Parallel/Serial conversion as well as buffering are necessary.
- NIC cards have a port meeting connector specs, and ROM allowing implementation of MAC standards. The NIC physical address is burned into the ROM: and consists of the vendor ID plus a unique number for the card specified by the vendor.

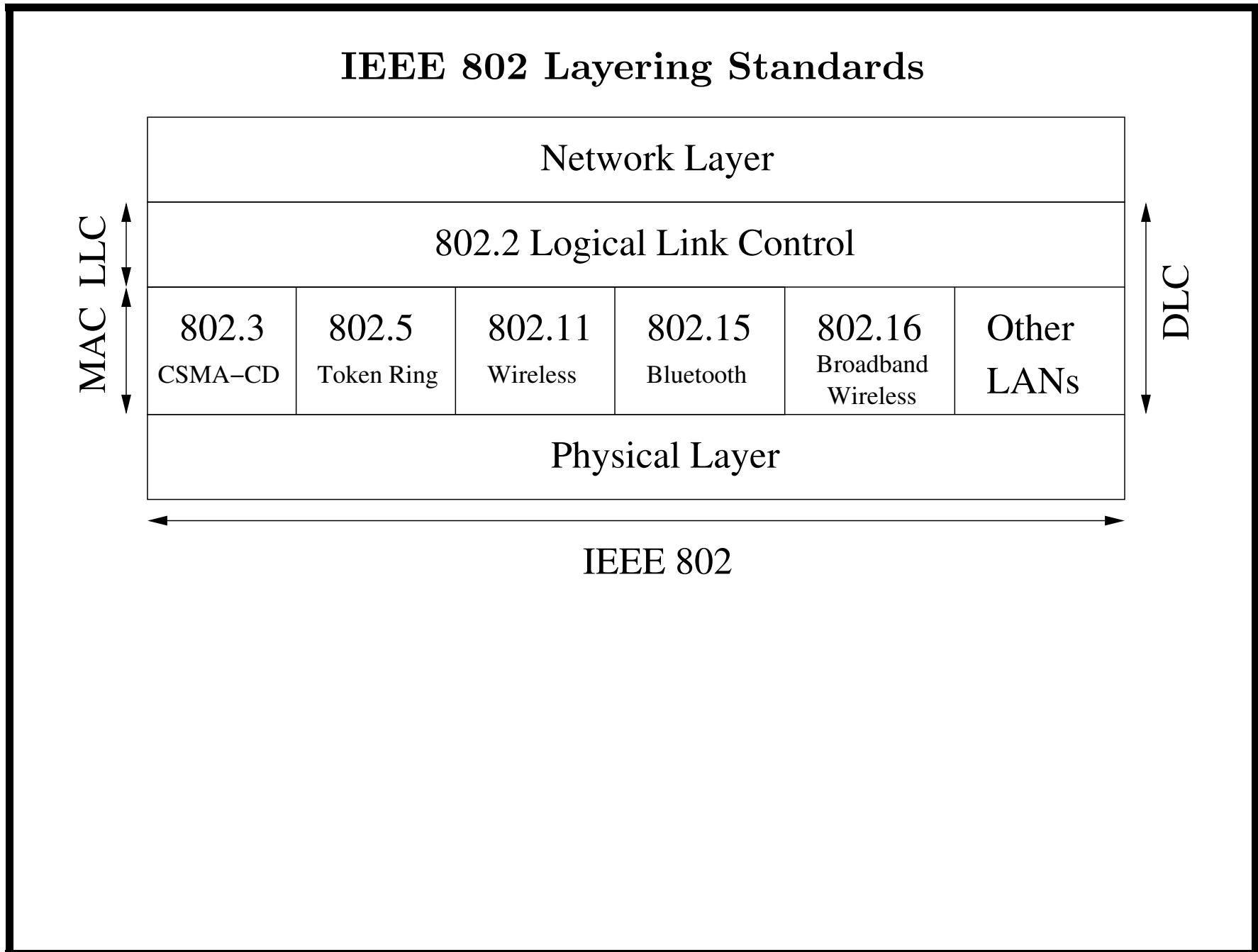


Network Adaptors

Network adaptors programmed by software running on host's CPU.



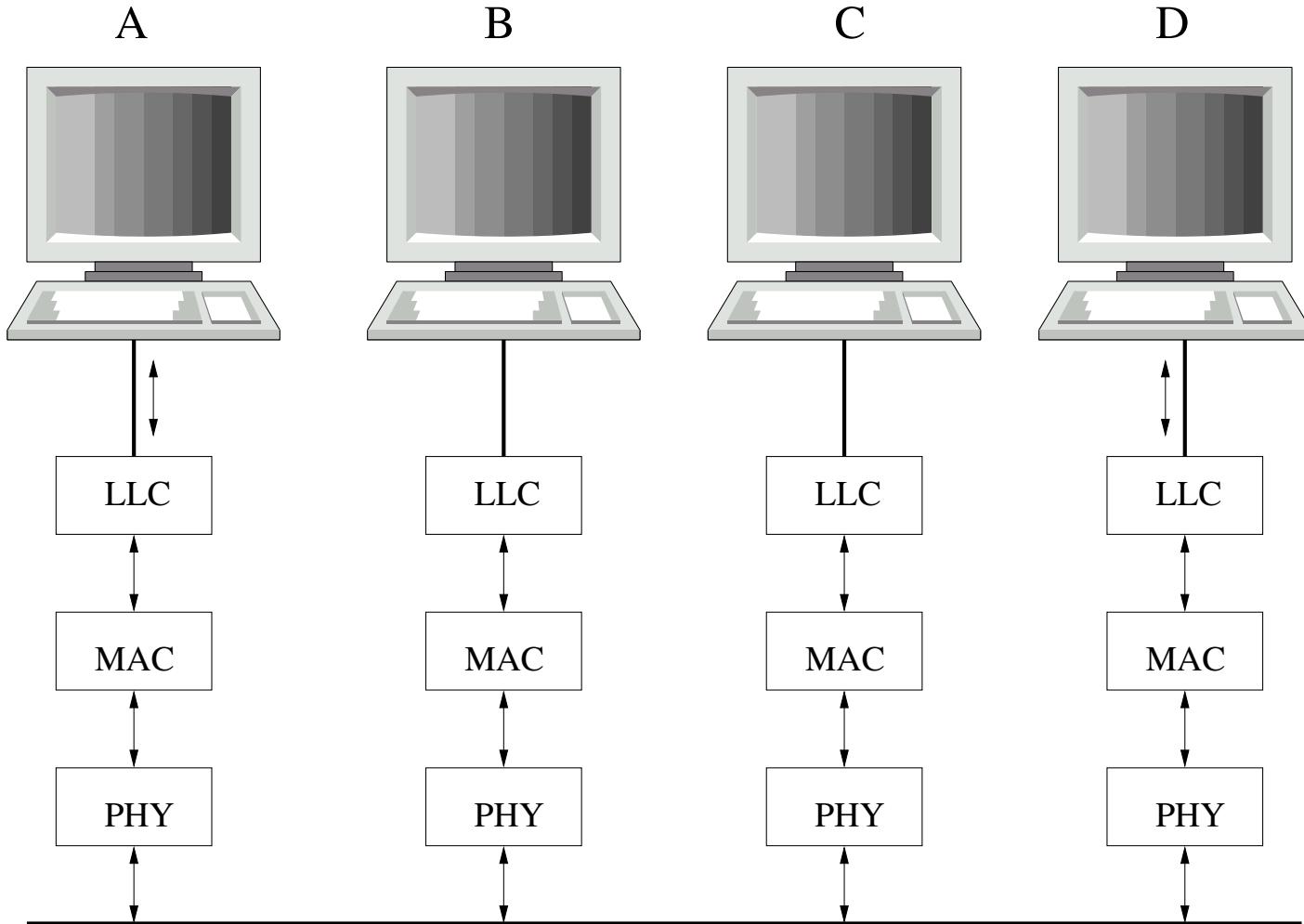
Adaptor exports Control Status Register (CSR) that can be read and written by CPU. e.g., CPU writes to CSR to instruct it to transmit and/or receive frame. CPU reads CSR to learn adaptor's status.



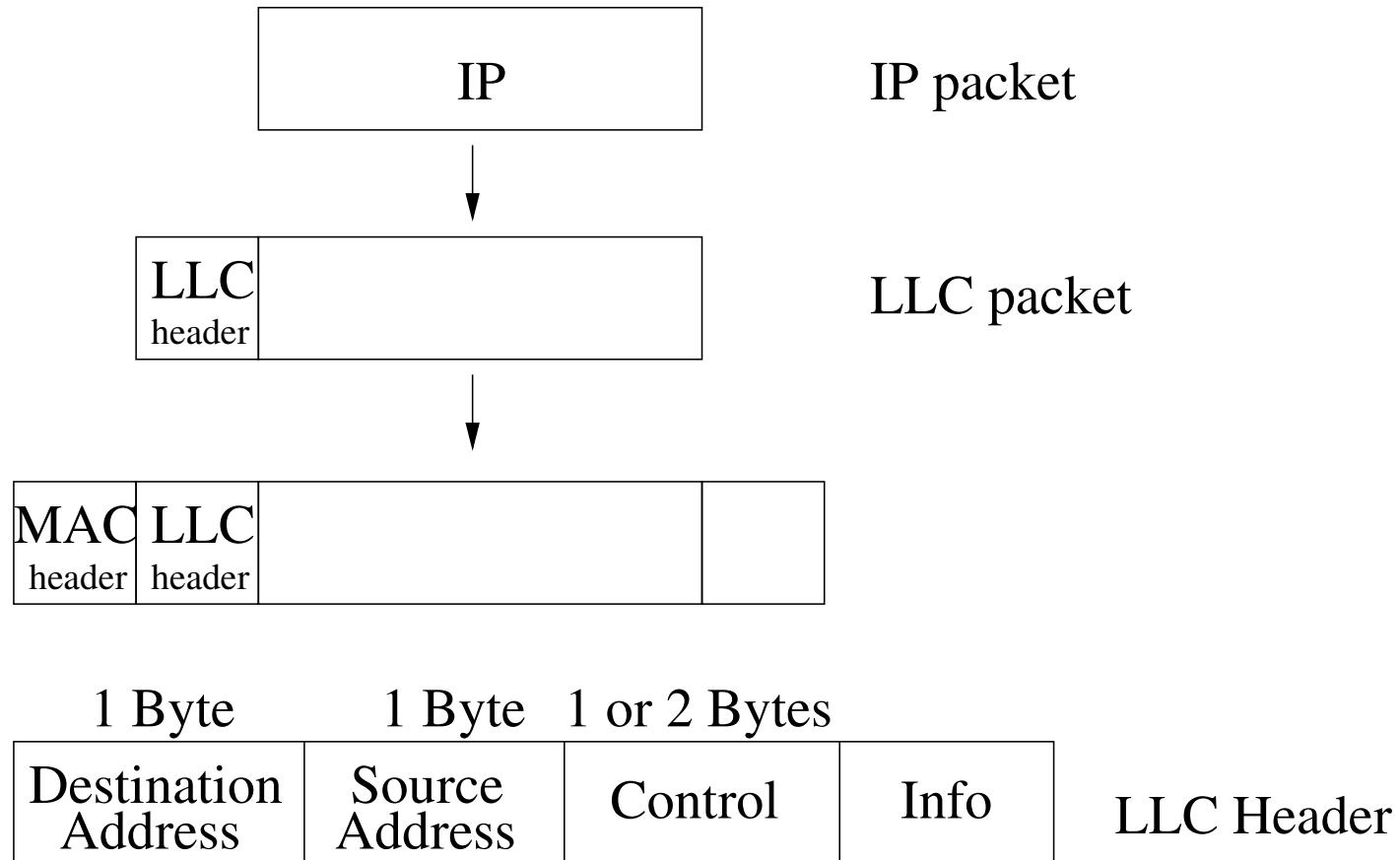
IEEE 802 Layering Standards (up to 802.25)

- 802.4 (Token Bus),
- 802.6 (Dual Queue, Dual Bus),
- 802.7 (Broadband Advisory Group),
- 802.8 (Fiber Optics Advisory Group),
- 802.9 (Isochronous Networks),
- 802.10 (VPNs, Security),
- 802.11 (Wireless),
- 802.12 (AnyLan from HP),
- 802.13 (Unlucky),
- 802.14 (Cable Modems),
- 802.17 (Resilient Ring).

IEEE 802 Layering Standards



IEEE 802 Packetization



IEEE 802.3: Ethernet Standard

- Developed by Xerox in the 1970s.
- In the 1980s, DEC, Intel, and Xerox completed the “DIX Ethernet” standard for a 10 Mbps LAN based on coaxial transmission.
- “DIX Ethernet” became the basis for the IEEE 802.3 Standard.
- IEEE 802.3 Standard frequently revised and expanded over the years.
- Specifications have been issued for running the protocol on coaxial cable, twisted pair, single-mode and multi-mode optical fiber.
- High-speed versions for Fast Ethernet (100 Mbps) and Gigabit Ethernet (1000 Mbps) have also been approved.

IEEE 802.3: Ethernet Protocol (1/2)

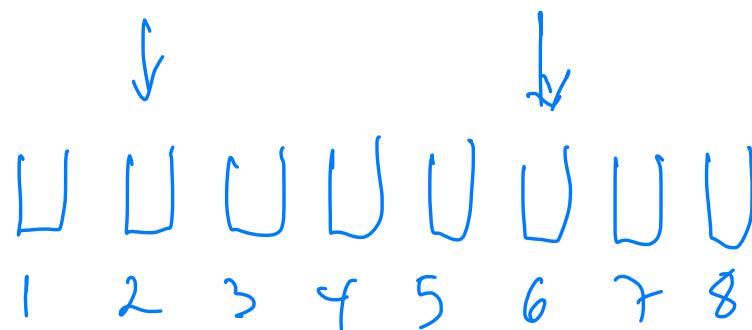
- CSMA/CD being used on bus based coaxial-cable.
- A node with a frame to transmit waits until channel is silent.
- When channel silent, node transmits but continues to listen for collisions.
- If collision occurs station aborts transmission and schedules a later random time when it will reattempt to transmit.
- If no collision occurs node knows it has captured the channel.

IEEE 802.3: Ethernet Protocol (2/2)

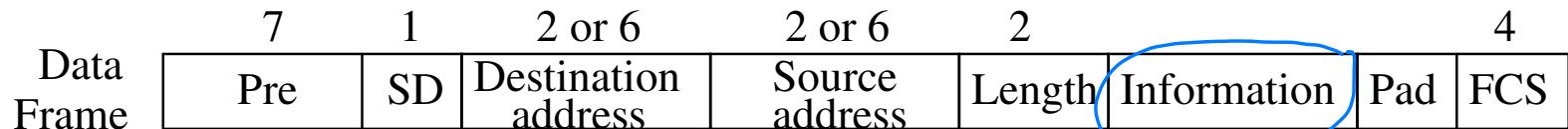
- Binary exponential backoff is being used for retransmission:
if a frame is about to undergo its n th retransmission attempt
then it reschedules transmission by selecting at random an
integer in the range

$$1 \dots 2^{\min\{n, 10\}} - 1.$$

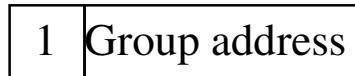
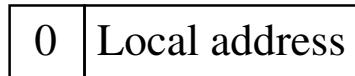
- Up to sixteen retransmissions will be attempted after which the system gives up.



IEEE 802.3: Frames

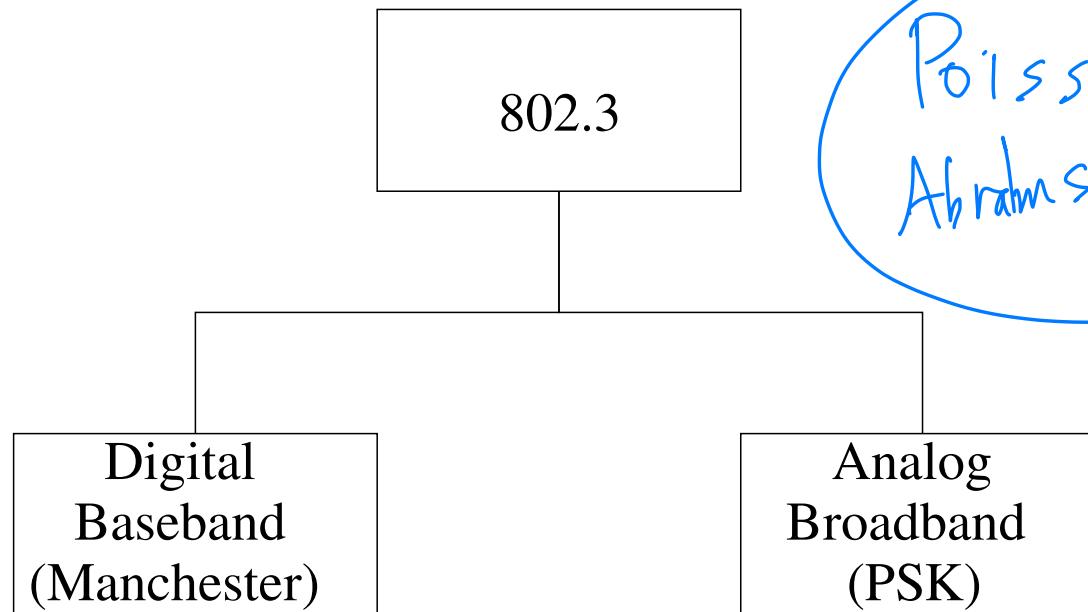


packet



Pre(amble) used for synchronization, SD is Frame Start. Addresses are either single or global, as well as local or universal. The number of possible global addresses is 2^{46} .

Ethernet (802.3)



Poisson
Abrams-Metcalf

Ethernet
traffic
resembles
Poisson
Distribution

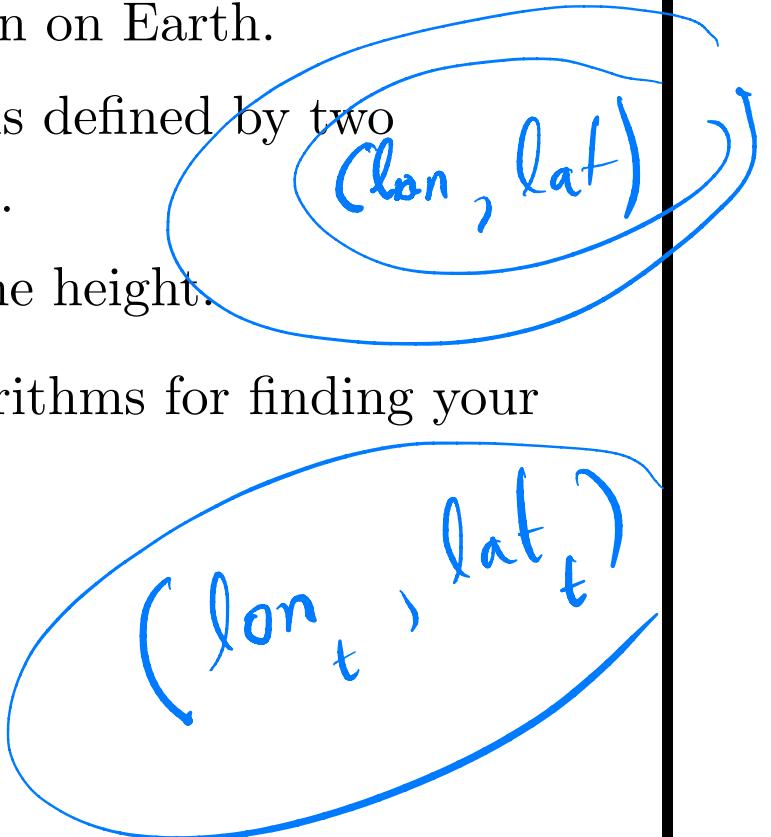
The baseband category has five varieties: 10Base5, 10Base2, 10Base-T, 1Base5, 100Base-T (first number is speed in Mbps, last number or letter is max cable length/type).

The broadband category has only one specification: 10Broad36 (10 is data rate, 36 max cable length).

Localization and GPS

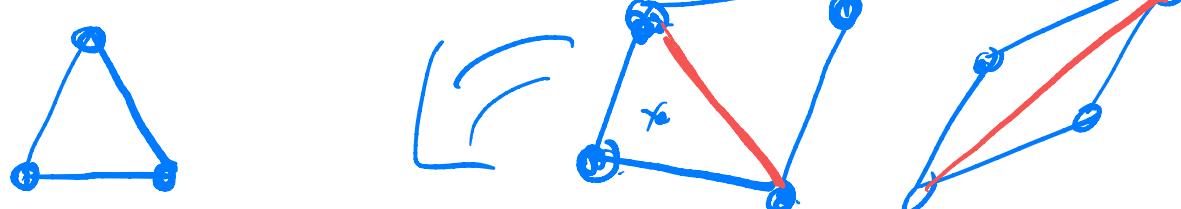
Location and Localizations

- Geographic location refers to a position on Earth.
 - Your absolute geographic location is defined by two coordinates: longitude and latitude.
 - For more accuracy you also need the height.
- Geographic Localization refers to algorithms for finding your geographic location.



Triangulation

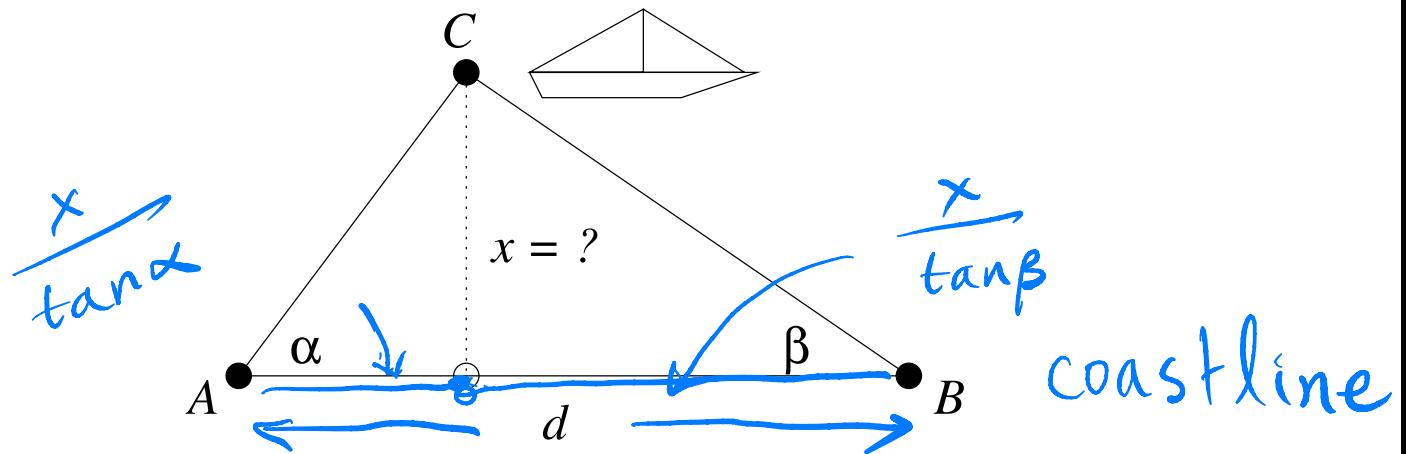
- Triangulation is the process of determining the location of a point by measuring angles to it from known points,
- It can also refer to the accurate surveying of systems of very large triangles, called triangulation networks.^a
 - Surveying error is minimised if a mesh of triangles at the largest appropriate scale is established first, so that points inside the triangles can all then be accurately located with reference to it.



^aWillebrord Snell in 1615-17, showed how a point could be located from the angles subtended from three known points, but measured at the new unknown point rather than the previously fixed points, a problem called re-sectioning.

Triangulation

- Assume a ship is being observed from two different locations.
- You want to measure its distance from coastline.



- Note that in the picture above
 - The coastline is formed by the line AB !
 - The coastline AB is perpendicular to the line formed by the observer and the ship!
 - You want to measure x .

trigonometry

Triangulation

- The unknown distance x can be computed from

$$d = \frac{x}{\tan \alpha} + \frac{x}{\tan \beta}$$

- It follows that

$$d = x \left(\frac{1}{\tan \alpha} + \frac{1}{\tan \beta} \right)$$

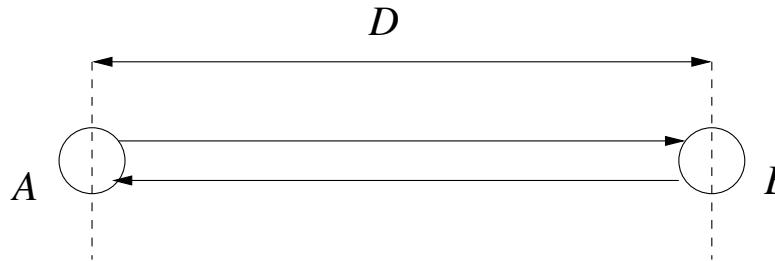
- Consequently

$$x = \frac{d}{\left(\frac{1}{\tan \alpha} + \frac{1}{\tan \beta} \right)}$$

- How do you compute α and β ?
- How do you compute d ?

Another Way to Measure the Distance

- Consider two sensors at unknown distance D .

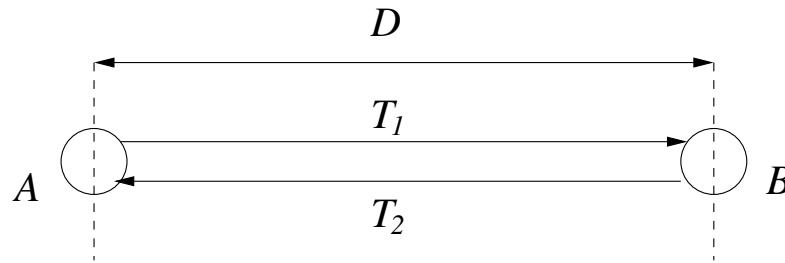


- **Algorithm**
 1. A sends a signal to B in medium1;
 2. B responds to A in a different medium2.
 3. Both A, B measure the roundtrip time, say T .
- From this they can determine the distance D !
- Why?

Another Way to Measure the Distance (1/2)

- Let v_1, v_2 be the propagation speeds in media medium1, medium2, respectively.
- Let T_1 (resp., T_2) be the time it takes from A to B (resp., B to A) in the first (resp., second) medium.

RTT



- They can both measure the roundtrip time $T (= T_1 + T_2)$.
- So we have a system with two equations: v_1, v_2 are known and T_1, T_2 are unknown quantities:

$$D = v_1 T_1$$

$$\left. \begin{array}{rcl} T & = & T_1 + T_2 \\ v_1 T_1 & = & v_2 T_2 \end{array} \right\}$$

v_1, v_2, T
known

Another Way to Measure the Distance (2/2)

- To solve the system observe that

$$\begin{aligned} T_1 + T_2 &= T \\ T_2 &= \frac{v_1}{v_2} T_1 \end{aligned}$$

- Substituting,

$$T_1 + \frac{v_1}{v_2} T_1 = T$$

- Therefore

$$T_1 = \frac{v_2 T}{v_1 + v_2}$$

$$T_2 = \frac{v_1 T}{v_1 + v_2}$$

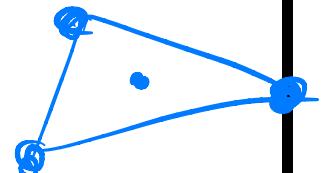
- So, $D = v_1 T_1 = v_2 T_2$.

A handwritten derivation shows the equation $D = v_1 T_1 = v_2 T_2$ being equated to $\frac{v_1 v_2 T}{v_1 + v_2}$. The term $v_1 T_1$ is circled in blue, followed by an equals sign, and then the fraction $\frac{v_1 v_2 T}{v_1 + v_2}$ is circled in blue.

$$D = \frac{v_1 v_2 T}{v_1 + v_2}$$

Location Awareness and GPS

- Location awareness has proven to be an important component in designing communication algorithms in ad hoc systems.
- The current Global Positioning System (GPS) is satellite based and determines the position of a GPS equipped device using the radiolocation method.
- However, there are instances where devices may not have GPS capability either because the signal is too weak (due to obstruction) or integration is impossible.
- Adding to these the fact that such devices are easy to jam and there have been calls to declare GPS critical infrastructure.



Modern Localization Techniques

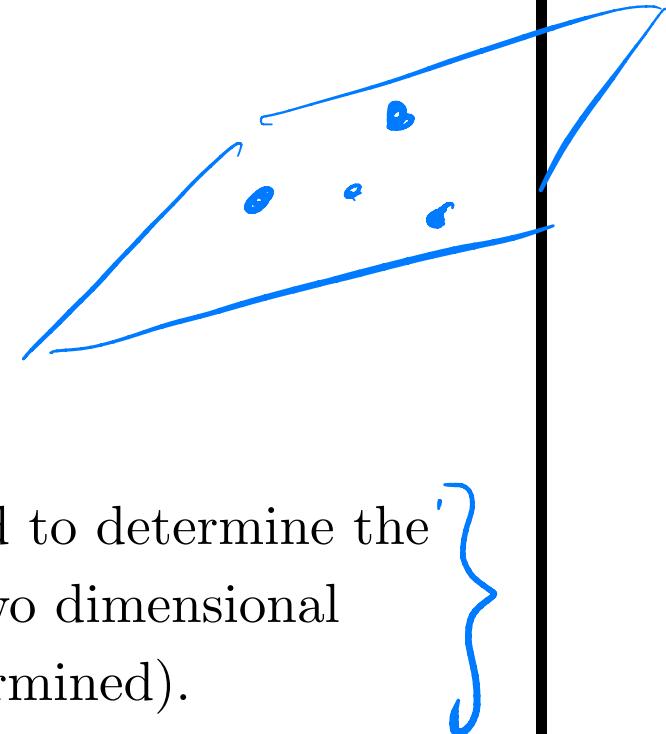
- **Network-based** techniques utilize the service provider's network infrastructure to identify the location of the handset.
- **Handset-based** technology requires the installation of client software on the handset to determine its location.
- **Hybrid-based** techniques use a combination of network-based and handset-based technologies for location determination (e.g., assisted-GPS, which uses both GPS and network information to compute the location).

Various Techniques

- Cell Identification:
accuracy depends on the known range of the particular network base station serving the handset at the time of positioning.
- Enhanced Cell Identification:
similar to Cell Identification, but for rural areas, with circular sectors of 550 meters.
- Distance Based:
TOA (Time of Arrival), TDOA (Time Difference of Arrival), AOA (Angle of Arrival).
- Assisted-GPS:
uses an operator-maintained ground station to correct for GPS errors caused by the atmosphere/topography.
- Many more . . .

Distance Based GPS Techniques

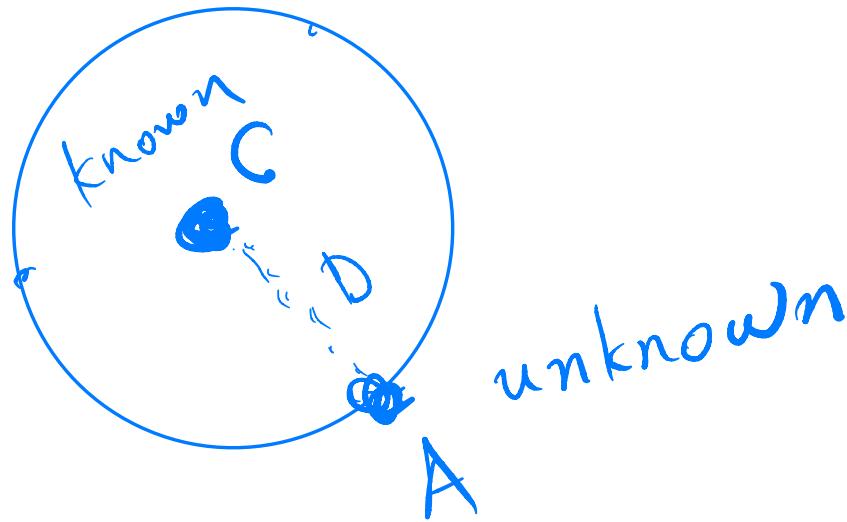
- Existing GPS techniques require line of sight propagation otherwise accuracy is affected.
 - TOA (Time of Arrival)
 - TDOA (Time Difference of Arrival).
 - AOA (Angle of Arrival)^a
 - Signal Strength
- Three position aware neighbors are required to determine the location of a position unaware node, in a two dimensional model (e.g. latitude and longitude are determined).
- Four neighbors are required in a three dimensional model (e.g. altitude is determined as well).



^aAOA won't discussed here

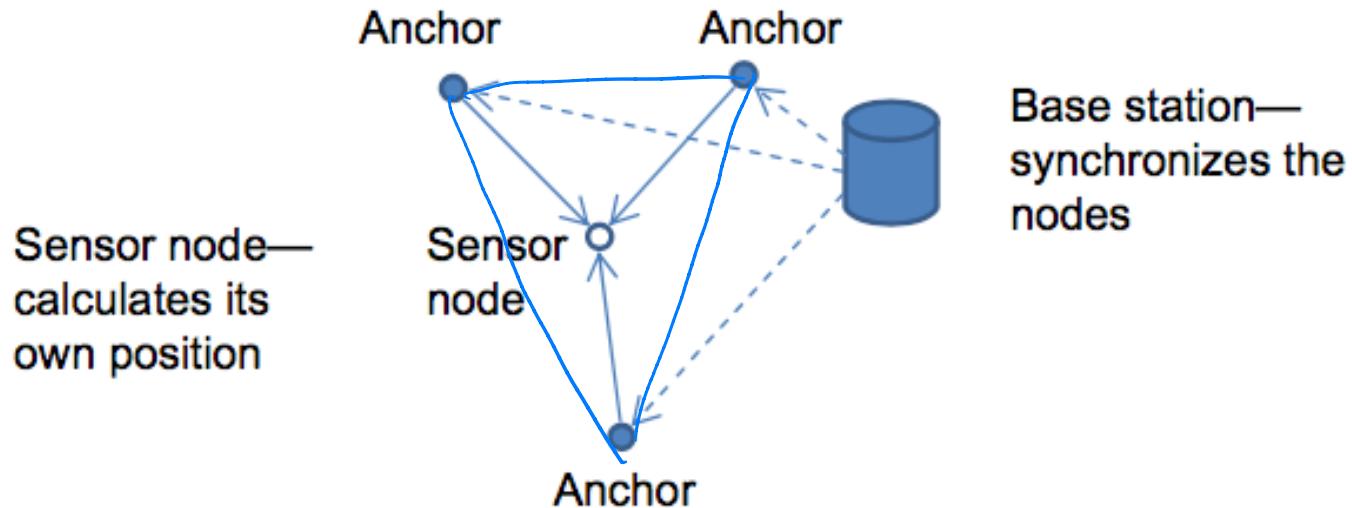
Common Features

- A sufficient number of nodes participate in a computation.
- Depending on the method: distances or angles are measured.
- Resulting system of equations is sufficient to determine locations.



Lateration

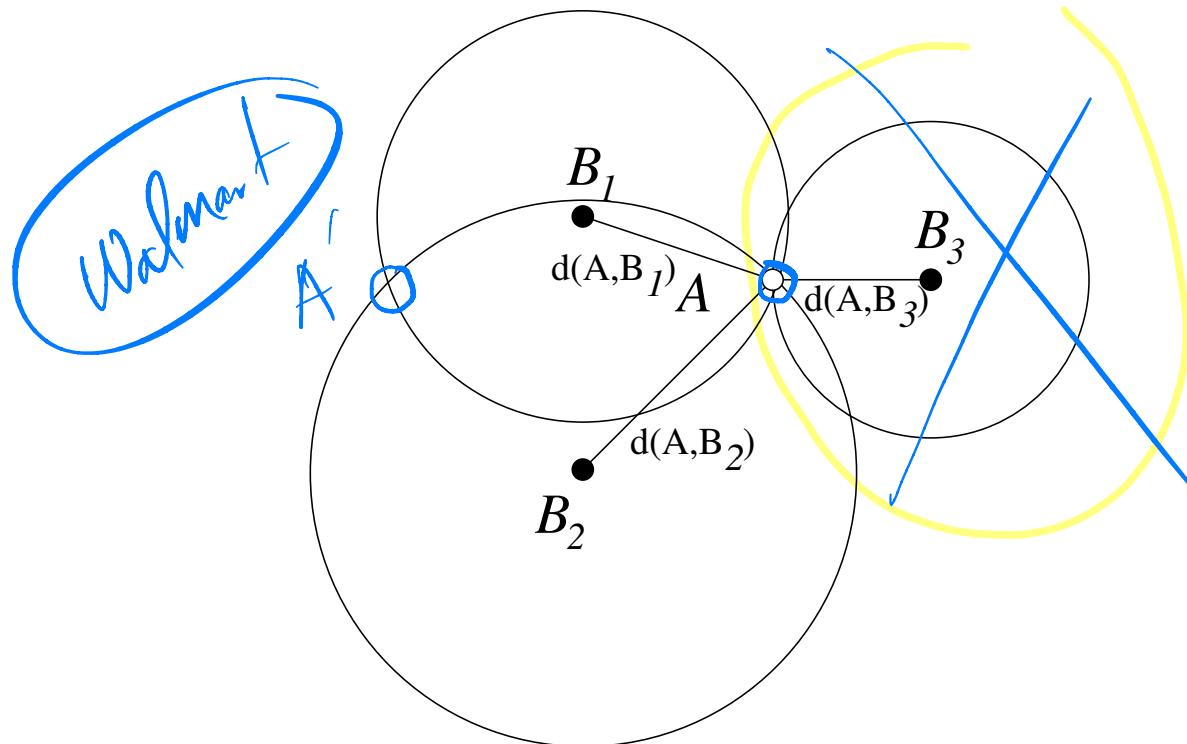
- Use a number of fixed anchor nodes at known positions:



- Anchors are synchronized to emit a signal at the same time.

The TOA Technique

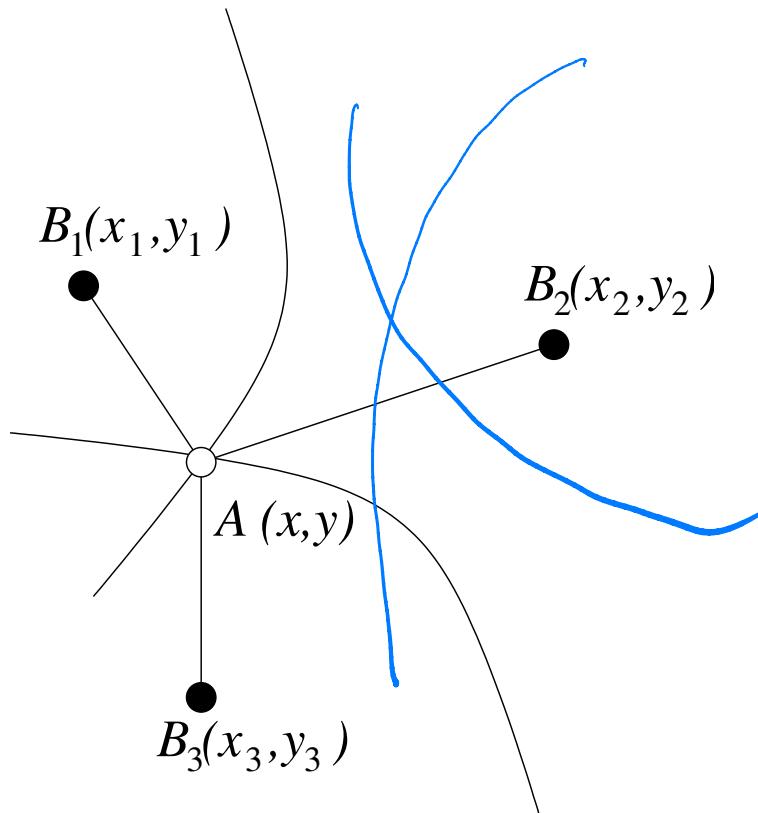
- “Vehicle” A computes its distance from fixed stations B_1, B_2, B_3 , resp.. A lies on circles centered at B_1, B_2, B_3 .



- A sensor at A not equipped with a GPS device can determine its position from the positions of its three neighbors B_1, B_2, B_3 .

The TDOA Technique

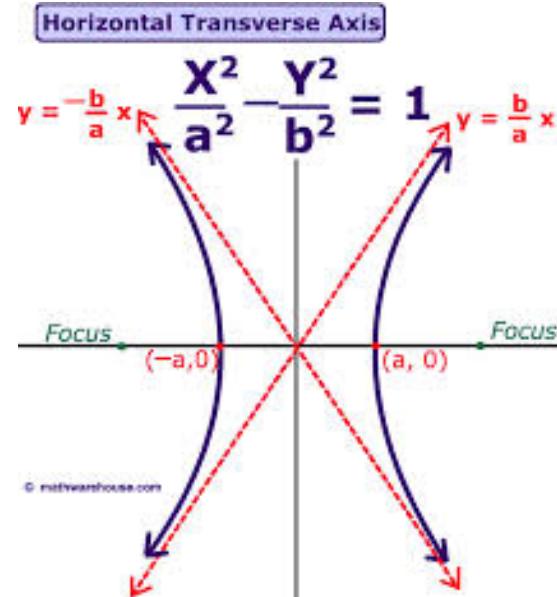
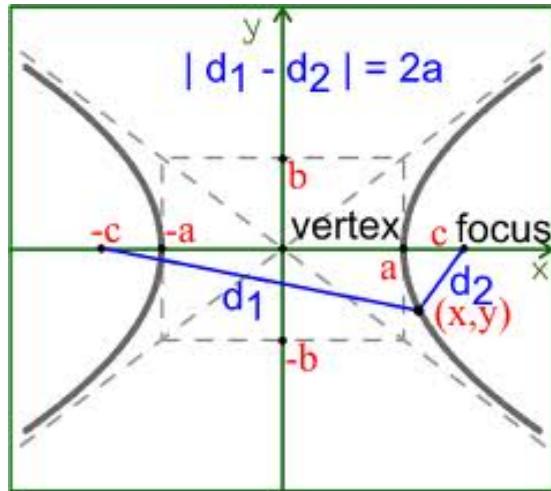
- Time difference $|t_1 - t_2|$ of arrivals t_1 and t_2 of signals from B_1 and B_2 , respectively, are measured by “vehicle” A .



- A lies on the hyperbola with foci the pair B_1 and B_2 .

The TDOA Technique: Why it works

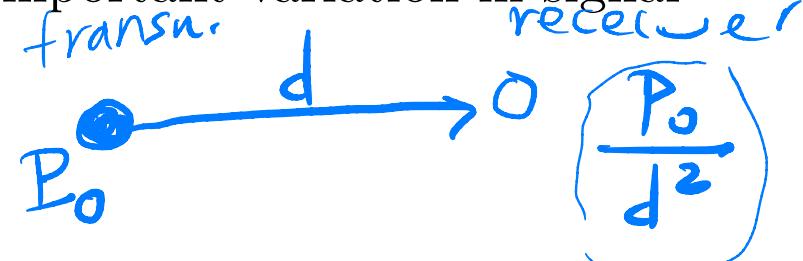
- Since the speed is known it is equivalent to measure difference in time and difference in distance!
- One measures the difference of arrivals



- This leads to a hyperbola!

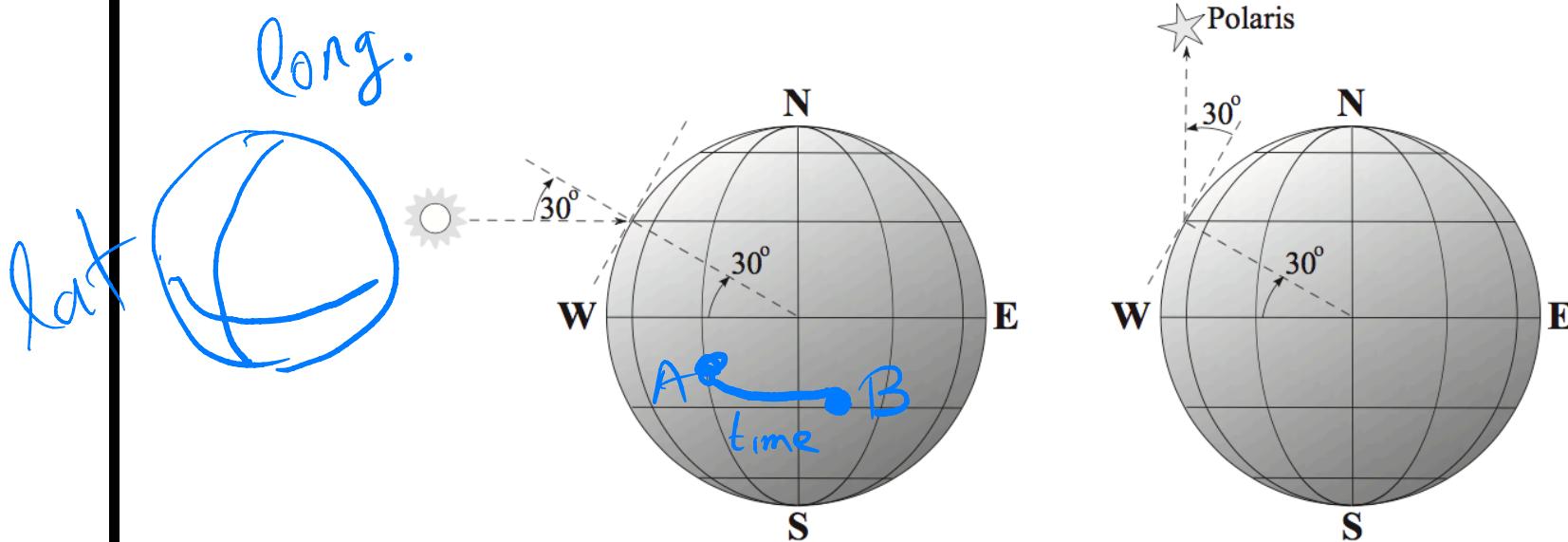
The Signal Strength Technique

- The signal strength based technique exploits the fact that a signal loses its strength as a function of distance.
 - Given the power of a transmitter and a model of free-space loss, according to the formula $\frac{P_0}{d^2}$ a receiver can determine the distance traveled by a signal.
 - If three different such signals can be received, a receiver can determine its position in a way similar to the TOA technique.
- The main criticism about the accuracy of the technique
 - is due to transmission phenomena such as multi path fading and shadowing that cause important variation in signal strength.



Finding your Latitude!

- Determining latitude from the sun or the Pole Star: Measured by ray shooting to the sun or to the Pole Star.



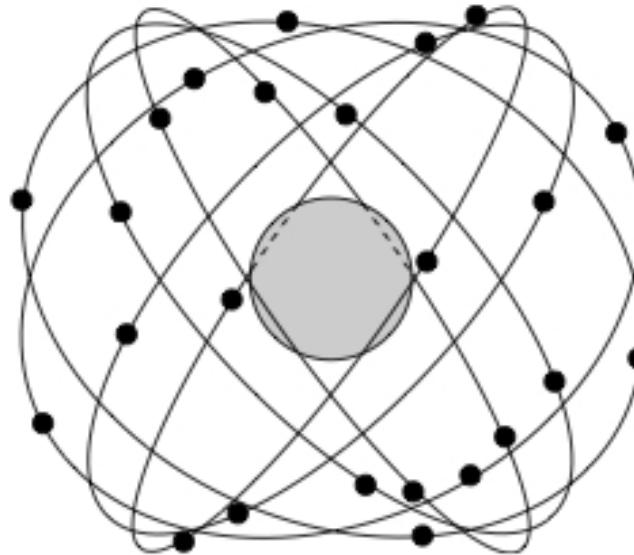
- At equinox, if the sun is due south at noon, a measured altitude of 60 (with the sun 30 from zenith) means that the latitude of the observer is 30.
- Measured altitude of Pole Star equals observer's latitude.

Finding your Longitude!

- That's not so easy!
- It was necessary to build accurate clocks!

GPS Satellite System

- Completed in July 1995 by the US Defense Department, and authorized for use by the general public.



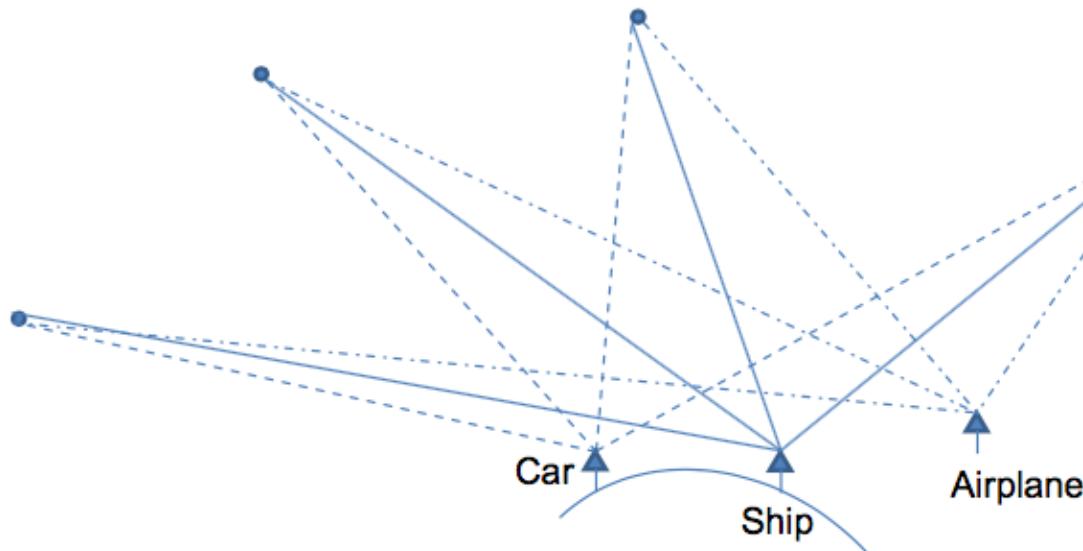
- When first deployed, it consisted of 24 satellites designed such that at least 21 would be functioning 98% of the time.
- There are several more GPS systems available today with varied accuracy in performance.

Uses of GPS

- Transportation
- Surveying
- Location Based services
- Map making
- Sports

GPS in Transportation

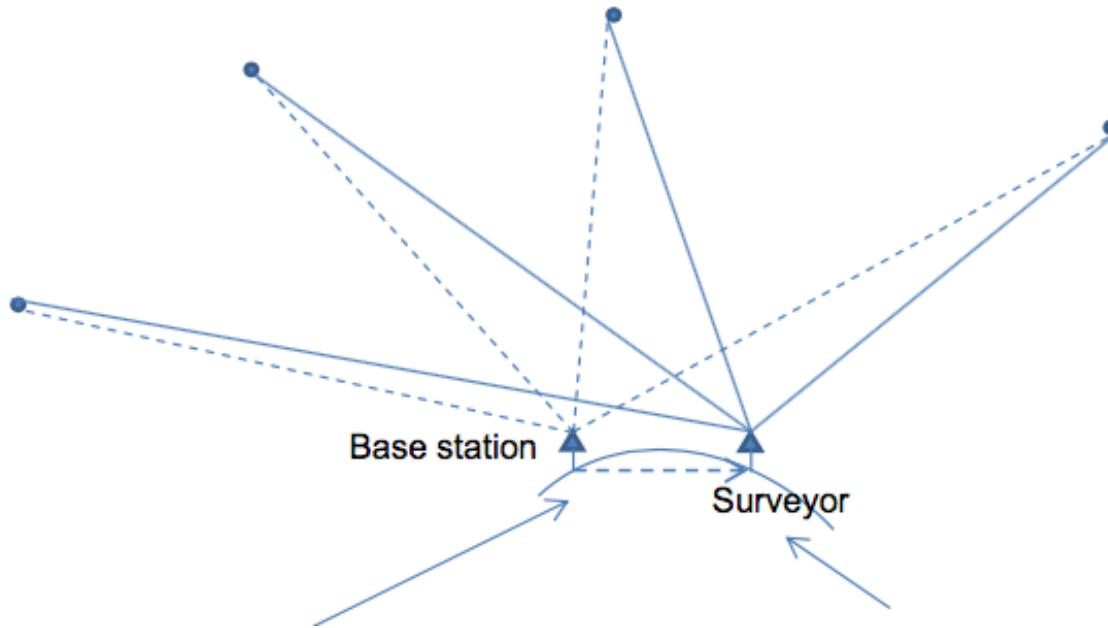
- Use of GPS in transportation



- Driving and other transportation uses—using devices installed in aircraft, cars, trucks, and ships.

GPS in Surveying

- Base station GPS receives satellite signals and hands them to a base station radio transmitter that broadcasts them.



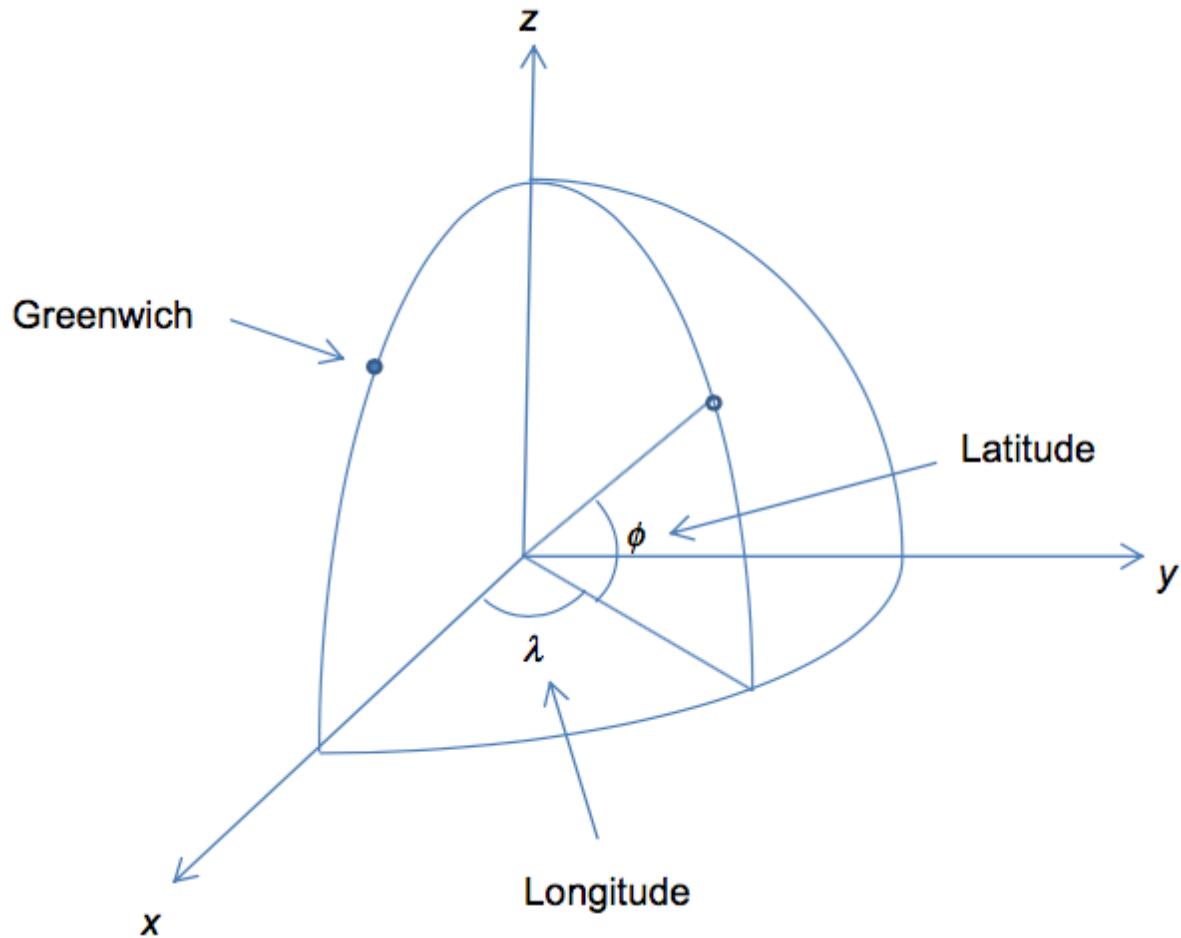
- Surveyor carries GPS antenna (for receiving satellite signals). Also carries a backpack containing a receiver connected to the antenna, as well as a radio receiver and radio (for receiving the base station's signals).

Satellites

- In 2005 the system had 32 satellites, of which at least 24 are to be functioning while the others are ready to take over in case a satellite fails.
- Satellites positioned 20,200 km from the surface of the Earth.
- Distributed across 6 orbital planes, each tilted at an angle of 55 degrees to the equatorial plane.
- At least 4 satellites per orbital plane, roughly equidistant from each other.
- Each satellite completes a circular orbit around the Earth in 11 hours and 58 minutes.
- Satellites are situated such that at any moment and at any location on Earth we may observe at least 4 satellites.

Geographic Coordinates

- Cartesian (x, y, z) and spherical (R, ϕ, λ) coordinates



Earth-Centered, Earth Fixed (ECEF) Coordinate System

- In practice the coordinate system used is geocentric but has fixed axes with respect to the Earth, and the axes rotate with Earth.
- It is a rotating frame of reference.
- The coordinates of any point on the Earth's surface are fixed.
- This coordinate system is called the ECEF frame.

How does the receiver calculate its position?

- Assume clocks of the receiver and all the satellites are perfectly synchronized.
- Receiver calculates its position through triangulation.
- The basic principle of triangulation methods is to determine where a person (object) is located by using some knowledge relating the position of the person (object) with respect to reference objects whose positions are known.
- In the case of the receiver of the GPS, it calculates its distance to the satellites, whose positions are known.

Receiver Measurements (1/2)

- The receiver measures the time t_1 it takes for the signal emitted from satellite P_1 to reach it.
- Given that the signal travels at the speed of light c , the receiver can calculate its distance from the satellite as $r_1 = ct_1$.
- The set of points situated at a distance r_1 from the satellite P_1 forms a sphere S_1 centered at P_1 with radius r_1 .
- So we know that the receiver is on S_1 . Consider these points as defined in a Cartesian coordinate system.
- If (x, y, z) is the unknown position of the receiver and (a_1, b_1, c_1) the known position of the satellite P_1 then (x, y, z) must satisfy the equation describing points on the sphere S_1 ,

$$(x - a_1)^2 + (y - b_1)^2 + (z - c_1)^2 = r_1^2 = c^2 t_1^2. \quad (1)$$

Receiver Measurements (2/2)

- This piece of information is insufficient to determine the precise position of the receiver.
- But the receiver can repeat the same procedure with two more satellites: P_2, P_3 having positions (a_2, b_2, c_2) and (a_3, b_3, c_3) .

$$(x - a_2)^2 + (y - b_2)^2 + (z - c_2)^2 = r_2^2 = c^2 t_2^2. \quad (2)$$

and

$$(x - a_3)^2 + (y - b_3)^2 + (z - c_3)^2 = r_3^2 = c^2 t_3^2. \quad (3)$$

- Equations (1, 2, 3) are a system of 3 equations with 3 unknowns.

$$(x - a_1)^2 + (y - b_1)^2 + (z - c_1)^2 = r_1^2 = c^2 t_1^2$$

$$(x - a_2)^2 + (y - b_2)^2 + (z - c_2)^2 = r_2^2 = c^2 t_2^2$$

$$(x - a_3)^2 + (y - b_3)^2 + (z - c_3)^2 = r_3^2 = c^2 t_3^2.$$

Reducing the System

- The equations of this system are quadratic, not linear, which complicates the solution.
- We can replace the system by an equivalent system obtained by replacing the first equation by the difference $(1) - (3)$ and the second equation by the difference $(2) - (3)$ and by keeping the third equation

$$2(a_3 - a_1)x + 2(b_3 - b_1)y + 2(c_3 - c_1)z = A_1, \quad (4)$$

$$2(a_3 - a_2)x + 2(b_3 - b_2)y + 2(c_3 - c_2)z = A_2, \quad (5)$$

$$(x - a_3)^2 + (y - b_3)^2 + (z - c_3)^2 = r_3^2 = c^2 t_3^2, \quad (6)$$

where

$$A_1 = c^2(t_1^2 - t_3^2) + (a_3^2 - a_1^2) + (b_3^2 - b_1^2) + (c_3^2 - c_1^2),$$

$$A_2 = c^2(t_2^2 - t_3^2) + (a_3^2 - a_2^2) + (b_3^2 - b_2^2) + (c_3^2 - c_2^2).$$

Non-Linearity

- By orbital design the satellites have been placed in such a manner that no three

$$(a_1, b_1, c_1), (a_2, b_2, c_2), (a_3, b_3, c_3)$$

will ever fall along a line.

- Using the system of Equations (4), (5), and (6) and linear algebra, this ensures that at least one of the 2×2 determinants

$$\begin{vmatrix} a_3 - a_1 & b_3 - b_1 \\ a_3 - a_2 & b_3 - b_2 \end{vmatrix}, \begin{vmatrix} a_3 - a_1 & c_3 - c_1 \\ a_3 - a_2 & c_3 - c_2 \end{vmatrix}, \begin{vmatrix} b_3 - b_1 & c_3 - c_1 \\ b_3 - b_2 & c_3 - c_2 \end{vmatrix}$$

is not zero.

- In fact, if all three determinants were zero, then the vectors (depicted in the determinants) would be collinear, implying that the three points (i.e., satellites) P_1, P_2, P_3 fall on a line.

Solution (1/2)

- Using Cramer's Rule in linear system (4), (5), and (6), we see

$$x = \frac{\begin{vmatrix} A_1 - 2(c_3 - c_1)z & 2(b_3 - b_1) \\ A_2 - 2(c_3 - c_2)z & 2(b_3 - b_2) \end{vmatrix}}{\begin{vmatrix} 2(a_3 - a_1) & 2(b_3 - b_1) \\ 2(a_3 - a_2) & 2(b_3 - b_2) \end{vmatrix}}$$

$$y = \frac{\begin{vmatrix} 2(a_3 - a_1) & A_1 - 2(c_3 - c_1)z \\ 2(a_3 - a_2) & A_2 - 2(c_3 - c_2)z \end{vmatrix}}{\begin{vmatrix} 2(a_3 - a_1) & 2(b_3 - b_1) \\ 2(a_3 - a_2) & 2(b_3 - b_2) \end{vmatrix}}$$

- Substituting x, y into Equation (3) yields a quadratic equation in z , which we solve to find the two solutions z_1, z_2 .

Solution (2/2)

- Back-substituting z for the values z_1 and z_2 into the two above equations yields the corresponding values x_1, x_2, y_1, y_2 .
- We could easily find closed forms to these solutions, but the formulas involved quickly become too large to offer any insight or convenience.

Relativistic Effects (1/3)

- Calculations relating to special relativity (SR) and general relativity (GR) effects have to be carried out.
- The speed of the satellites is sufficiently large that all of the calculations must be adapted to account for the effects of special relativity.
- The clocks on the satellites are traveling very fast compared to those on Earth.
- SR theory predicts that these clocks will run slower than those on Earth.
- The satellites are in relatively close proximity to the Earth, which has significant mass.
- GR predicts a small increase in the speed of the clocks on board the satellites.

Relativistic Effects (2/3)

- While the ECEF frame is useful for navigation, many physical processes are easier to describe in the inertial reference frame.
- A point in the inertial frame is denoted by cylindrical space-time coordinates (t, r, ϕ, z) .
- The point in ECEF is denoted by (t', r', ϕ', z') .
- The coordinates are related to one another as follows:

$$t- = t', r = r', \phi = \phi' + \omega_E t', z = z',$$

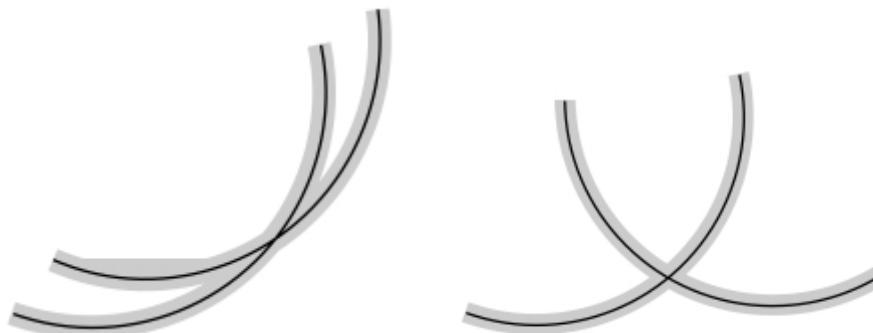
where ω_E is the uniform angular velocity of the Earth.

Relativistic Effects (3/3)

- The velocity of a satellite clock is relatively small and the gravitational fields near the Earth are relatively weak: Both these aspects, however, cause significant relativistic effects.
- The fundamental concept upon which GPS is based is that the speed of light, c , is constant.
- Satellites contain clocks stable to about 4 ns over one day. At the speed of light, a 1 ns error is about 30 cm . If speed of light varied then a GPS measurement would be out by $\geq 30\text{ cm}$.
- Calculations, take account of the gravitational fields near the Earth due to the Earth's own mass. The relevant expression in the amended version of the solution of Einstein's field equations involves a number of components, including the Earth's quadrupole moment coefficient and centripetal potential.

Many Other Issues

1. Layers that surround the Earth: Ionosphere, Troposphere.
 - Refraction, Reflection effects.
2. Satellites and receivers may not be perfectly in sync.
 - Satellites are laid out such that no four of them that are visible from a given point on the Earth will ever lie in the same plane. Use extra equation to correct clock offsets!
3. Which satellites should I choose if I can see more than four?
 - Choose the spheres that minimize the errors, i.e., that intersect each other at as large an angle as possible



Exercises^a

1. The power of a signal attenuates according to the inverse cubic law $P(d) = P(0)/d^3$, where $d > 0$ is the distance, $P(d)$ is the power at distance d , and $P(0)$ is its power at the start. How far can a signal reach if its power at distance d has to be at least $1/8$ its power at the start?
2. Due to the presence of obstacles, the power of a signal attenuates according to the inverse a th power law $P(d) = P(0)/d^a$, where $d > 0, a > 1$ is the distance, $P(d)$ is the power at distance d , and $P(0)$ is its power at the start. If the power at distance $d = 1$ is 8, up to what distance d is the power of the signal at least $1/10$ its power at the start?
3. Two stations located at A and B transmit wireless signals simultaneously and against each other. The signal at station A

^aDo not submit!

has speed u and the signal at station B has speed v . Determine the point at which the two signals collide.

- (a) Do the same exercise as above when the signals are transmitted with a time difference $\Delta t > 0$.

Location Awareness (Route Discovery in Ad-Hoc Networks)

Outline

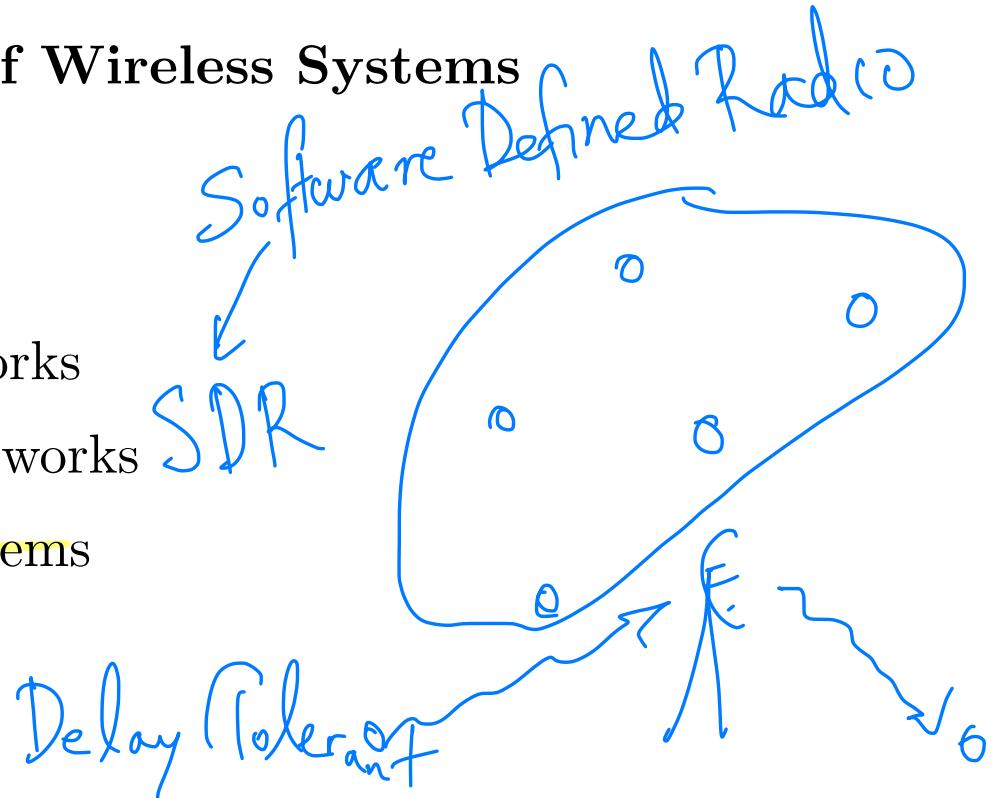
- Introduction
- Models
- Gabriel Test
- Geometric Routing
 - Compass Routing
 - Face Routing

⌚ Bluetooth networks

Introduction

Lots of Wireless Systems

- Wireless systems from
 - Piconets
 - Home/Office Networks
 - (Packet) Radio Networks
 - Cellular phone systems
 - Sensor Networks
 - Satellite networks



have become all too pervasive in our everyday lives.

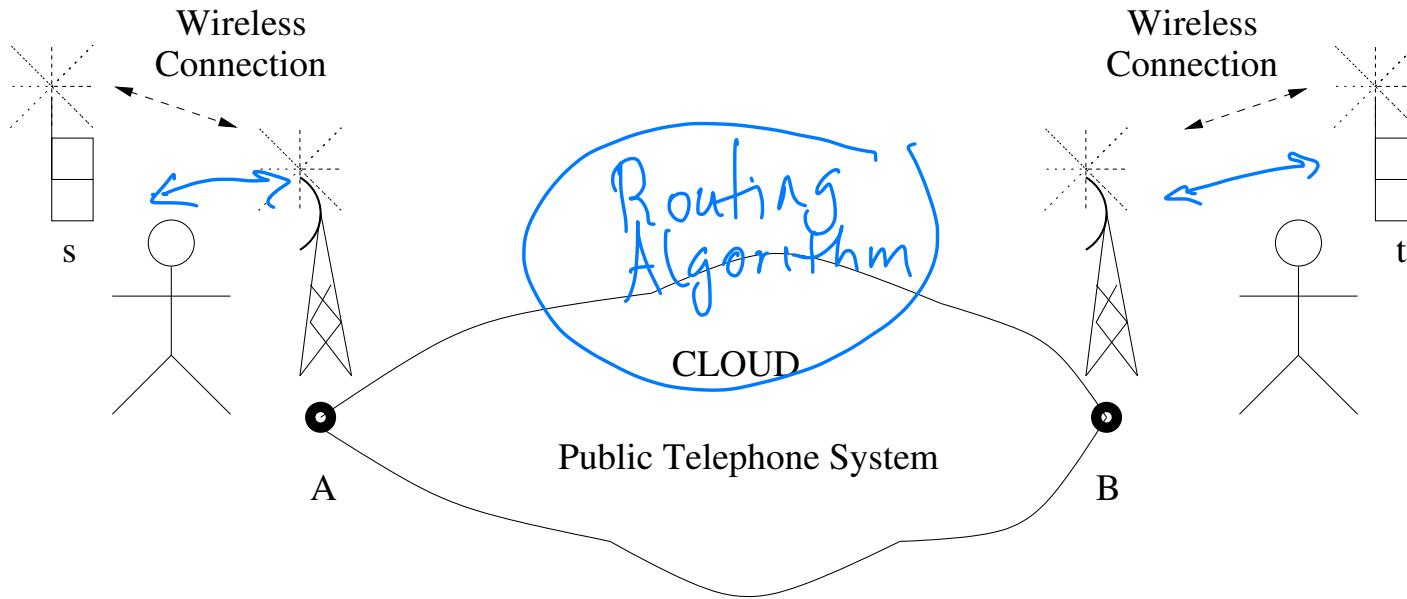
- **Questions:**

- How do you discover a **route** in such a wireless system?
 - Is there a general method to find a route?

We will see later Routing Principles

The Way it is

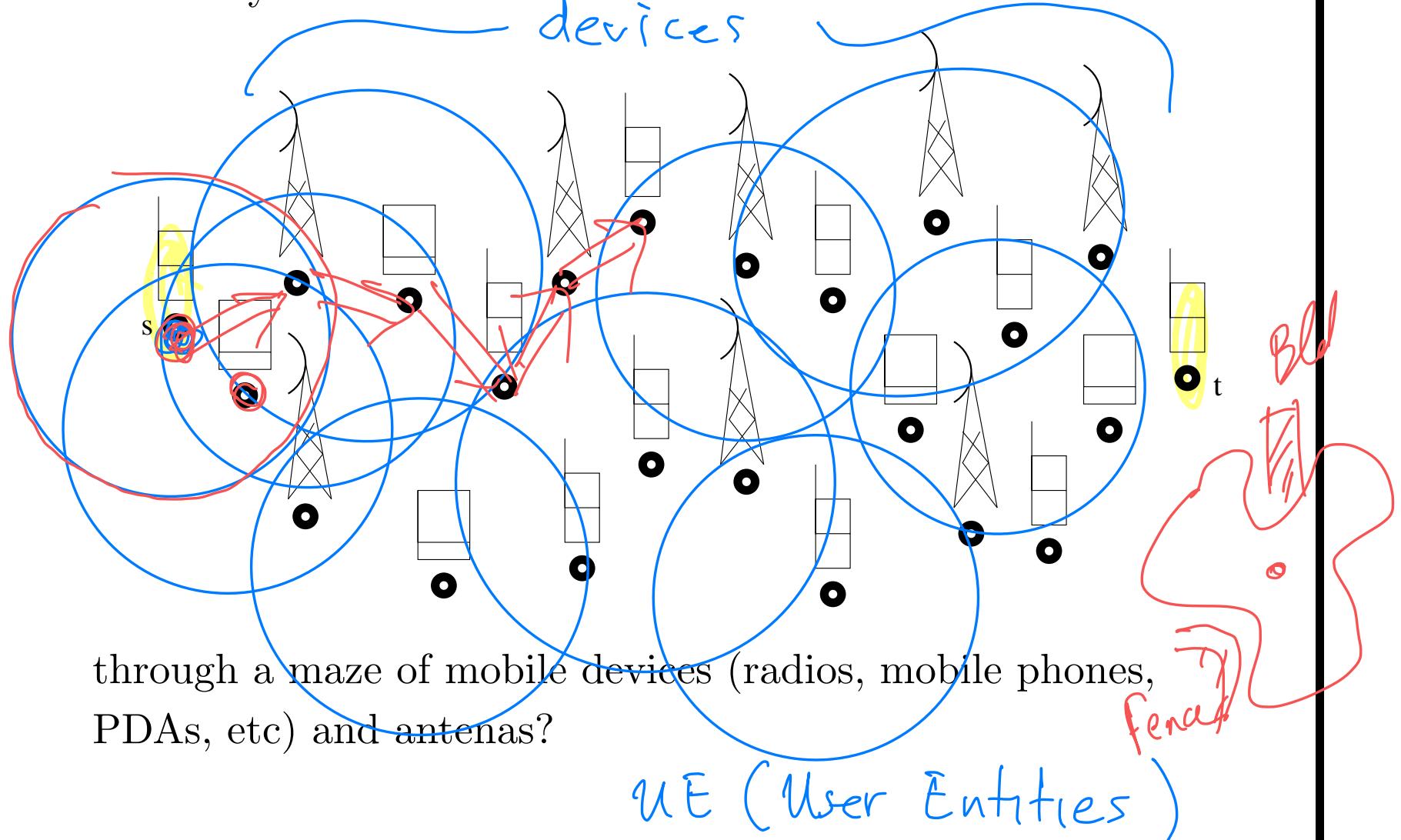
- How do you discover a route from a source s to a destination t ?

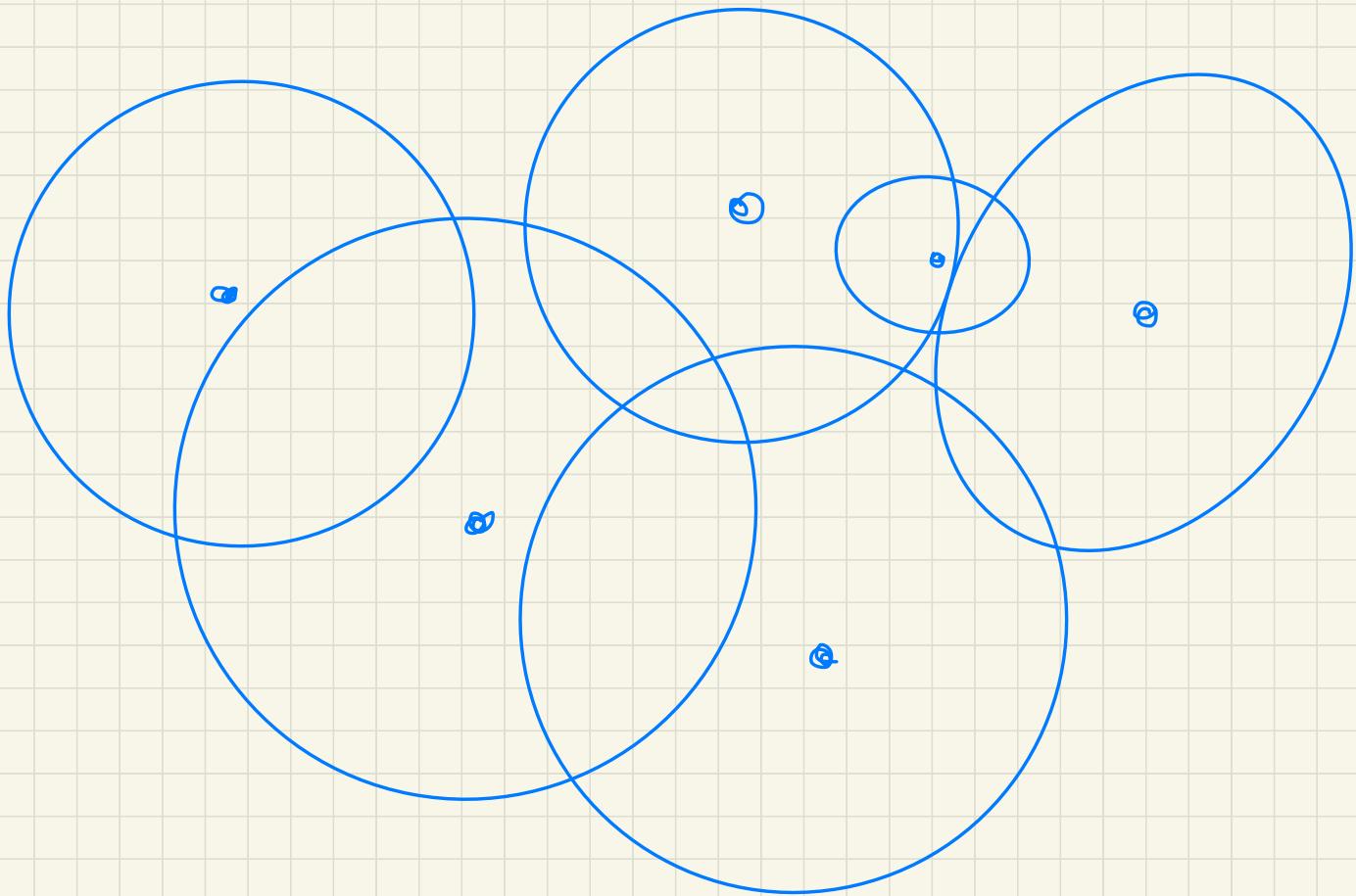


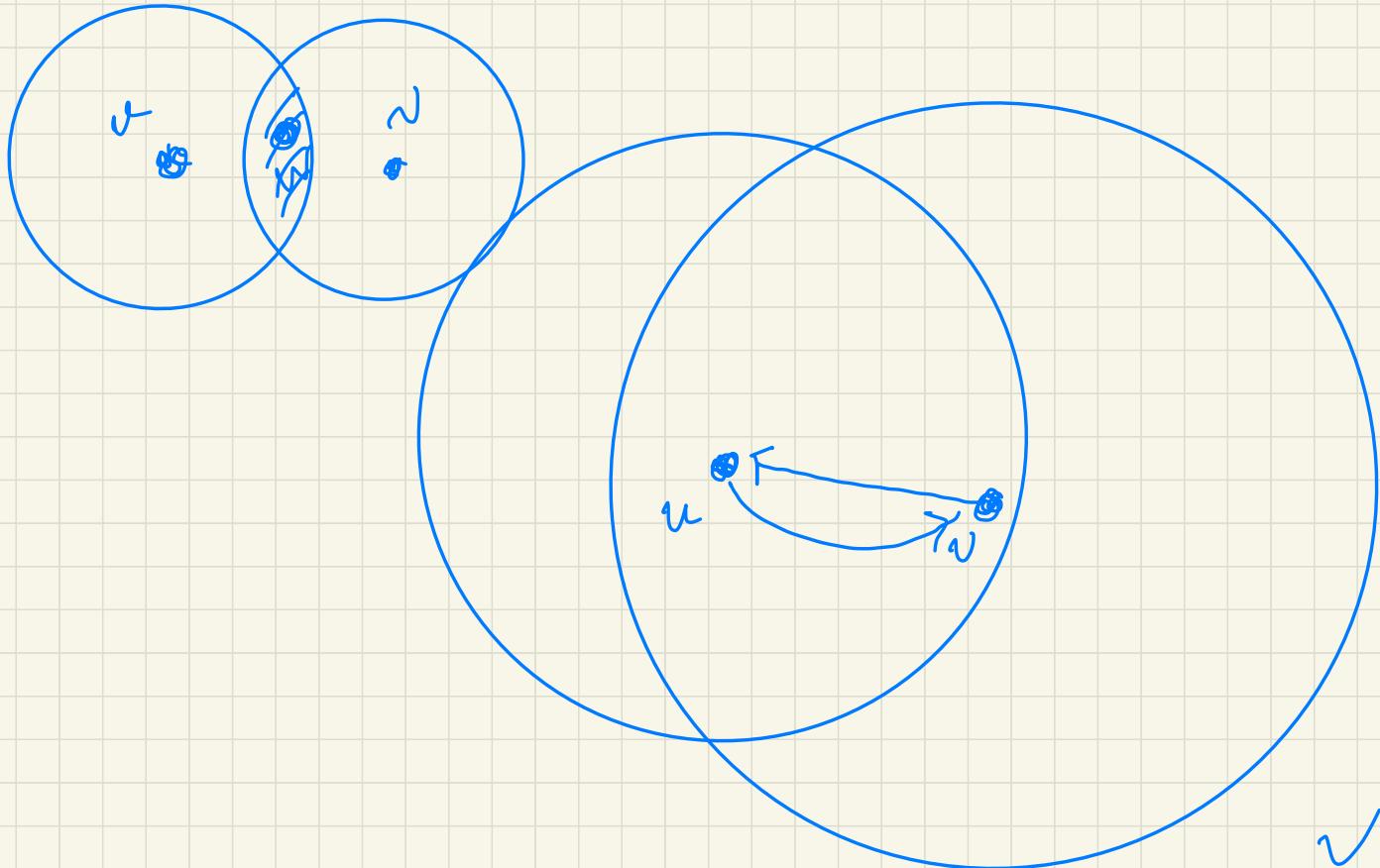
- **Simple:** 1. s sends message to base station A . 2. The public phone system transmits it to base station B . 3. Base station B transmits it to t .

Routing Data from s to t in a Wireless Ad-Hoc System

- How do you discover a route from a source s to a destination t

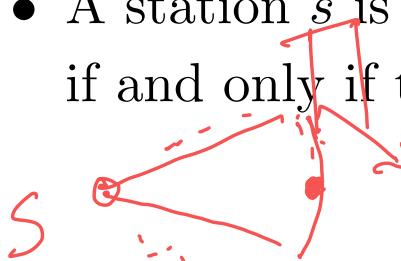






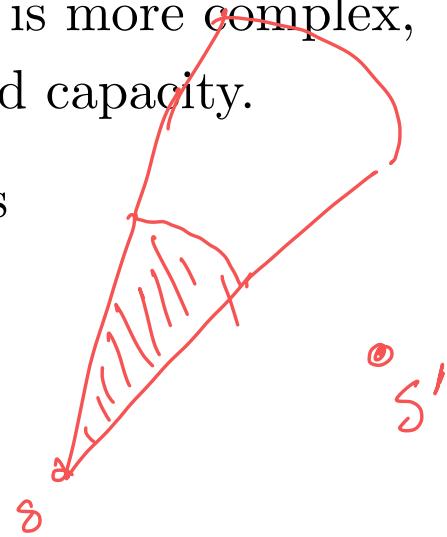
Realistic Models

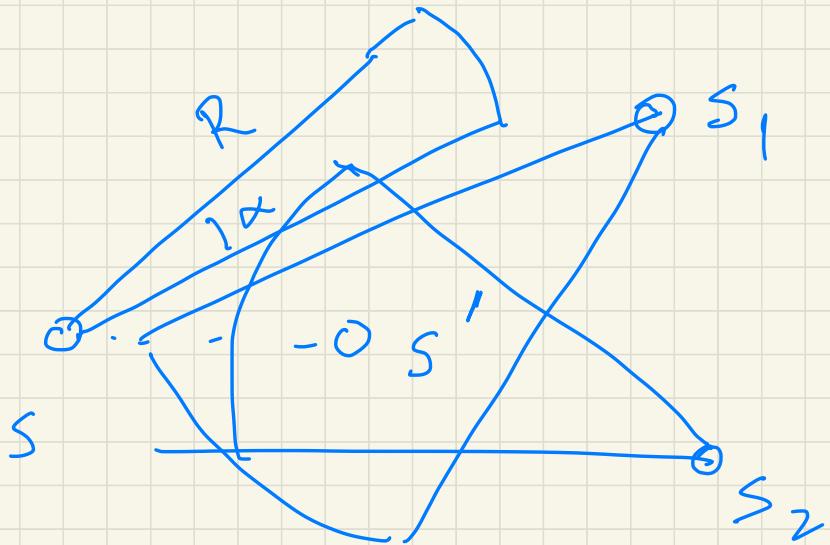
Complexity in Models for Wireless Communication

- Traditional (wired, point-to-point) communication networks can be described satisfactorily using a graph representation.
- A station s is able to transmit a message to another station s' if and only if there is a wire connecting the two stations.
- Accurately representing a wireless network is considerably harder, since it is nontrivial to decide whether a transmission by a station s is successfully received by another station s' .
- This may depend on the positioning and activities of s and s' , and on other nearby stations, whose activities might interfere with the transmission and prevent its reception

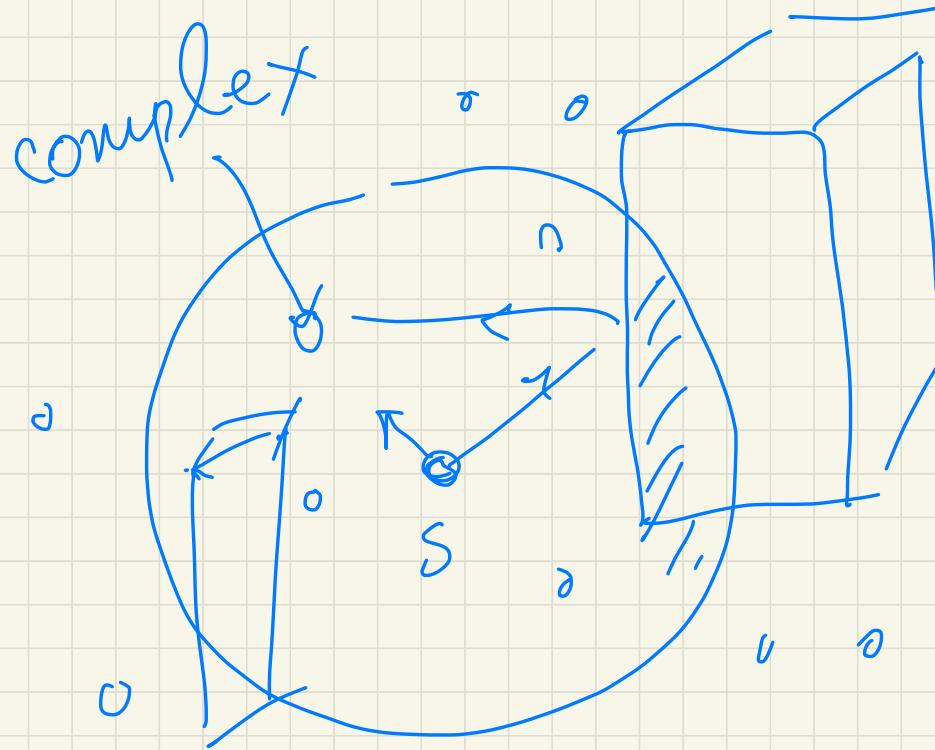
A Lot of Factors

- This means that a transmission from s may reach s' in some settings but fail to reach it under other settings.
- Moreover, the question of successful reception is more complex, since connections can be of varying quality and capacity.
- There are many other relevant factors, such as
 - the presence of physical obstacles,
 - the directions of the antennae at s and s'
 - the weather, and more.
- Obtaining an accurate solution taking all of those factors into account involves solving the corresponding Maxwell equations.
- Since this is usually far too complicated, the common practice is to resort to approaches based on approximation models.





S' may not have control
of who is transmitting
to S'



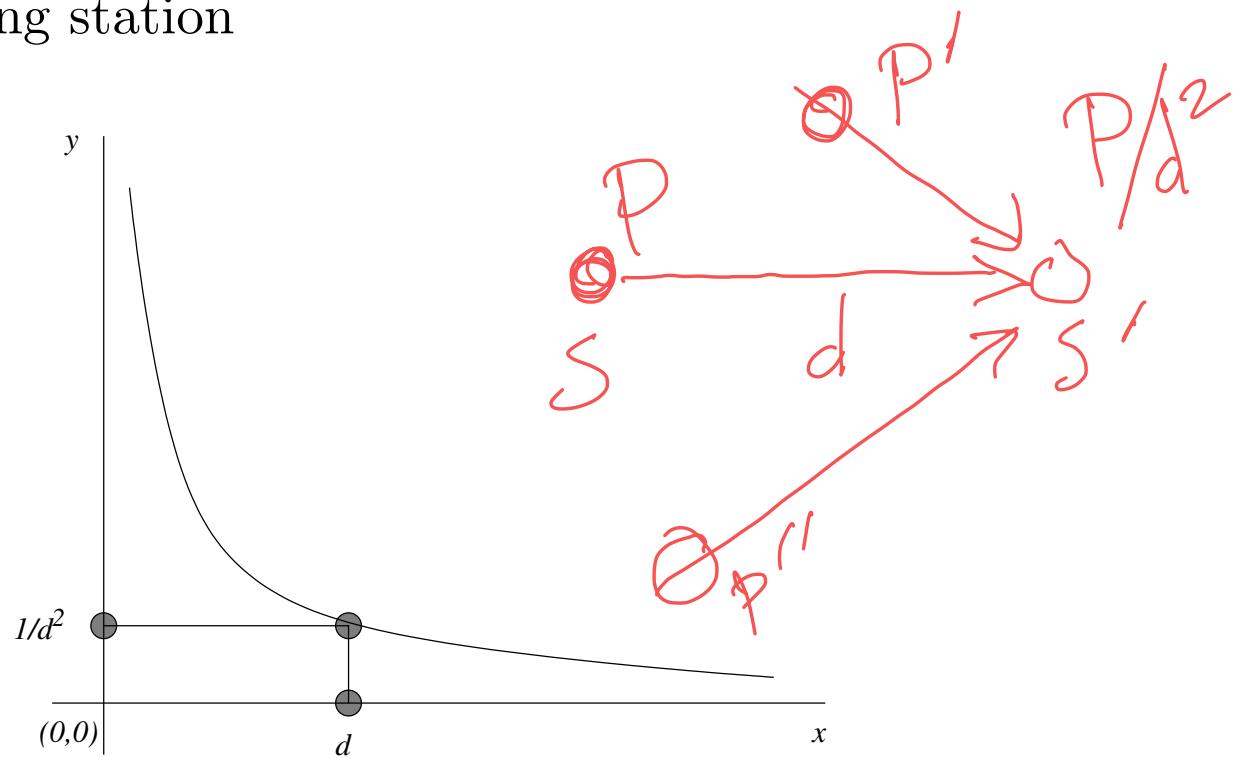
Noises:

Gaussian noise

Other noise

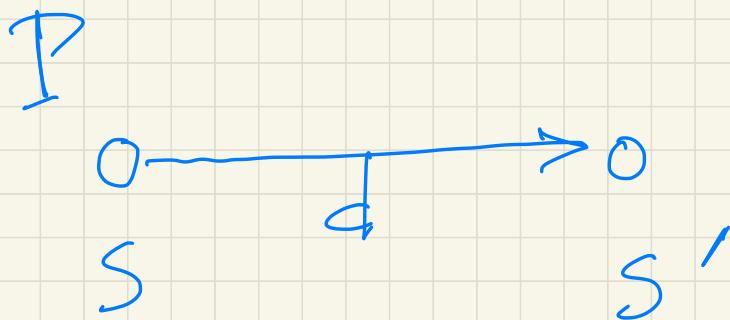
A Fact of Life: Power Assignments

- When a sensor transmits to another sensor located at distance d from the transmitting sensor, the power of the signal at the receiving station is P/d^2 , where P is the power of the signal at the transmitting station



- What does it mean when $d = 0$?

s : sends a signal to s'



s' : receives a signal whose strength is $\frac{P}{d^a}$

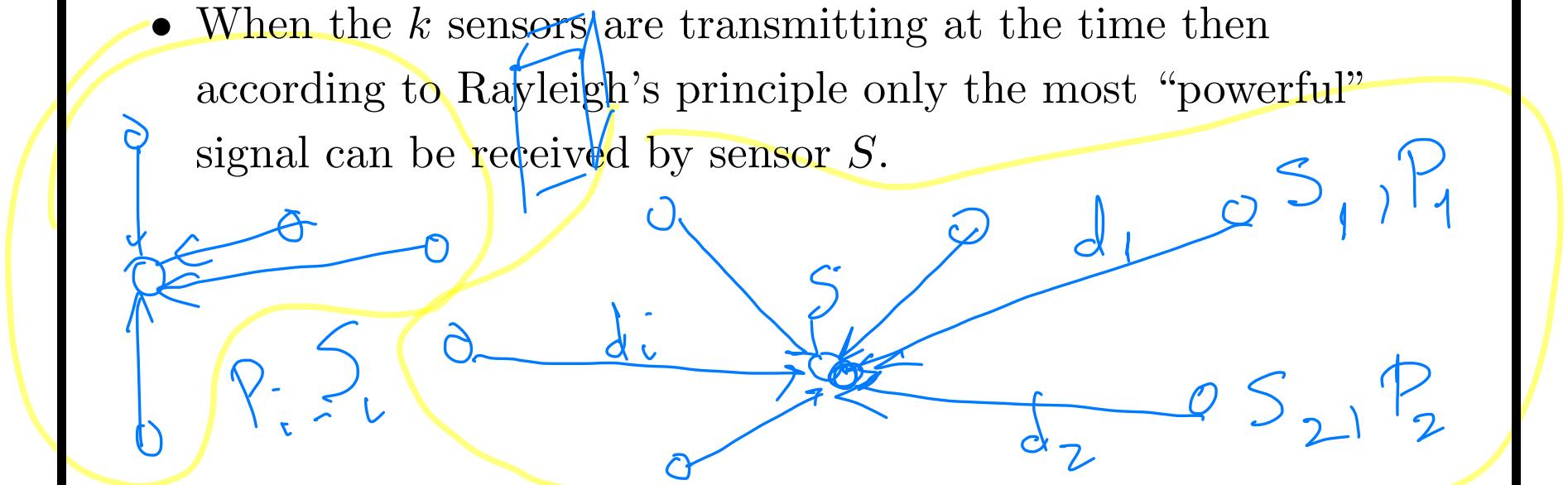
$d = \text{distance}(s, s')$

$a \geq 2$

$\frac{P}{d^a}$

Rayleigh's Principle: Physical Model

- Consider the setting whereby sensors S_1, S_2, \dots, S_k and S are located in the plane and suppose that sensor S_i is at distance d_i from S .
- When the signal from a transmitting station S_i reaches the sensor S it will have power P_i/d_i^2 , where P_i is the power of the transmitted signal at S_i .
- When the k sensors are transmitting at the time then according to Rayleigh's principle only the most “powerful” signal can be received by sensor S .



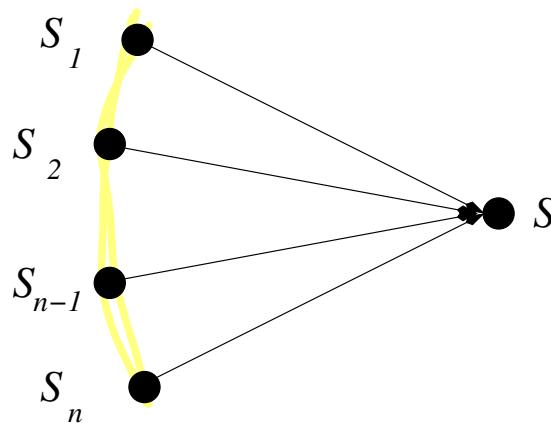
Rayleigh's Principle

- The signal from a sensor S_i , for some i , will be received by sensor S if and only if there is a threshold $\lambda > 0$ s.t.

$$\frac{P_i}{d_i^2} > \lambda \left(N + \sum_{j=1, j \neq i}^n \frac{P_j}{d_j^2} \right),$$

ambient noise N

which depends on technical considerations, like, sensor equipment sensitivity, and N is ambience noise.

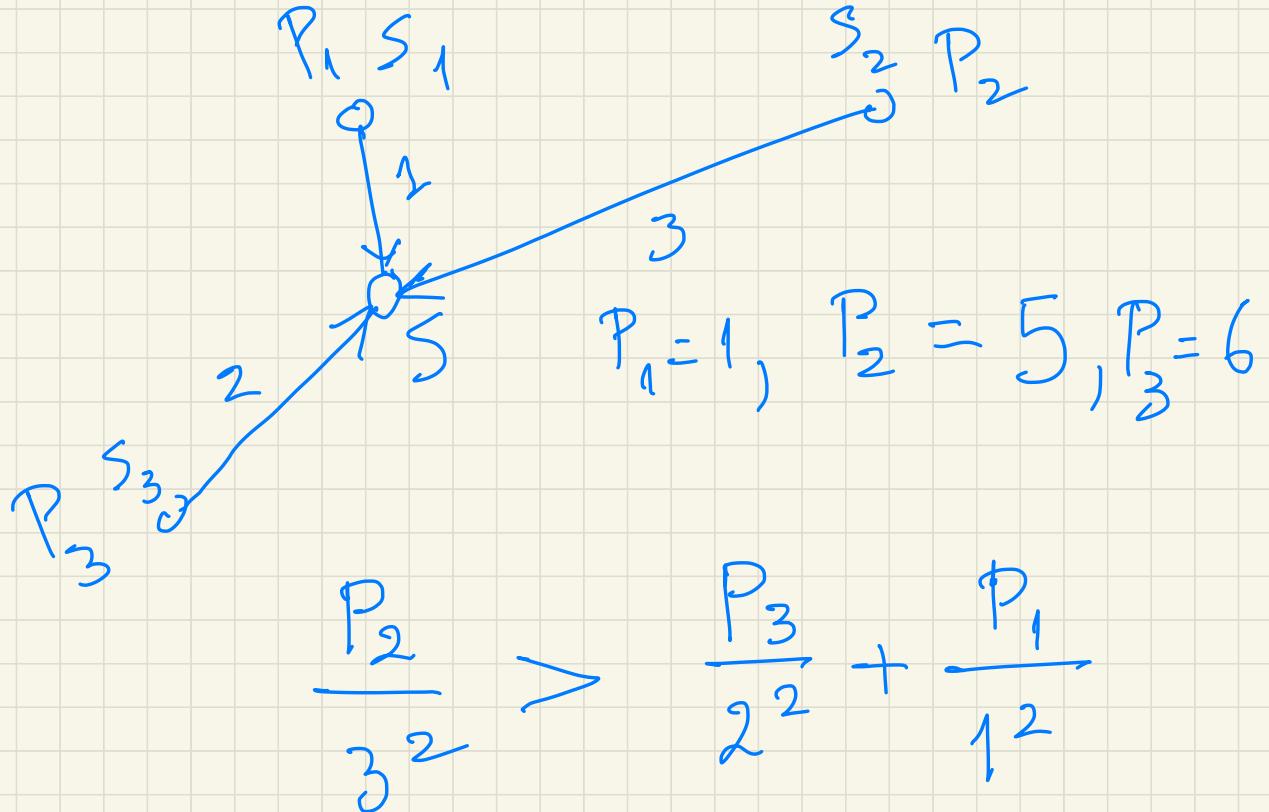


- Usually, to simplify notation, we assume that $\lambda = 1, N = 0$.

Example

$$\lambda = 1$$

$$N=0$$



$$\frac{P_3}{g^2} > \frac{P_2}{3^2} + \frac{P_1}{1^2}$$

$$\frac{5}{9} > \frac{6}{4} + \frac{1}{1}$$

$$\frac{6}{4} > \frac{5}{9} + 1$$

$$P_1 = 1, P_2 = 20, P_3 = 4$$

$$\frac{20}{9} > \frac{4}{4} + 1$$

S_3

S_2

S_1

SINR (Signal-to-Interference & Noise Ratio)

- This formula represents a rather general model concerning the allowed transmission power, referred to as the **power control model**, in which each station can control the power with which it transmits.
- A simpler (and weaker) model is the uniform wireless network model, which assumes that all transmissions use the same transmission power, i.e., $P_i = 1$ for every i .

If all MEs are identical,
say $P_i = 1$, H_i

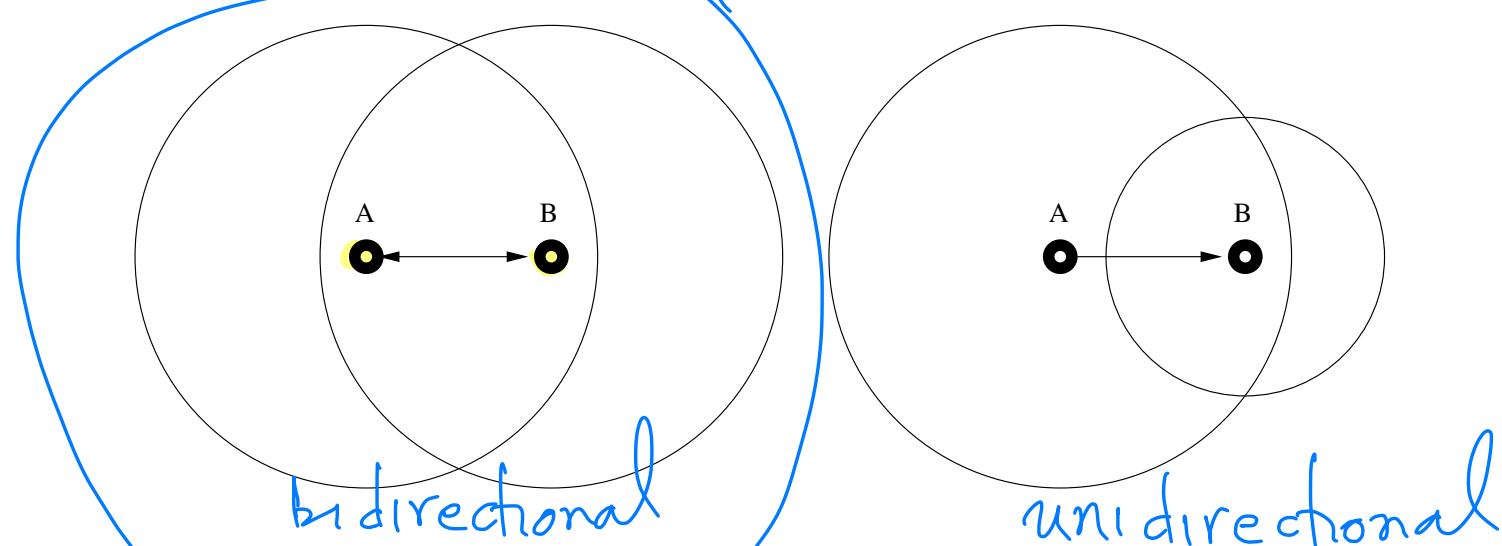
Rayleigh's
Formula

$$\frac{1}{d_i^2} > \sum_{k \neq i} \frac{1}{d_k^2}$$

Idealized Models

Protocol Model: Equal Power Assumption!

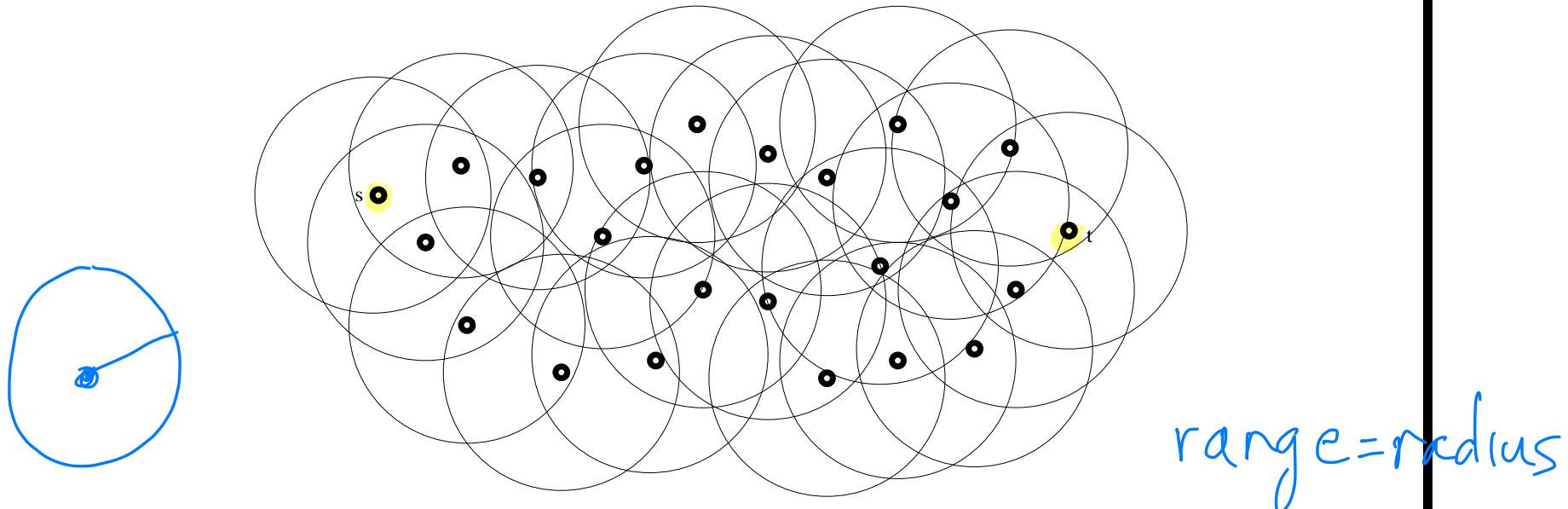
- The disk indicates an omnidirectional antenna!
- It is rare for two stations to have the same signal power!



- To simplify things stations are assumed to have equal power!
- In the left picture A can reach B and B can reach A .
- In the right picture A can reach B but B cannot reach A .

From Mobile Devices to Circles

- Assuming the equal power assumption for the signals...



...a group of circles is formed that determines network connectivity, i.e. who can reach whom!

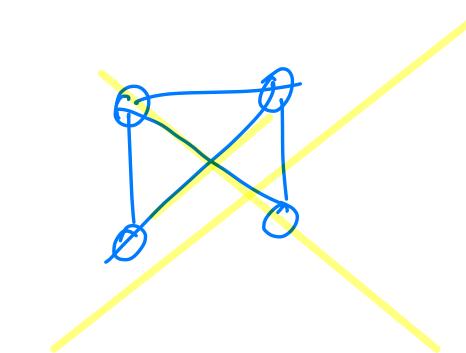
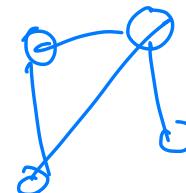
- Note that the circles have equal radius, say r : two hosts can communicate with each other if and only if their distance is at most r .

For the sake of discovering routes...

We will show...

1. ...how from a network of circles we can produce a simplified network in which edges have no crossings, and
2. how to discover routes in networks with non-crossing edges.

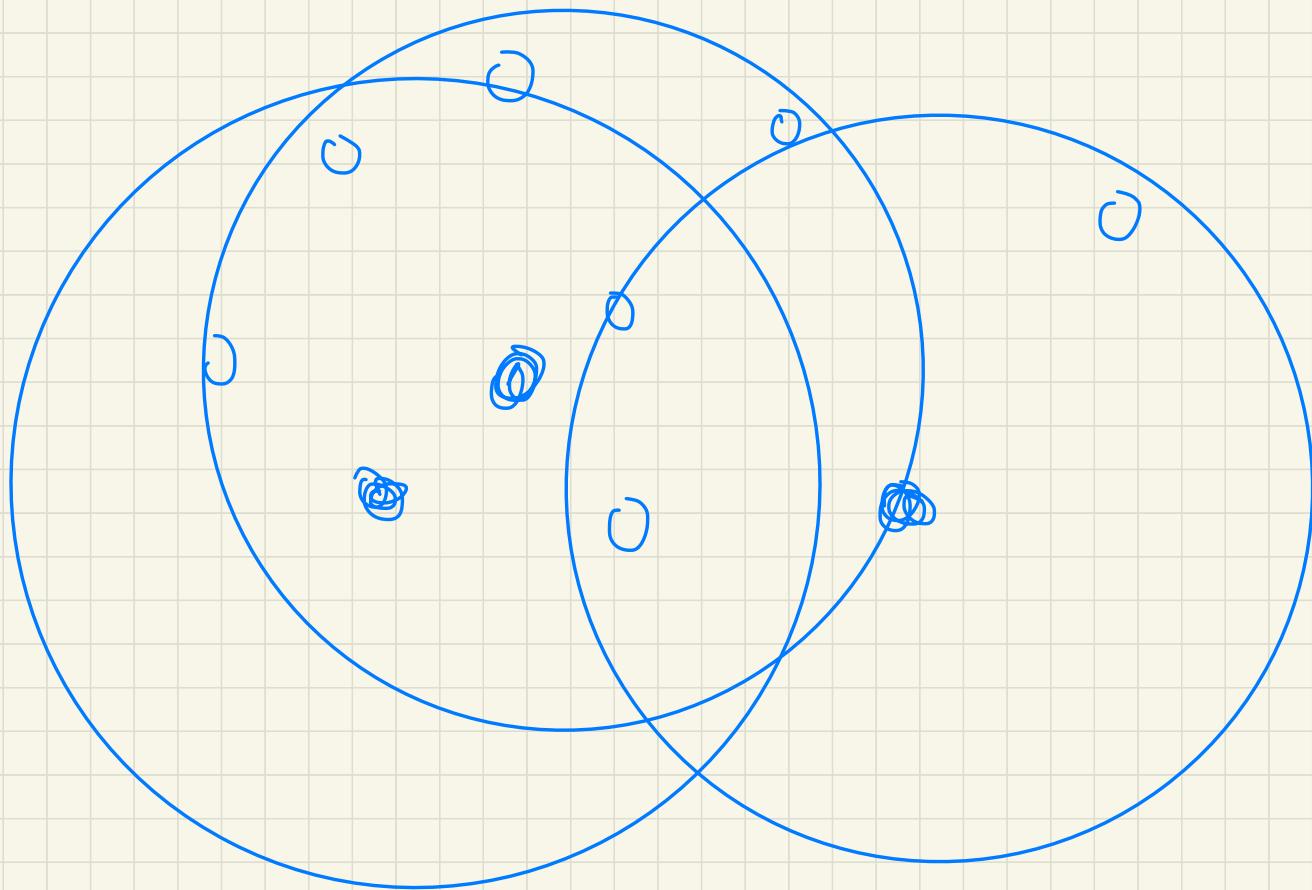
planar graph



protocol

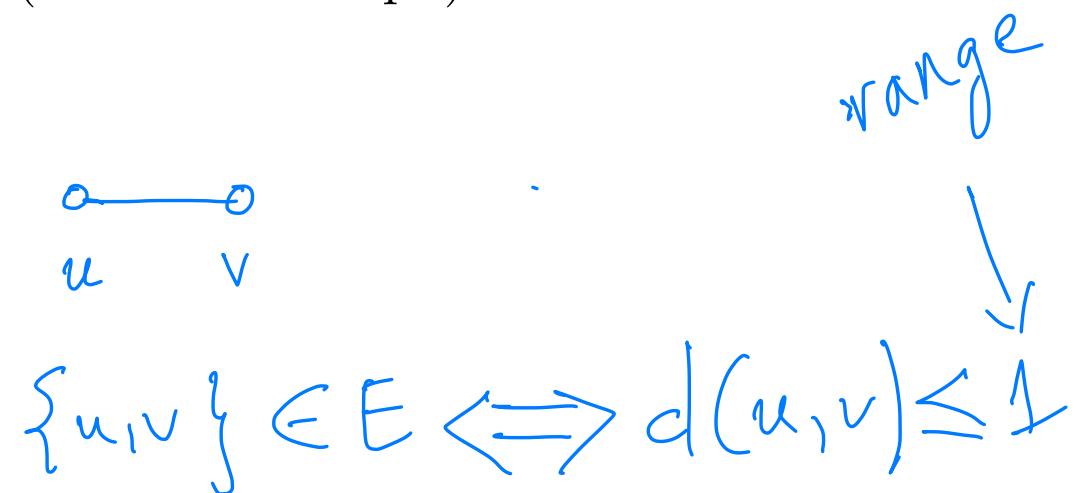
UDGs and Wireless

- Unit Disk Graphs (UDGs) are used in computer science to model the topology of ad hoc wireless communication networks.
- Nodes are connected through a direct wireless connection without a base station. It is assumed that all nodes are homogeneous and equipped with omnidirectional antennas.
- Node locations are modeled as Euclidean points, and the area within which a signal from one node can be received by another node is modelled as a circle.
- If all nodes have equal transmission power, the circles are equal.
- Random geometric graphs, formed as unit disk graphs with randomly generated disk centers, have also been used as a model of percolation and various other phenomena.



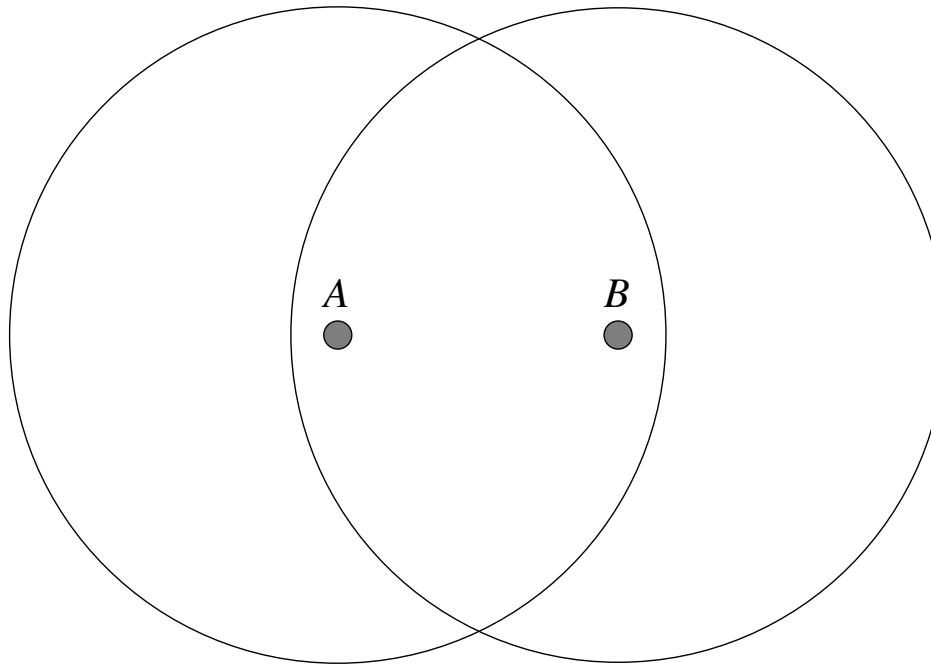
UDGs: Vertices and Edges

- The UDG is an abstract model of an ad hoc network.
 - It is a graph $G(V, E)$ with V the set of vertices and E the set of its edges.
 - V : Vertices are the sensor nodes.
 - E : Edges between vertices represent connectivity, i.e., whether or not they can communicate.
- Why the name UDG (Unit Disk Graph)?



Why UDG?

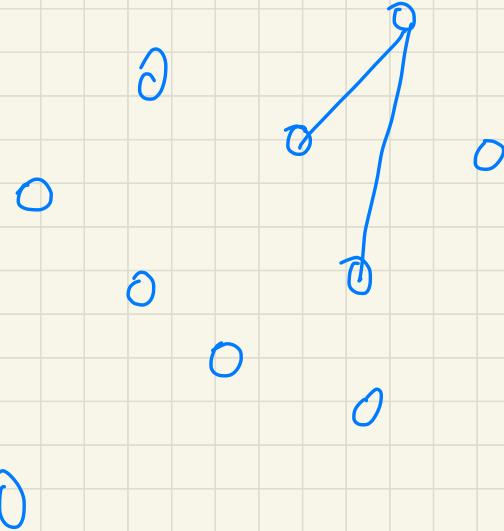
- Two (mobile) hosts A, B are adjacent if they are within reach of each other:



- There is an edge between A, B if and only if $d(A, B) \leq 1$.

\downarrow
 r

All "sensors"
have equal
range



$r=1$

(a) (b)

(c) (d)

(e)

(f)

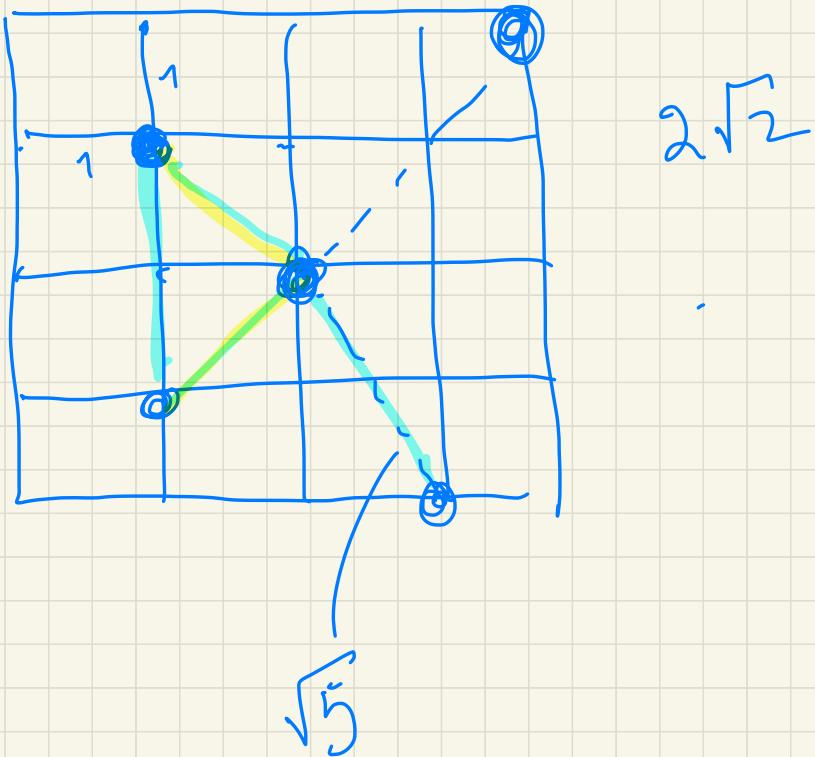


$$r = 1$$

$$r = \sqrt{2}$$

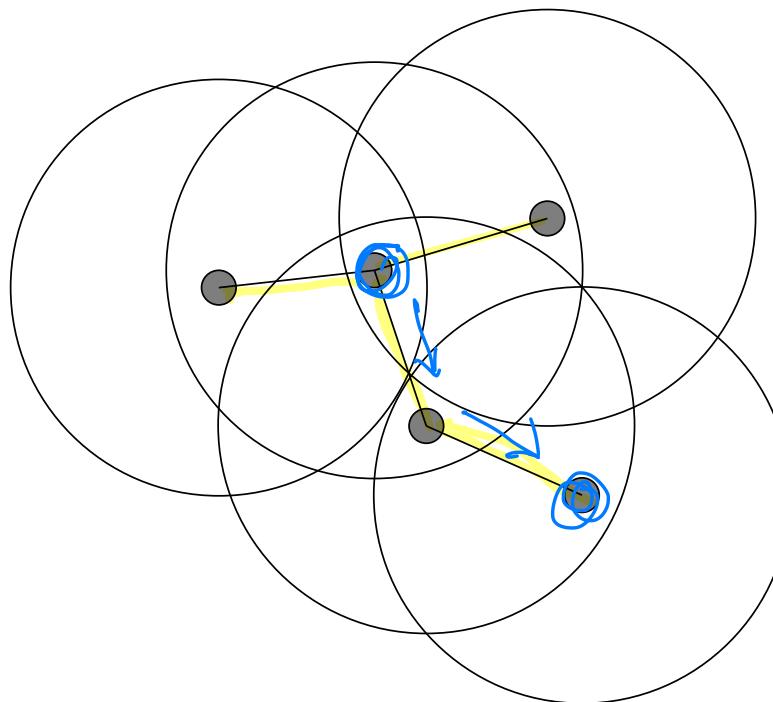
$$r = 2.5$$

$$r = 3$$



Example of UDG

- Underlying graph.



The transmission ranges of the sensors define a communication graph.

- The disks determine an “underlying” graph.



UDGs and Mobility

- The previous UDG model is static.
- If you want to include mobility then time t must be incorporated in the model.
- $G_0, G_1, \dots, G_t, \dots$ is a sequence of UDGs whereby G_t is the “state of the ad hoc network” at time t .
- Given G_t the new network G_{t+1} is obtained from G_t by the addition/deletion of nodes/links.

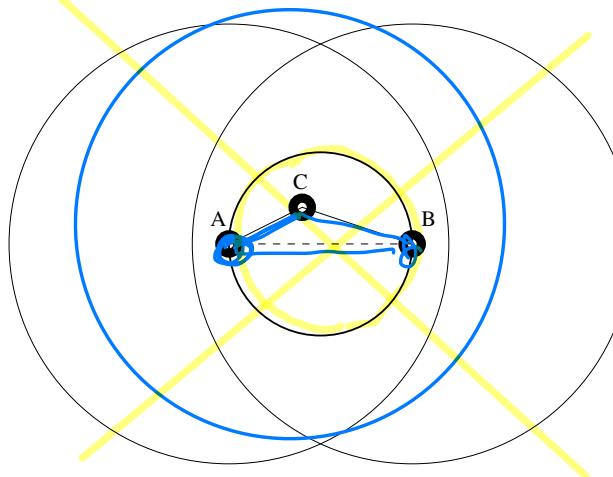
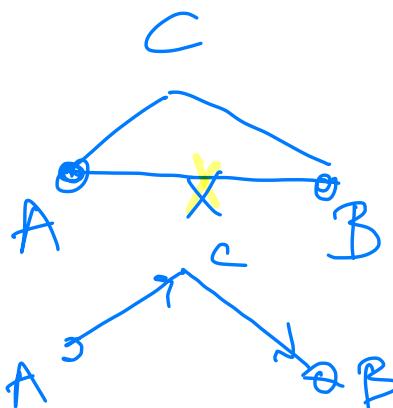
G_0, G_1, G_2, \dots

$G_t \rightarrow G_{t+1}$

Gabriel Test

Gabriel Test (Algorithm)

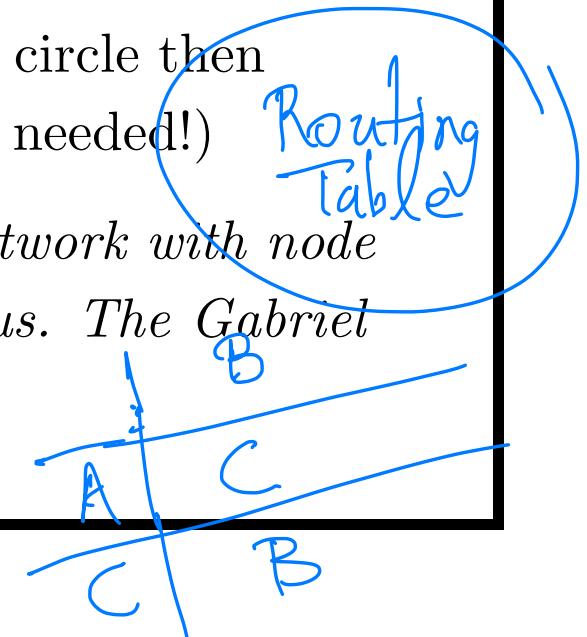
- Assume points A and B are within range of each other.



A, B, C
determine
disk

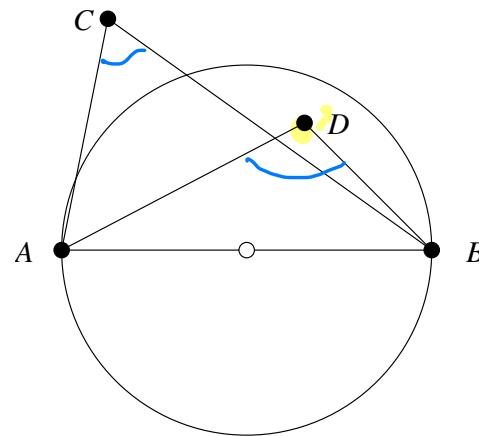
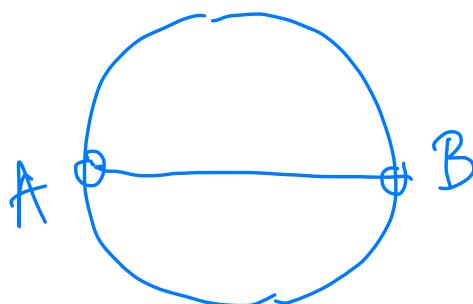
- Draw circle with diameter AB .
- If there is another point, say C inside this circle then remove the link connecting A to B (is not needed!)

- Theorem 1** Assume a connected wireless network with node ranges represented as circles of identical radius. The Gabriel algorithm removes all edge crossings!



Gabriel Graph: Observations

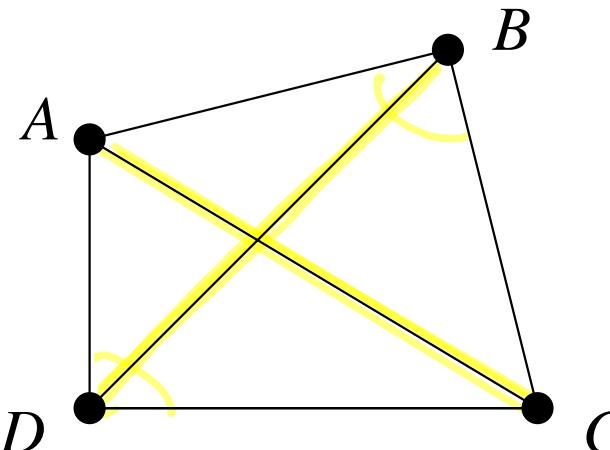
- Call AB a Gabriel edge if the circle with diameter AB contains no other points.



- A point X is inside the circle with diameter AB if and only if the angle AXB is bigger than $\pi/2$.
- A point X is inside the circle with diameter AB if and only if its distance from the center of the circle is bigger than $|AB|/2$.

Gabriel Graph is Planar

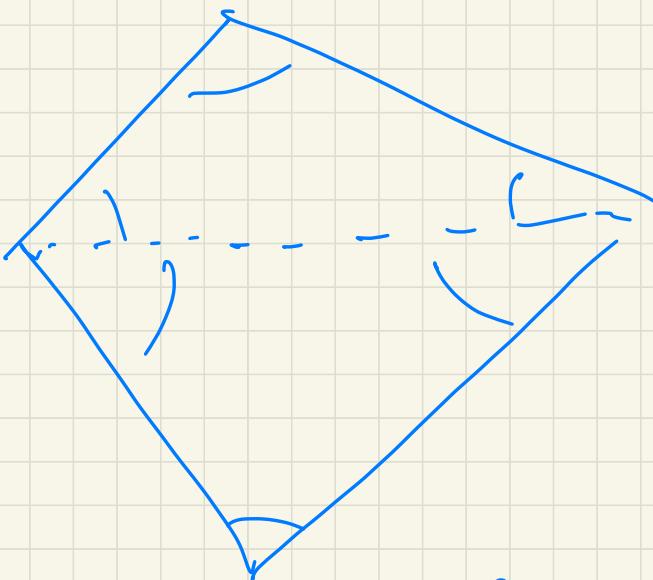
- If the Gabriel edges AC and BD intersect, A, B, C, D must form a convex quadragon!



$\cancel{B} > \pi/2$
 $\cancel{D} > \pi/2$
 $\cancel{A} > \pi/2$
 $\cancel{C} > \pi/2$
sum $> 2\pi$

AC is Gabriel edge because
there is no point inside circle
of diameter $|AC|$

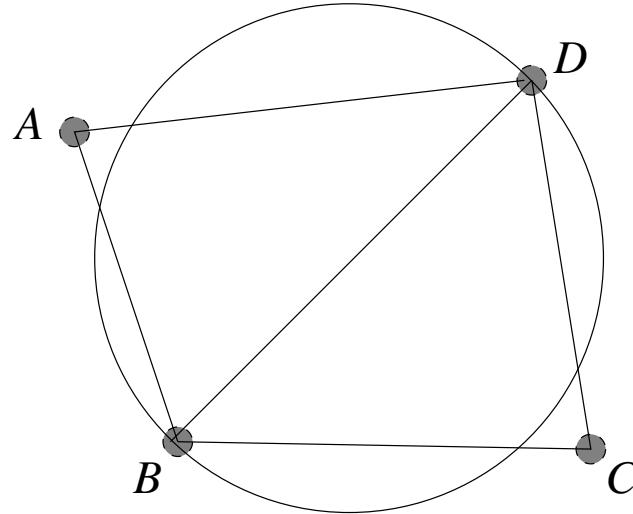
BD is a Gabriel edge because
there is no point inside
circle of diameter $|BD|$



sum of angles is 2π

Why is edge BD preserved?

- Edge BD is preserved because it is a Gabriel edge.

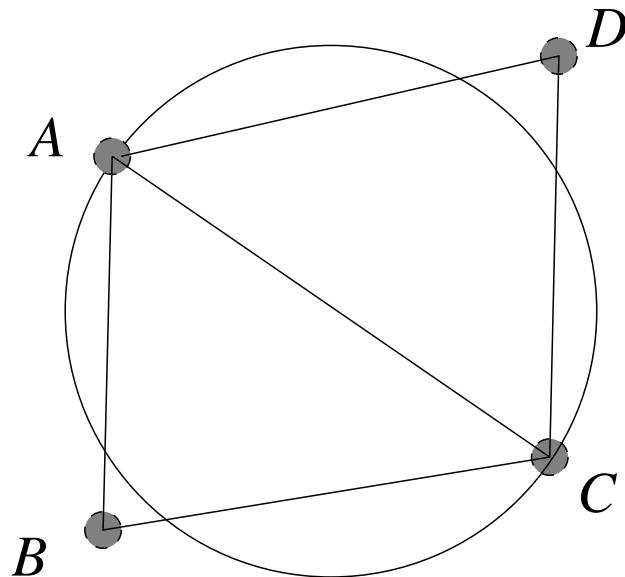


- Therefore both A and C lie outside the circle with diameter BD .

Repetition

Why is edge AC preserved?

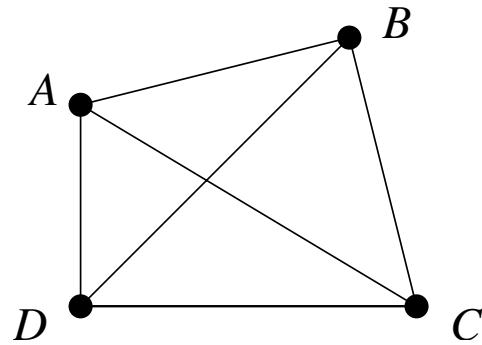
- Edge AC is preserved because it is a Gabriel edge.



- Therefore both B and D lie outside the circle with diameter AC .

Gabriel Test Removes Edge Crossings

- If the Gabriel edges AC and BD intersect, A, B, C, D must form a convex quadragon!

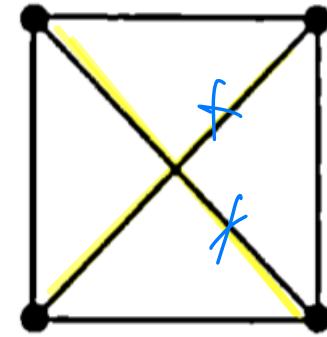


- Hence $\angle ABC, \angle BCD, \angle CDA, \angle DAB < \pi/2$, contradicting the fact that $\angle ABC + \angle BCD + \angle CDA + \angle DAB = 2\pi$.

Examples: Gabriel Graph Depends on the Drawing!

- Find the Gabriel graph in each case:

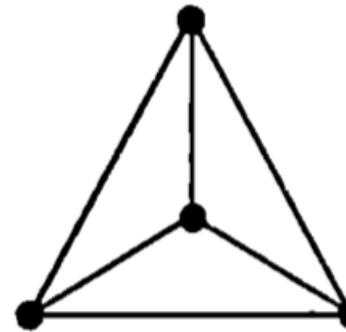
A



f

or

B
 K_4

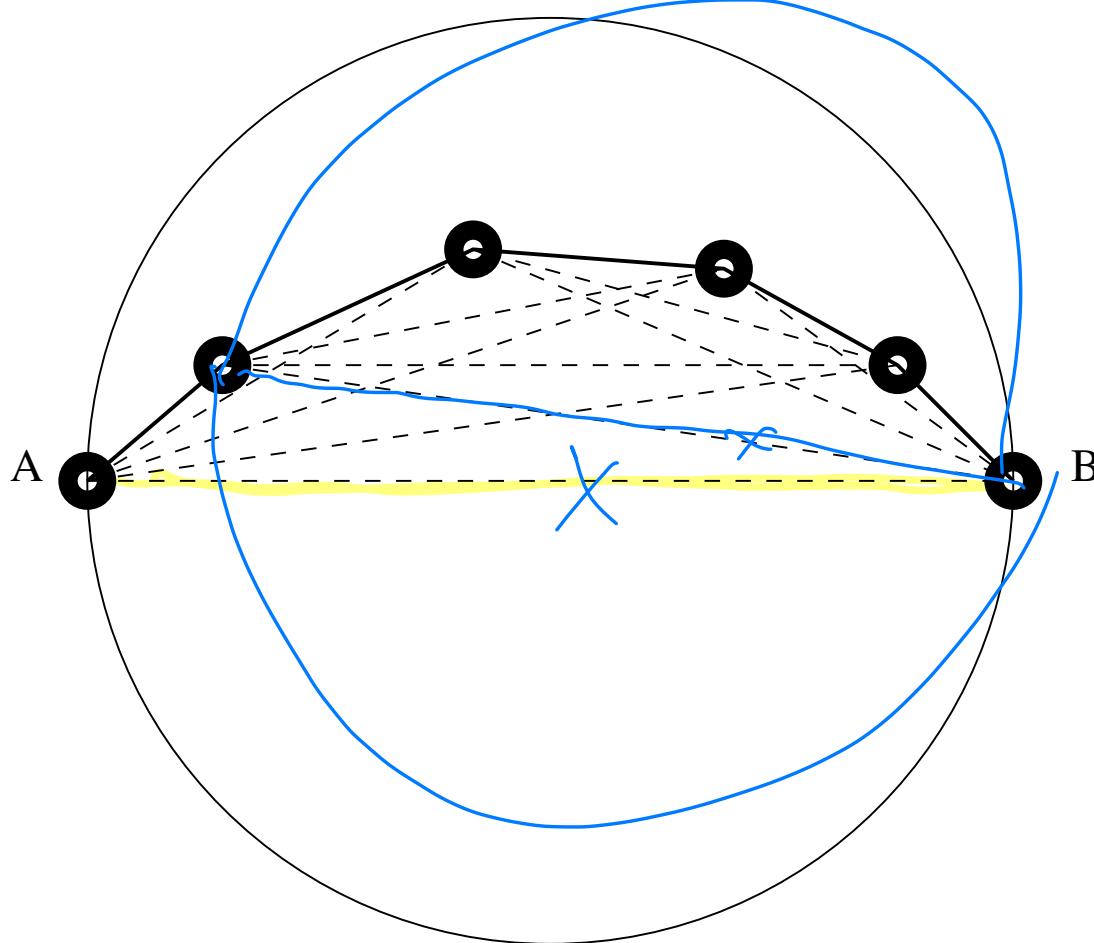


Crossings

No - crossings

Example: Gabriel Test and Shortest Paths

- How well does the Gabriel graph perform?



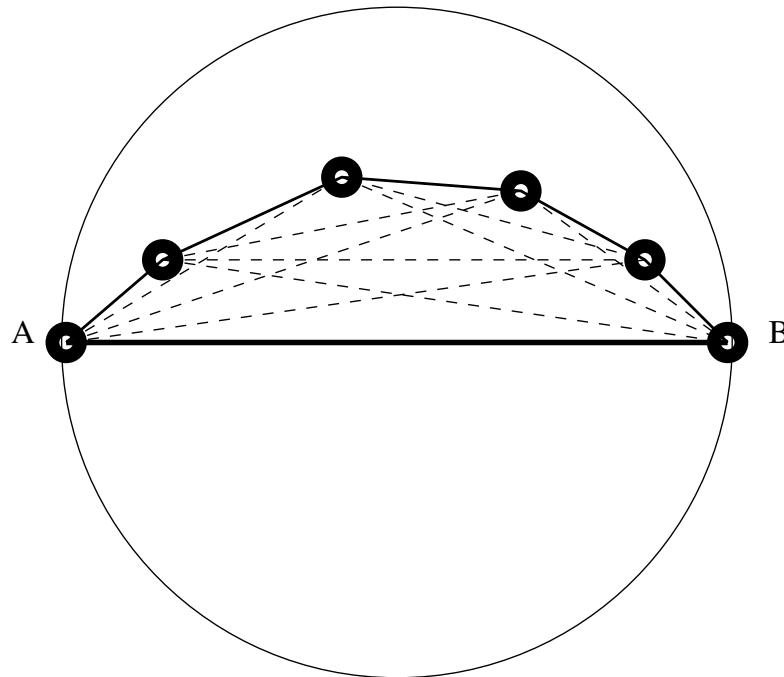
- For more details, see Appendix

How about Deleted Edges?

- You maintain a *Routing Table*.
 - A data base that when you are at A you ask:
How do I reach B ?
 - It gives you the answer: Go to C .
 - And when you reach C you ask again:
How do I reach B ?
 - It gives you the answer: Go to B .
- Standard routing table contains an entry for each possible destination with the out-going link to use for destination
- Message delivery proceeds in the obvious manner one link at a time, looking up the next link in the table.

Too many hops spoil the batteries!

The Gabriel test creates a planar graph but removes long links.



I could have reached B directly from A in one hop.

Instead it takes me five hops!

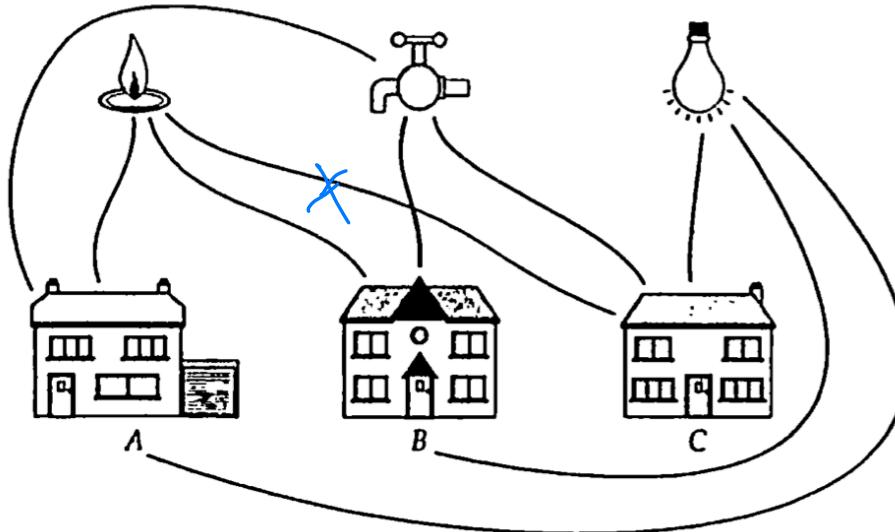
What is important: the Gabriel test will remove all crossings!

Planarity

- Planar Graph
 - A graph G is planar if it can be drawn in the plane in such a way that no two edges meet except at a vertex with which they are both incident.
 - Any such drawing is a plane drawing of G .
 - A graph G is non-planar if no plane drawing of G exists.
- The Gabriel test produces a planar network!
 - It was done by removing edges!

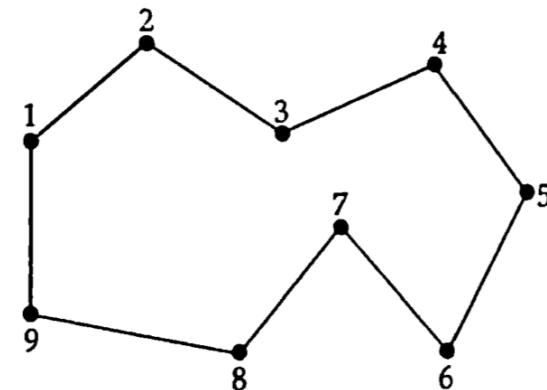
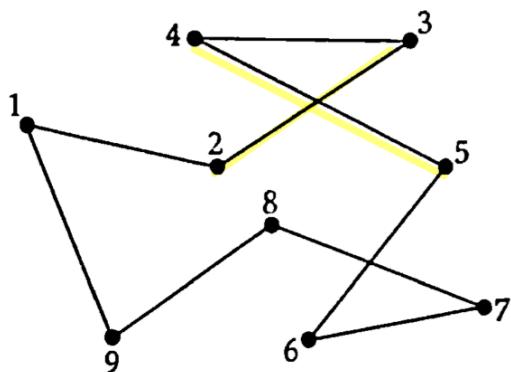
Planarity: Sometimes it is not Possible

- Impossible to draw as a planar graph

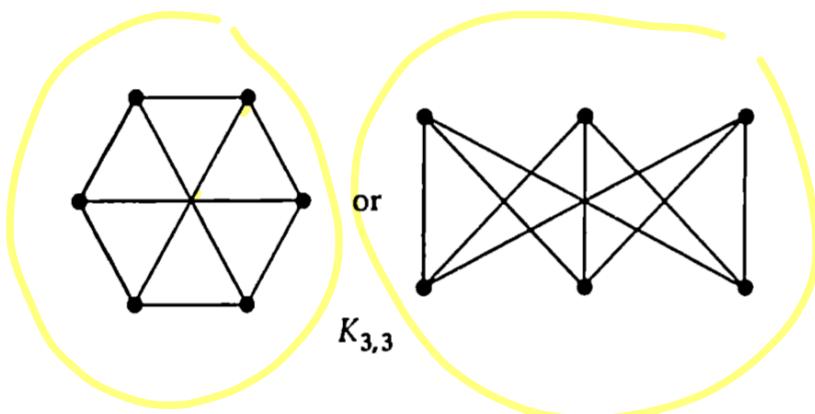
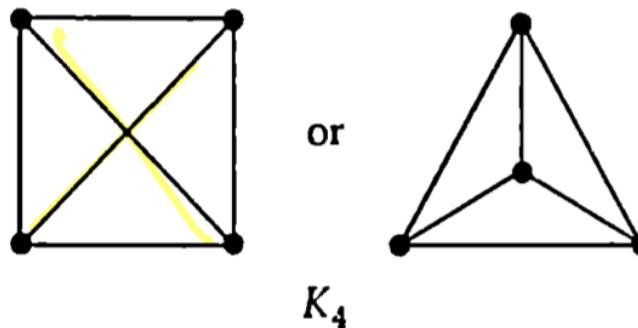


Planarity: Depends on the Drawing

- Just redraw the graph:

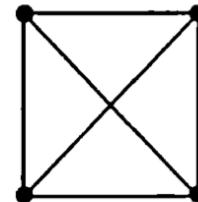


- Just redraw the graph:



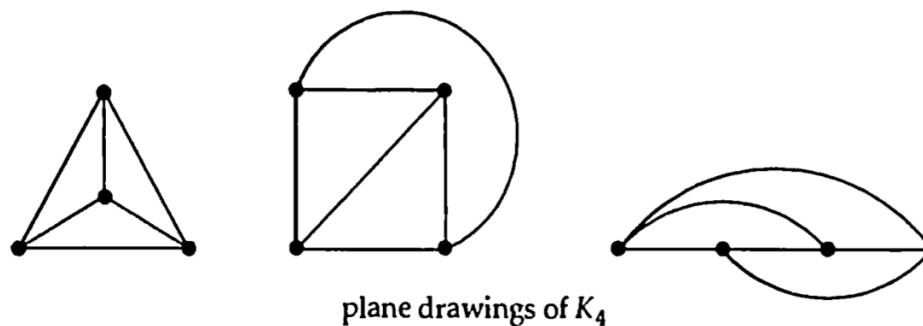
Planarity: Depends on the Drawing

- K_4



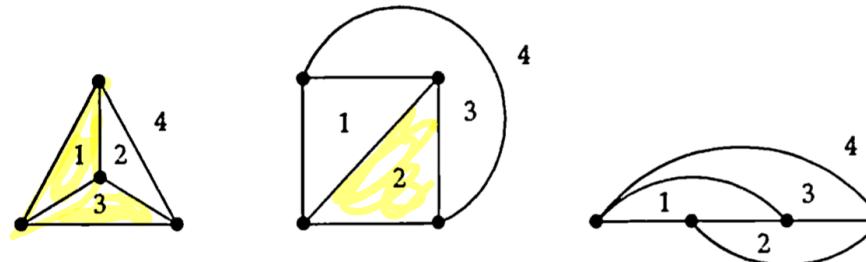
K_4

- Different ways to draw K_4

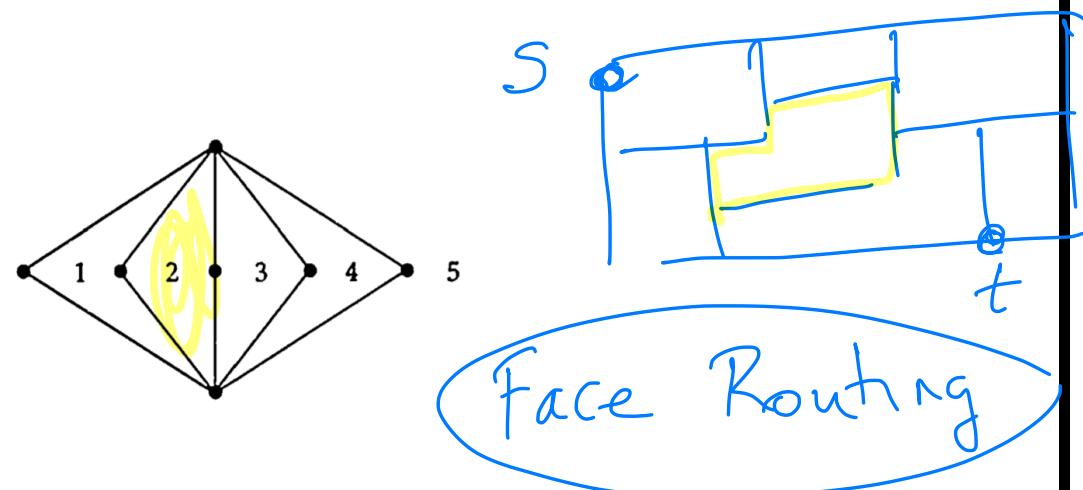


Faces of a Planar Graph (1/2)

- Every plane drawing of a planar graph divides the plane into a number of regions.
- For example, any plane drawing of K_4 divides the plane into four regions: three triangles (3-cycles) and one *infinite region*

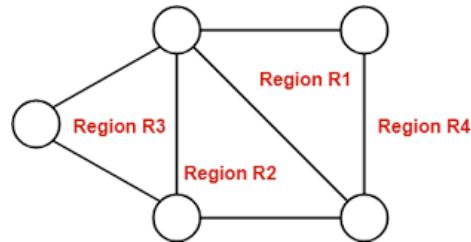


- Another example

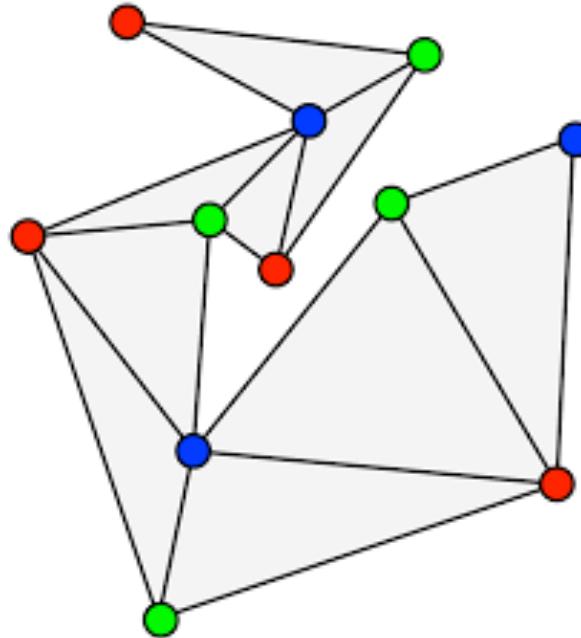


Faces of a Planar Graph (2/2)

- What are the faces of the planar graph?



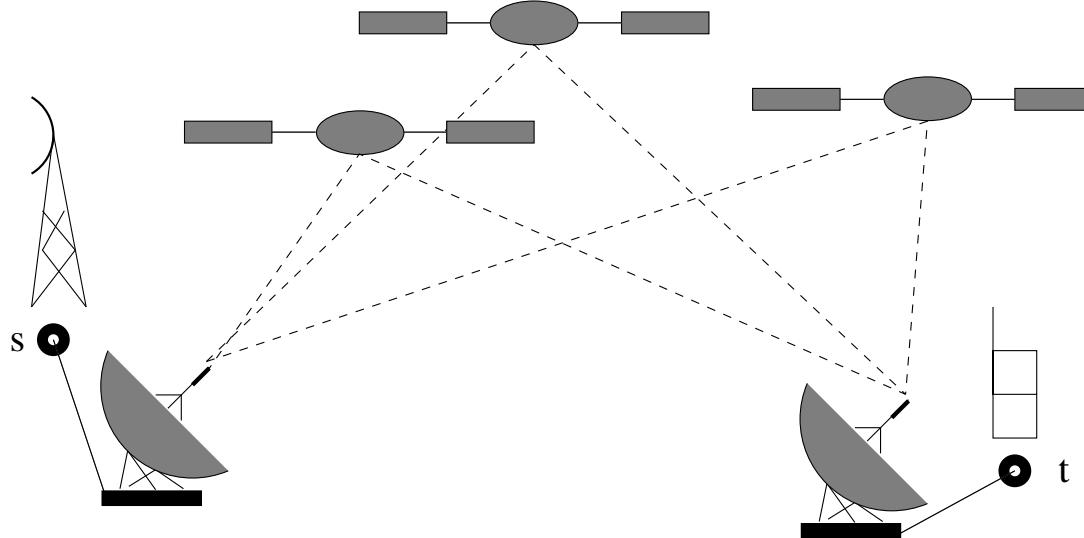
- What are the faces of the planar graph?



Geometric Routing

Global Positioning Systems (GPS)

Can use GPS to discover the (x, y) coordinates of the target node.



GPS uses three satellites in line of sight.

It determines location by **time-of-arrival** differences (temporal delays of several signals).

Can always construct an undelying geometric planar graph using the Gabriel test!

Routing in a Geometric Planar Network

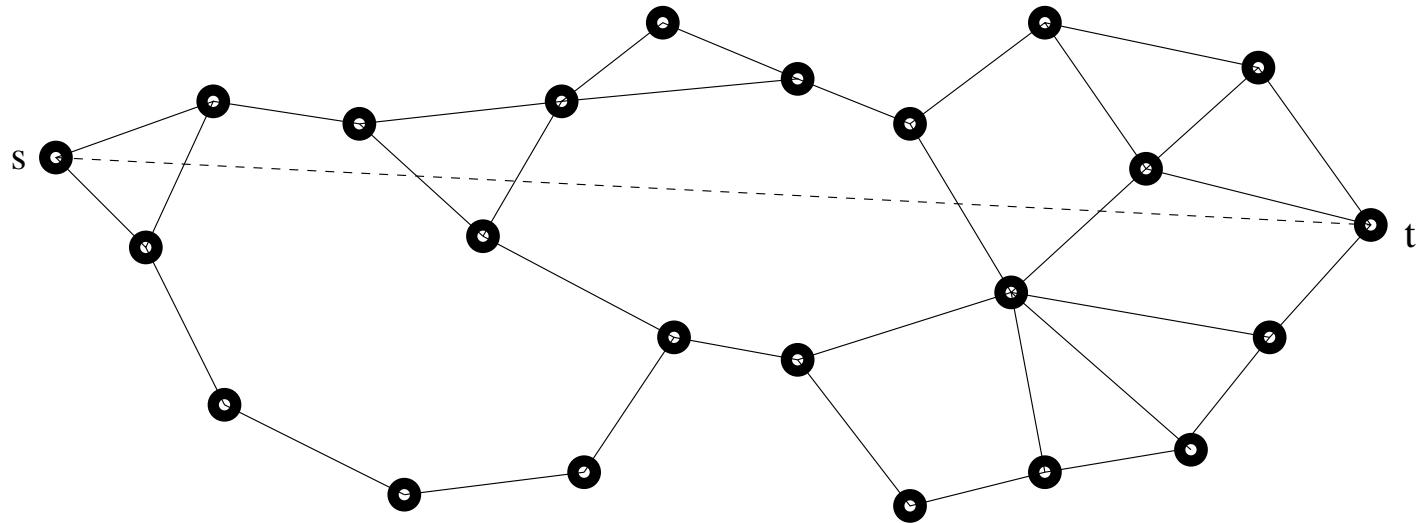
Input: A geometric graph.

Goal. Go from source node s to target node t .

- We need some kind of “capabilities” in order to move towards the target t . This may include the following
 - Updating coordinates of current position c .
 - Must know the coordinates of t .
 - If c is our current position we need to be able to determine the slope of the line \vec{ct} .
- We need to be able to determine the slopes of the edges incident to our current position.

Back to Route Discovery

After applying the Gabriel test we have a planar graph.



Using GPS we can find out the (x, y) coordinates of s and t .

Hence, we can compute the slope of the line \vec{st} .

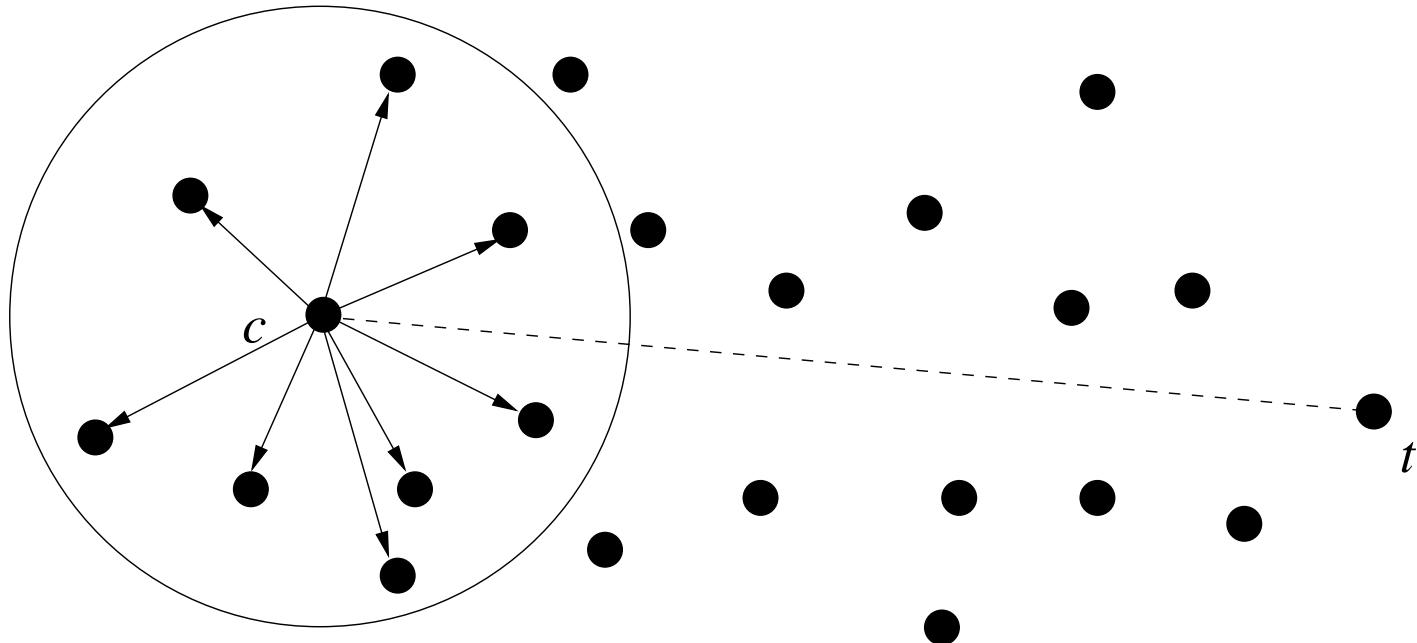
But how do we use this information in order to discover a route?

Compass Routing Algorithm

1. Start at source node $c := s$.
2. in a recursive way:
 - (a) Choose edge of our geometric graph incident to our current position and with the smallest slope to that of the line \vec{ct} .
 - (b) Traverse the chosen edge.
 - (c) Go back to (a) and repeat until target t is found
- **Theorem 2** *Compass routing requires GPS and works in many cases (like, random graphs with high probability) and is the basis of tiny OS.*

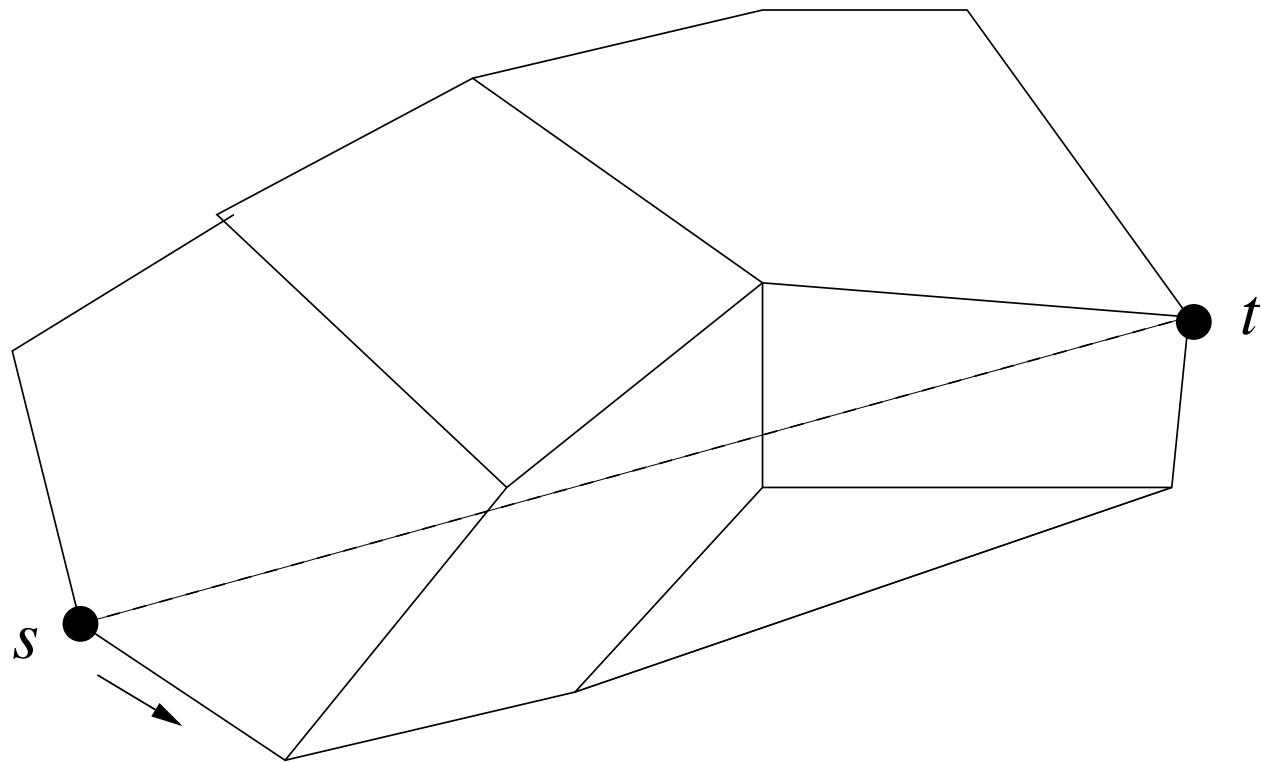
Compass Routing: Next Move

- Choose the smallest angle!

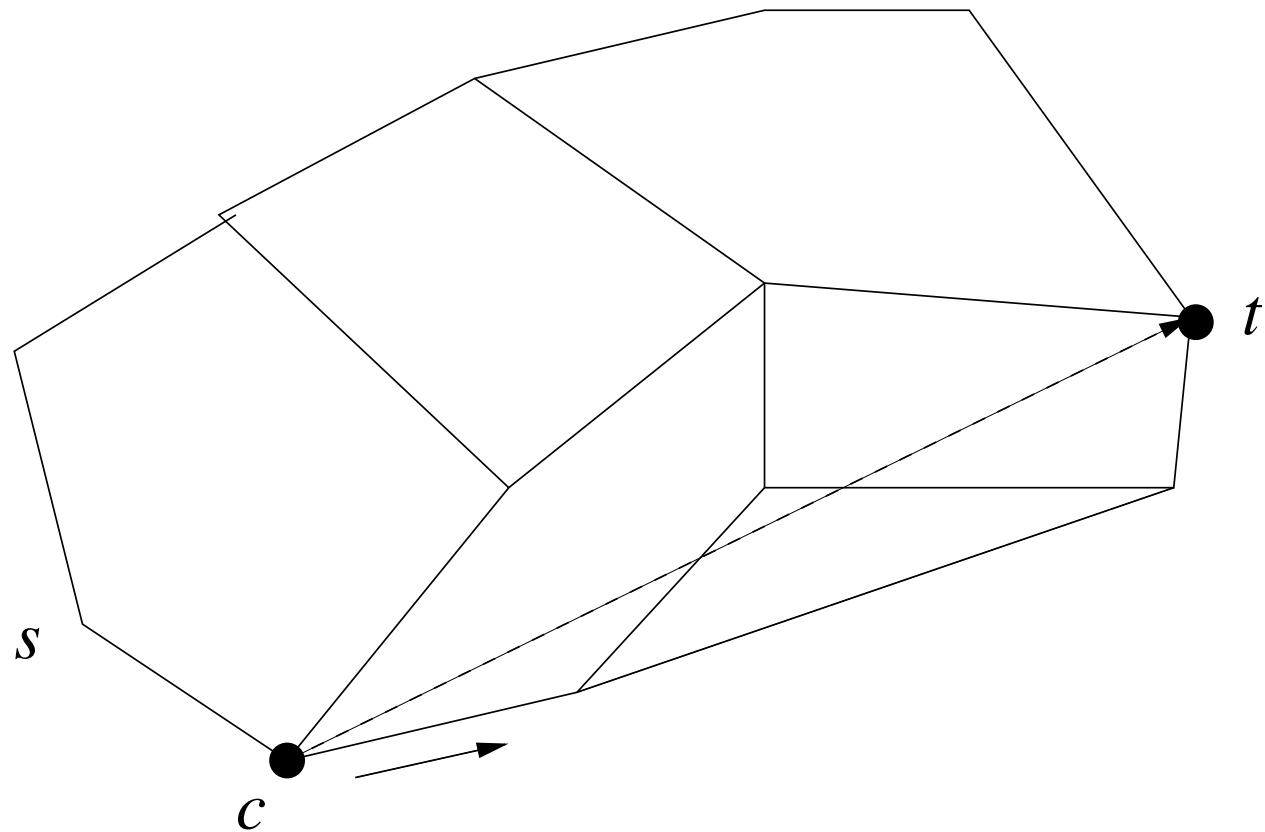


- **Problem:** Compass routing can fail to reach destination!

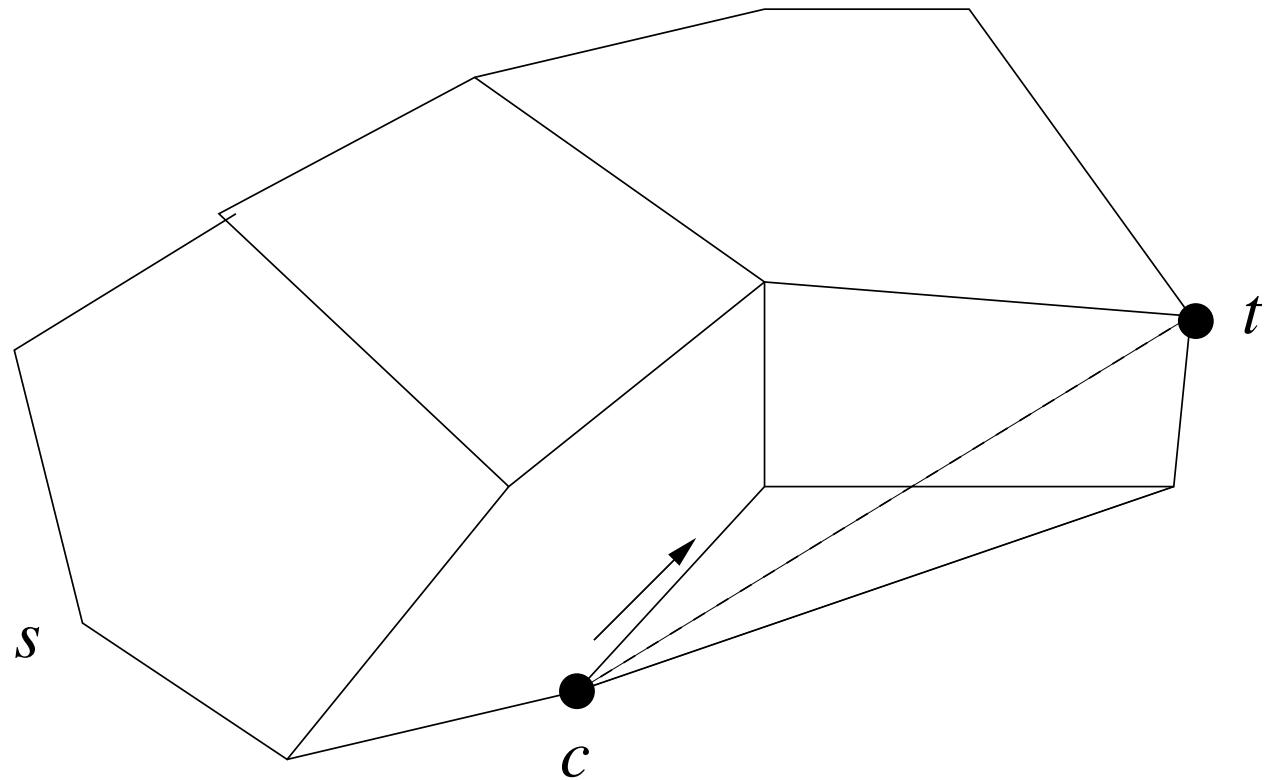
Compass Routing (1/5)



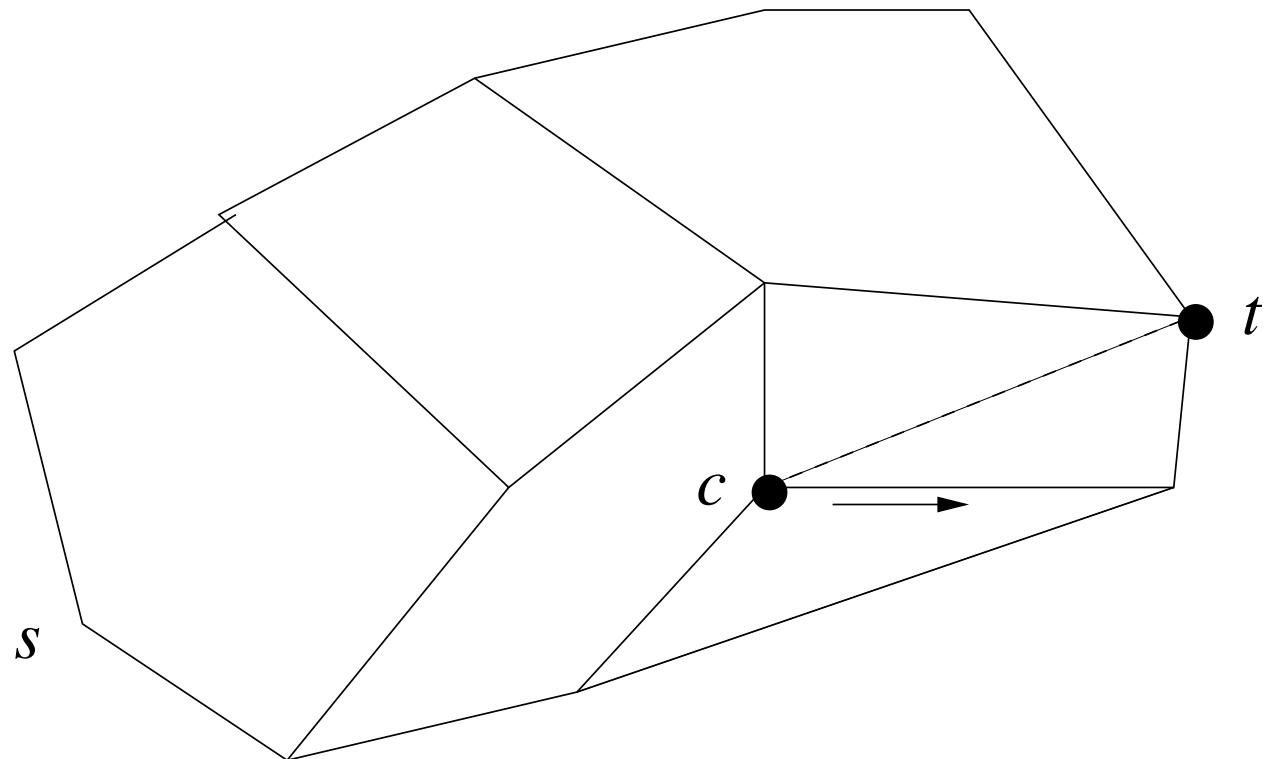
Compass Routing (2/5)



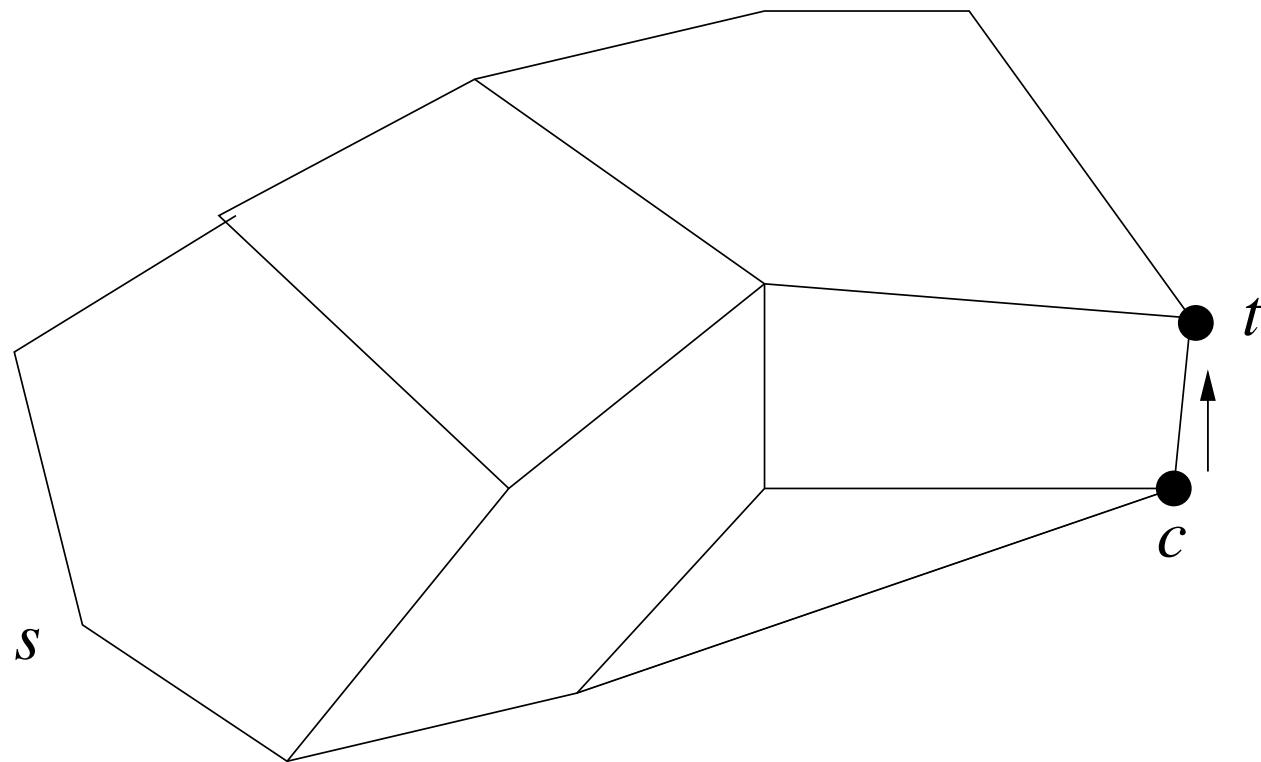
Compass Routing (3/5)



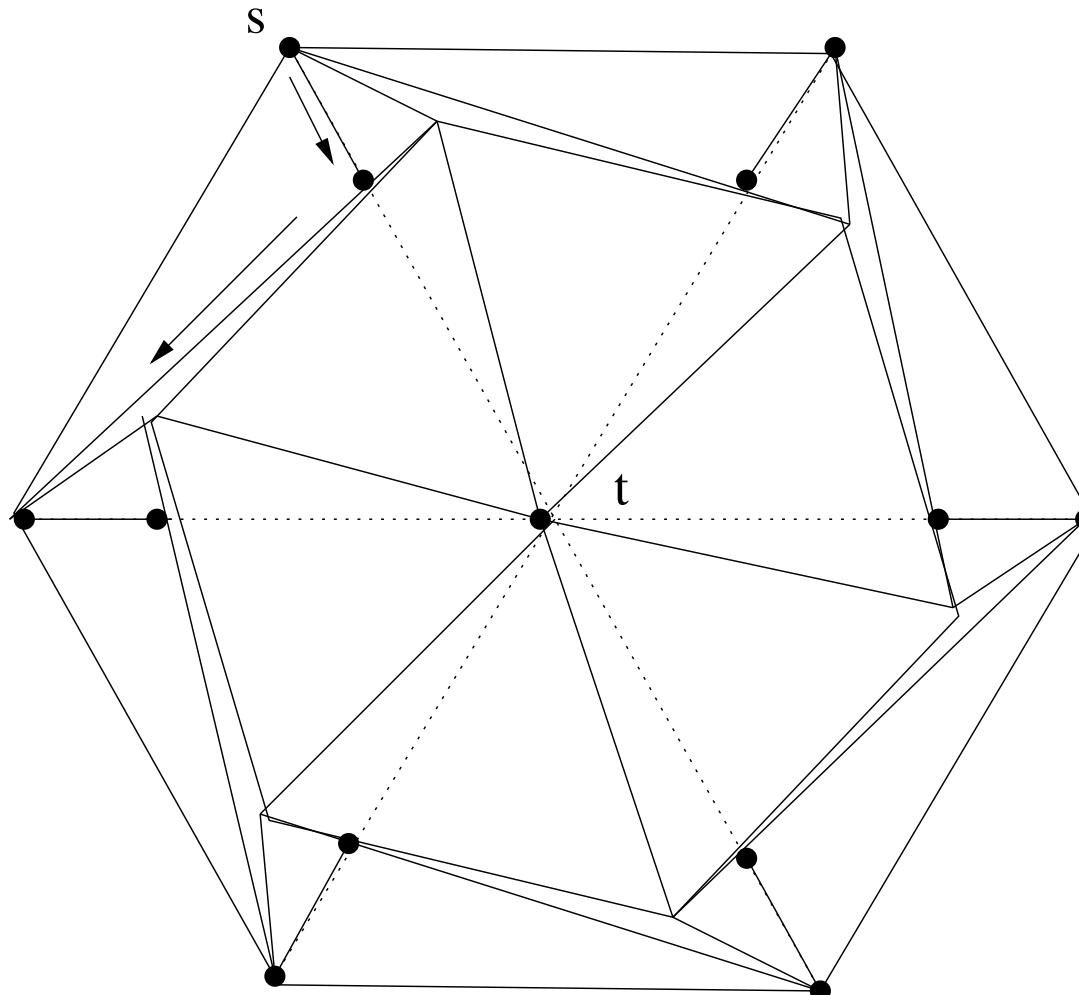
Compass Routing (4/5)



Compass Routing (5/5)



Compass Routing May not always Work!



Face-Routing Algorithm (1/3)

- Starting Phase
 1. Let s, t be the source and target nodes in a geometric graph.
 2. Determine the straight line \vec{st} and remember it.
 3. Start with $c := s$ as the current node.
- Note that one must remember the straight line \vec{st} , which remains the same throughout the algorithm.

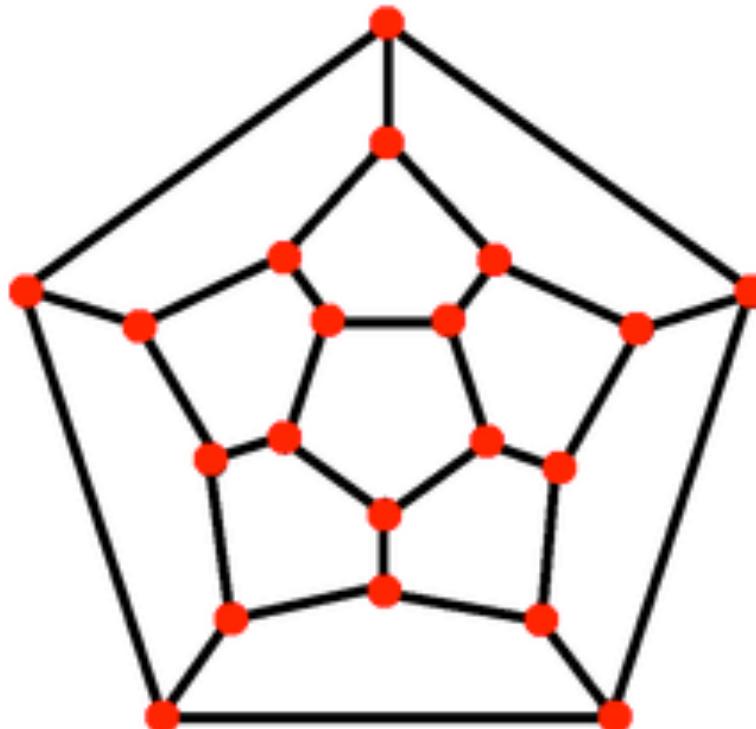
Face-Routing Algorithm (2/3)

- Face selection and traversal phase:
 1. Determine the face F incident to c such that F intersects the straight line \vec{st} . This determines an edge one of whose endpoints is c .
 2. Select a direction of movement (Left or Right) and move along the edges of the face F .
 3. In this traversal, eventually you hit an edge, say $\{u, v\}$, which crosses the straight line \vec{st} . If neither u nor v is equal to t then select the first vertex u and update the current vertex $c \leftarrow u$.
 4. Iterate: Go back to Item 1.
- Notice that you have the choice to go either Left or Right. It does not matter which direction you select.

Face-Routing Algorithm (3/3)

- Final phase:
 1. Stop when t is found.
- Why does the algorithm terminate correctly?
- **Theorem 3** *Face routing requires GPS and works in all planar graphs (and is the basis of route discovery in many ad hoc networks).*

Example: Go from s to t

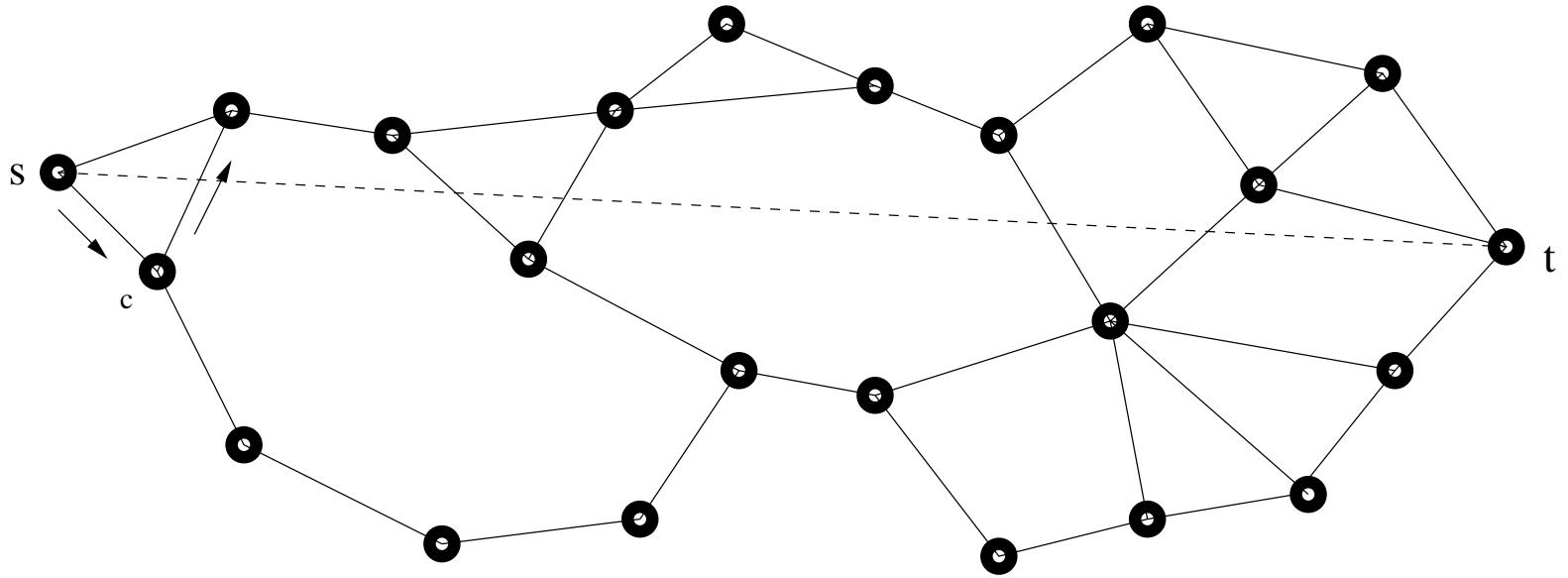


Initially $c := s$.

Update c and repeat.

If the last edge you cross in a round is (a, b) then in the next round start from b

Example: Go from s to t



Initially $c := s$.

Update c and repeat.

If the last last edge you cross in a round is (a, b) then in the next round start from b

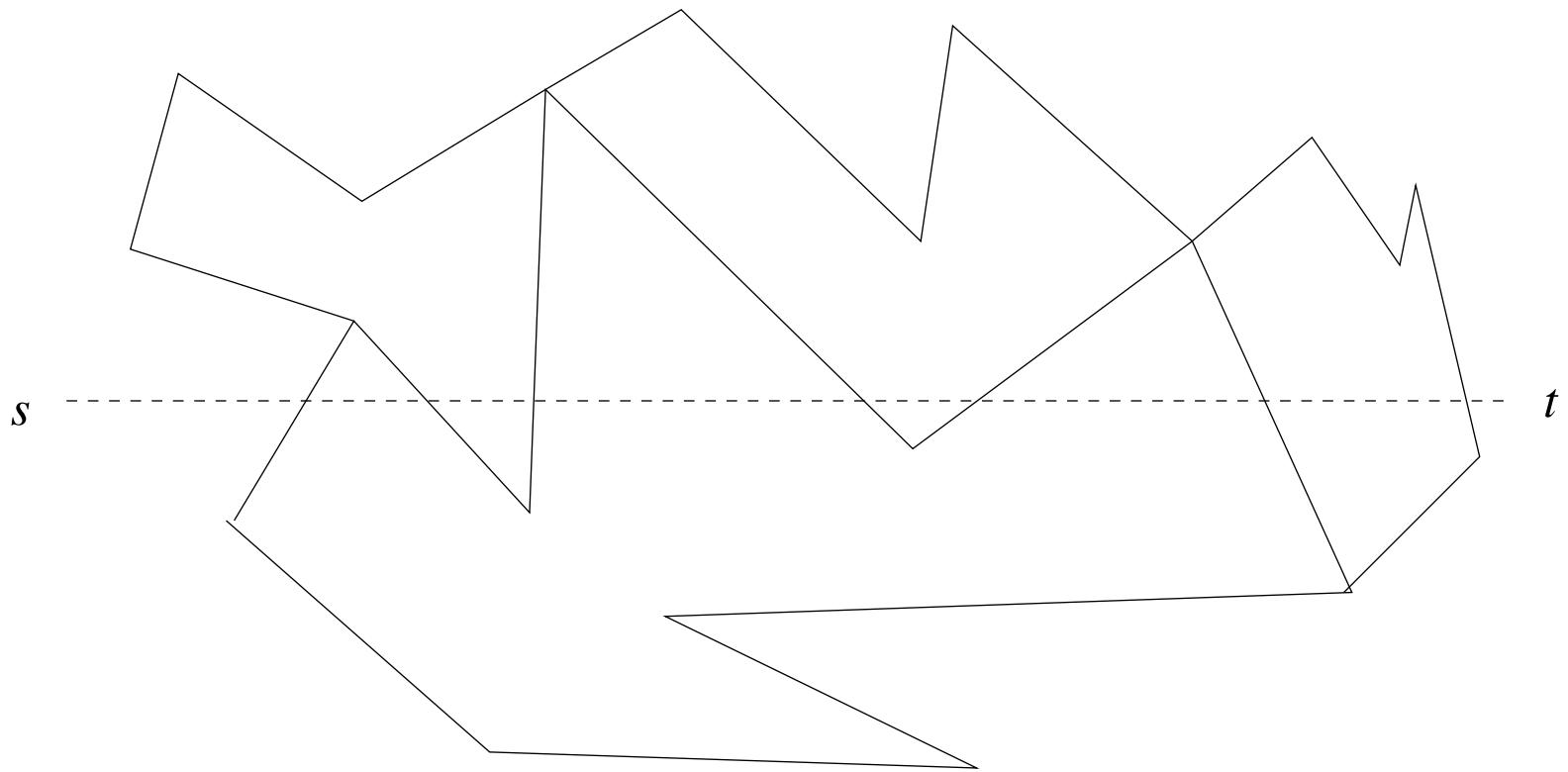
Analysis of Face-Routing

- Face routing always advances to a new face. We never traverse the same face twice.
- The distance from the current position c to t gets smaller with each iteration.
- Each link is traversed a constant number of times. Since the graph is planar face routing traverses at most $O(n)$ edges.

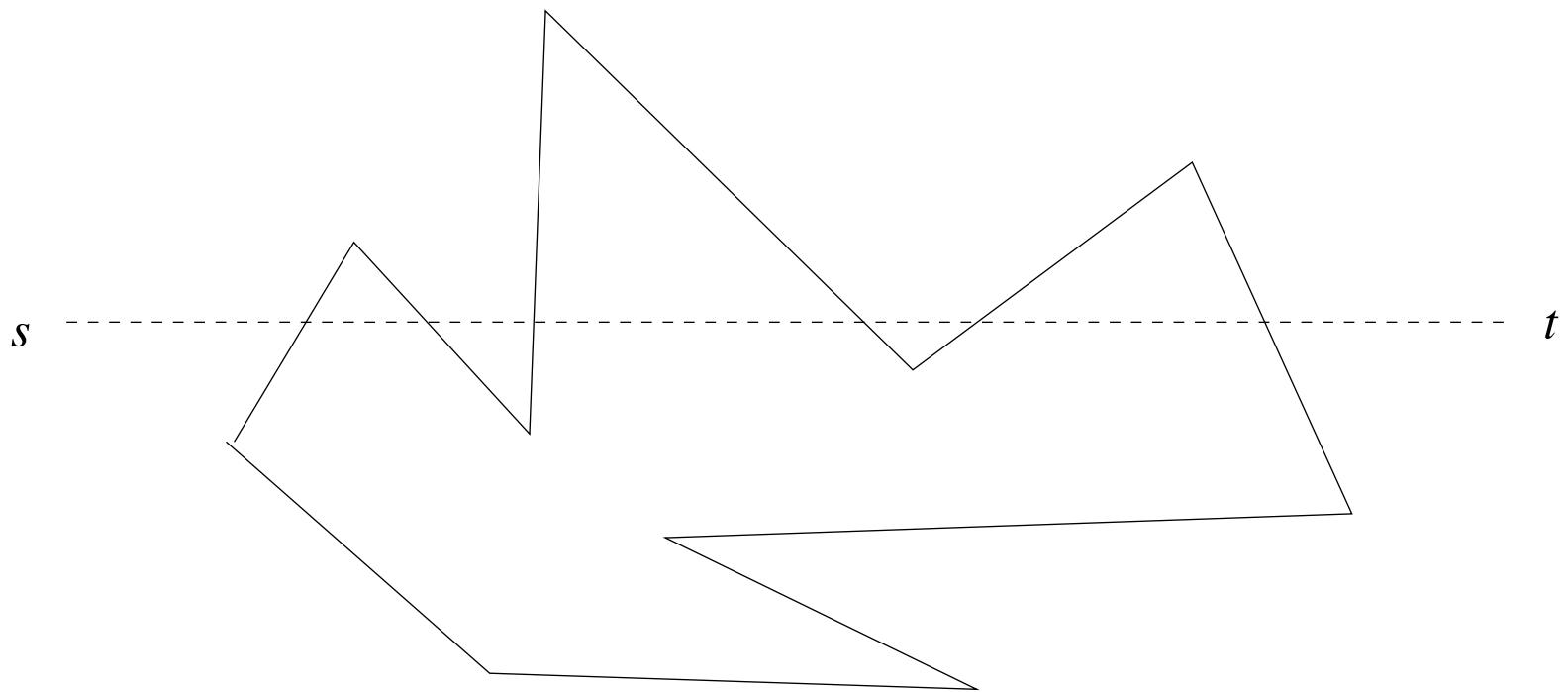
Problems with Face-Routing

- No indication how long is the Euclidean distance traveled!
- But does it matter? All we wanted was to discover a route!

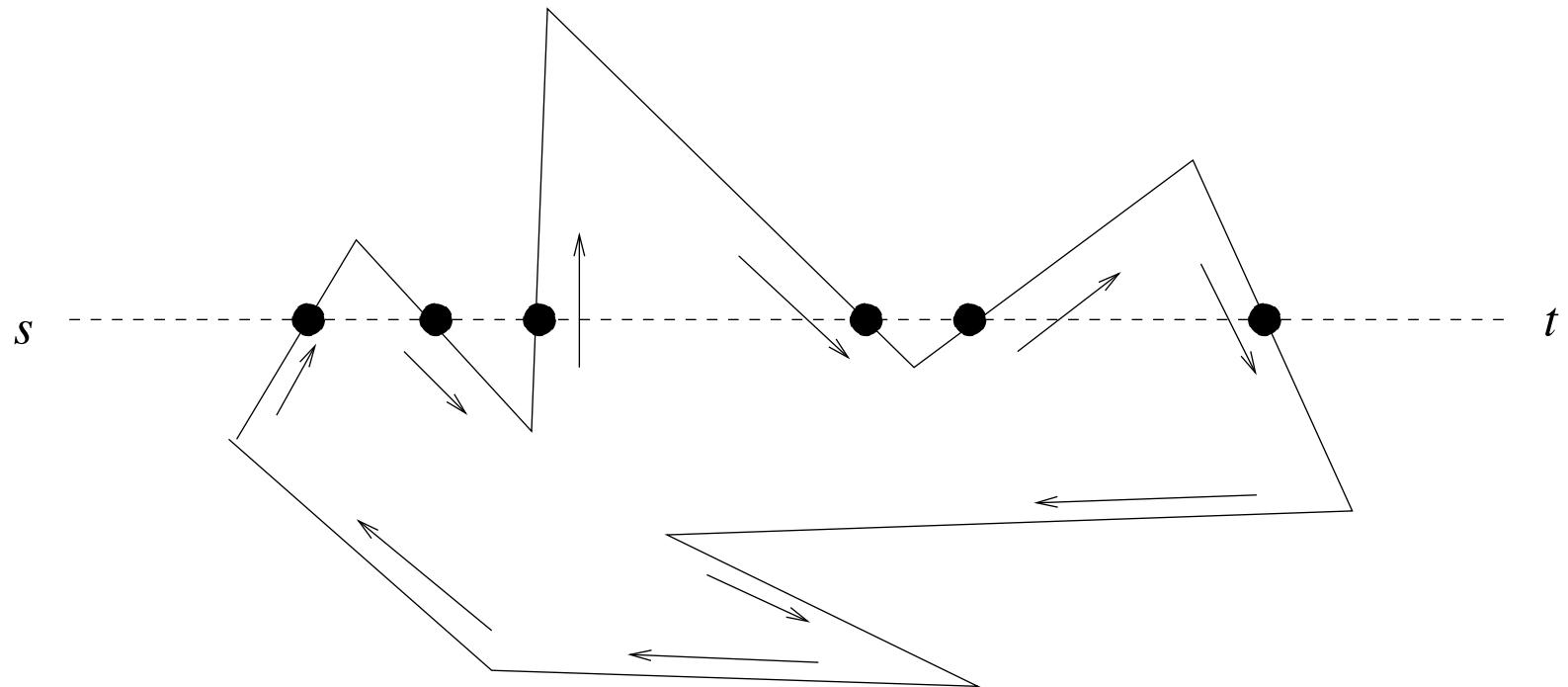
Example 1: How do you Traverse Faces?



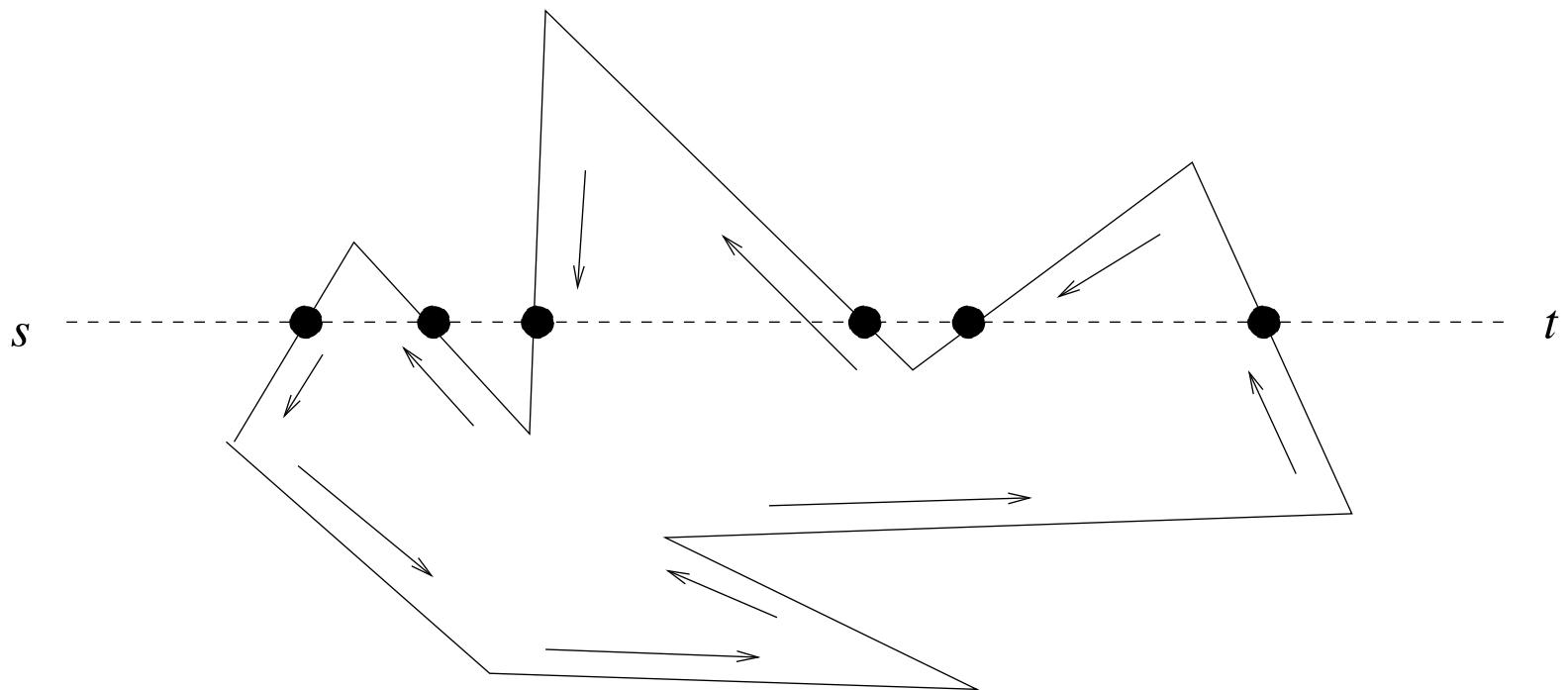
Example 1: When do you Flip Face?



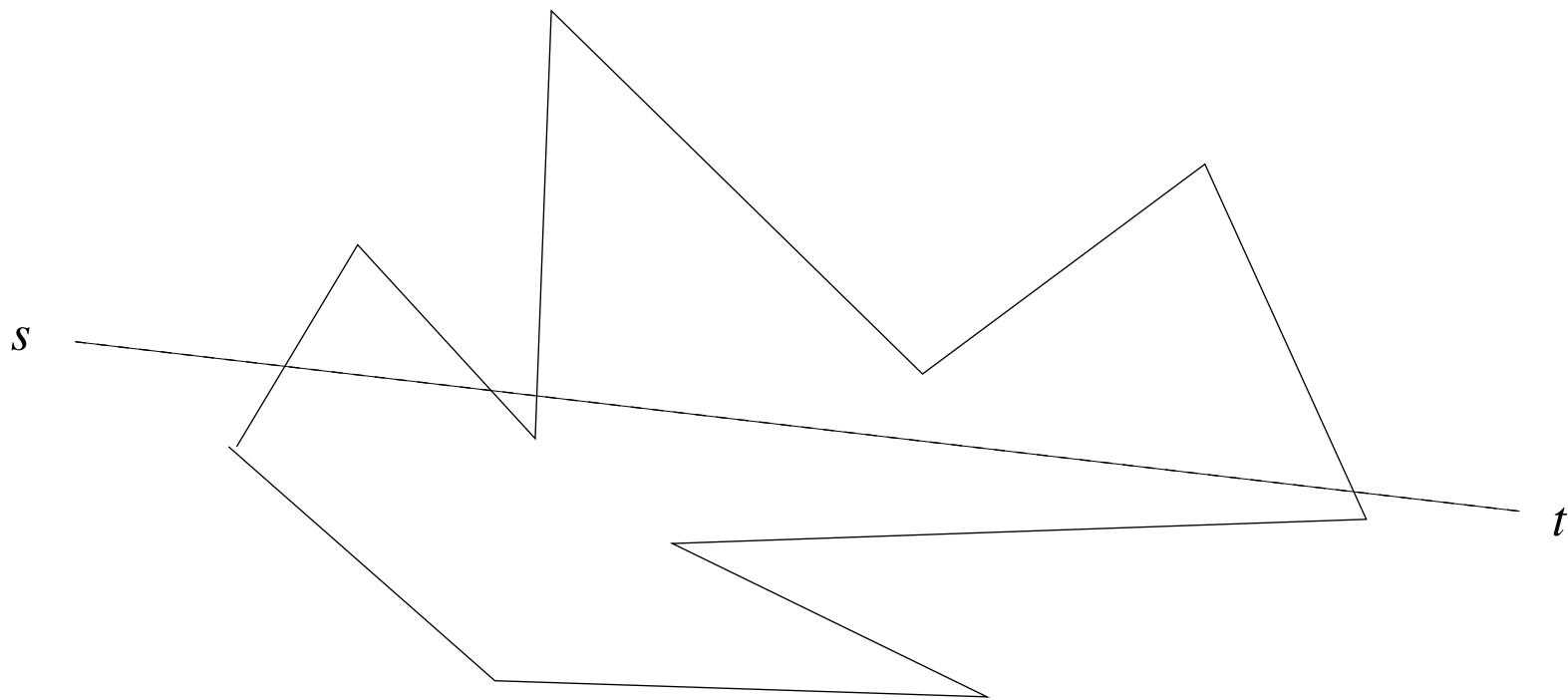
Example 1: Go to the Farthest Crossings!



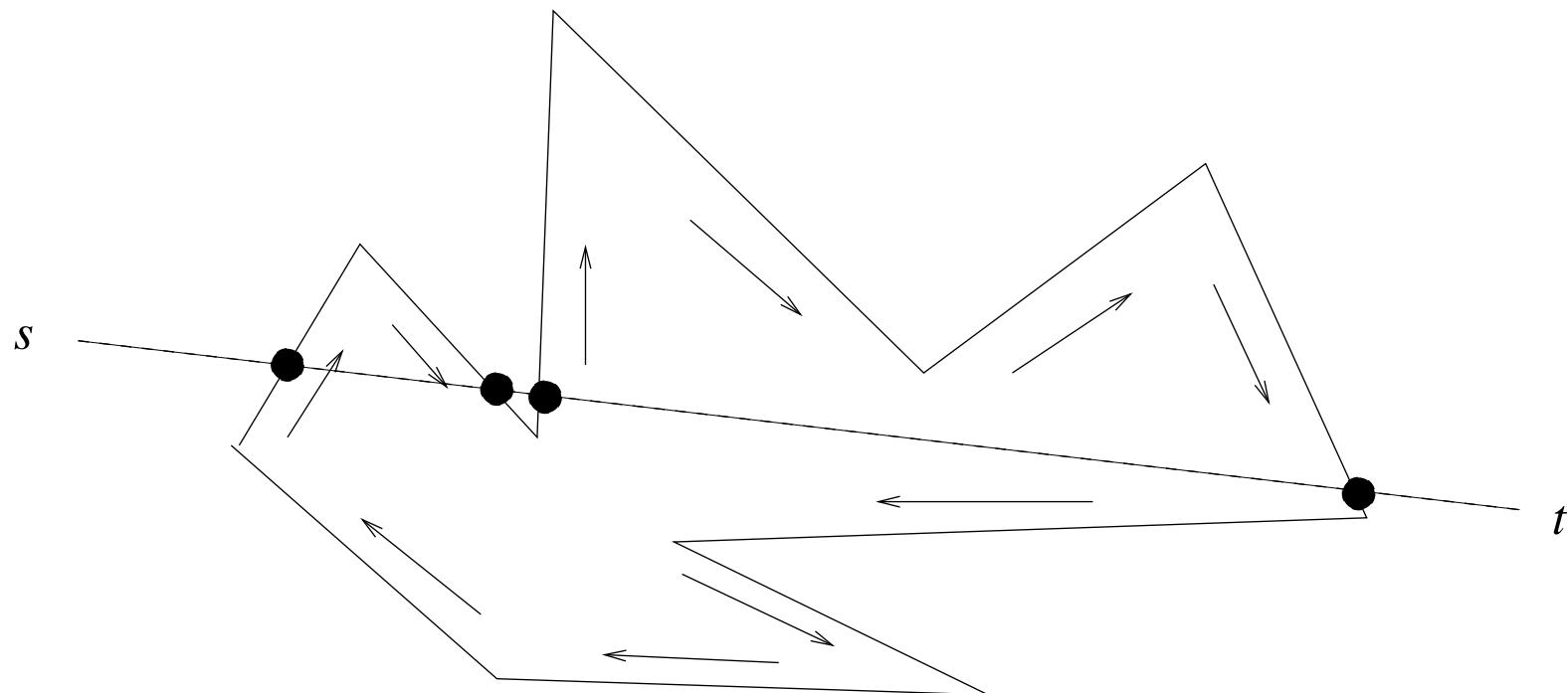
Example 1: Can Choose Right Hand Rule!



Example 2: Different Target, Different Direction?

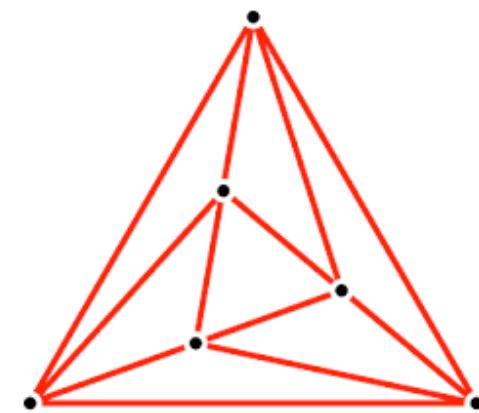
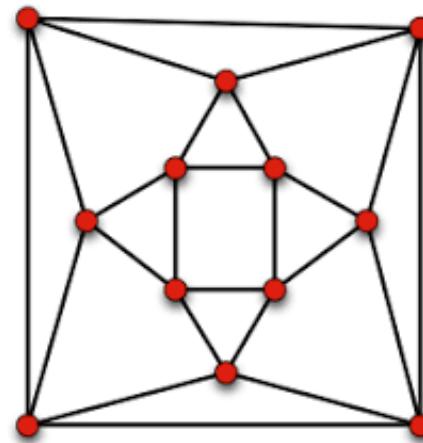
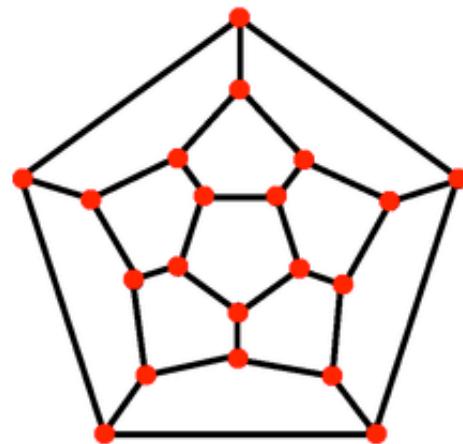


Example 2: Different Intermediate Crossings!



Exercises^a

1. Why in face routing after choosing the new face you may select either the left or right wall on this face?
2. In the planar graphs below execute face and compass routing between any two pairs of nodes.



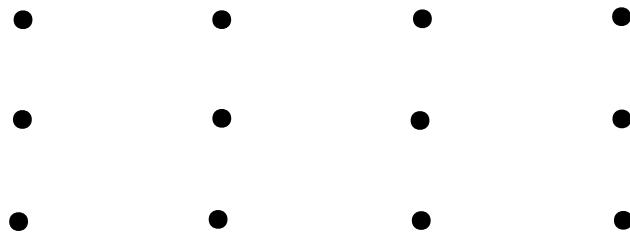
3. Indicate the outer face in each of the graphs in Exercise 2.
4. How many faces (including the outer face) do each of the

^aDo not submit

graphs in Exercise 2 have?

5. There is a formula in graph theory (due to Euler) that says $V + F = E + 2$, where V, E, F is the number of vertices, edges, and faces of a planar graph. Verify that this formula is valid in each of the graphs in Exercise 2.
6. Can you give an example of a planar graph in which face routing from a node to a node t will have to employ the outer face?
7. Consider Rayleigh's principle. Four sensors S_1, S_2, S_3, S_4 which are at distance 1, 2, 3, 4, respectively, from a sensor S broadcast simultaneously towards S with powers 2, 4, 8, 16, respectively. Assuming the threshold $\lambda = 1$ and the external noise $N = 0$ will S be able to hear the signal of any of the four sensors? If yes, which one?
8. Consider the 12 sensors depicted below. Assume the sensors

have identical range $r > 0$

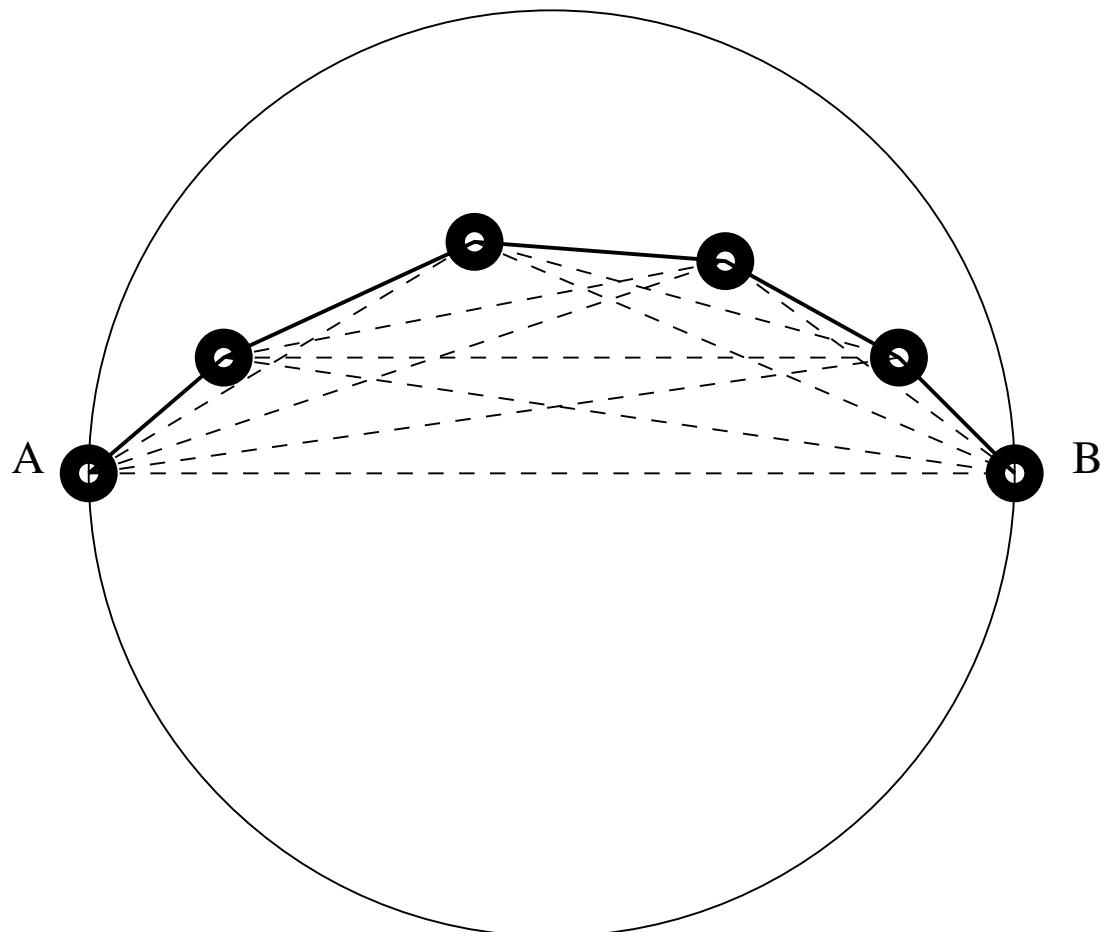


and horizontal distances are 2 units while vertical are 1 unit.

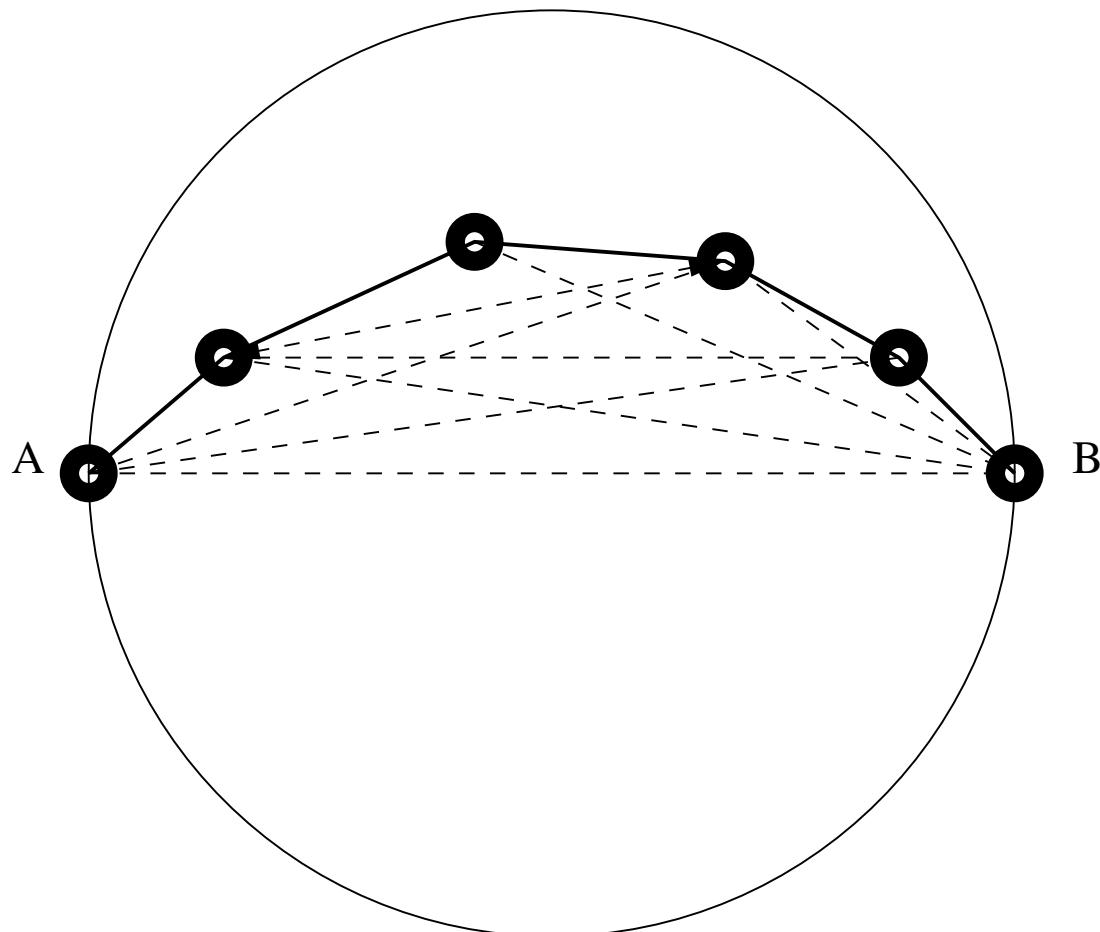
- (a) Draw the resulting UDG when $r < 1$.
 - (b) Draw the resulting UDG when $r = 1$.
 - (c) Draw the resulting UDG when $r = 1.5$.
 - (d) Draw the resulting UDG when $r = 2$.
 - (e) Draw the resulting UDG when $r = \sqrt{5}$.
 - (f) Draw the resulting UDG when $r = 4$.
9. Apply the Gabriel Test in any of the UDGs of Exercise 8 and draw the resulting graph.

Appendix

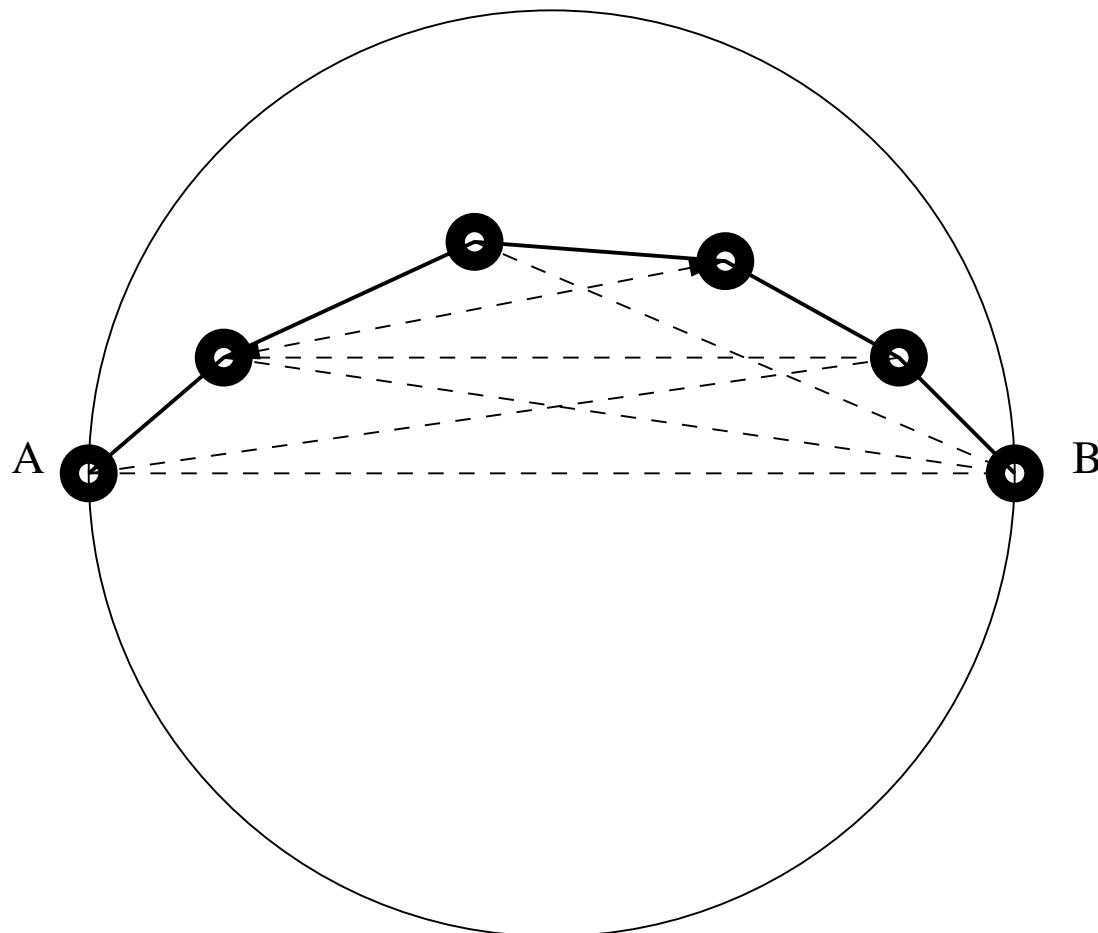
Applying the Gabriel Test (1/9)



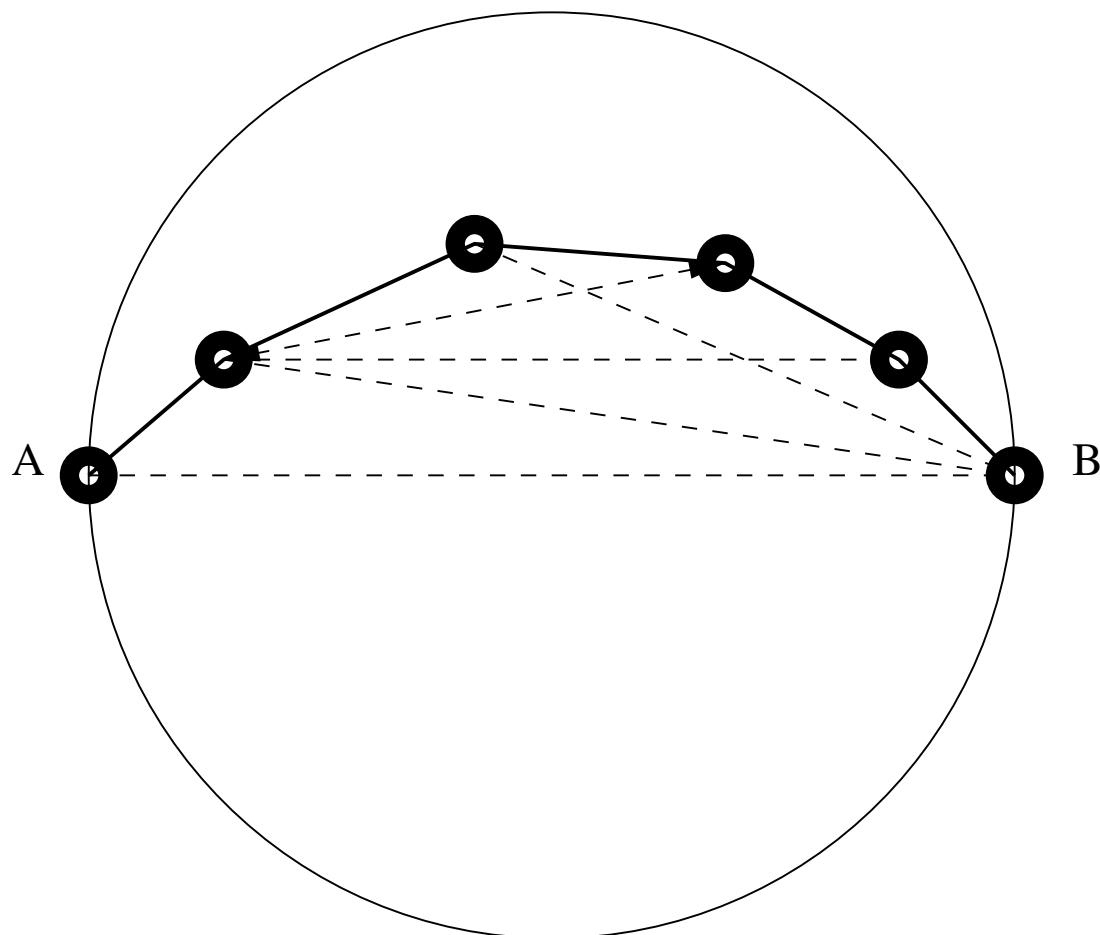
Applying the Gabriel Test (2/9)



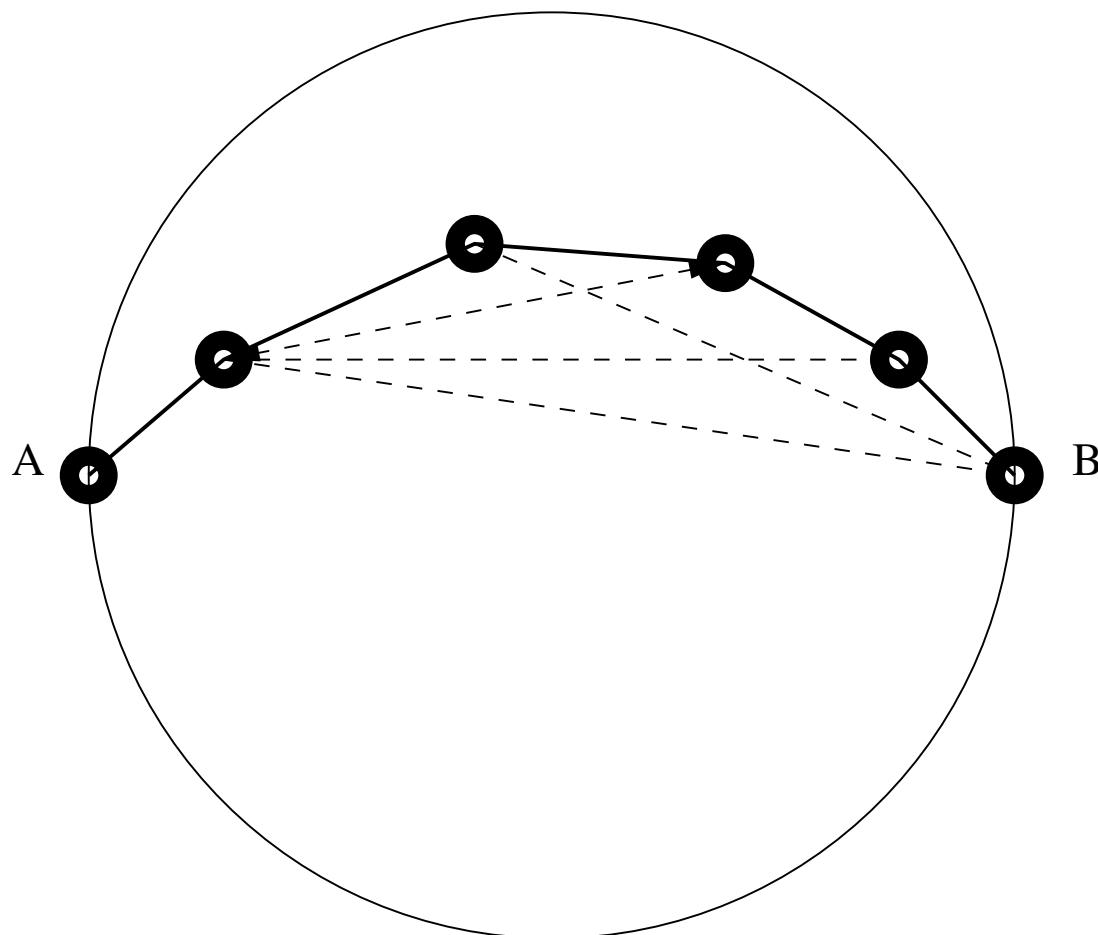
Applying the Gabriel Test (3/9)



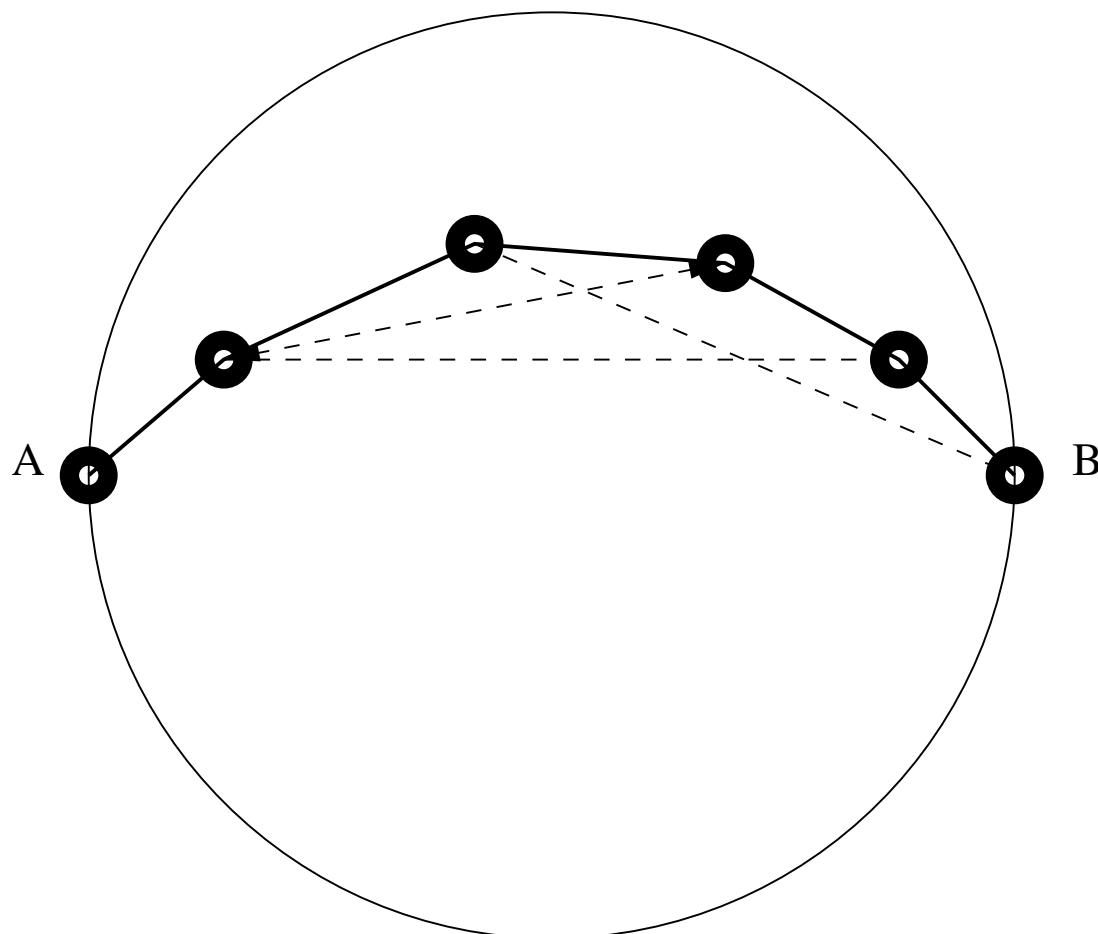
Applying the Gabriel Test (4/9)



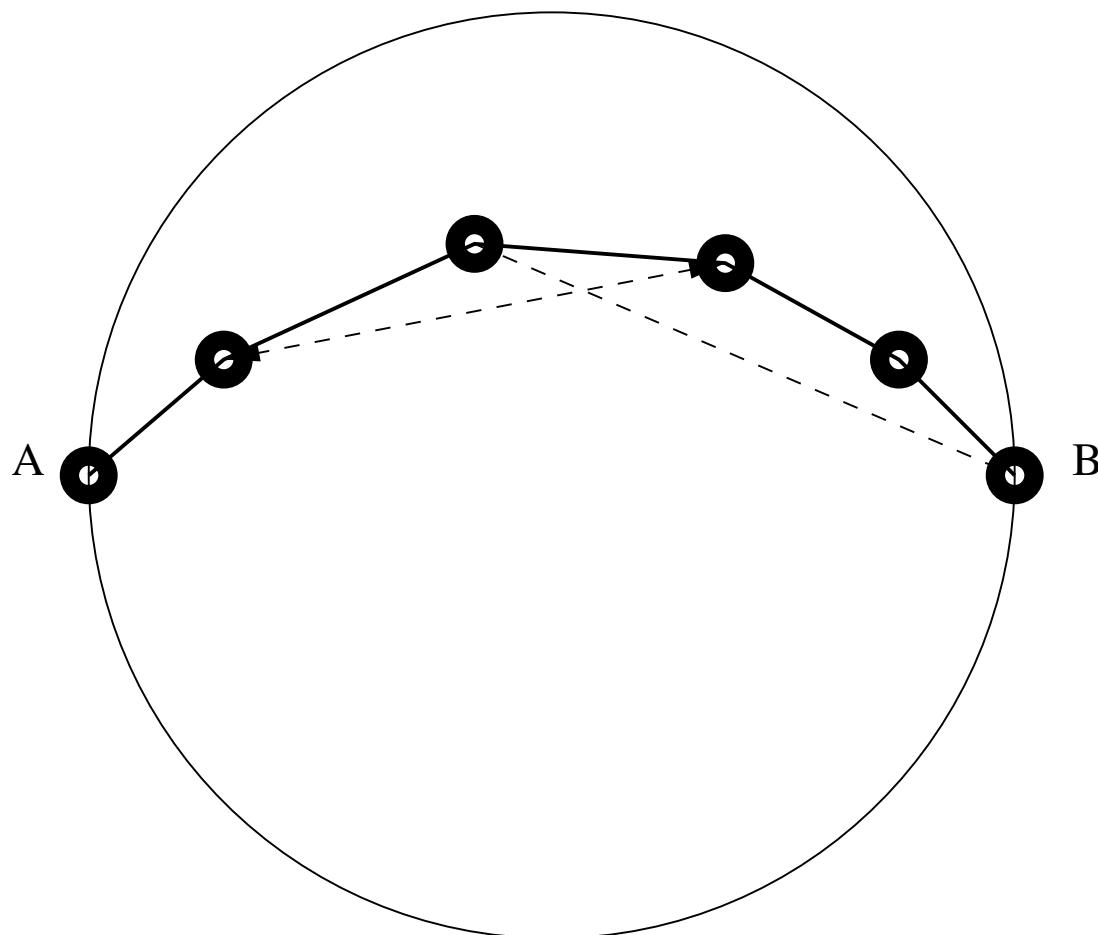
Applying the Gabriel Test (5/9)



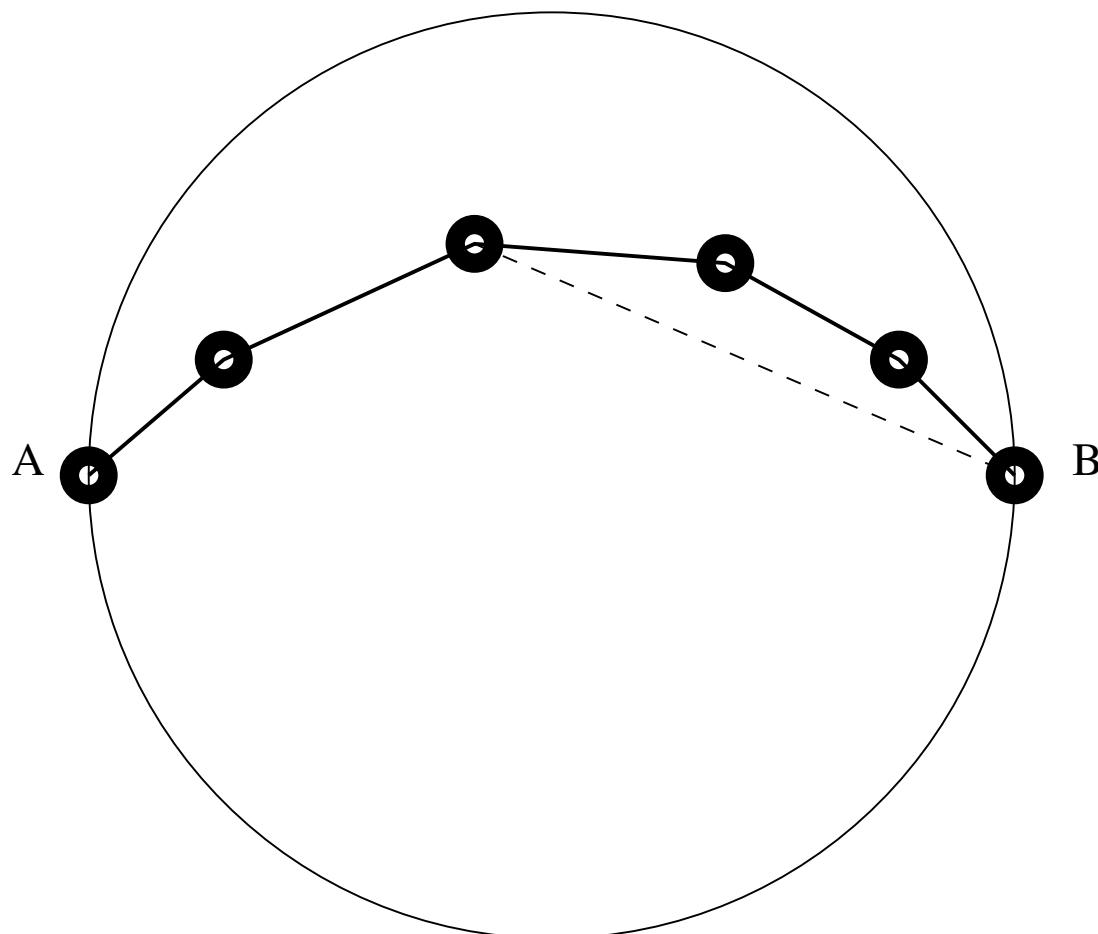
Applying the Gabriel Test (6/9)



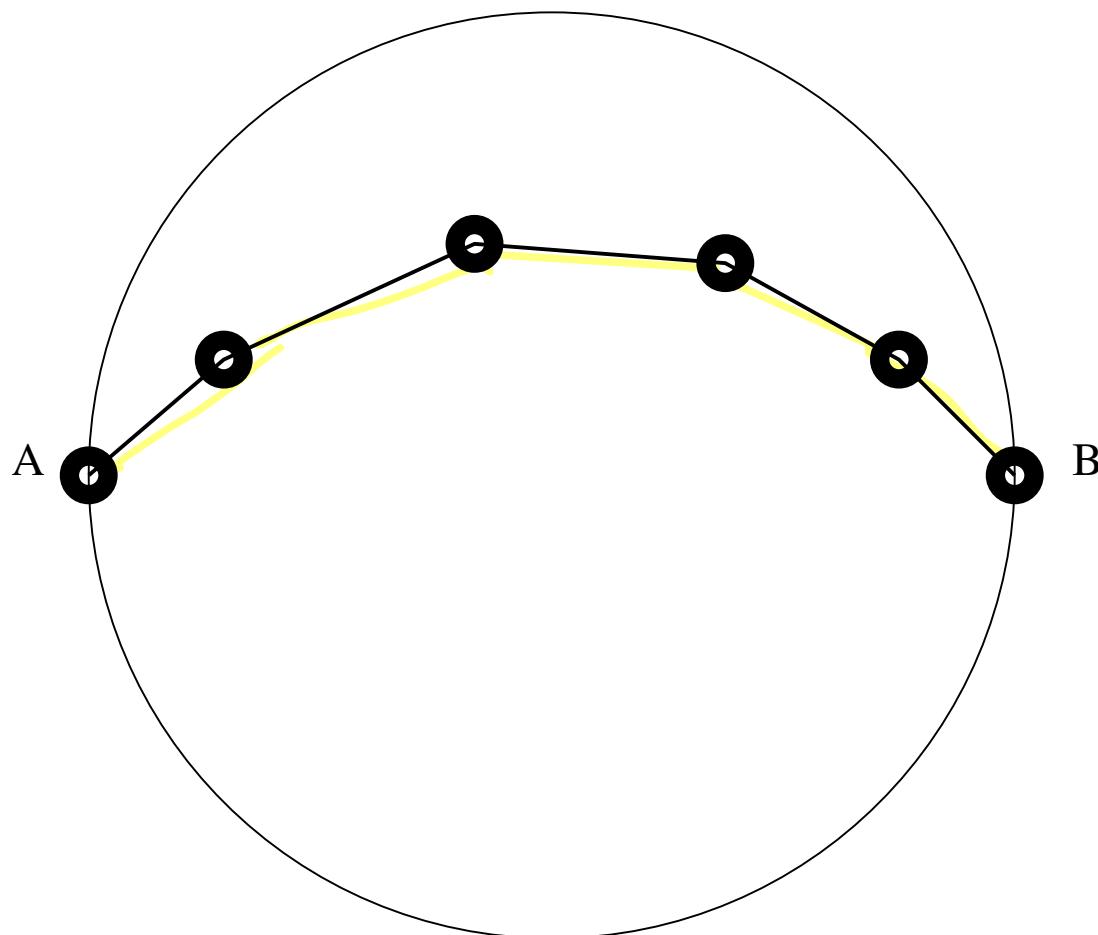
Applying the Gabriel Test (7/9)

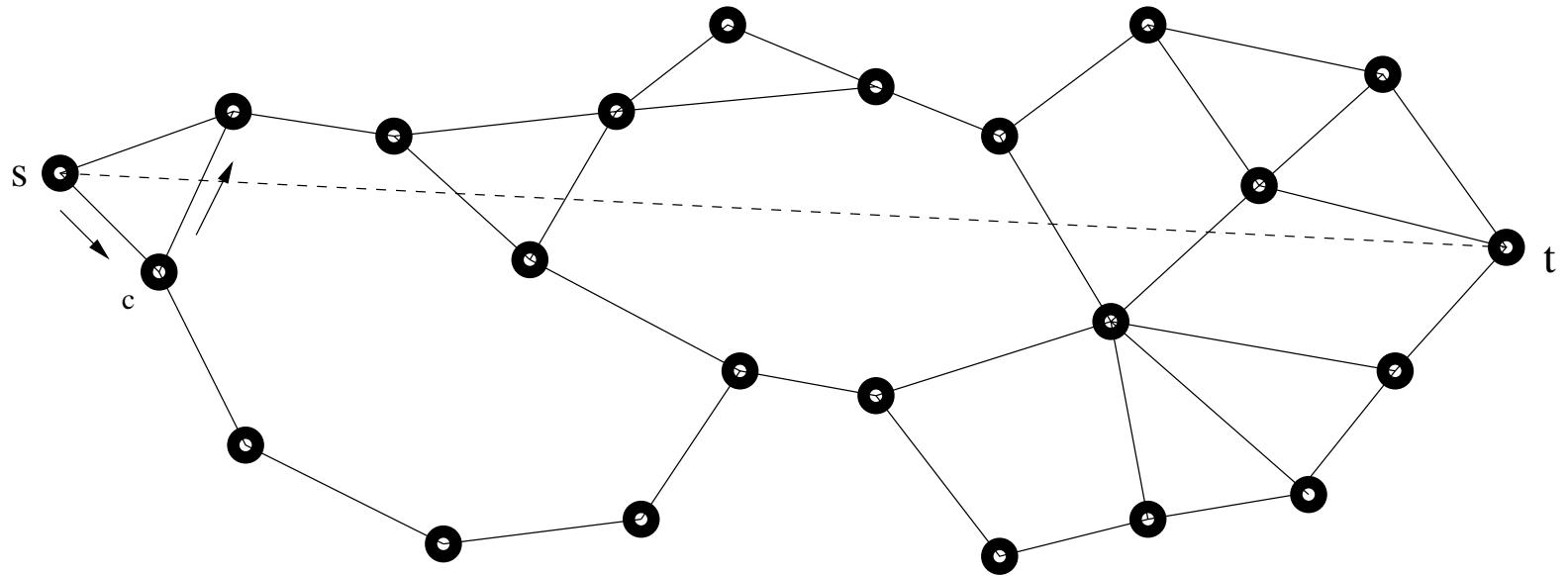


Applying the Gabriel Test (8/9)



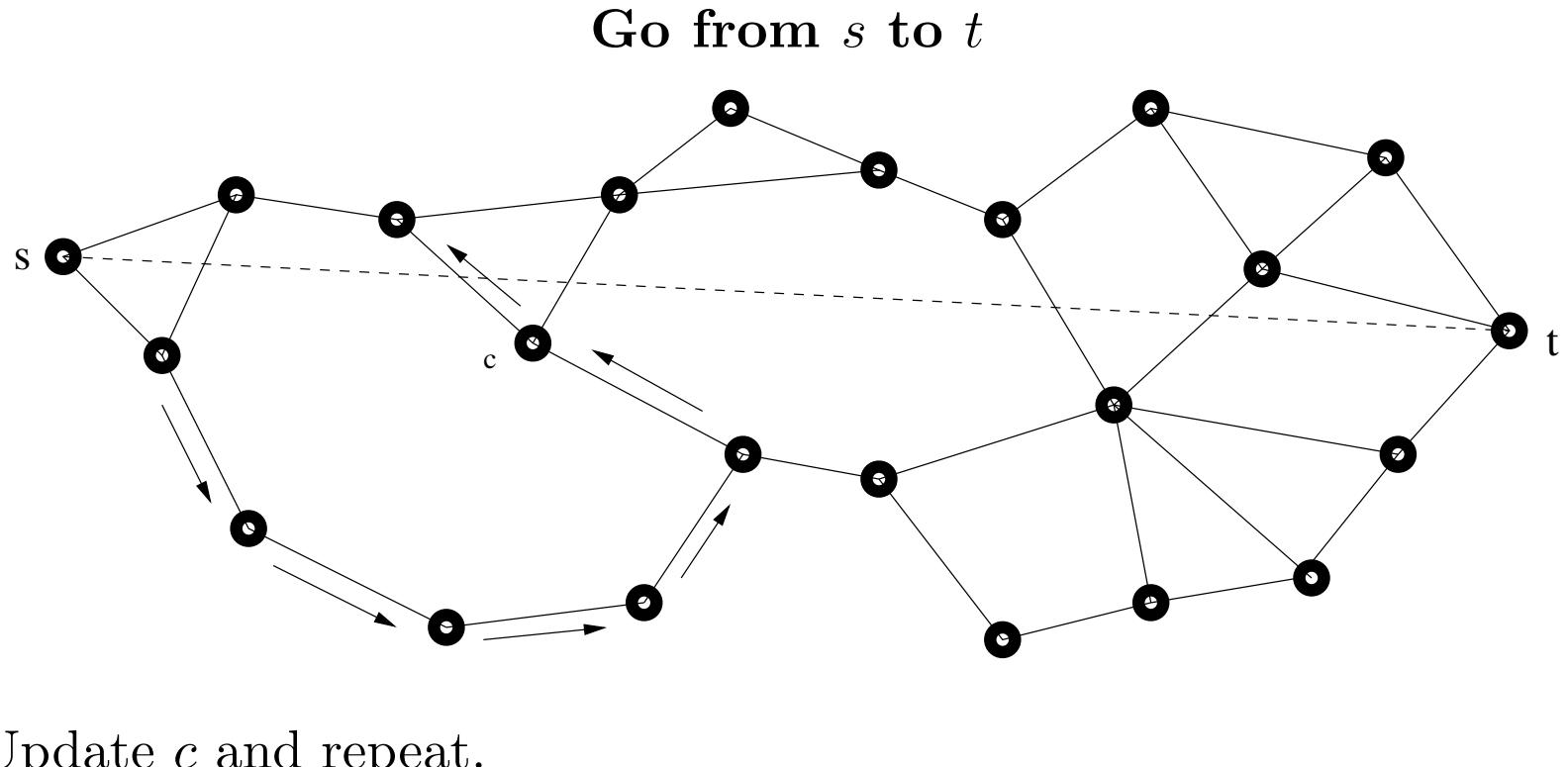
Applying the Gabriel Test (9/)9

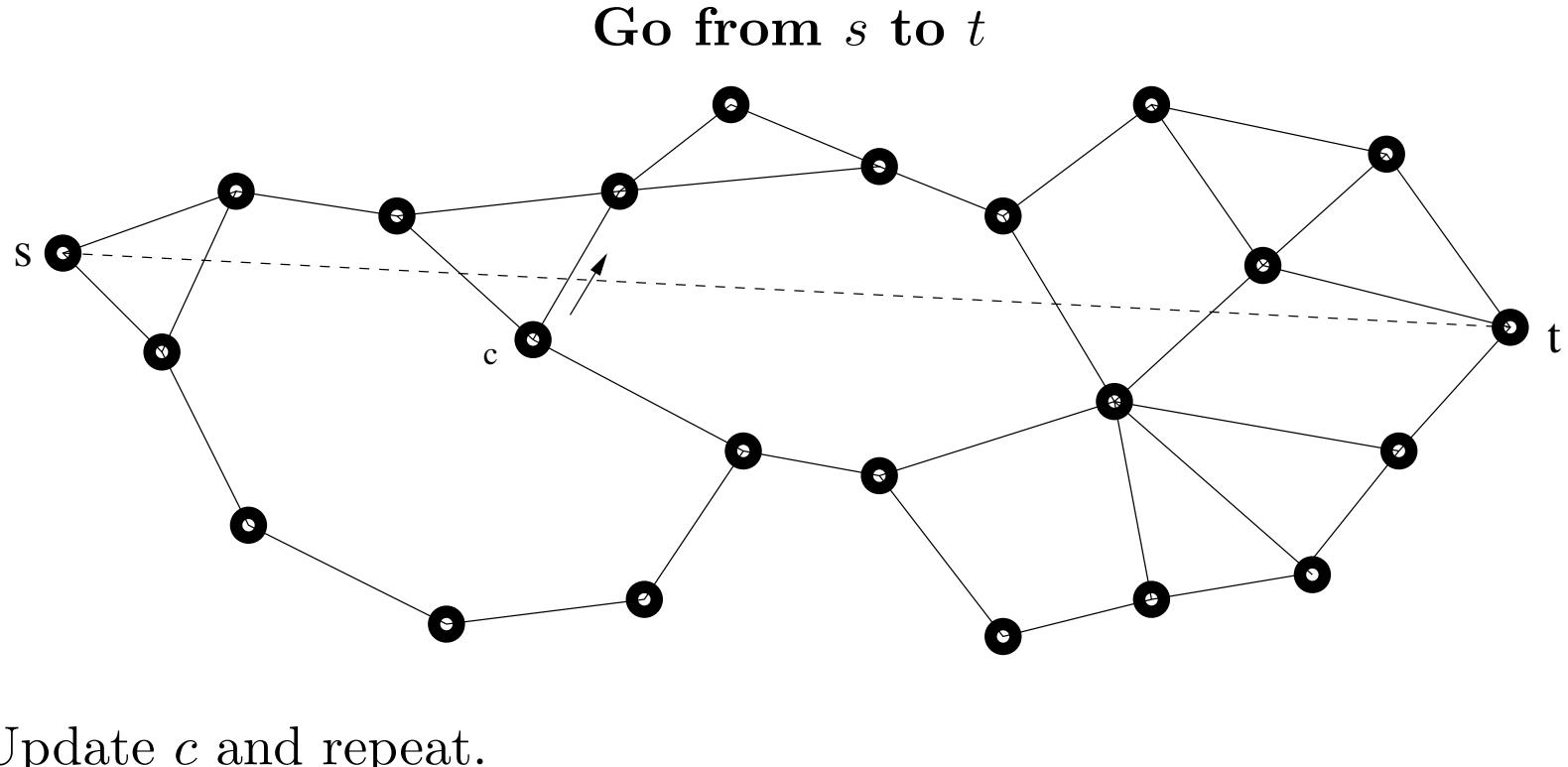


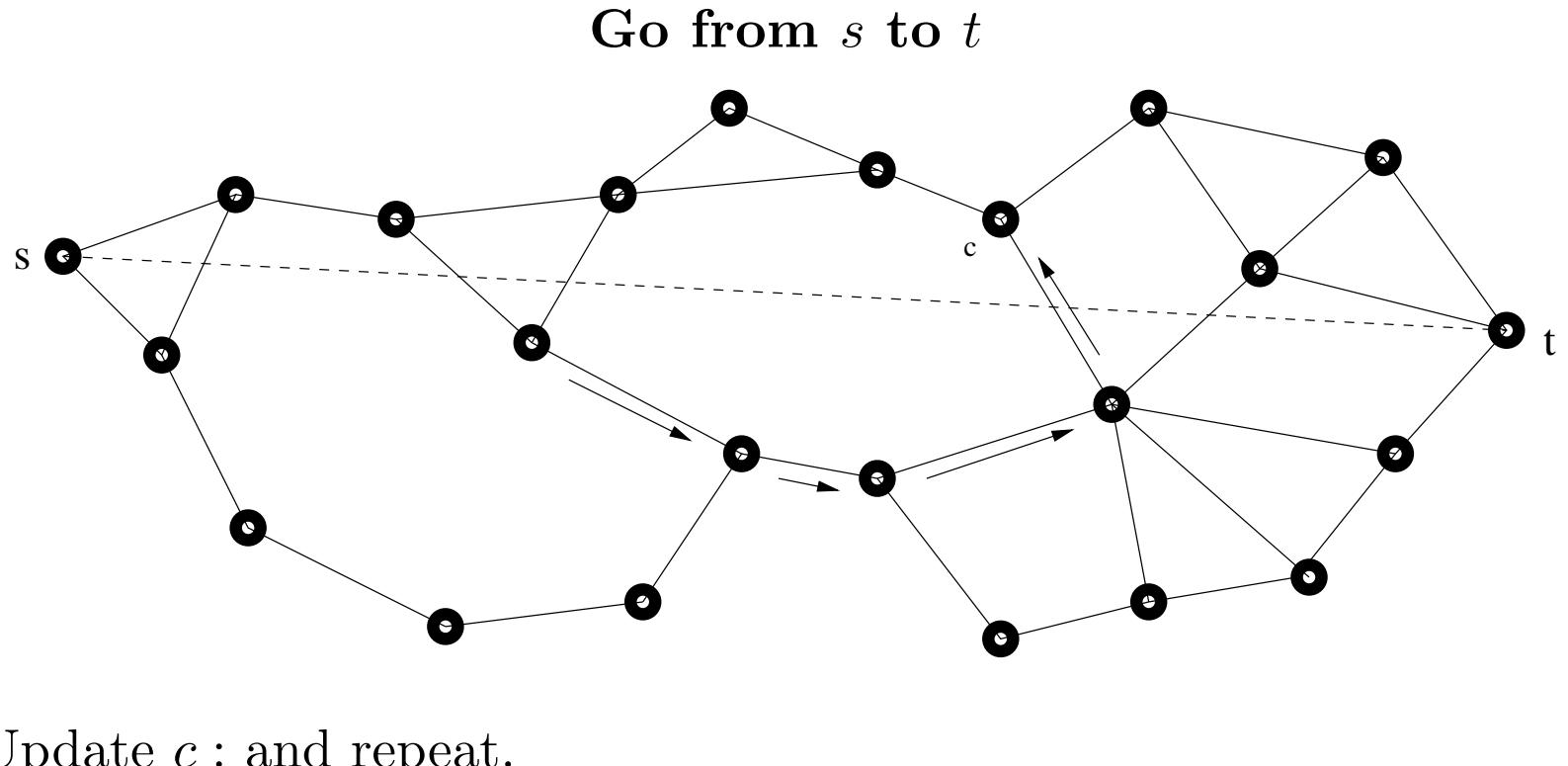
Example: Go from s to t 

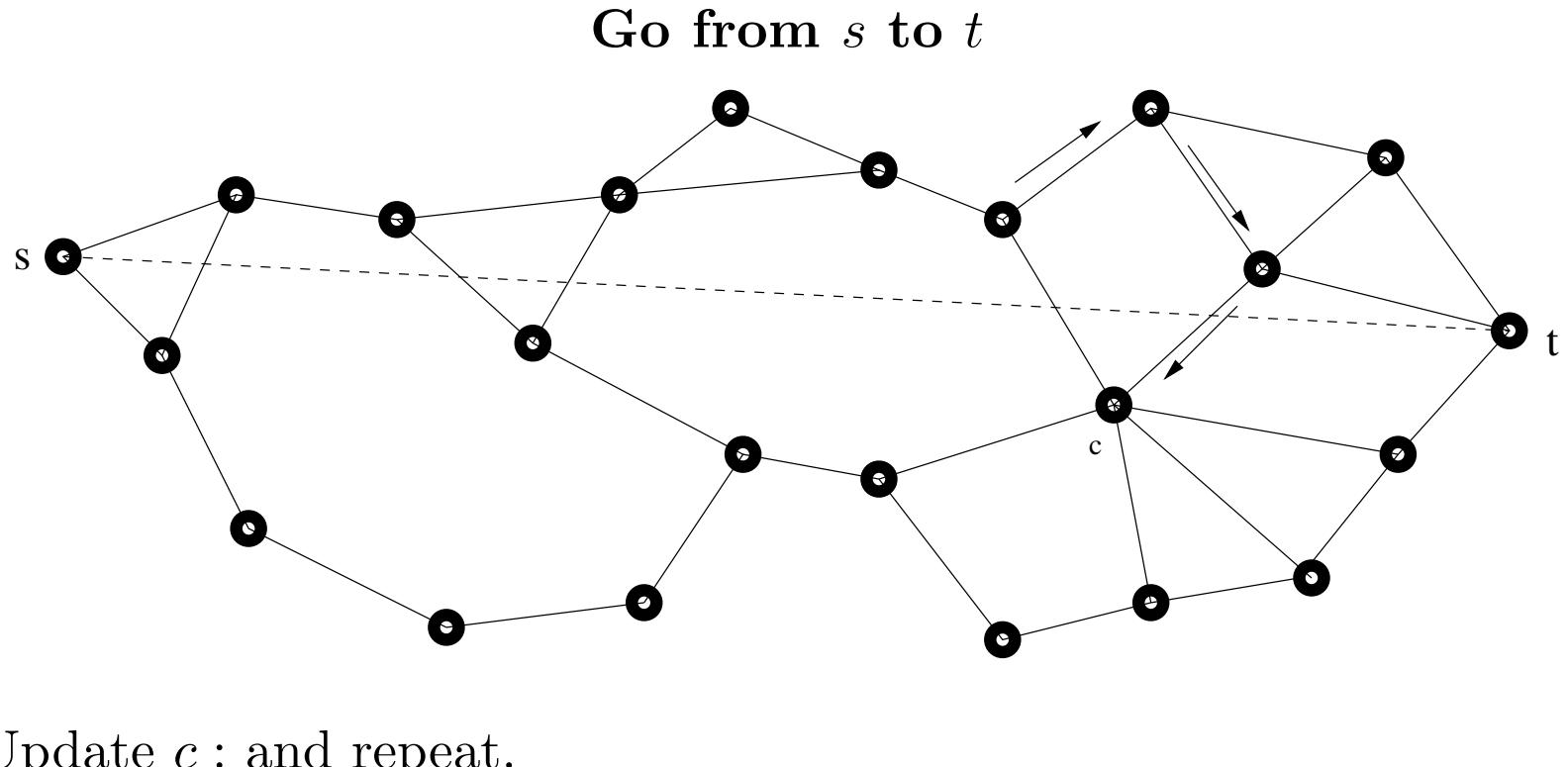
Initially $c := s$.

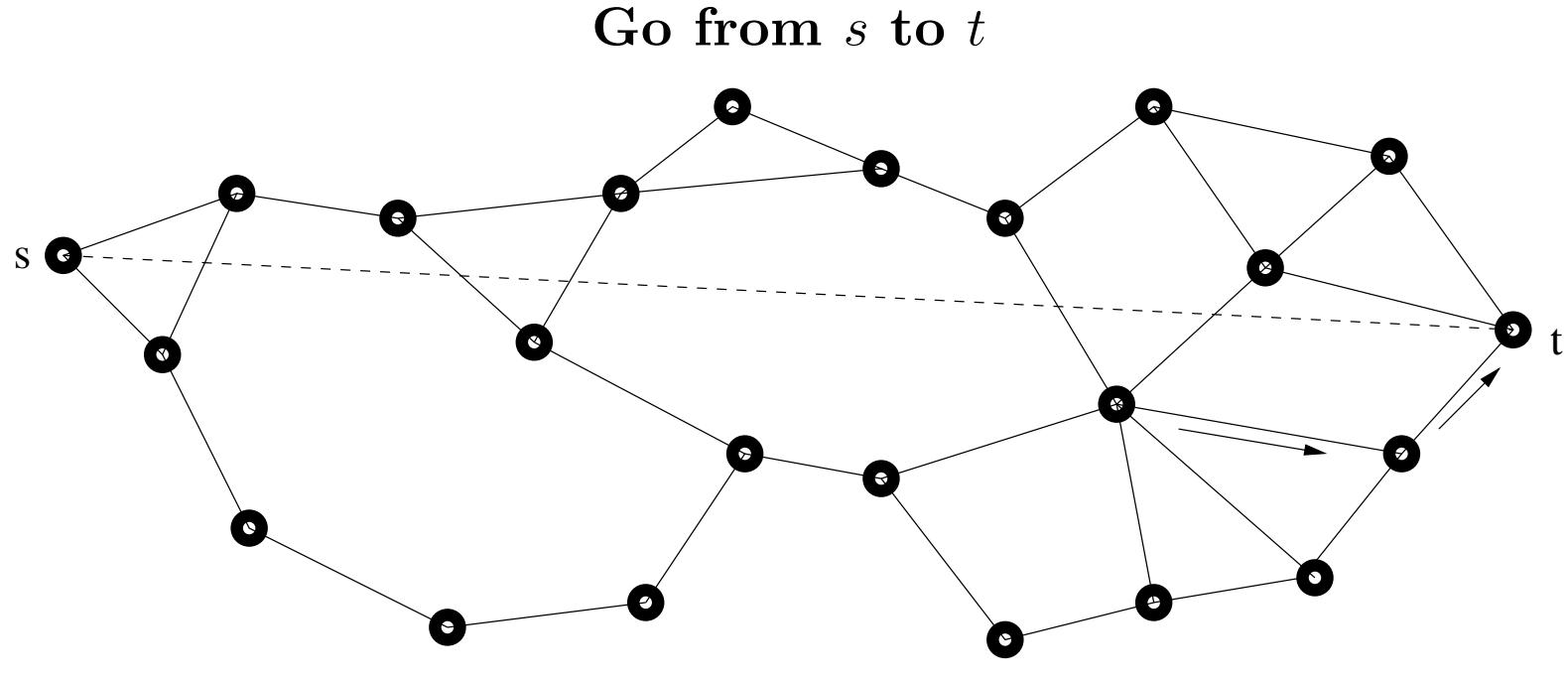
Update c and repeat.











Now t is found.

LANs: WIRELESS

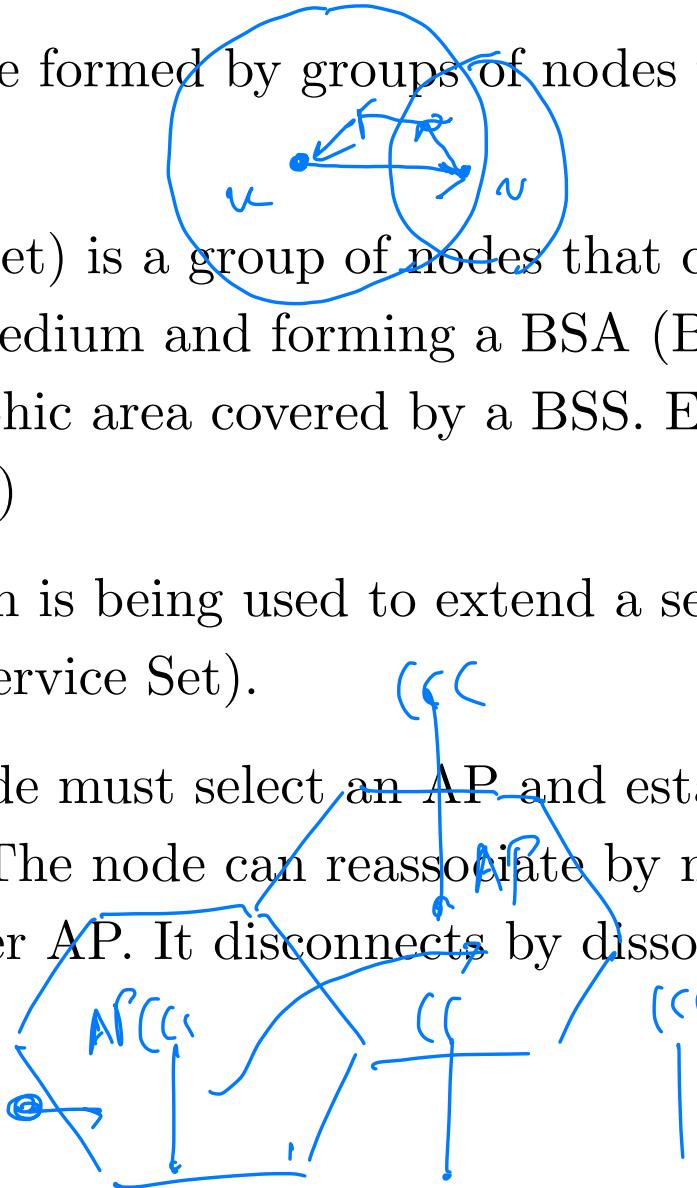
Outline

- Frequency Hoping
- CDMA
- Communication Issues
- MACA
- Bluetooth
- nG
- Satellite

Frequency Hoping

Dynamic LANs

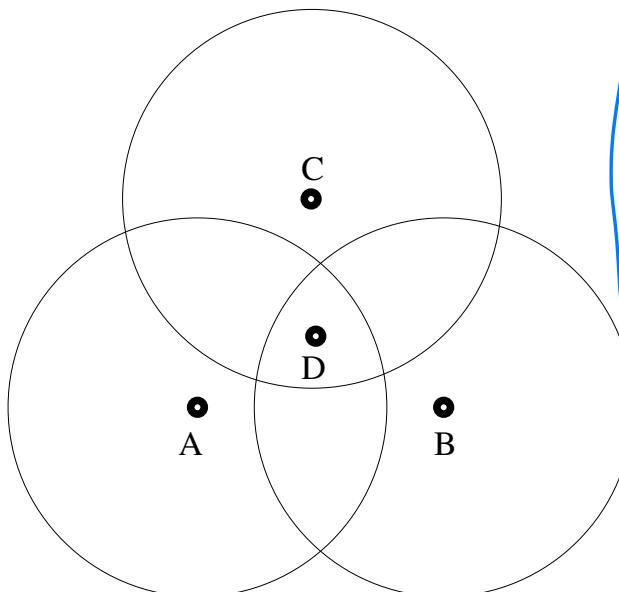
- Wireless networks are formed by groups of nodes within “range of each other”.
- BSS (Basic Service Set) is a group of nodes that coordinate their access to the medium and forming a BSA (Basic Service Area), i.e., a geographic area covered by a BSS. Each BSS has an AP (Access Point)
- A distribution system is being used to extend a set of BSSs to an ESS (Extended Service Set).
- To join an ESS a node must select an AP and establish an association with it. The node can reassociate by moving association to another AP. It disconnects by dissociating.



Collisions in Wireless Networks

- Let D be a point at the intersection of three disks centered at A, B, C .
- If A, B, C transmit at the same time then D will not be able to “hear” the message and may not even know who attempted to talk to it.

In Ethernet you can hear collisions!



In wireless when a collision occurs you may not know it!

- The issue is: How do you avoid collisions?

Spread Spectrum (1/3)

- Spread-spectrum is a radio transmission technique which refers to any method that widens the frequency band of a signal.
- Frequency hopping is the simplest version of Spread-spectrum.
- Radio stations broadcast on a single carrier frequency, which makes eavesdropping deliberately easy: You tune your radio to the correct frequency and receive the programming.
- Frequency hopping prevents the interception and decipherment of a transmission by shifting the carrier frequency in a predetermined, usually pseudorandom fashionin other words, in a way that appears random but is produced by a deterministic algorithm.

Spread Spectrum (2/3)

- A receiver hopping around in synchrony with the transmitter can pick up the message, but an eavesdropper tuned to a single frequency will hear only a blip as that bit of message flashes by.
- Frequency hopping is largely jam-proof as well. If the frequencies are spaced widely enough, any jamming signal will interfere with only a small part of the message.

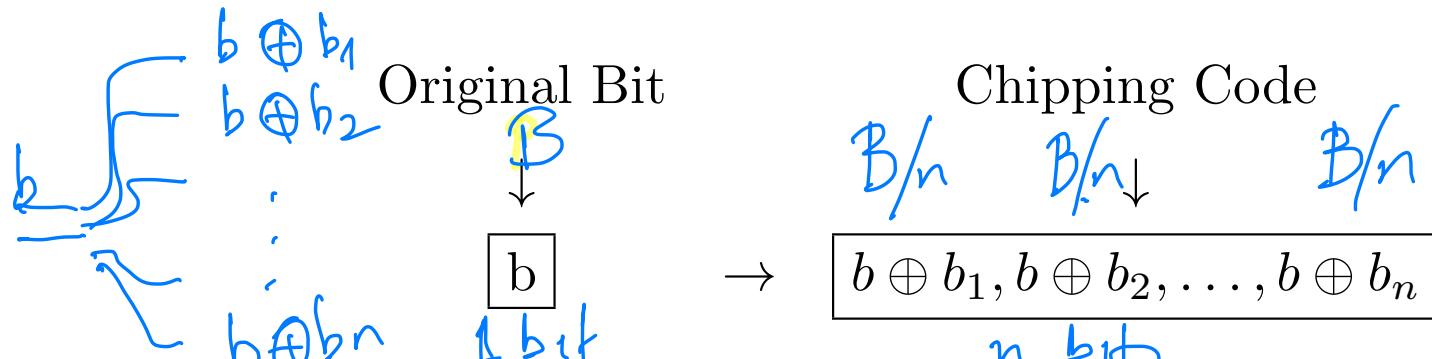
Spread Spectrum (3/3)

- Essential idea:
“spread the information signal over a wider bandwidth to make jamming (and interception) more difficult.”
- There are two types of spread spectrum techniques:
 1. Direct Sequencing.
 2. Frequency hopping.
- The advantage of doing this is to
 1. hide or encrypt signals,
 2. avoid various kinds of noise,
 3. use independently the same bandwidth (CDMA).

CDMA: Use orthogonal signals

DSSS (Direct Sequence Spread Spectrum)

- DSSS
 1. For chosen n , each transmitted bit is represented by a sequence of n bits.
 2. The n bit sequence is generated as follows: sender uses a pseudorandom generator to produce n bits, b_1, b_2, \dots, b_n and XORs b with each bit of the sequence.

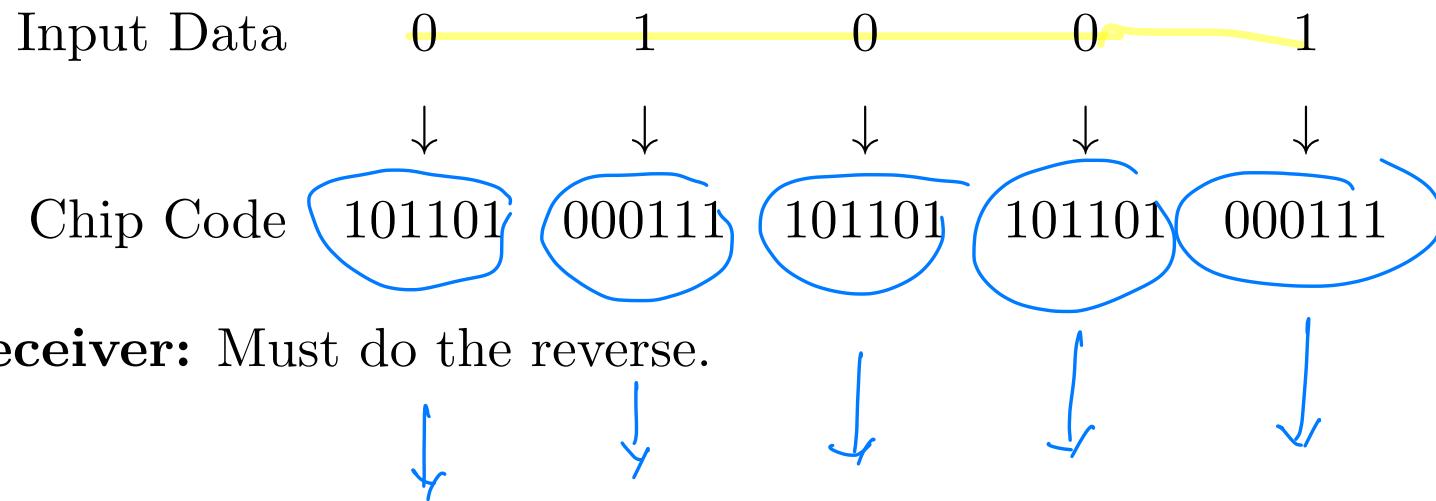


- DSSS looks like CDMA but it is implemented in the physical layer! It is not a multiple access method.

$\cdots | b \oplus b_i | \cdots$

DSSS Example

- If chip code for 0 is 101101 and for 1 is 000111 then
- **Sender:**



- **Receiver:** Must do the reverse.

FHSS (Frequency Hopping Spread Spectrum)

- This is a technique whereby the sender hops from frequency to frequency sending bits at each frequency for the same amount of time.
- After n hops the *cycle* is repeated.
- If B is the total bandwidth of the spectrum allocated then each hop must be allocated bandwidth B/n , this is the bandwidth of the subband.
- Sender and receiver must agree in advance on the subbands allocated for each hopping.
- The time a user stays in a subband is called the dwell time.

CDMA: **Code Division Multiple Access**

- CDMA is a multiplexing technique used with spread spectrum.
- Looks like DSSS but is on a different layer.
- The scheme works in the following manner.
 - Start with a data signal with rate R , called the bit data rate.
 - Break each bit into k chips according to a fixed pattern that is specific to each user, called the user's code.
 - The new channel has a “chip data” rate of kR chips per second.



channel
has a
central
bandwidth

$$R_0$$

0

$$R_0 = \frac{R}{k}$$

chip(0)

$\frac{k}{b_1 b}$

$$R = k R_0$$

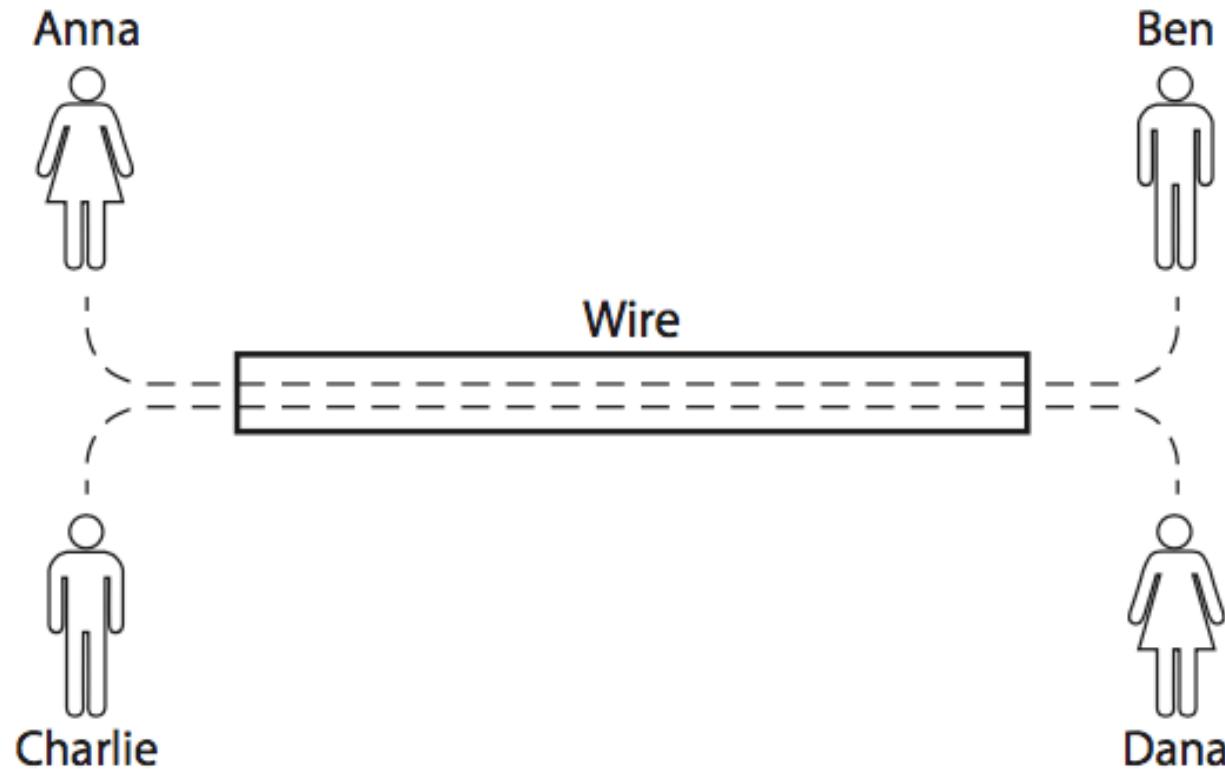
There is a limited rate R that we have; depends on the bandwidth.

To have a k -chip code, I will achieve a rate of R/k .

CDMA

Sharing a Channel

- A familiar Question:



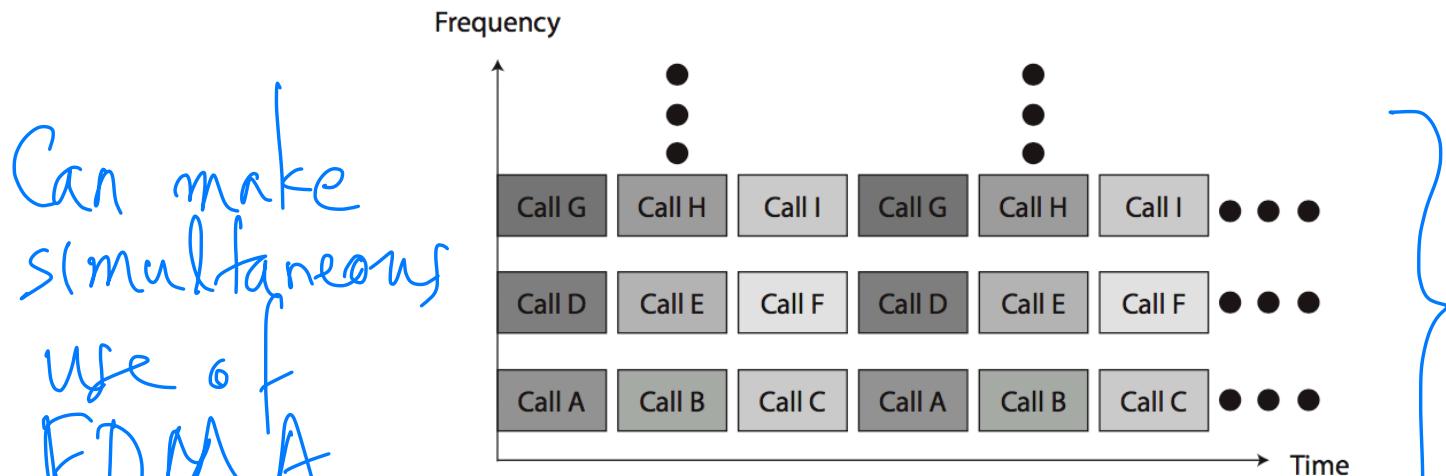
- How is it possible for both pairs to use the same wire without interfering with each other?

Sharing Methods

- Exclusive use of FDMA or TDMA



- Simultaneous use of FDMA and TDMA

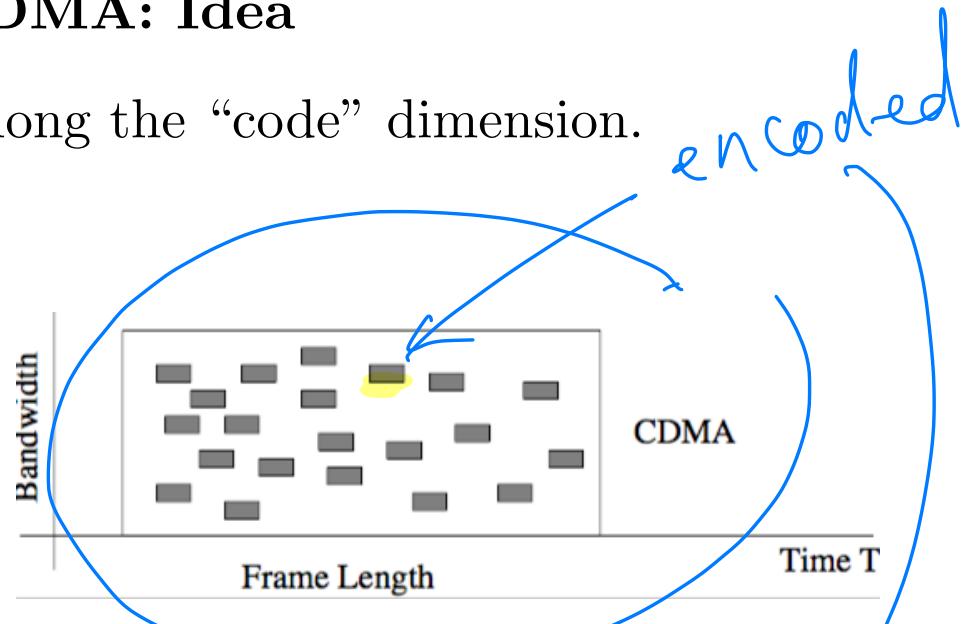
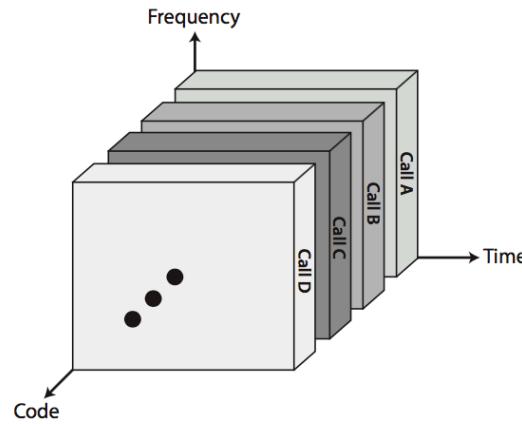


Can make
simultaneous
use of
FDMA
and TDMA

Software Radio

CDMA: Idea

- Calls are distinguished along the “code” dimension.



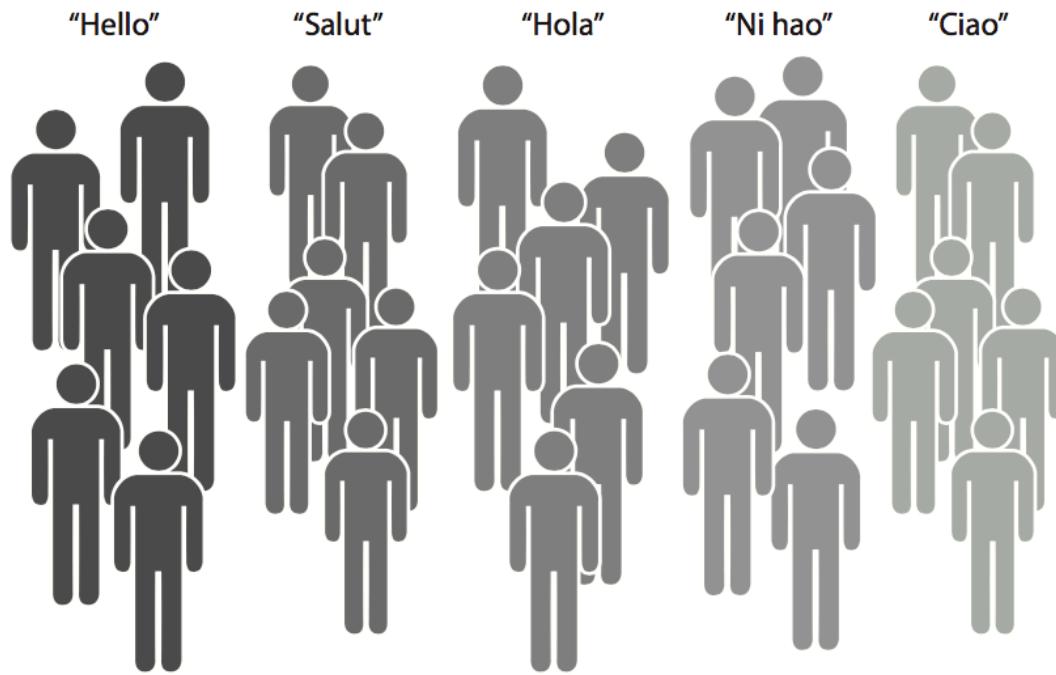
- All calls may operate over the same frequencies and at the same times, because each transmission in the network is assigned a unique code

Each user has
a specific code
to "decode" the
signal.

CDMA algorithm

CDMA: Cocktail Party

- With CDMA, each code is like a separate language.



- In the cocktail party analogy, multiple conversations can occur in a room if they use different languages.
- The issue then becomes controlling speaking volume levels.

CDMA

- Consider a simple example with $k = 6$.
- It is simplest to write a code as a sequence of
 - (+1)s and (-1)s.
- For three users, A, B, C , each of which is communicating with the same base station receiver, say R . let the codes be

$$c_A = (+1, -1, -1, +1, -1, +1) \quad \text{111}$$

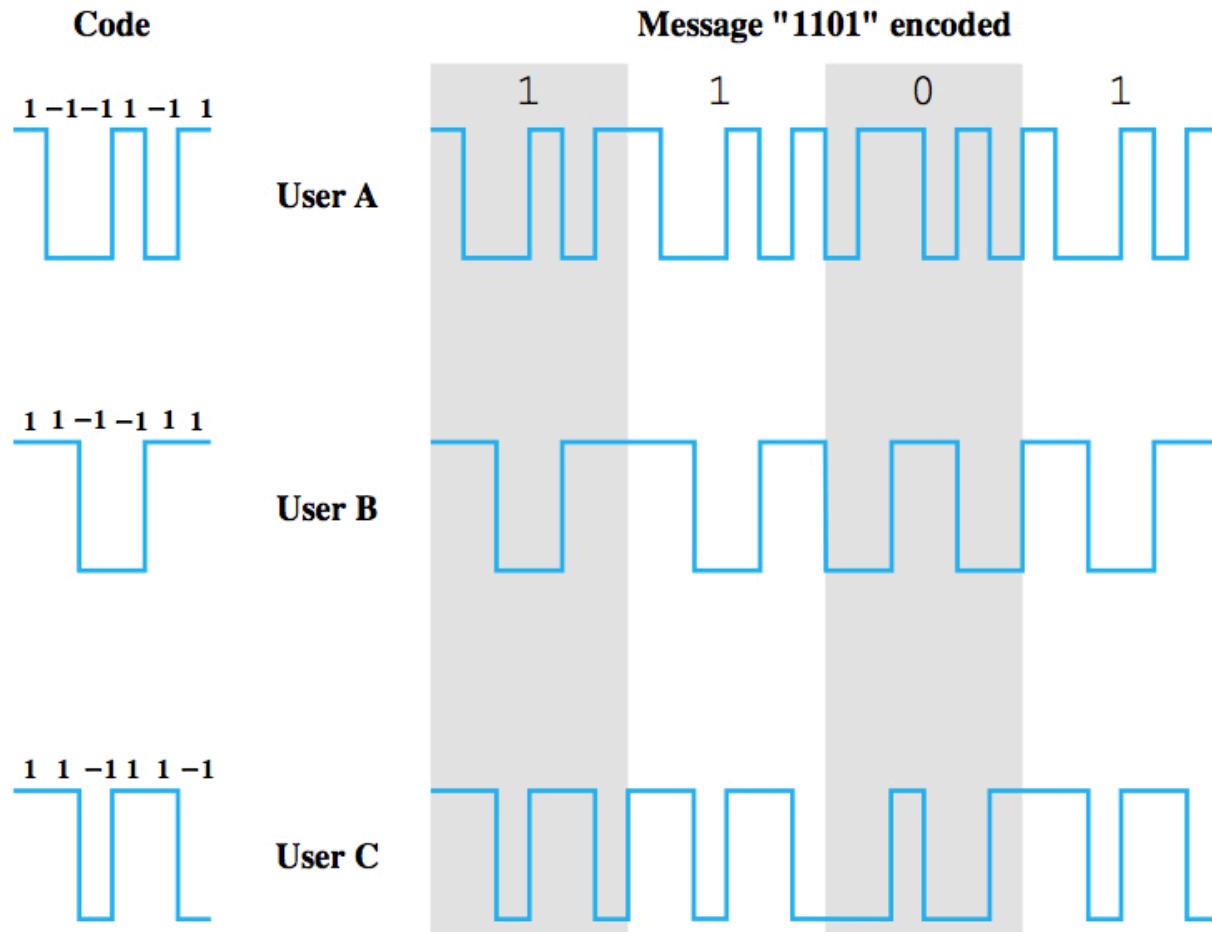
$$c_B = (+1, +1, -1, -1, +1, +1) \quad \text{111}$$

$$c_C = (+1, +1, -1, +1, +1, -1). \quad \text{111}$$

CDMA Example

- If A wants to send a **1** bit:
 A transmits its code as a chip pattern $(1, -1, -1, +1, -1, +1)$.
- If A wants to send a **0** bit:
 A transmits the complement (+1s and -1s reversed) of its code, $(-1, +1, +1, -1, +1, -1)$
- Something analogous happens with B and C .

CDMA Example



CDMA: Code Division Multiple Access

- CDMA works as follows:

Every user U owns a specific bit pattern consisting of n bits:

$$\textcolor{blue}{C}_U = (b_1, b_2, \dots, b_n).$$

- The main question is

How are patterns selected?

Patterns are selected to be pairwise mutually orthogonal.

CDMA (1/2)

- Each of n users, U , is assigned a vector $\mathbf{u} \in \{-1, +1\}^n$.

$$U \leftarrow \mathbf{u} = (u_1, u_2, \dots, u_n)$$

u_i are the components of the vector \mathbf{u} . $\bar{\mathbf{u}} = (-u_1, -u_2, \dots, -u_n)$

- Let $\bar{\mathbf{u}} = (-u_1, -u_2, \dots, -u_n)$ denote the bit-complement of $\mathbf{u} = (u_1, u_2, \dots, u_n)$.
- Note that^a

$$\underbrace{\langle \mathbf{u}, \mathbf{u} \rangle}_{\sim n} = \frac{1}{n} \sum_{i=1}^n u_i u_i = 1$$

$$\underbrace{\langle \mathbf{u}, \bar{\mathbf{u}} \rangle}_{\sim n} = \frac{1}{n} \sum_{i=1}^n u_i (-u_i) = -1$$

Division
by n is
to normalize
the values

^aThe notation $\langle \cdot, \cdot \rangle$ means inner product of vectors.

The inner product of

$$u = (u_1, u_2, \dots, u_n)$$

$$v = (v_1, v_2, \dots, v_n)$$

$$\langle u, v \rangle = \frac{1}{n} \sum_{i=1}^n u_i \cdot v_i$$

(from linear algebra)!

CDMA (2/2)

- **Orthogonality Condition:** The vectors assigned to the users are pairwise orthogonal, i.e. for any users $U \neq V$,

$$\langle \mathbf{u}, \mathbf{v} \rangle := \frac{1}{n} \sum_{i=1}^n u_i v_i = 0$$

$$\langle \bar{\mathbf{u}}, \mathbf{v} \rangle = 0$$

- Hence, also

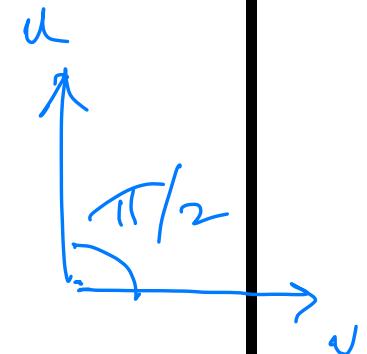
$$\langle \mathbf{u}, \bar{\mathbf{v}} \rangle := \frac{1}{n} \sum_{i=1}^n u_i (-v_i) = 0$$

- **Transmission:** To transmit a bit 0 or 1 user U sends a vector as follows:

To transmit 1 user U sends its vector: \mathbf{u}

To transmit 0 user U sends complement of its vector: $\bar{\mathbf{u}}$

$$\begin{array}{ll} U & \xleftarrow{} u, \bar{u} \\ V & \xleftarrow{} v, \bar{v} \end{array}$$



CDMA Example

- Code assignment

User	$\{0, 1\}$ -Vector	$\{-1, +1\}$ -Vector
A	00011011	$\mathbf{a} := -1-1-1+1+1-1+1+1$
B	00101110	$\mathbf{b} := -1-1+1-1+1+1+1-1$
C	01011100	$\mathbf{c} := -1+1-1+1+1+1-1-1$
D	01000011	$\mathbf{d} := -1+1-1-1-1+1+1$

To transmit data sequence 1011:

- A sends: $\mathbf{a}\bar{\mathbf{a}}\mathbf{a}\bar{\mathbf{a}}$ = 00011011111001000001101100011011
- B sends: $\mathbf{b}\bar{\mathbf{b}}\mathbf{b}\bar{\mathbf{b}}$ = 00101110110100010010111000101110
- etc



CDMA: Additivity

- Let $\{b_U : U \text{ is a user}\}$ be the vectors transmitted by the users on a given transmitted bit $b = 1$ or $b = 0$.
- According to our assumptions/definitions this means that

$$b_U = \mathbf{u} \text{ if } b = 1$$

$$b_U = \bar{\mathbf{u}} \text{ if } b = 0$$

- When a subset S of the set of users transmits simultaneously then the vector sum

$$\sum_{U \in S} b_U$$

is being transmitted.

- How does a user recover the bit from this sum?

CDMA: Decoding (1/2)

- If a station wants to recover the message transmitted by user U from a set S of users then it computes the inner product

$$\langle \mathbf{u}, \sum_{V \in S} b_V \rangle$$

A user can do this because \mathbf{u} is known!

- Also note that^a

$$\langle \mathbf{u}, \sum_{V \in S} b_V \rangle = \sum_{V \in S} \langle \mathbf{u}, b_V \rangle$$

by the linearity of the inner product.

Additivity
of inner-product!

^aThis is called additivity property of the inner product.

CDMA: Decoding (2/2)

- But for each user V we have that

$$\begin{aligned} \langle \mathbf{u}, b_V \rangle &= \begin{cases} \langle \mathbf{u}, \mathbf{v} \rangle & \text{if } b_V = \mathbf{v} \\ \langle \mathbf{u}, \bar{\mathbf{v}} \rangle & \text{if } b_V = \bar{\mathbf{v}} \end{cases} \\ &= \begin{cases} 0 & \text{if } U \neq V \\ +1 & \text{if } U = V \text{ and } b_V = \mathbf{v} \\ -1 & \text{if } U = V \text{ and } b_V = \bar{\mathbf{v}} \end{cases} \end{aligned}$$

- This is because:
 - if $U \neq V$ then $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \bar{\mathbf{v}} \rangle = 0$;
 - if $U = V$ and $b_V = \mathbf{v}$ then $\langle \mathbf{u}, b_V \rangle = \langle \mathbf{u}, \mathbf{v} \rangle = 1$;
 - if $U = V$ and $b_V = \bar{\mathbf{v}}$ then $\langle \mathbf{u}, b_V \rangle = \langle \mathbf{u}, \bar{\mathbf{v}} \rangle = -1$;
- In other words, each user V will recover the bit that was sent to it in encoded form!

CDMA: Walsh Matrices

- The Walsh matrices of dimension 2^k are given by the recursive formula

$$W(2^1) = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}$$

and

$$W(2^{k+1}) = \begin{bmatrix} W(2^k) & W(2^k) \\ W(2^k) & -W(2^k) \end{bmatrix}$$

Gramm-Schmidt
Orthogonalization Procedure

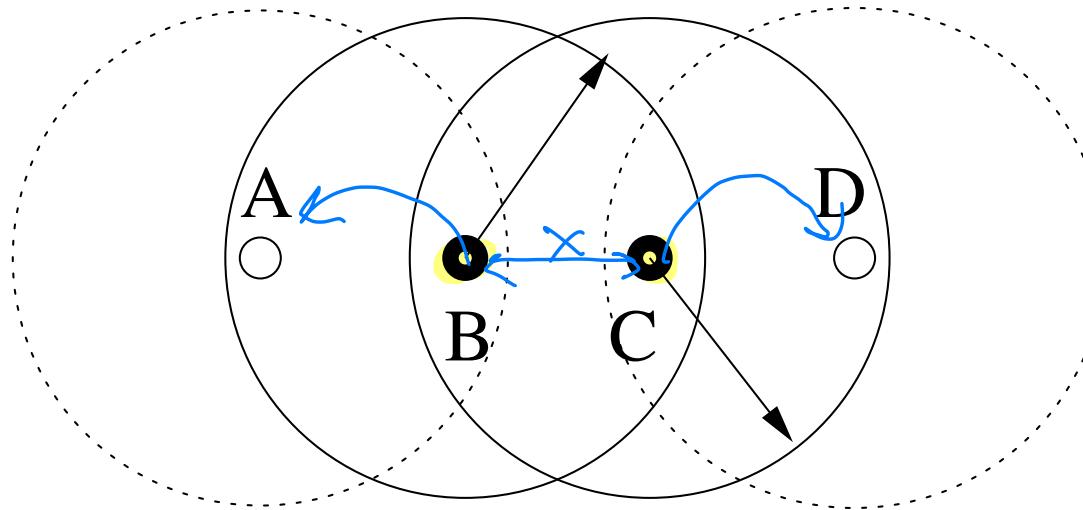
Communication Issues

Communication Problems

- Collision Avoidance
- Exposed Node
- Communication Paths
- Asymmetry
- Attenuation
- Power Level
- Interference
- SIR

Collision Avoidance

- B and C will collide if they transmit at the same time.

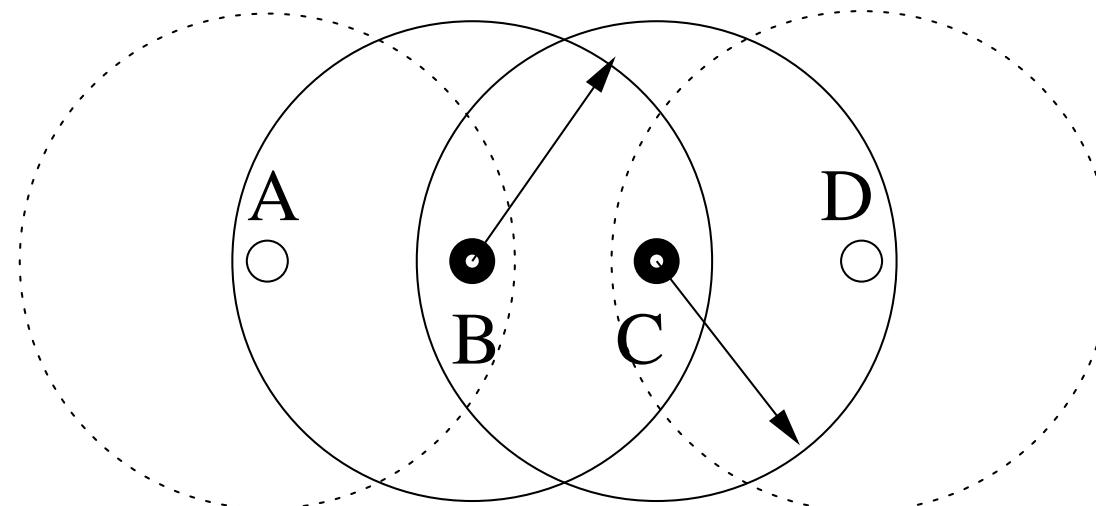


- A can reach B but is unaware of C.
- C can reach B but is unaware of A.

Exposed Node

Exposed Node

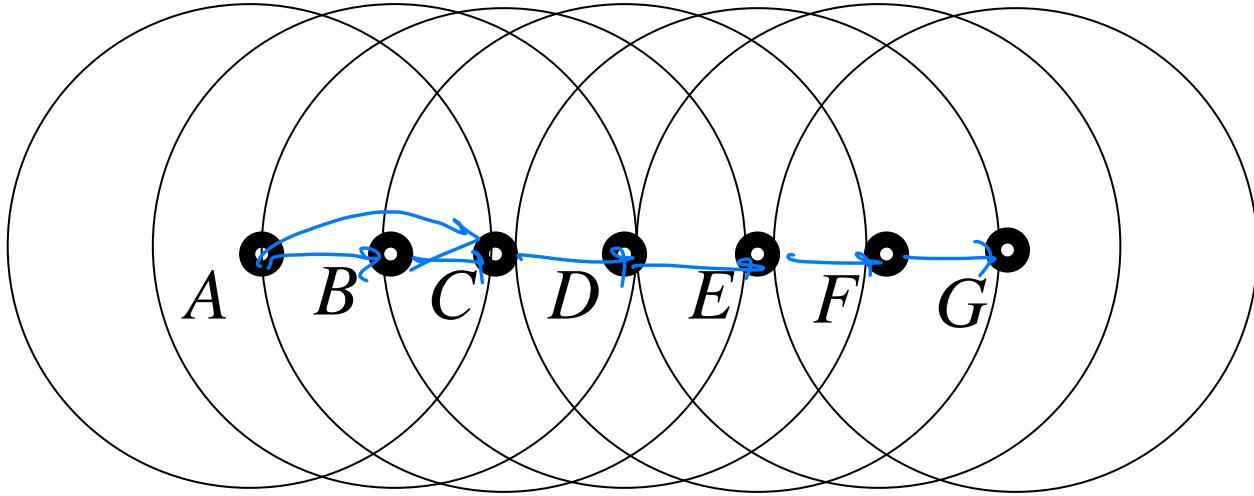
- C can hear if B sends to A.
- It is a mistake for C to assume that it cannot transmit to D.



- In fact: C can transmit to D and simultaneously B can transmit to A.

Communication Paths in Wireless

- Each node forwards to a node within its range:



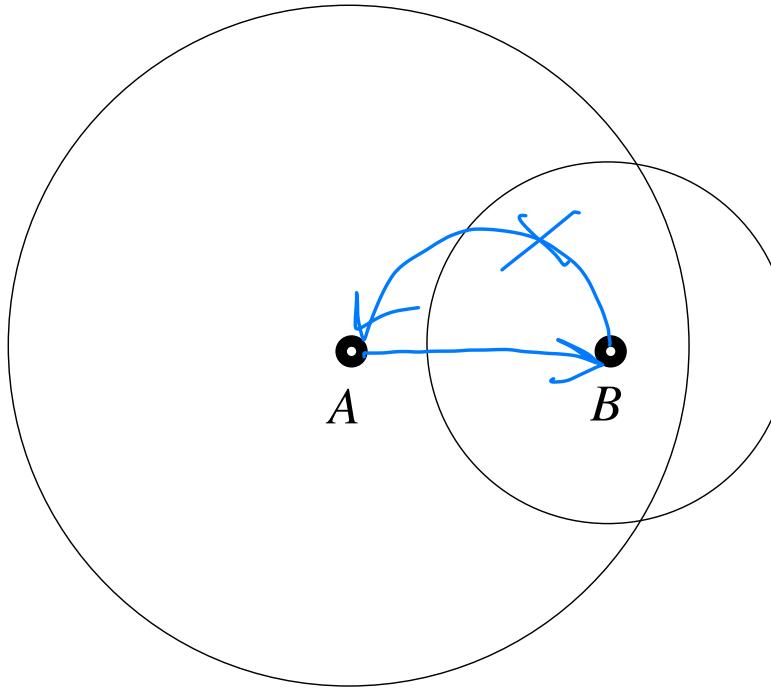
- This gives a communication path:

$$A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow F \rightarrow G$$

You are limited by your range

Communication Asymmetry in Wireless

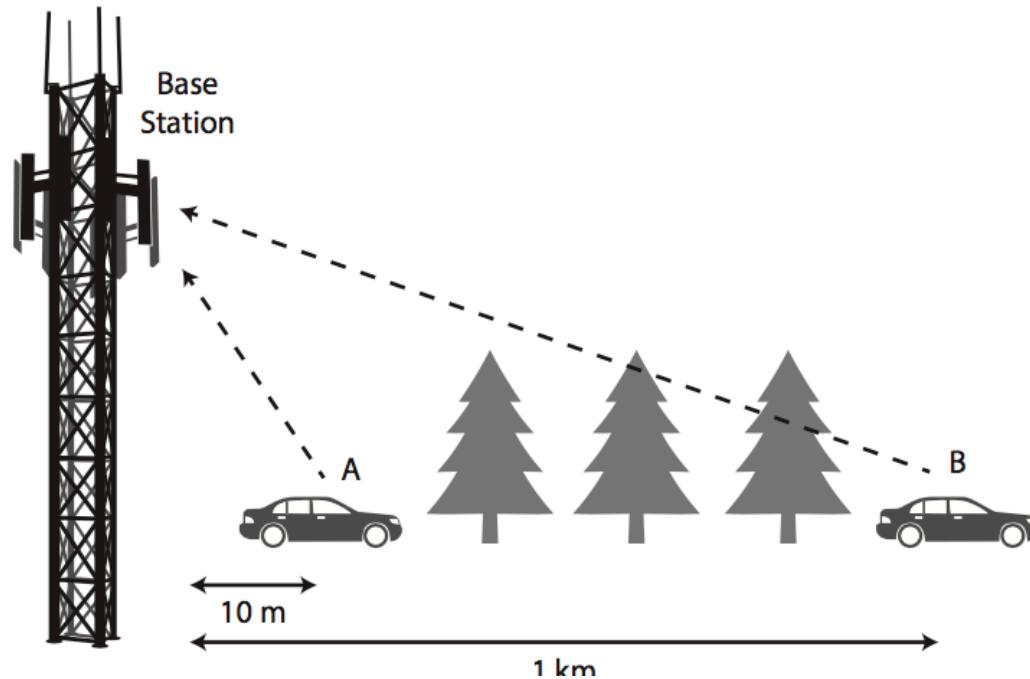
- In the real world, there is asymmetry:



- A can reach B but B cannot reach A .

Attenuation

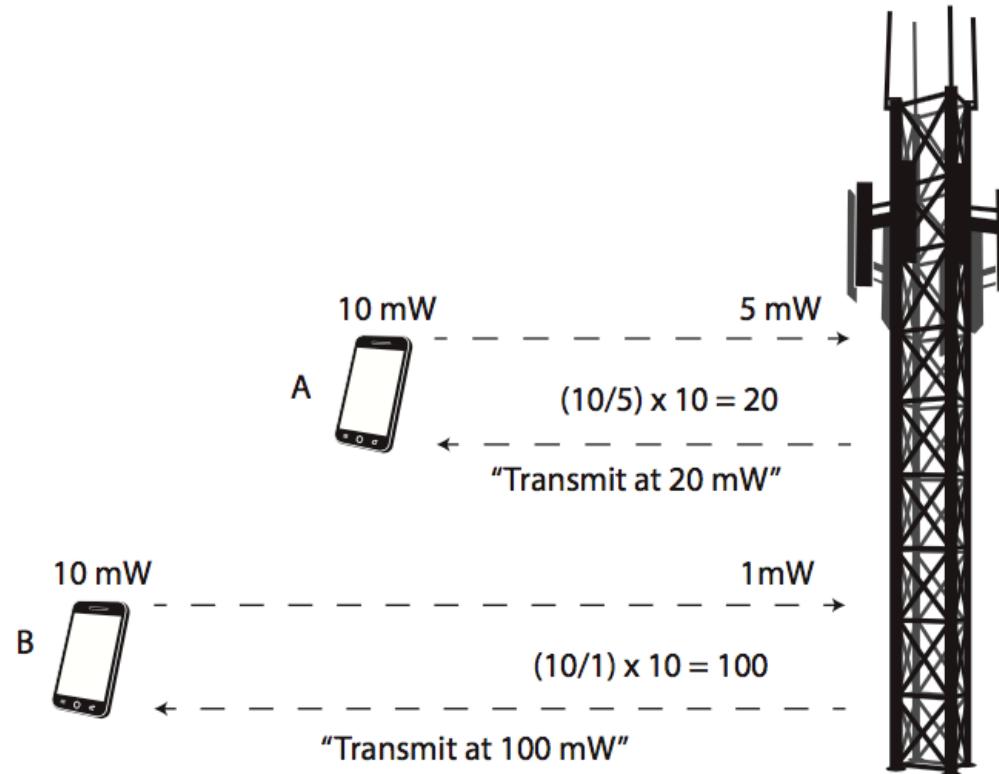
- The farther a transmitter is from its receiver, the higher the attenuation is, and the more objects there are to obstruct the path.



- Here, A has a short, clear path to the tower, while B has a long path that is obstructed by objects (e.g., trees)

Power Levels

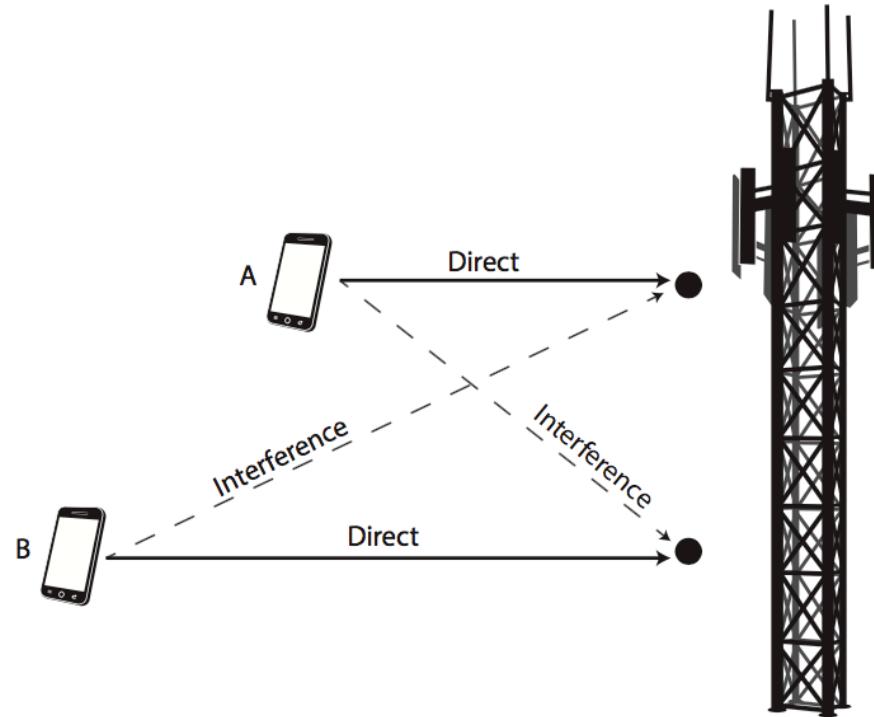
- Transmission power control (TPC) algorithm.



- Attempts to equalize received signal powers.

Interference

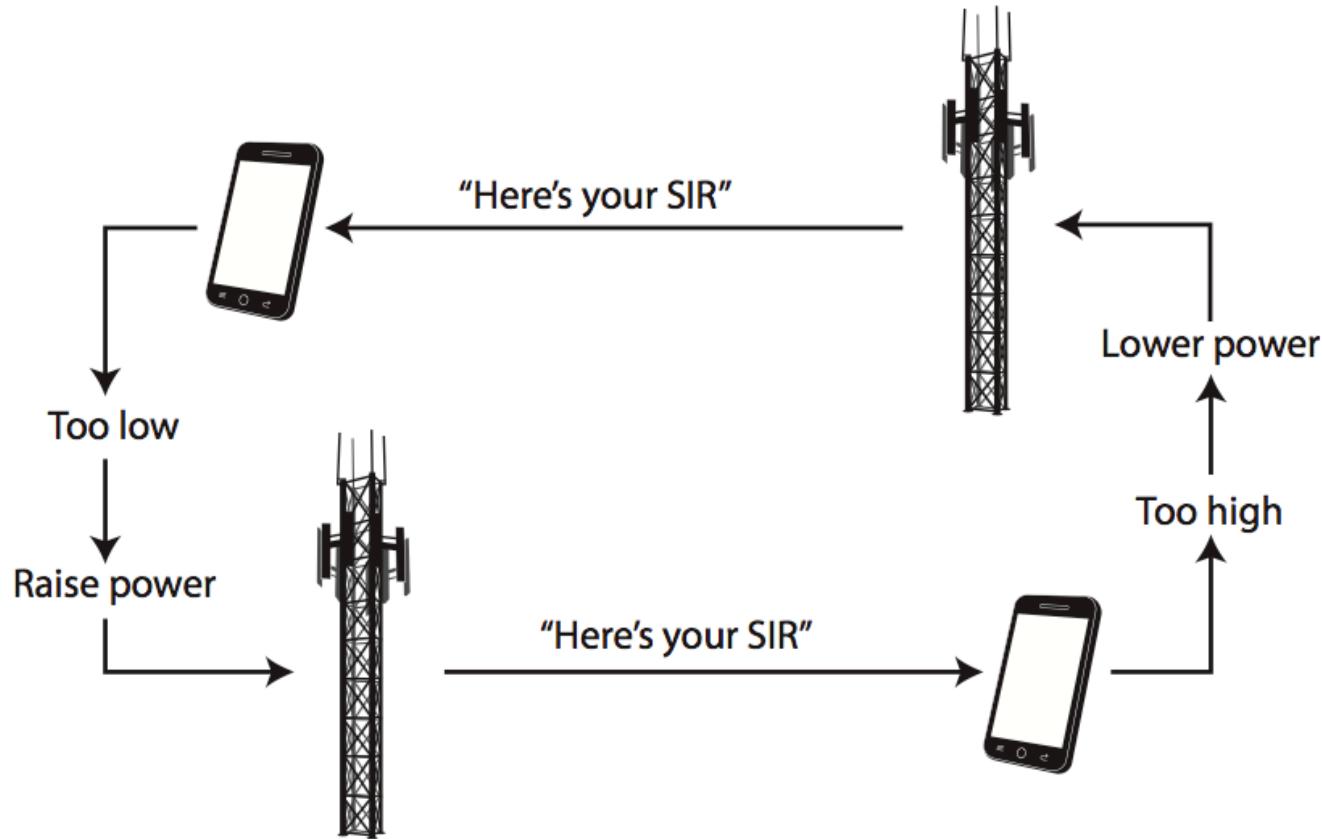
- Ideally, only the power from the transmitter of a link would be present at its receiver.



- But this is not the reality: here, some of A's transmission will be coupled into B's receiver, and vice versa.

Signal to Interference Ratio (SIR)

- Tower tells a device its current received signal-to-interference ratio (SIR), which serves as a negative feedback signal.



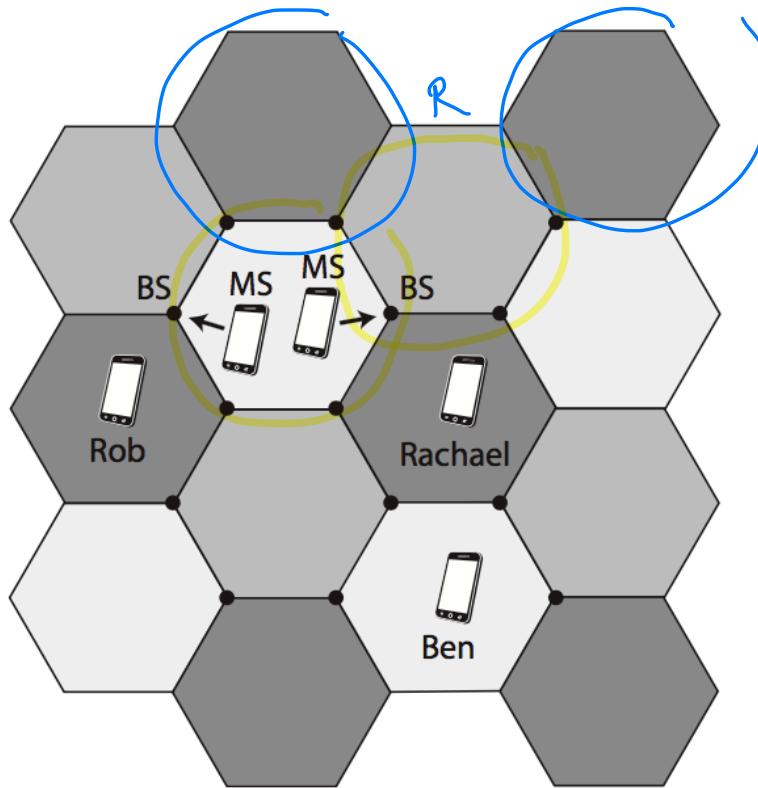
- Each device can update its transmission power independently.



Hexagonal Cell Organization

Cells

- Multiple mobile stations (MSs) & base stations (BSs).



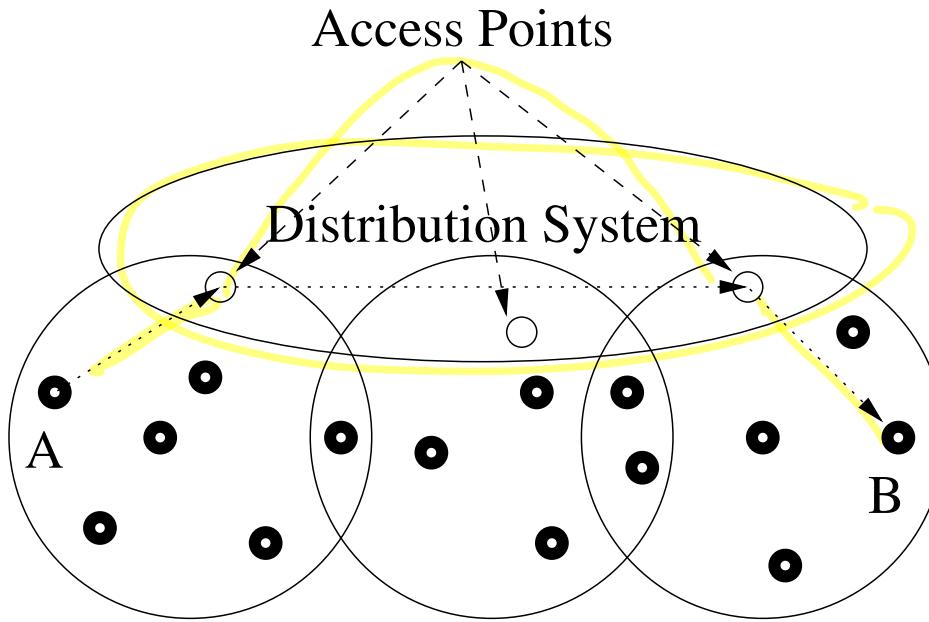
- Shading of a cell indicates frequency band that the cell is using.
- Neighboring cells have different frequency bands.

MACA Algorithm

- 802.11 uses Multiple Access Collision Avoidance (MACA):
 1. Sender sends Request To Send (RTS) message to receiver that includes how long it wants to hold medium.
 2. Receiver responds with Clear To Send (CTS) message.
 3. If CTS not received nodes realize after a period of time that collision occurred, in which case a backoff algorithm is being used.
 4. Receiver sends ACK after receiving.
 5. All other nodes must wait for ACK prior to transmitting.
- In practice, it is much more complicated than this!

Access and Distribution: Not All Nodes are Equal!

- Nodes are associated to access points.
- A sends frame to B as follows:



- A sends to A's Access Point., A's Access Point sends to B's Access Point that forwards to B.

Scanning for Access Points

- Stations select access points by scanning:
 1. Station sends **Probe** frame.
 2. All Access Points within reach of station reply with **Probe Response** frame.
 3. Station selects access point and responds with **Association Request** frame.
 4. Access Point responds with **Association Response** frame.
- 802.11 frames include a control field indicating whether or not frame is data, RTS or CTS.
- It has four addresses to account for the fact that it must be transmitted through the distribution system.

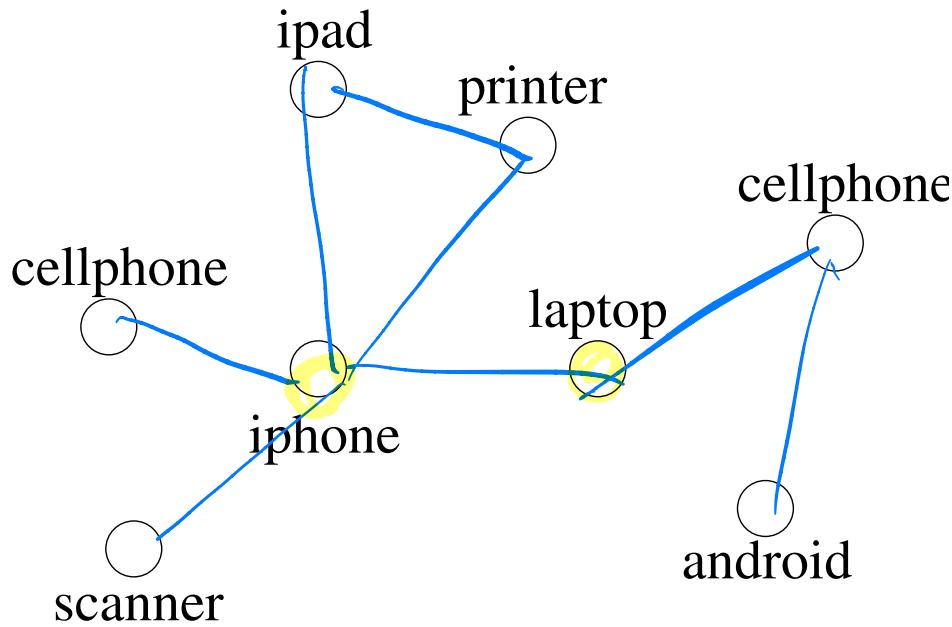
IEEE 802.11: Frames

- It has three types of frames: management frames (MF), control frames (CF), and data frames(DF).
- **MF:** Used for node association, disassociation, timing, synchronization, authentication and deauthentication.
- **CF:** Used for Handshaking and positive ACKs during an exchange.
- **DF:** Used for data transmission.
- **IEEE 802.11: MAC**
 - MAC protocol is specified in terms of a coordination function that determines when a node in a BSS is allowed to transmit and when it may be able to receive.

Bluetooth

How to Establish a Link

- A set of nodes wants to establish a connected network.



- What protocol should they follow to get connected?

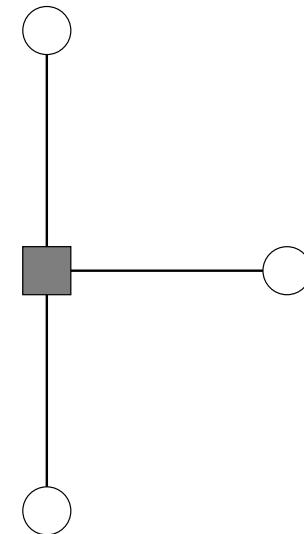
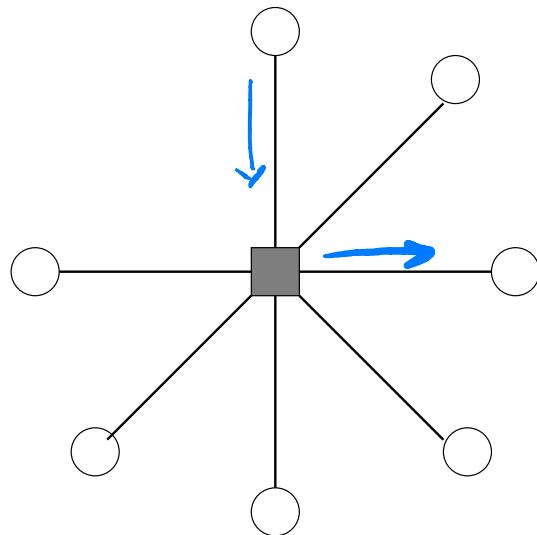
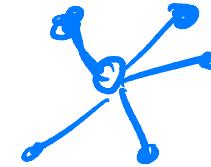
"wire replacement
technology"

Bluetooth

- Originally conceived as cable replacement technology.
- Is the first defacto standard for ad hoc networking brought about by several companies.
- Its particular design is less suited for other applications.

Organization: Piconets

- They are **star** networks.
- In the leftmost piconet the master has **seven** slaves, in the rightmost it has three.



■ = Master

○ = Slave

diameter = 2

Piconet

Organization: Piconets

- They are managed by a single **master** that implements centralized control over channel access.
- All other participants are called **slaves**.
- Communication is strictly
 - { master → slave,
 - and
 - slave → master.
- Direct slave-to-slave communication is impossible.
- A master has at least one and at most seven slaves.
- Piconets can be enlarged to form scatternets.

Double Personalities

- **Roles of Master and Slave:**

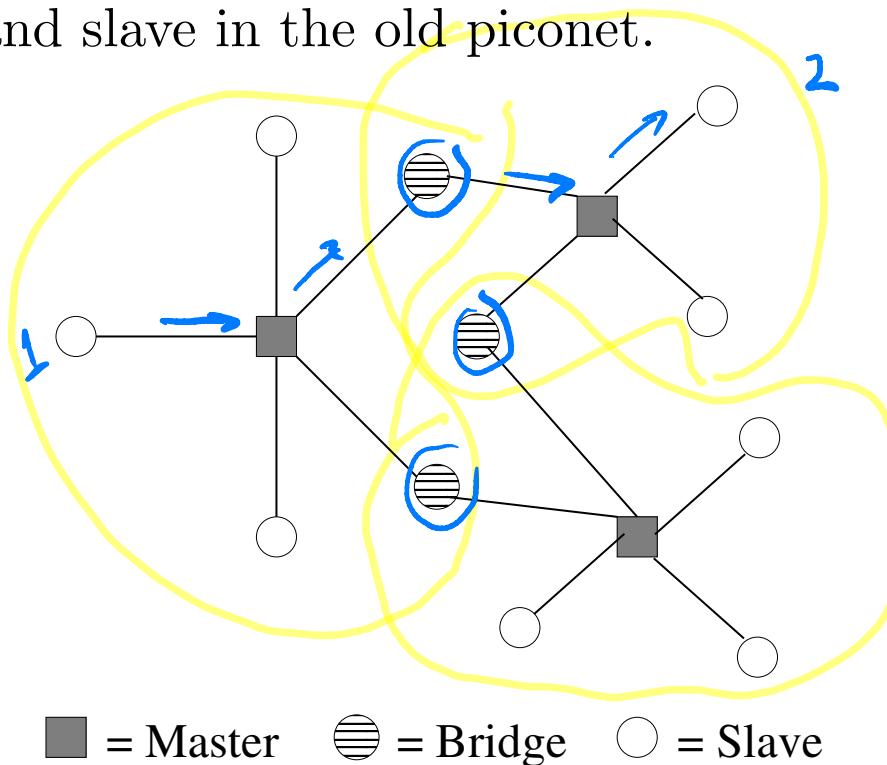
- Since a slave may want to set up a new piconet or take over an existing piconet, during the existence of a piconet the role of master and slave can be switched.
- This is done by employing a different frequency.

Bluetooth enforces a network topology

- piconet
- scatternet

Example of a 14-node Scatternet

- Piconets are joined to form **scatternets**.
- A node can be slave in two piconets, or become master in a new piconet and slave in the old piconet.



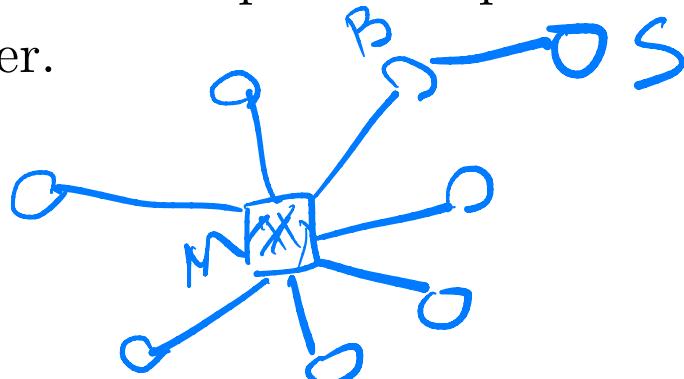
You enlarge the number of nodes: not too much

Bluetooth Communication

- If you want to communicate with more than eight nodes at the same time multiplexing is required. Moreover, nodes would need to alternate between their respective piconets.
- Bluetooth does not provide for slave-to-slave communication (maybe technology will improve in the future).

To solve this problem one has

1. either to channel traffic through a master (this increases communication and power consumption)
2. or one of the two slaves could setup its own piconet or even switch roles with a master.



Scatternet

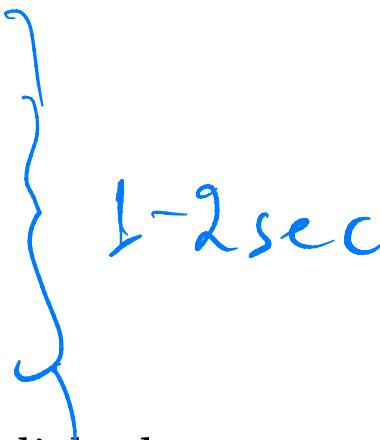
Scatternets are collections of piconets satisfying the following rules.

- 
1. The scatternet is a connected network formed from piconets.
 2. It has masters and slaves. Slaves are of two types:
“pure” slaves (i.e., slaves belonging to a single piconet), and
“bridge” slaves (i.e.. slaves that belong to multiple piconets).
 3. Two masters can share only a single slave.
 4. A bridge may connect only two piconets.
 5. A piconet can have at most seven slaves.

How to Establish a Link

Bluetooth nodes want to establish a connected network.

They follow the protocol below:

1. Start
 2. Synchronization
 3. Discovery
 4. Paging
 5. Connection established.
- 
- 1-2 sec

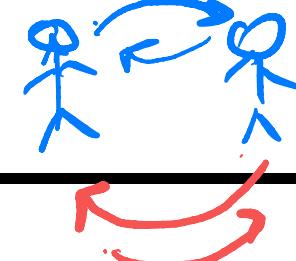
Discovery Delay Procedure

Bluetooth supports the paradigm of **spontaneous connectivity**. The procedure used for node discovery is called **Inquiry** and connections are established based on information exchange.

1. Bluetooth node is set into **Inquiry** mode by the application.
2. Then sends **Inquiry** messages to probe for other nodes.
3. Other Bluetooth nodes (within the range) only listen.
4. They reply to **Inquiry** messages only when they have been set explicitly to **InquiryScan** mode.

To prevent “collisions” and since **Inquiry** needs to be initiated periodically then some type of randomness must be employed in order to determine the time interval between two **Inquiries**. This technique is called **Collision Avoidance**.

Backoff protocol

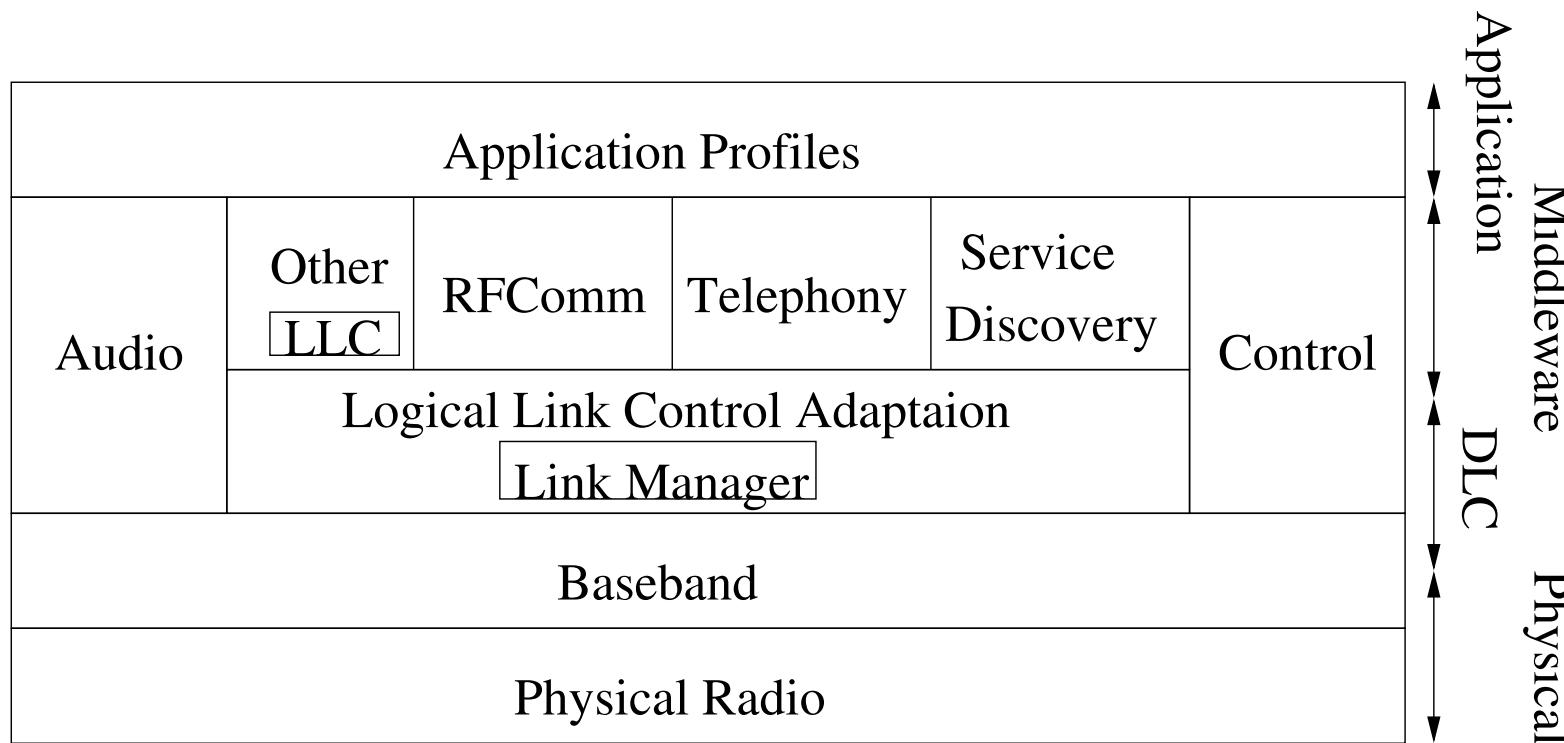


Connection Establishment

- Once a unit has discovered another unit, connection establishment is very fast.
- In an ideal scenario, the expected delay for link formation (Discovery plus Connection) is about 1 sec when both nodes follow the uniform distribution between the **Inquiry** and **InquiryScan**.
- In practice this takes several seconds.

IEEE 802.15 (Bluetooth Protocol Architecture)

Does not follow any of the OSI, TCP/IP or 802 models.



The Radio layer moves bits from master to slave. The baseband resembles the MAC sublayer.

IEEE 802.15 (Bluetooth Frames)

The Bluetooth frame includes an access code identifying the master so that slaves can tell which traffic belongs to them.

72	54	0-2744
Access Code	Header	Data

In the Header, Addr identifies which of the active devices frame is intended for. Type identifies frame type.

3	4	1	1	1	8
Addr	Type	F	A	S	Checksum

{ NFC (Near Field Communication)
Tag technology

Comm. between



two entities

$n\mathbf{G}$

nG (n-th Generation Wireless)

- Provides high quality, reliable communication and each new generation of services represents a big leap in that direction.

Features	1G	2G	3G	4G	5G
Start/Development	1970/1984	1980/1999	1990/2002	2000/2010	2010/2015
Technology	AMPS, NMT, TACS	GSM	WCDMA	LTE, WiMax	MIMO, mm Waves
Frequency	30 KHz	1.8 Ghz	1.6 - 2 GHz	2 - 8 GHz	3 - 30 Ghz
Bandwidth	2 kbps	14.4 - 64 kbps	2 Mbps	2000 Mbps to 1 Gbps	1 Gbps and higher
Access System	FDMA	TDMA/CDMA	CDMA	CDMA	OFDMA/BDMA
Core Network	PSTN	PSTN	Packet Network	Internet	Internet

- Each Generation defined by telephone network standards.
- Evolution started in 1979 with 1G and is still ongoing with 5G.
- Each of the Generations has standards that must be met to officially use the G terminology.
- There are institutions in charge of standardizing each generation of mobile technology.

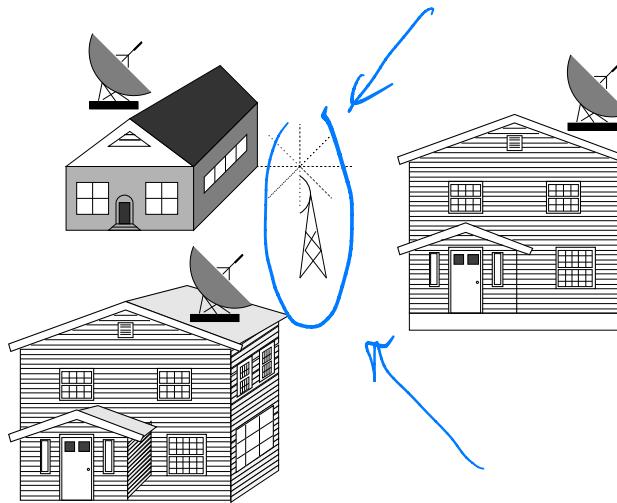
5G has wider bandwidth

$n\mathbf{G}$ (n -th Generation Wireless)

- **(1G)** Introduced in the late 70's with fully implemented standards being established throughout the 80's
- **(2G)** Main difference between 1G and 2G, is that the radio signals used by 1G are analog, while 2G networks are digital.
- **(3G)** Utilizes UMTS (Universal Mobile Telecommunications System); combines aspects of 2G with new technology and protocols to deliver a significantly faster data rate.
- **(4G)** Currently in use. Made possible by MIMO (Multiple Input Multiple Output) and OFDM (Orthogonal Frequency Division Multiplexing). Important 4G standards are Broadband Wireless, WiMAX (fizzling out) and LTE (has seen widespread deployment).

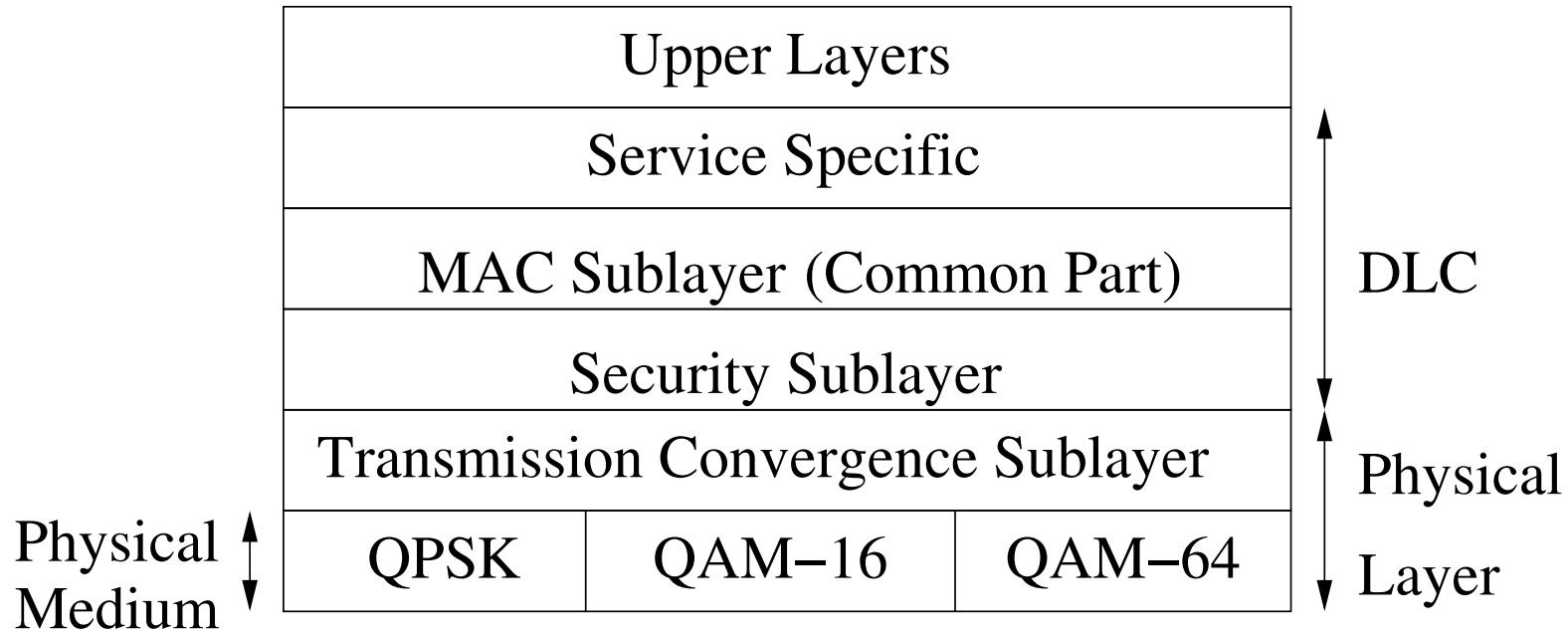
Broadband Wireless (IEEE 802.16): Wireless Last Mile

- Running fiber optic, coaxial cable, etc, to millions of homes is rather expensive. A big antenna on a hill solves last mile transmission problems.



- 802.16 is different than 802.11: 1) it provides service to static buildings not nomadic devices, 2) buildings can have more than one computer, 3) uses full-duplex 4) more spectrum in the range 10-66 GHz is used, 5) provides QoS.

Broadband Wireless (IEEE 802.16): Protocol Stack



Broadband Wireless (IEEE 802.16): Wireless Last Mile

- 802.16a, 802.16b planned: to operate on different frequency ranges.
- Service specific sublayer interfaces with the network layer.

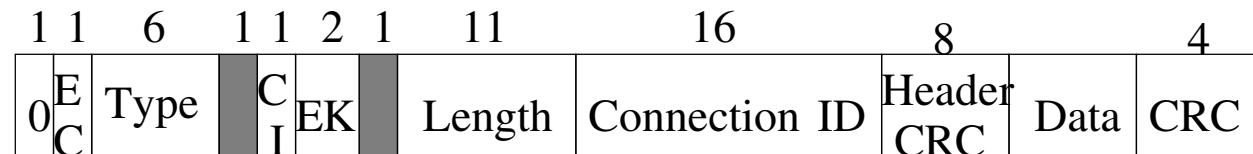
Modulation	Mbps	Bits/Baud
QAM-64	150	6
QAM-16	100	4
QPSK	50	2

Broadband Wireless (IEEE 802.16): MAC Sublayers

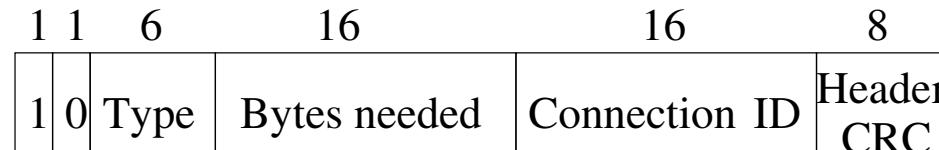
- The DLC has three sublayers.
- **Security:** Only the payloads are encrypted with symmetric DES (or triple DES). Authentication is done with RSA using X.509 certificates. Integrity uses SNA-1.
- **MAC Sublayer (Common Part).** In downstream: the base station decides what to put in which subframe. In upstream: four classes of service are defined. CBR (for uncompressed voice), RTVariable Bit Rate (for compressed multimedia), Non-RTVariable Bit Rate (for large file transfers), Best Effort (for everything else).
- Bandwidth allocation is either per station (subscriber station makes collective requests for all users in a building) or per connection (base station manages each connection directly). 

Broadband Wireless (IEEE 802.16): Frames

- MAC frames begin with a generic header: EC (tells if payload is encrypted), Type (gives frame type), CI (indicates presence or absence of final checksum), EK (tells which encryption key is being used), Connection ID (gives connection frame belongs to), Header CRC uses polynomial $x^8 + x^2 + x + 1$.



- Frames requesting bandwidth have different header type and do not carry payload. Instead of 0 they start with bit 1.



WiMax (Fizzling Out)

- The name "WiMAX" was created by the WiMAX Forum, which was formed in June 2001 to promote conformity and interoperability of the standard.
- WiMAX (Worldwide Interoperability for Microwave Access) is a trademark for a family of telecommunications protocols that provide fixed and mobile Internet access.
- The 2005 WiMAX revision provided bit rates up to 40 Mbit/s with the 2011 update up to 1 Gbit/s for fixed stations.
- WiMAX is a standards-based technology enabling the delivery of last mile wireless broadband access as an alternative to cable and DSL.
- WiMax requires special antennae and Network Interface Cards.

LTE Wireless (Currently in Use)

- Long Term Evolution (LTE) is a standard for wireless communication of high-speed data.
- Goal of LTE is to increase the capacity and speed of wireless data networks utilizing cutting-edge hardware and DSP techniques that have recently been developed.
- Its wireless interface is incompatible with 2G and 3G networks, and so it must be operated on separate wireless spectrum.
- LTE includes an all-IP flat network architecture, end-to-end QoS including provisions for low-latency communications, peak download rates nearing 300 Mbps and upload rates of 75 Mbps, capacity exceeding 200 active users per cell, the ability to manage fast-moving mobiles, and support for multi-cast and broadcast streams.

5G (5-th Generation Wireless)

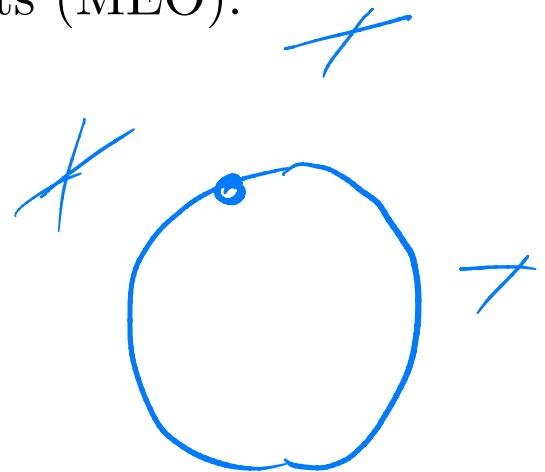
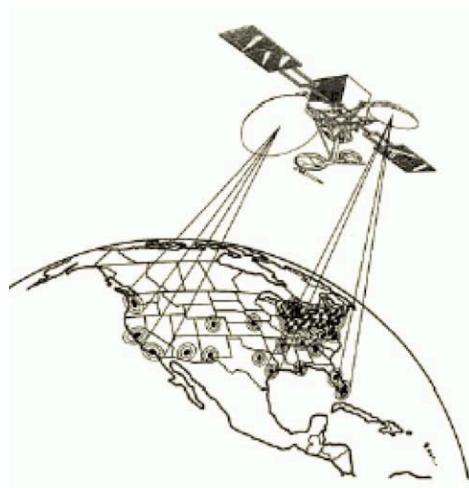
- Currently being rolled out.
- Includes device-to-device communication, better battery consumption, and improved overall wireless coverage.
- May include Massive MIMO, Millimeter Wave Mobile Communications, small cells, Li-Fi
- New technologies could be used to give 10Gb/s to a user
- Expected low latency, and allows connections for billions of devices.

Satellite

Satellite Based

- Three types of satellite orbits: geostationary orbits (GSO), low earth orbits (LEO) and medium earth orbits (MEO).

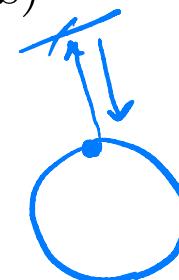
Delay Tolerant



- MEO is mainly utilized for navigation services such as GPS and Galileo, while GSO and LEO orbits are used for point?to?point and point?to?multipoint satellite communications
- Several providers available and currently in development.

Satellite Based

- Satellite based services deliver broadband to customers in the US and Canada.
- ViaSat (a recent satellite) with a total data throughput of some 140 Gbps, the satellite has more capacity than all other commercial communications satellites over North America combined.
- This is a wave of new satellites operating in the Ka-band, a part of the satellite-apportioned radio spectrum that allows high data-rates (download range of 8-12 Mbps)



Satellite Based: Latency

- Satellite signals travel near the speed of light.
- The basic time delay formula is

$$\text{Time Delay} = \frac{\text{Distance to Satellite}}{\text{Speed of Light}}$$

- Example:

A Distance to Satellite = 35,786 km,

and Speed of Light = 300,000 km/s

yields a Time Delay = 120 ms.

- The total delay for one-way communication between two ground stations is between 250 and 300 ms. For two-way communications (when one satellite customer communicates with another satellite customer), the round-trip time would typically be between 500 and 600 ms.

Planetary
TCP/IP

Exercises^a

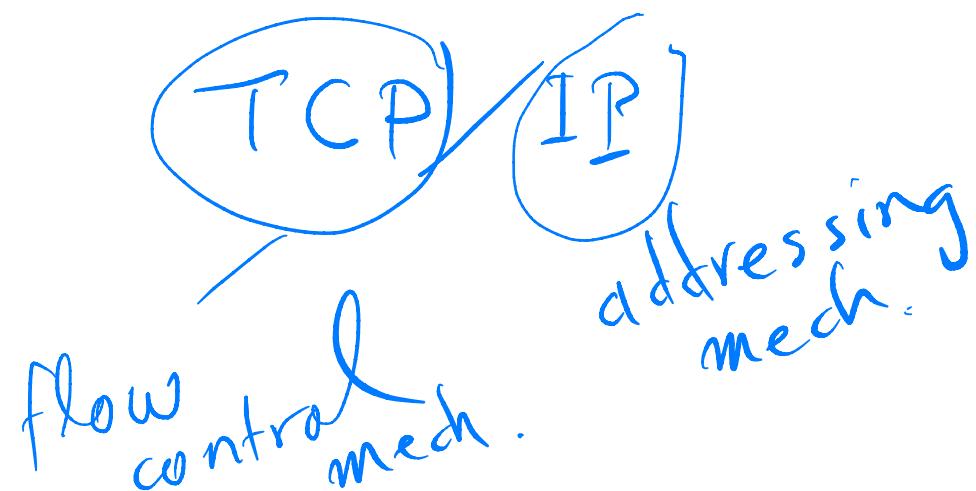
1. What are the differences between TDMA, FDMA, and CDMA.
2. Can you devise an algorithm for sharing a given bandwidth both with TDMA and FDMA? Give details of how it would work.
3. What are the difference between Multiple Access Collision Avoidance and Multiple Access Collision Detections? Why do they have to be different?
4. Compute
 - (a) the inner product of the vectors
 $\mathbf{u} := (-1, +1, +1, +1, -1, -1),$
 $\mathbf{v} := (+1, +1, -1, +1, -1, +1),$
 - (b) and the complements $\bar{\mathbf{u}}, \bar{\mathbf{v}}$, where \mathbf{u}, \mathbf{v} are as above.

^aNot to submit!

5. Generate two vectors $\mathbf{u} = (u_1, o_2)$ and $\mathbf{v} = (v_1, v_2)$ such that u_1, u_2, v_1, v_2 are $+1$ or -1 with probability $1/2$ independently at random. What is the probability that the inner product of \mathbf{u} and \mathbf{v} is 0 ?
6. The power of a signal attenuates according to the inverse square law $P(d) = P(0)/d^2$, where $d > 0$ is the distance, $P(d)$ is the power at distance d , and $P(0)$ is its power at the start. How far can a signal reach if its power at distance d has to be at least $1/4$ its power at the start?
7. Due to the presence of obstacles, the power of a signal attenuates according to the inverse cubic law $P(d) = P(0)/d^3$, where $d > 0$ is the distance, $P(d)$ is the power at distance d , and $P(0)$ is its power at the start. If the power at distance $d = 1$ is 8 , up to what distance d is the power of the signal at least $1/10$ its power at the start?

8. Two stations located at A and B transmit wireless signals simultaneously and against each other. The signal at station A has speed u and the signal at station B has speed v . Determine the point at which the two signals collide.
 - (a) Do the same exercise as above when the signals are transmitted with a time difference $\Delta t > 0$.
9. Why is the number of slaves of a piconet limited to a small number (in our case seven)?
10. Consider bluetooth networks.
 - (a) How many bluetooth networks with exactly one master are possible? Describe them all.
 - (b) Recall that two masters can share only a single slave. How many bluetooth networks with at most two masters are possible? Draw one with a total of 13 nodes.

ROUTING



Outline

- Introductory concepts
- Distance Vector Routing (RIP)
- Link State Routing (OSPF)
 - Flooding
 - BFS and Dijkstra (Appendix)
- Miscellaneous
 - Distance Vector vs Link State Routing
 - Routing in Mobile IP
 - Inter Domain Routing
- Spanning Trees (see Appendix)
 - Prim (Outline)
 - Kruskal (Outline)

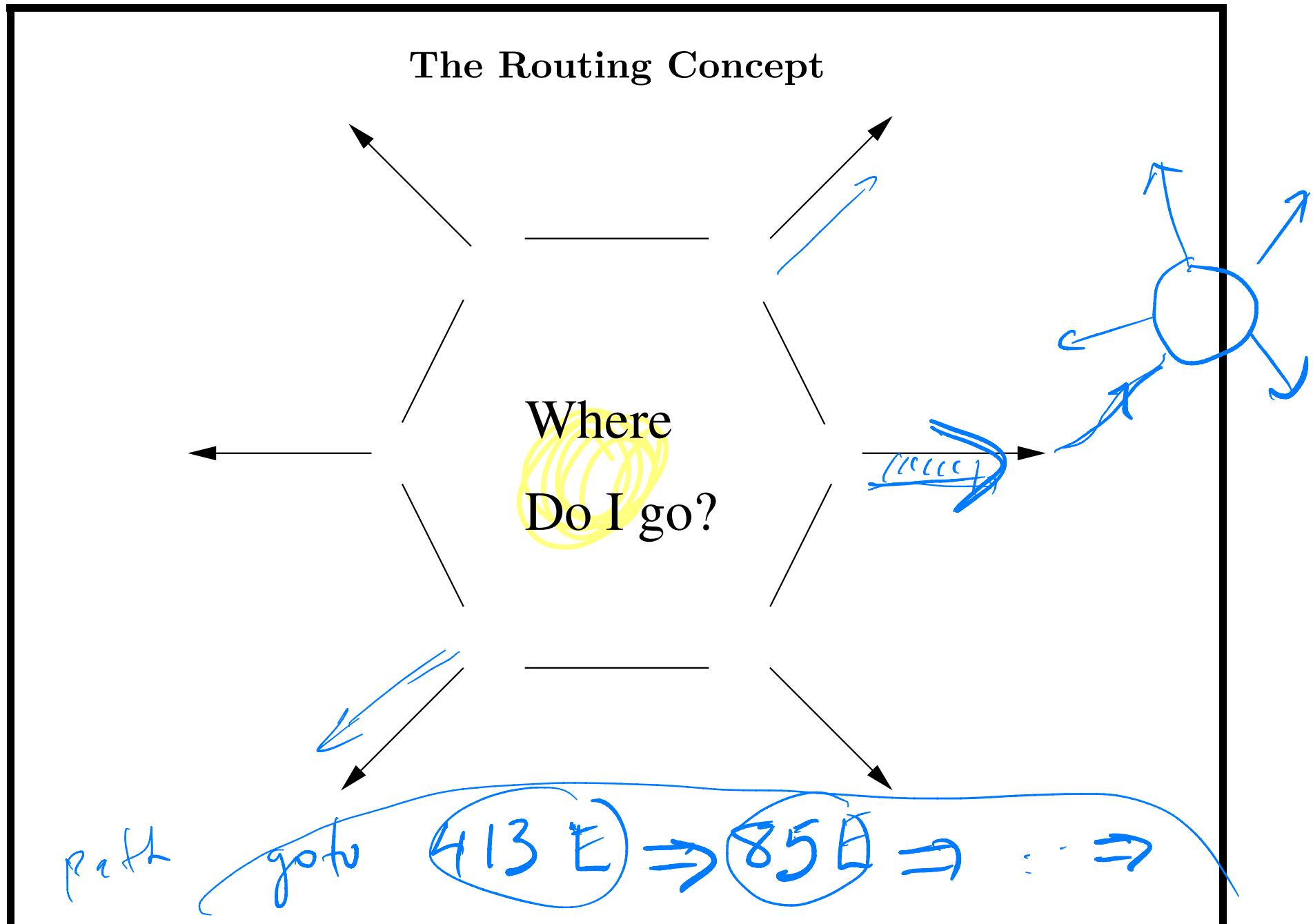
{

used in the full range of networks in any network which requires multi-hop transfers of packets.

{

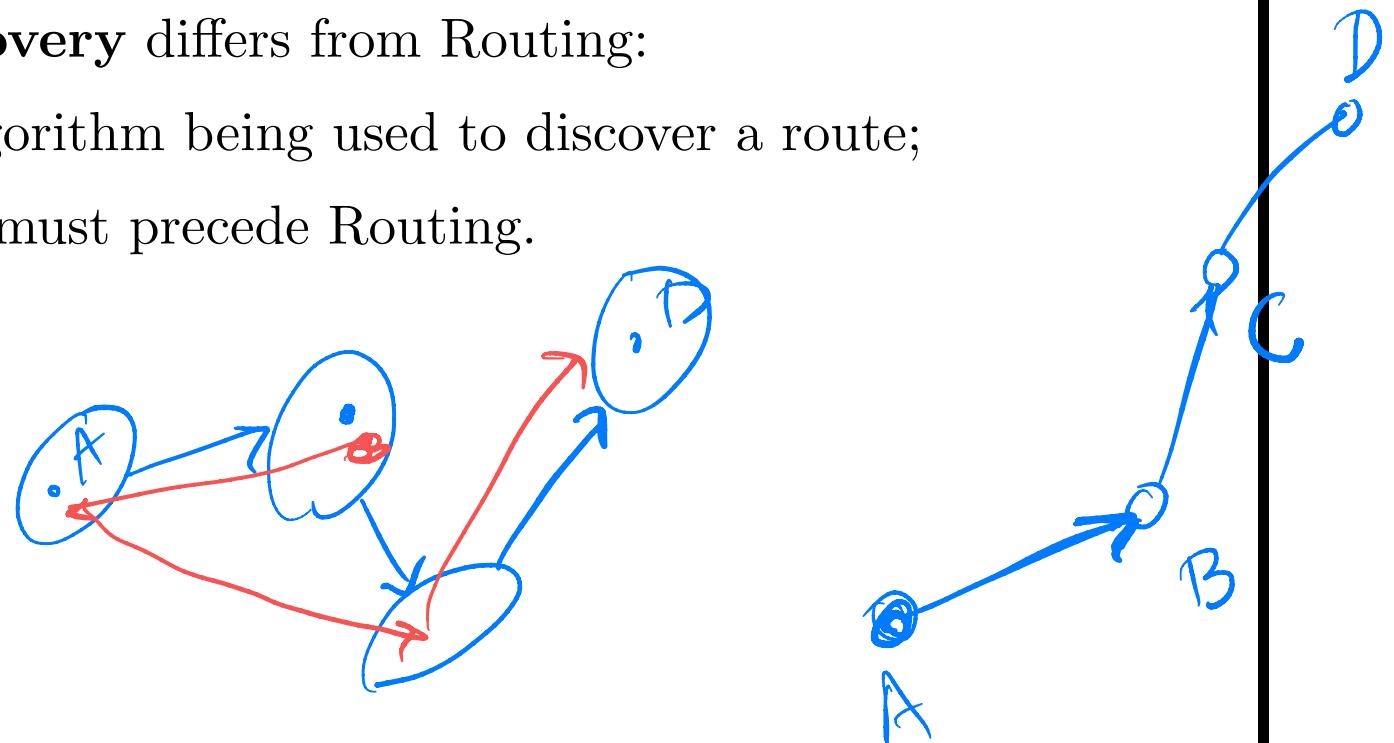
Used in Backbone networks.

Introduction



Routing vs Route Discovery

- **Routing** is a procedure (or algorithm)
 - being used to deliver packets between **nonadjacent** nodes in a point-to-point network and/or between subnetworks of a network.
- **Route Discovery** differs from Routing:
 - it is an algorithm being used to discover a route;
 - as such it must precede Routing.



Routing Table

next_hop

- A fundamental ingredient of routing is the routing table.
 - Standard routing table contains an entry for each possible destination with the out-going link to use for destination.
- Message delivery proceeds node-to-node in the obvious manner one link at a time, looking up next link in the table
- Variations on the above standard scheme exist:
 - Source routing: entries in the table contain the complete path from source to destination
 - Virtual circuit routing: routing tables are used to maintain virtual circuits between communicating nodes

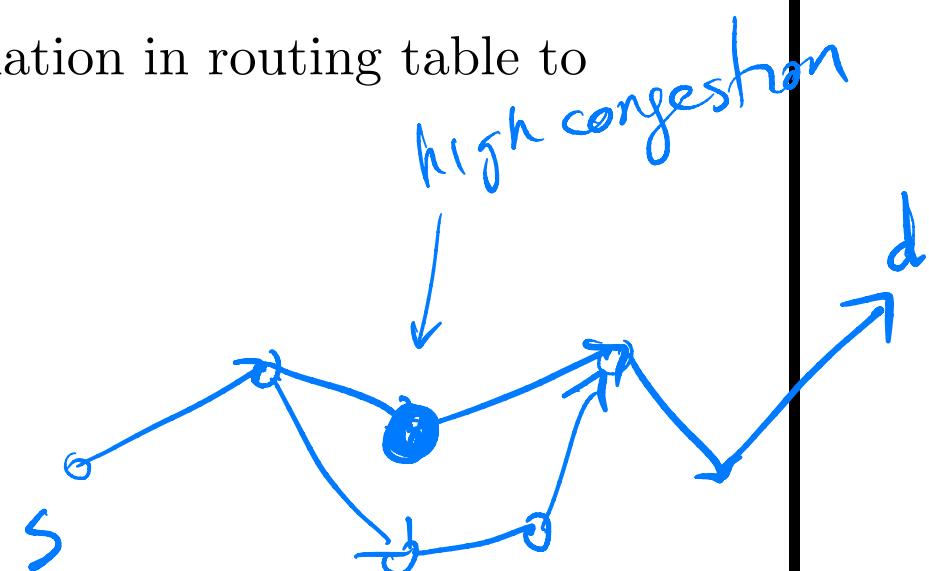
VPN

source
routing:

$P: v_0, v_1, v_2, \dots, v_n$

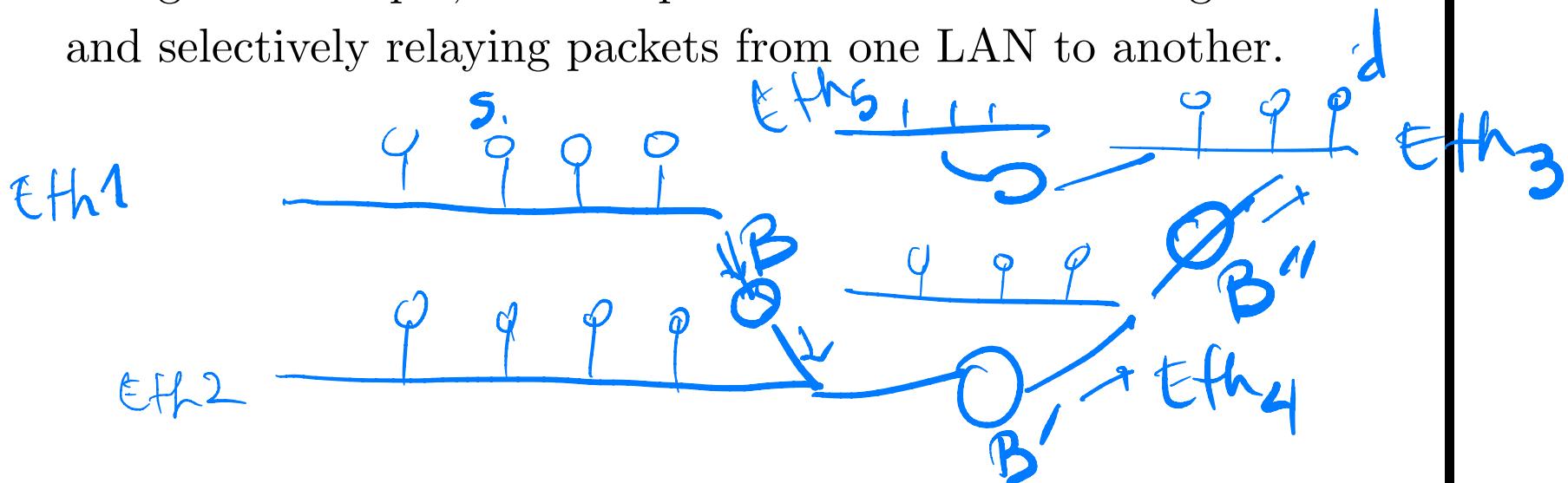
Routing

- Two main components of routing problem are
 - **Route selection:** determination of paths in the network from each source to each destination, i.e., construction and maintenance of *routing tables*
 - **Message delivery:** protocol for converting information in routing table to active packet forwarding



Need for Relaying and Optimization

- Standard LANs like ethernet, token-ring, etc., have fixed topology (e.g., bus, star, ring, or arbitrary graph) and are limited to some maximum number of hosts
- LANs with the same MAC sublayer can be easily interconnected by using *bridges* to form a bigger LAN also called bridged LAN
- Bridges are simple, fast inexpensive routers connecting LANs and selectively relaying packets from one LAN to another.

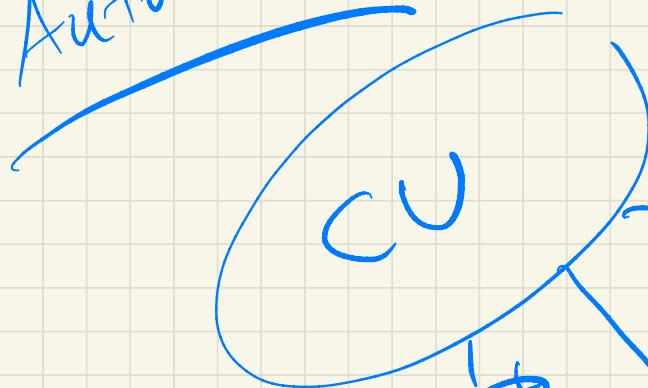


Network Units: Autonomous Systems (AS)^a

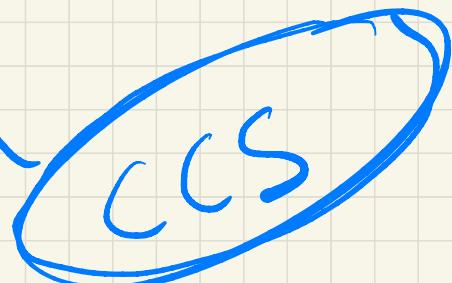
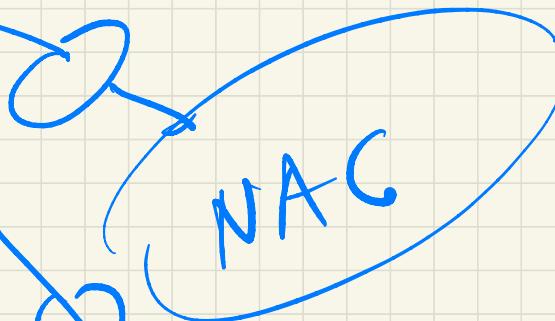
- Subnetworks are aggregated to form larger networks.
- Routing becomes more complex and there is a need to optimize routing in aggregation consisting of smaller “network units”.
- An autonomous system (AS)
 - consists of a number of subnets exchanging packets via routers that are using the same routing protocol, and
 - its routers are managed by a single or cooperating organizations
- Routing protocol used by an AS called *interior* routing protocol (IRP)
- Since all routers are managed by one organization the protocol can be optimized to best serve the users of the AS

^aAn AS may be a network used by a large company or organization.

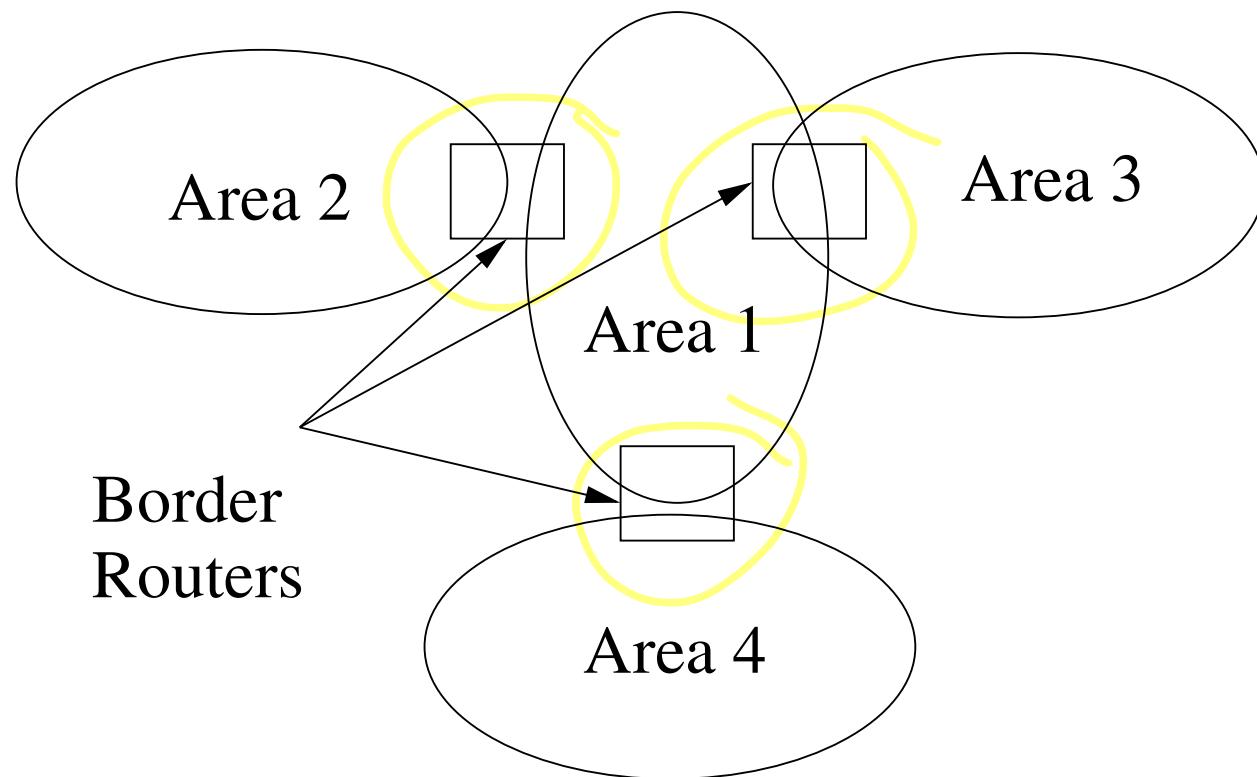
Autonomous



Rent



Example



Internets

- In general, an internet will connect a number of different autonomous systems (ASs) run by many different organizations and running many different interior routing protocols
- Routers used to connect different ASs often called *gateways*
- Protocols used by gateways are called *exterior routing* protocols (ERP).
- Routing protocols operating at the “network frontier” are called *border gateway* protocols (BGP),

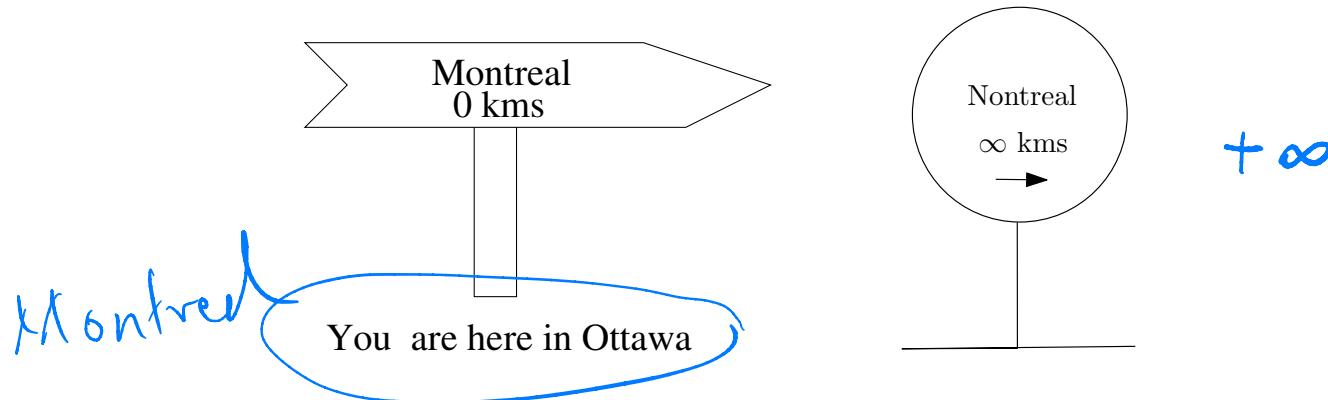
Two Routing Algorithms

- All popular network layer protocols one way or the other are based on two types of distributed routing algorithms:
 - **Distance Vector**
Also known as RIP (Routing Information Protocol) it is based on “Bellman-Ford” algorithm and is sometimes referred to as “old arpanet routing”,
 - **Link State**
Also known as LS Routing, it is based on Dijkstra’s “shortest paths” algorithm.
- These algorithms are applicable in the entire range of networks: from wireless to wireline and optical.
- We will describe both of these algorithms, though in practice they are way more sophisticated.

Distance Vector (RIP)

Distance Vector Routing

- You have been told to post distance signs (one for each city).
You may not even know all the city names!
- If you are in Montreal, initially, you post sign depicted left.

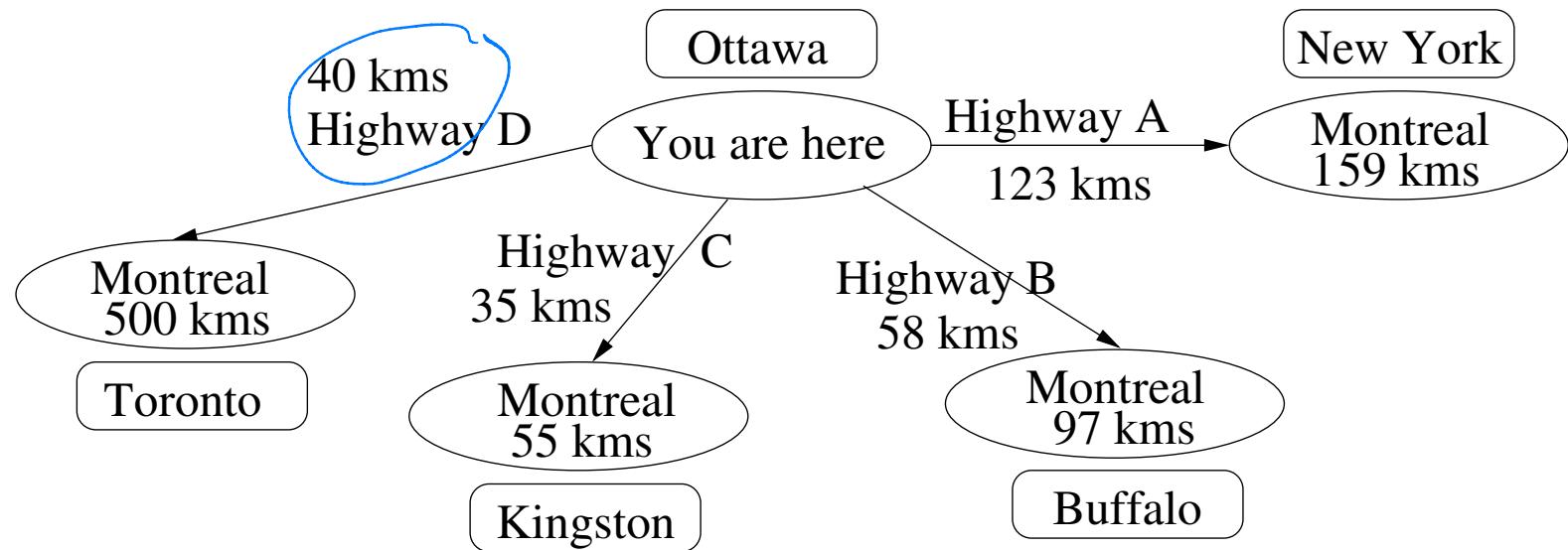


If you are in Ottawa and don't know you post a sign marked ∞ .

- At every intersection there is someone doing the same thing!
- Measure distance to nearest intersection for each of the roads in your intersection and update your sign.
- Keep track of signs posted at each other intersection.

Distance Vector Routing

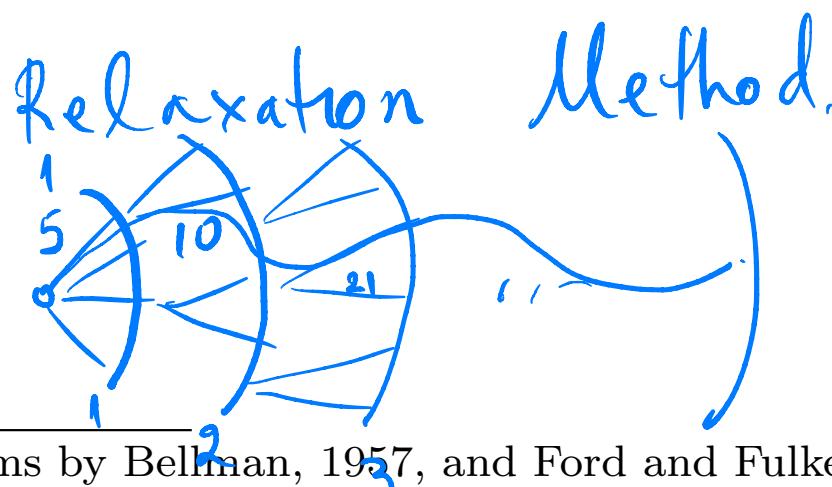
- Now calculate distance to each city by determining which direction gives the smallest distance.



- For example, to go to Montreal if you follow Highway D, it will require 40 kms to next intersection, and there you find an intersection with a new sign posted that says 500 kms to Montreal..

Distance Vector Routing: Idea^a

- Initializes distance to the source to 0 and all other nodes to ∞ .
- Now update distances.
 - For all edges, if the distance to the destination can be shortened by taking the edge, the distance is updated to the new lower value.
- At i th iteration when edges are scanned, the algorithm finds all shortest paths of at most length i (in at most edges).



^aBased on algorithms by Bellman, 1957, and Ford and Fulkerson, 1962

Distance Vector Routing: Details (1/2)

- Uses the basic idea of shortest path routing, but handles topology changes. Forms Routing table as an array of triples $(\text{destination}, \text{distance}, \text{nexthop})$.
- To send a packet to a given destination, it is forwarded to the process in the corresponding nexthop field of the tuple.
- In a graph with n nodes, the distance vector D for each node i contains n elements $D[i, 0]$ through $D[i, n - 1]$, where $D[i, j]$ defines the distance of node i from node j . Initially,

$$D[i, j] = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } j \text{ is a neighbor of } i \\ \infty & \text{otherwise} \end{cases}$$

$w[i, j]$
 $w[i, j] = 1$
 $\{i, j\}$ link
one hop

- Can also use weighted link distances $w[i, j]$.

$$\left(D[\zeta, 0], D[\zeta, 1], \dots, D[\zeta, n-1] \right)$$

↓
-
i

vertex
you're
now

$$D[\zeta, i]$$

if
 $+ \infty$

$$D[\zeta, i] = 0$$

Distance Vector Routing: Details (2/2)

- Each node j periodically broadcasts its distance vector to its immediate neighbors.
- Every neighbor i of j , after receiving the broadcasts from its neighbors, updates its distance vector as follows:

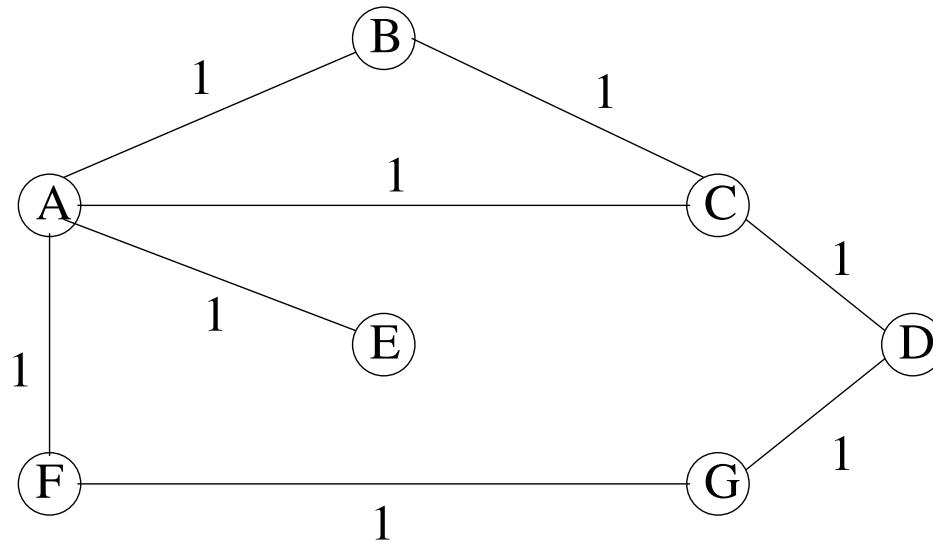
$$D[i, k] = \min_j \{w[i, j] + D[j, k]\}, \forall k \neq i$$

Relaxation

- When a node j or a link crashes, some neighbor k of it detects the failure, and sets the corresponding distance $D[j, k]$ to ∞ .
- When new node joins, or an existing node is repaired, the neighbor detecting it sets corresponding distance to 1.
- Following this, the distance vectors are corrected, and routing table is eventually recomputed.

Example of Distance Vector Routing

- Each node maintains a one dimensional array (vector) of distances to all other nodes.



- For simplicity, in this example all link weights are equal to 1. Similar argument even when the weights are arbitrary positive numbers.
- Next you form a distance matrix between nodes of the network.

Distance Vector Routing: Vectors

- Each node maintains a one dimensional array (vector) of distances to all other nodes.

Node	Vector (of distances)
A	$(d_A^A, d_A^B, d_A^C, d_A^D, d_A^E, d_A^F, d_A^G)$
B	$(d_B^A, d_B^B, d_B^C, d_B^D, d_B^E, d_B^F, d_B^G)$
C	$(d_C^A, d_C^B, d_C^C, d_C^D, d_C^E, d_C^F, d_C^G)$
D	$(d_D^A, d_D^B, d_D^C, d_D^D, d_D^E, d_D^F, d_D^G)$
E	$(d_E^A, d_E^B, d_E^C, d_E^D, d_E^E, d_E^F, d_E^G)$
F	$(d_F^A, d_F^B, d_F^C, d_F^D, d_F^E, d_F^F, d_F^G)$
G	$(d_G^A, d_G^B, d_G^C, d_G^D, d_G^E, d_G^F, d_G^G)$

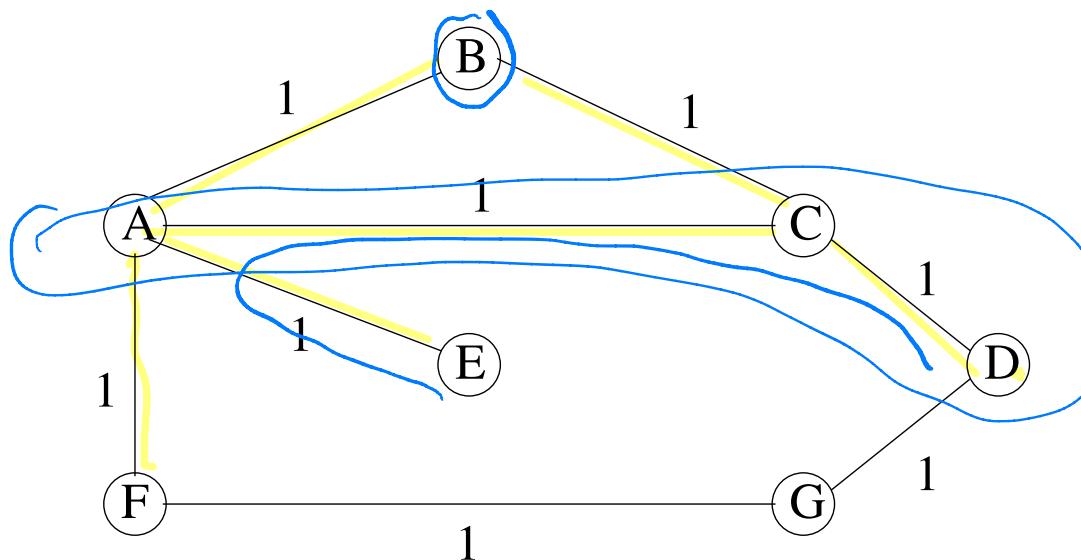
- E.g., where d_u^v is u 's view of its current distance from v .

Distance Vector Routing Algorithm

- Algorithm
 1. Nodes initialize the routing tables.
 2. Every node sends its vector to all its directly connected neighbors.
 3. Every node after receiving this information updates its distance vector.
- Nodes maintain locally their own information: No node has all the information.
- **Convergence:** is the process of getting consistent routing information.
- Periodic updates are required in order to deal with failures.
- It is simple to implement this routing algorithm in software

Initial Routing Cost Tables

Desination (from A)	B	C	D	E	F	G
Distance (Cost)	1	1	∞	1	1	∞
NextHop	B	C	-	E	F	-



$$\text{dist}(E, D) = 3 \text{ hops}$$

Initial Routing Cost Tables

- All nodes do the same.

Source/Destination	A	B	C	D	E	F	G
B	1	0	1	∞	∞	∞	∞
C	1	1	0	1	∞	∞	∞
D	∞	∞	1	0	∞	∞	1
E	1	∞	∞	∞	0	∞	∞
F	1	∞	∞	∞	∞	0	1
G	∞	∞	∞	1	∞	1	0

- Remaining nodes “try to” communicate their vector to A.

Computing Routing Cost Tables

- Routing algorithm is in rounds of
 - Send → Receive → Update
- In each round each node
 1. transmits its vector to all its neighbors
 2. receives vectors from all its neighbors
 3. updates its vector on the basis of what it receives

Computation at A

Look at the first round at A :

	A	B	C	D	E	F	G
A' s current vector	0	1	1	∞	1	1	∞
A sees B' s vector	1	0	1	∞	∞	∞	∞
A sees C' s vector	1	1	0	1	∞	∞	∞
A sees D' s vector	∞	∞	1	0	∞	∞	1
A sees E' s vector	1	∞	∞	∞	0	∞	∞
A sees F' s vector	1	∞	∞	∞	∞	0	1
A sees G' s vector	∞	∞	∞	1	∞	1	0
<hr/>							
A computes new vector	0	1	1	2	1	1	2

Final Routing Cost Tables

Desination (from A)	B	C	D	E	F	G
Cost	1	1	2	1	1	2
NextHop	B	C	C	E	F	F

Source/Destination	A	B	C	D	E	F	G
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

hops, delay, bandwidth

Routing Information Protocol (RIP)

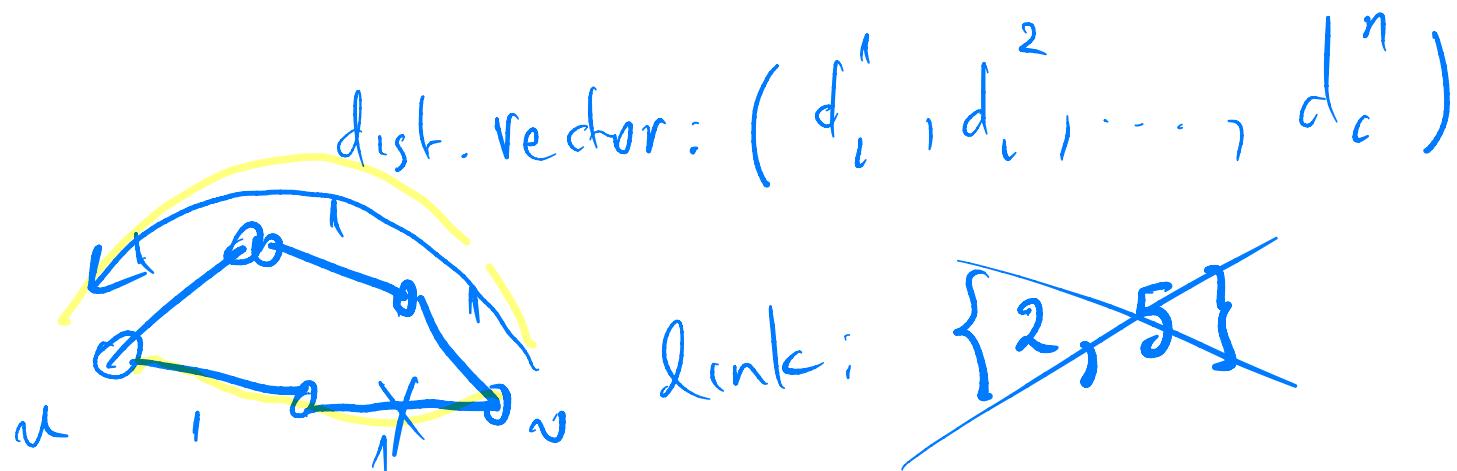
- RIP is based on distance vector routing, was distributed with UNIX BSD. *Berkeley Distribution*

0	8	16	31
Command	Version	Must be zero	
Family of Net1		Address of Net1	
Address of Net1			
Distance to Net1			
Family of Net2		Address of Net 2	
Address of Net 2			
Distance to Net2			

- RIP runs advertisements every 30 sec, and is limited to 16 hops.

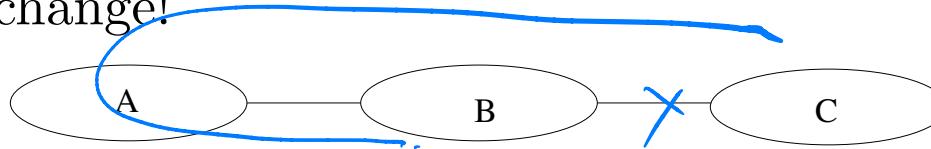
Fault Tolerant

- RIP is not fault tolerant.
- Can fail to give the correct answer even after a single topological change (e.g., link or vertex disruption).
- Sometimes this is due to the time it takes to converge to the correct answer.



Counting to Infinity!

- Distance vector can take a long time to converge after a single topological change!



- Metric is # of hops: C calculates distance to C as 0, B calculates distance to C as 1, A calculates distance to C as 2.
- Now assume C dies or that the link between B and C is broken!
 - B does not conclude immediately that C is unreachable.
 - Instead, it reports its distance to C as 3: because it knows its distance to A is 1, and that A's distance to C is 2!
- In the next iteration, B reports its distance to C as 4: because it knows its distance to A is 1, and that A's distance to C is 3!
- And so on...counting to infinity! In practice *infinity* is set to a fixed value, like 20!

Some Solutions

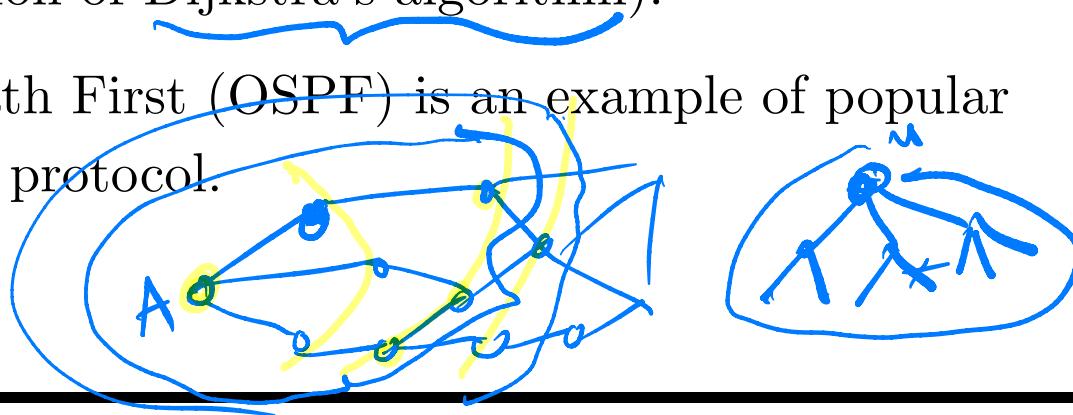
- **Hold Down:** If the path you are using goes down, you wait for some time before switching to another path.
- **Report Entire Path:** You do not report only cost of distances but also path to destination.
- **Multiple Metrics:** Compute routes based on more metrics, e.g. # of hops, link bandwidth, host speed, etc. Used by DECnet, IPX RIP.
- **Triggered Updates:** Periodically updates tables not when vector changes.
- **Split Horizon:** Look at previous example. If A forwards traffic for destination C through B then B reports to A that its distance is ∞ .

Link State

Link-State Protocols

- First introduced for ARPANET to overcome slow response problem of RIP algorithms for use by interior routing protocols
- Each node maintains information on the state of all links of the network (i.e., global information).
 - This is formed from the Link State Advertisements (LSAs).
- When state of out-going link changes, node broadcasts this information to all nodes in the network using **flooding**.
- Each node computes locally its routing table (usually using single source version of Dijkstra's algorithm).
- Open Shortest Path First (OSPF) is an example of popular interior link-state protocol.

Satan



Calculating Maps and Shortest Paths

- The *first stage* in the link-state algorithm is to give a map of the network to every node.
 - Done with several simple subsidiary steps. 1) Determine the neighbours of each node, 2) Distribute the information for the map, and 3) Create the map.
- In the *second stage*, each node independently runs an algorithm over the map to determine the shortest path from themselves to every other node in the network; generally some variant of Dijkstra's algorithm is used.
- A node maintains two data structures: a tree containing nodes which are “done”, and a list of candidates.
- The algorithm starts with both structures empty; it then adds to the first one the node itself.

Aggregating
information

Calculating Shortest Paths: Repeat:

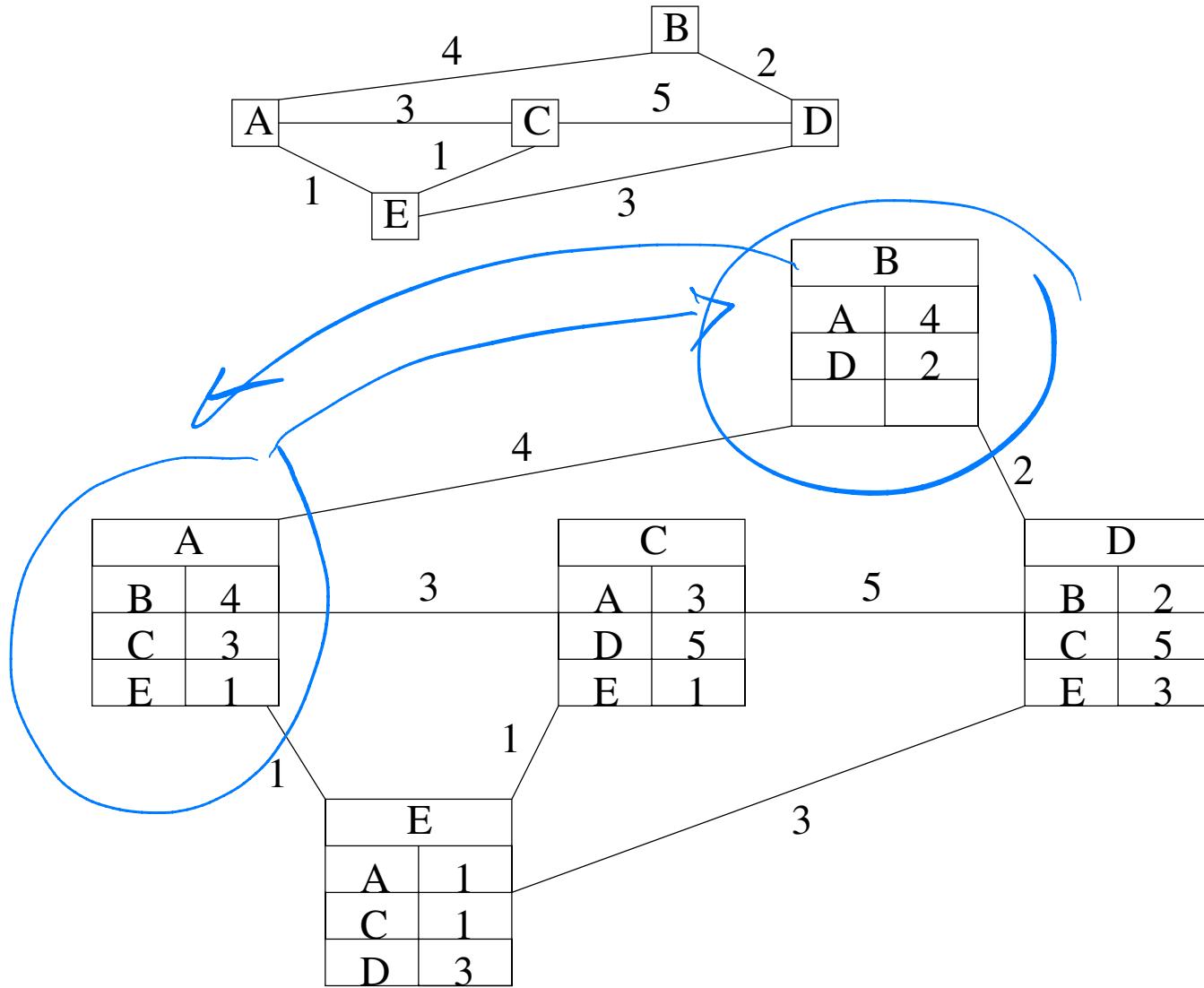
- Adds to the second (candidate) list all nodes which are connected to the node just added to the tree (excepting of course any nodes which are already in either the tree or the candidate list).
- Of the nodes in the candidate list, moves to the tree (attaching it to the appropriate neighbour node already there) the one which is the closest to any of the nodes already in the tree.
- Repeat as long as there are any nodes left in the candidate list.
- This procedure ends with the tree containing all the nodes in the network, with the node on which the algorithm is running as the root of the tree. The shortest path from that node to any other node is indicated by the list of nodes one traverses to get from the root of the tree, to the desired node in the tree.

Link State Protocol

1. Each router is responsible for information on its neighbors.
2. Each router constructs a Link State Packet (LSP) containing
 - (a) ID of node that created packet
 - (b) list of all neighbors together with cost
 - (c) sequence number
 - (d) a TTL (time to live) for this packet
3. LSP is transmitted to all other routers.
4. Each router (now knows the complete map of the network) and computes routes to each destination.

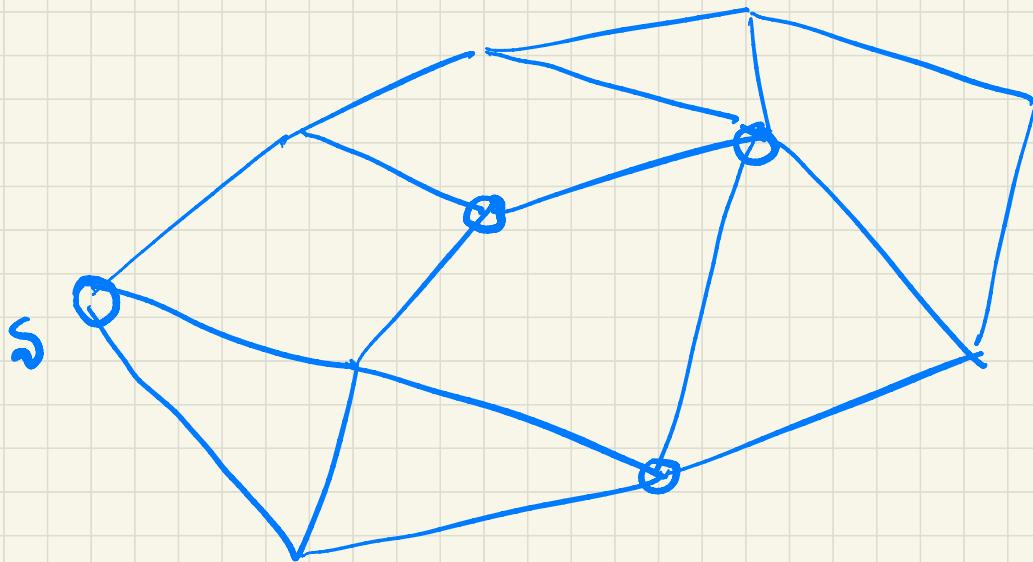
Route Updating/Calculation is done according to Dijkstra's algorithm.

Link-State Protocols

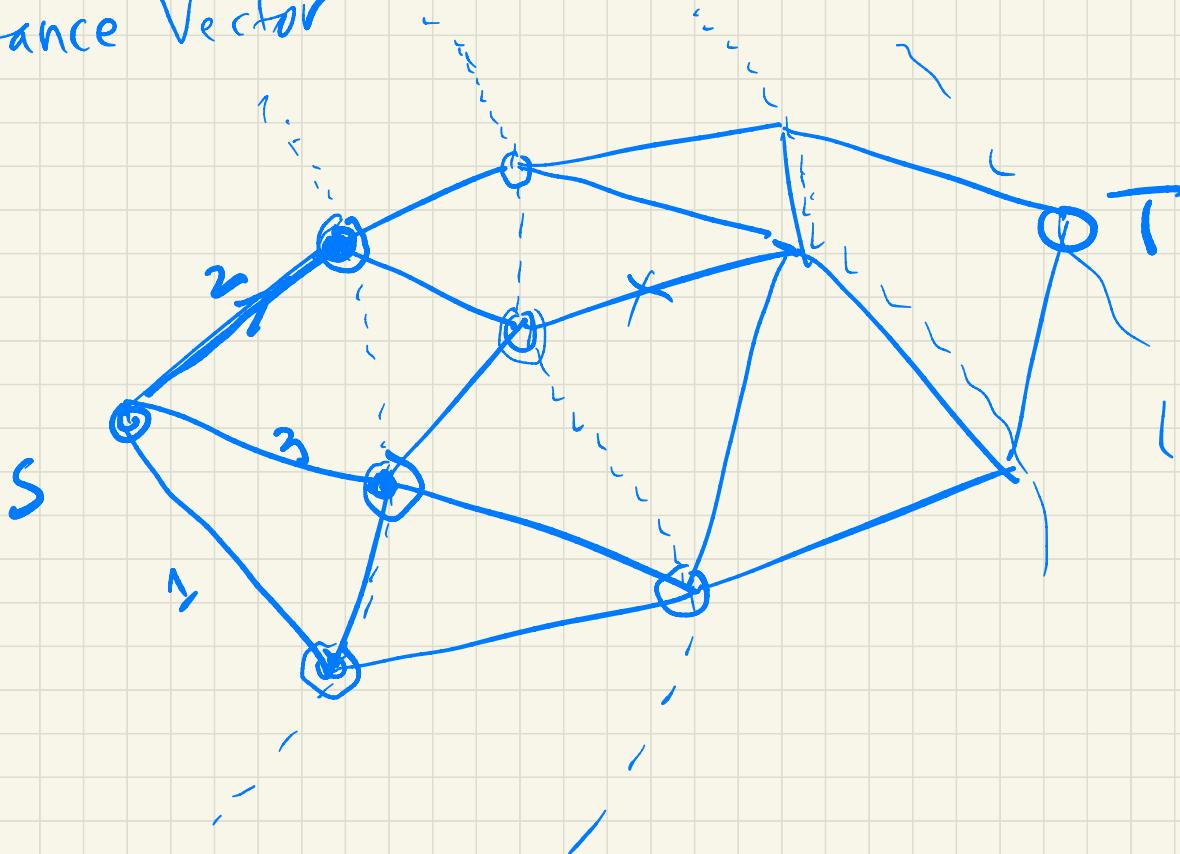


Two Routing Protocols

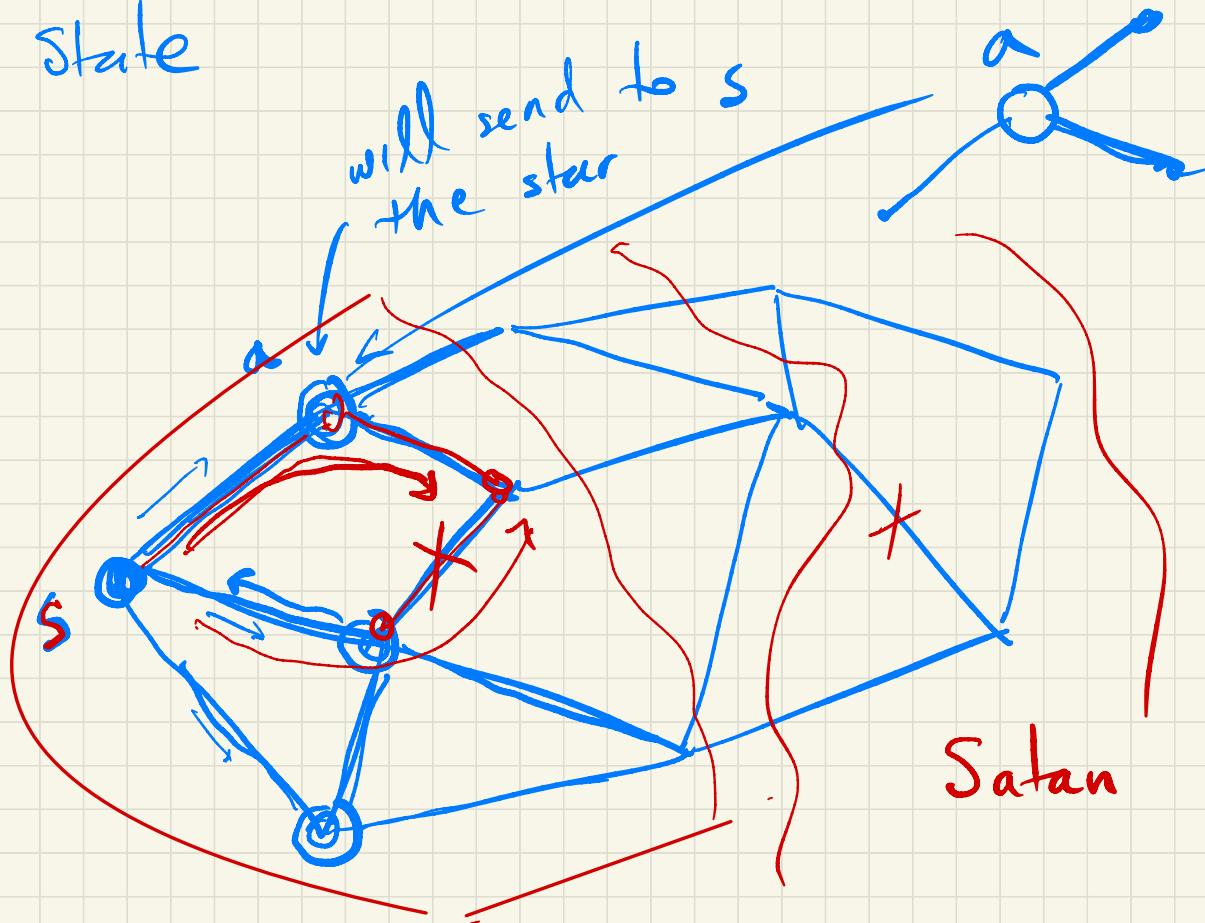
- 1) Distance Vector (Floyd's alg.)
- 2) Link State (Dijkstra's alg.).



Distance Vector



Link State



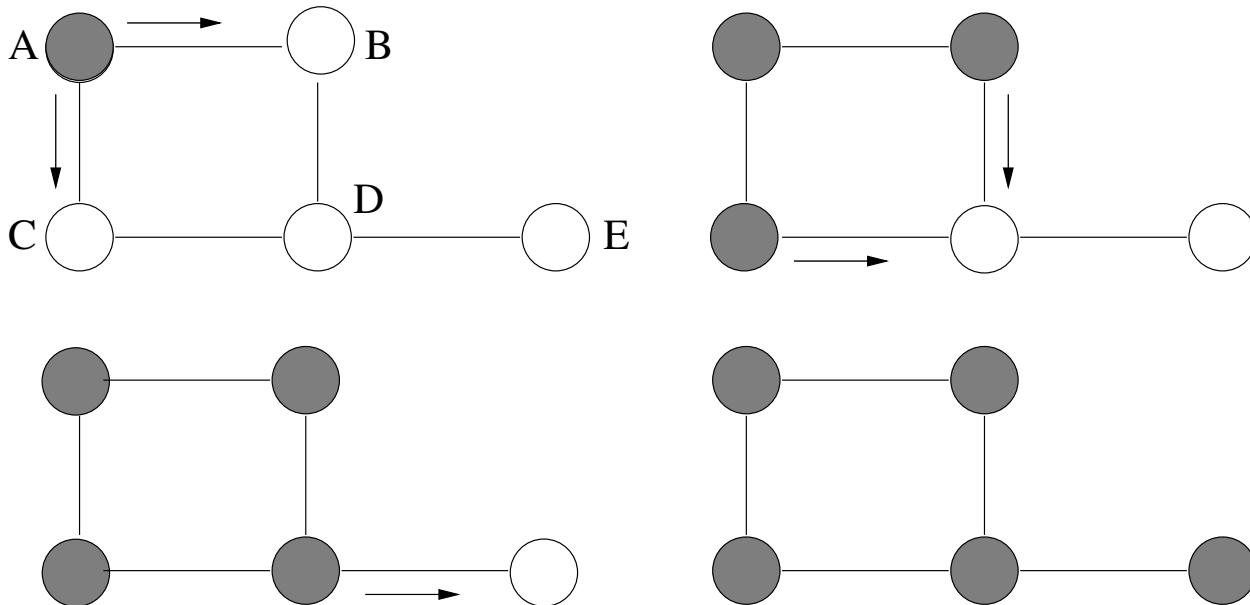
Issues in Link State Protocol

- A router generates an LSP when either it has new neighbors or the cost of a link to a neighbor has changed or link has gone down.
 1. Constructing the LSP
 2. Disseminating the LSP
 3. Timestamping LSPs (helps discover most recent LSP).
 4. Sequence Number/Age Schemes (each router keeps track of sequence number it used last time).
 5. The Arpanet LSP contains: source, sequence number, age, list of neighbors.
- LSP generation uses *Flooding*.

Dijkstra's Algorithm

Flooding

- Flooding sends a packet to all edges but the one you received the packet from.



- When change in state of a link occurs node sends a Link State Packet (LSP) to all of its neighbors
- When a node receives an LSP it forwards it to each of its neighbors except the one it received it from

New Information and Problems

- New Information
 - Newest information must be flooded as quickly as possible while old information be removed.
 - To minimize traffic one uses timers (in the order of hours) for LSP generation. LSPs carry sequence numbers up to 64 bits long.
 - LSPs also carry a time to live (TTL) information.
- Some Problems With Flooding *(Is not good for Backbone)*
 - As stated generates an infinite number of messages except in acyclic network
 - Slow dissemination and/or disconnections can lead to inconsistent view of the network
 - Spurious LSPs (possibly malicious) can introduce loops



Some Solutions

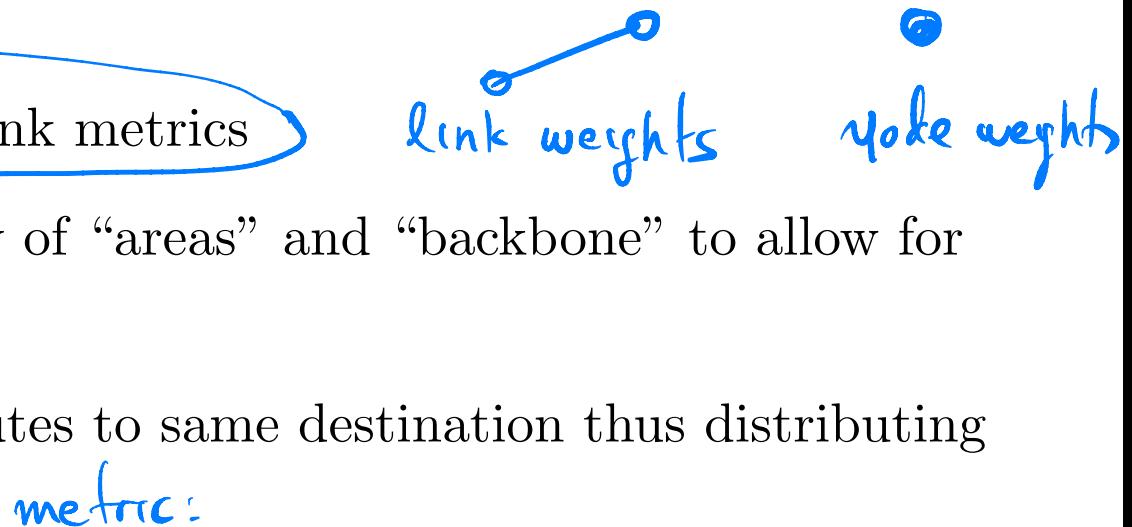
- Sequence numbers: LSP are given numbers and a database is maintained of highest number LSP received. Lower numbered LSPs are ignored.
- Wrapped sequence numbers:^a If sequence has n possible values then a is older than b if (a) $a < b \& |b - a| < n/2$ or (b) $a > b \& |b - a| > n/2$
modulo arithmetic
- Aging: Sequence numbers are insufficient to deal with nodes that come up and start at 0. LSPs also have an age field which is incremented to max age at which point it gets discarded
- Authentication: LSPs are authenticated to ensure they were not created by malicious agent

^aModulo arithmetic.

Open Shortest Path First (OSPF)

Created under auspices of IETF, OSPF is an Open Standard for link-state interior protocol with:

- Sequence numbers to control flooding
- Age fields to deal with failures
- Authentication for security. (E.g., misconfigured host may advertise it can reach every node at cost 0, thus causing wrong LSP updating).
- Includes multiple link metrics
- Two level hierarchy of “areas” and “backbone” to allow for route aggregation
- Allows multiple routes to same destination thus distributing traffic evenly.



metric

(distance, delay, bandwidth)

average:

$$\frac{\text{dist} + \text{del.} + \text{band}}{3}$$

OSPF Frames

0	8	16	31
Version	Type	MsgLen	
	SourceAddr		
	AreaID		
Checksum	AuthenticationType		
Authentication			

LSAge	Options	Type
	LS-ID	
	Advertising Router	
	LS Sequence Number	
LS Checksum	Len	
0	Flags	Number of Links
	Link-ID	
	Link Data	
LinkType	#-TOS	Metric
	Optional TOS Information	
	More Links	

Service
Management

Service
Agreement

Distance-Vector vs Link-State

- Most researchers favor link-state because:
 - Faster convergence after single change
 - Supports multiple paths for routing
- But some prefer distance-vector because:
 - Much simpler structure
 - Less storage space required
- There are techniques available for improving space: Interval Routing, Compact Routing, etc.

Nodes need
to transmuf
their Views
of the topology

BFS

BFS Algorithm^a

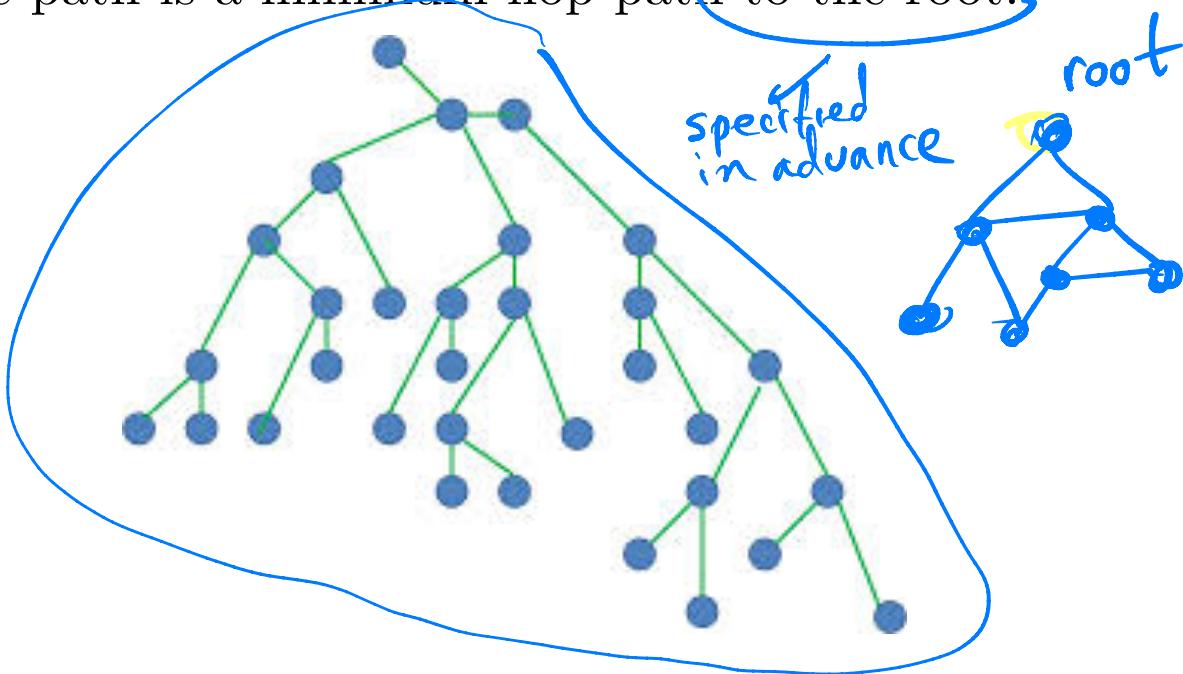
- LSP is based on Dijkstra's Tree Distributed algorithm
- Dijkstra's Tree Distributed algorithm is based on Dijkstra's algorithm (presented on the appendix).
- Dijkstra's algorithm is based on BFS algorithm.

Breadth First Search

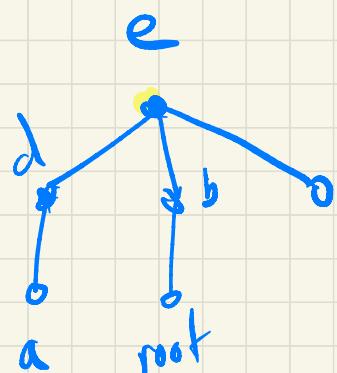
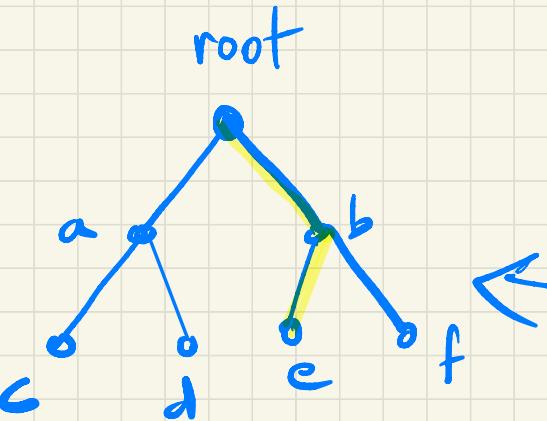
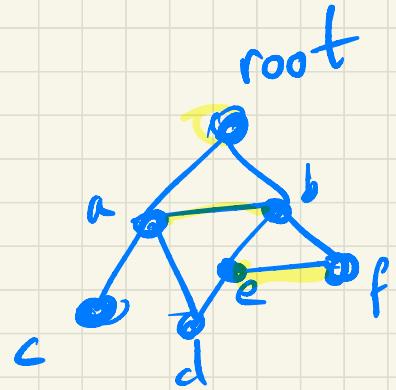
^aInvented in 1945 by Konrad Zuse

BFS Spanning Trees

- Traversal of a graph is performed by visiting all of its vertices in some predefined order.
- **Breadth-First-Search Tree.** A breadth-first-search tree T of a graph G is a spanning tree of G such that for every node of G , the tree path is a minimum-hop path to the root.



- Of course the root must be specified!



Shortest path
from e to
the root is

~~Shortest path
from e to
the root is~~

BFS Algorithm^a

- **BFS Algorithm:** Input a graph $G = (V, E)$
Proceed by layers,
 1. mark the root r ;
 2. mark all neighbor vertices that are one hop away from r ;
 3. mark vertices that are one hop away from these neighbors,
which are two hops away from r ;
 4. and so on.

have not been included before
- It uses a FIFO queue
- It checks whether a vertex has been discovered before enqueueing the vertex rather than delaying this check until the vertex is dequeued from the queue

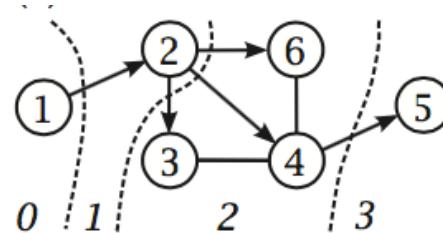
^aInvented in 1945 by Konrad Zuse

BFS (Distance Computation (1/2))

- It starts by placing the source node s at distance $d(s) = 0$; the distance of all other nodes starts as $d(i) = \infty$. *undiscovered*
- At the k th step (starting at $k = 0$), all nodes i at distance $d(i) = k$ are examined, and any neighbors j with $d(j) = \infty$. *undiscovered* have their distance $d(j)$ set to $k + 1$.
- The process halts when step k finds no such neighbors; $d(j)$ is then the length of the shortest path from s to j , or $d(j) = \infty$ if there is no such path.

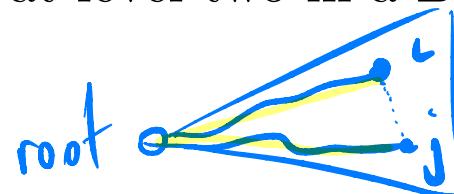
BFS (Distance Computation (2/2))

- BFS is the simplest way to search a graph.
- It is suited only for unweighted graphs: ignores edge weights.
- **Example:**



- **Example:**

In a social network, your friends are at level one and your friends of friends are at level two in a BFS starting at your node.



What is BFS Tree Used for?

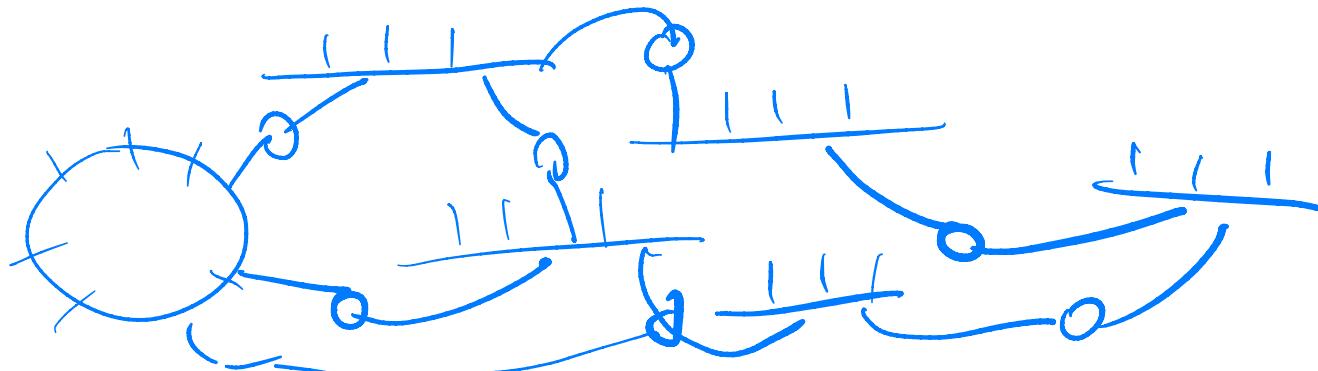
- Finding all nodes within one connected component
 - BFS by itself is not enough: some message passing is needed!
- Finding the shortest path between two nodes u and v (with path length measured by number of edges)
 - u and v could be the nodes initiating a BFS tree.
- Doing efficient broadcast
 - from any any node.

Spanning Trees

Bridged LANs

Bridges connect LANs at the MAC sublayer

- The resulting network is nonhierarchical
- Two standard methods of performing routing on bridged LANs are:
 - Spanning tree routing
 - Source routing
- Both schemes assume unique IDs for nodes and allow nodes to be turned off/on and to move from one location to another



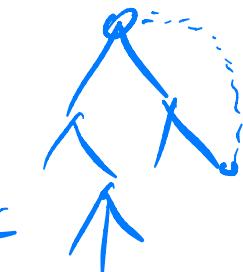
Spanning Tree Routing

- Bridge has *ports* corresponding to each of the LANs it is connected to
- For each port it maintains a Forwarding DataBase (FDB) which lists nodes with which it communicates via that port
- Bridges listen on each port and forward packets from one LAN to another using the FDBs
- Each node appears in exactly one FDB resulting in a *spanning tree* of the network with unique paths between each pair of nodes

What is a spanning tree?

There is a connection to each original node of the network but there are no cycles

Gabriel Test



Dynamic ST routing

- Due to nodes going down, coming up and/or changing location the FDBs of bridges are not always complete (i.e., the spanning tree may change)
- FDBs are updated using *bridge learning*:
 - FDB initialized to empty
 - Add source IDs to FDB
 - Delete inactive nodes
 - Initiate search and respond for unknown IDs

Failures

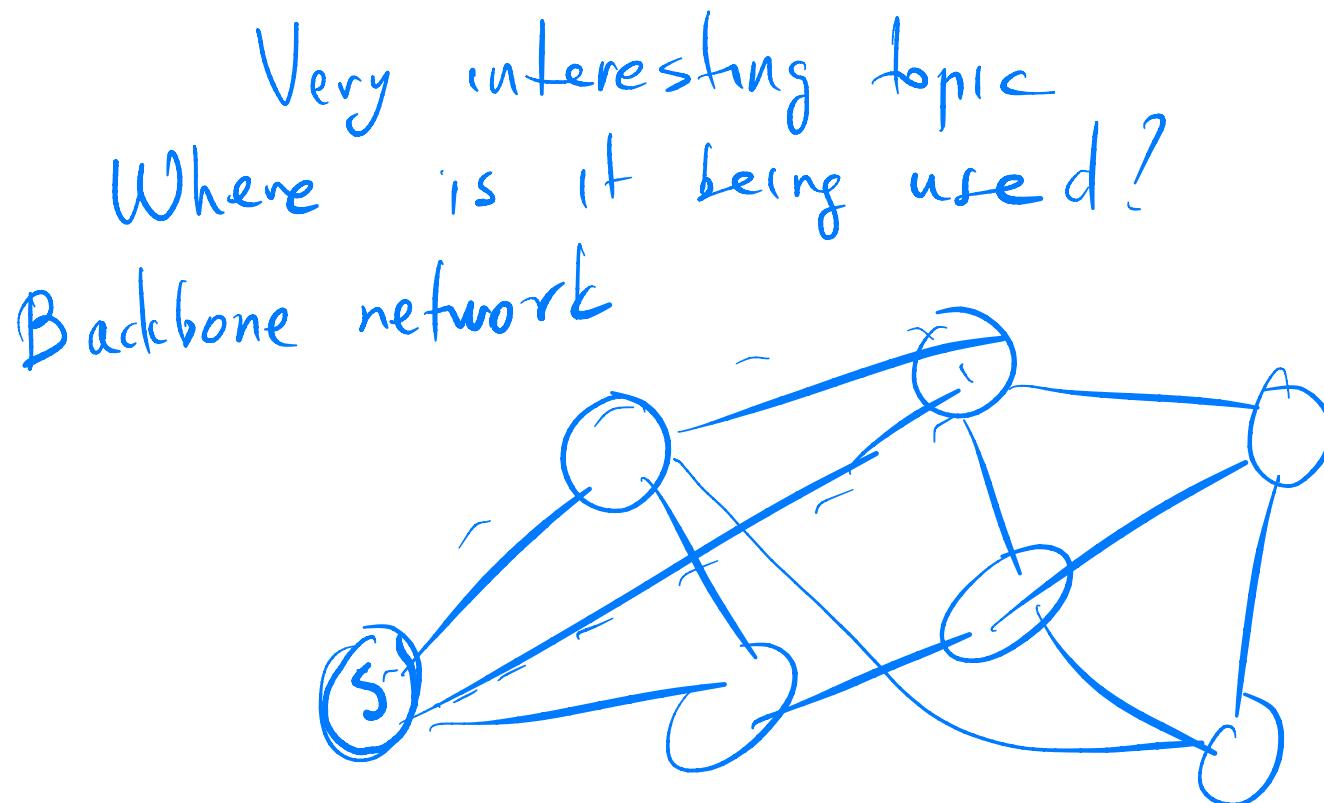
- Methods are required to deal with failures (and recoveries) of bridges or LANs
- Centralized schemes use MST algorithm to compute new spanning tree and update FDBs
 - Distributed schemes use spanning tree to a fixed leader
 - If leader goes down then distributed leader election algorithm is invoked

Kruskal's
Tarjan's
Gallagher's

MST : a spanning tree
s.t. the sum of weights
of all its edges is
minimized.

MSTs

- Minimum spanning tree algorithms are presented in the appendix



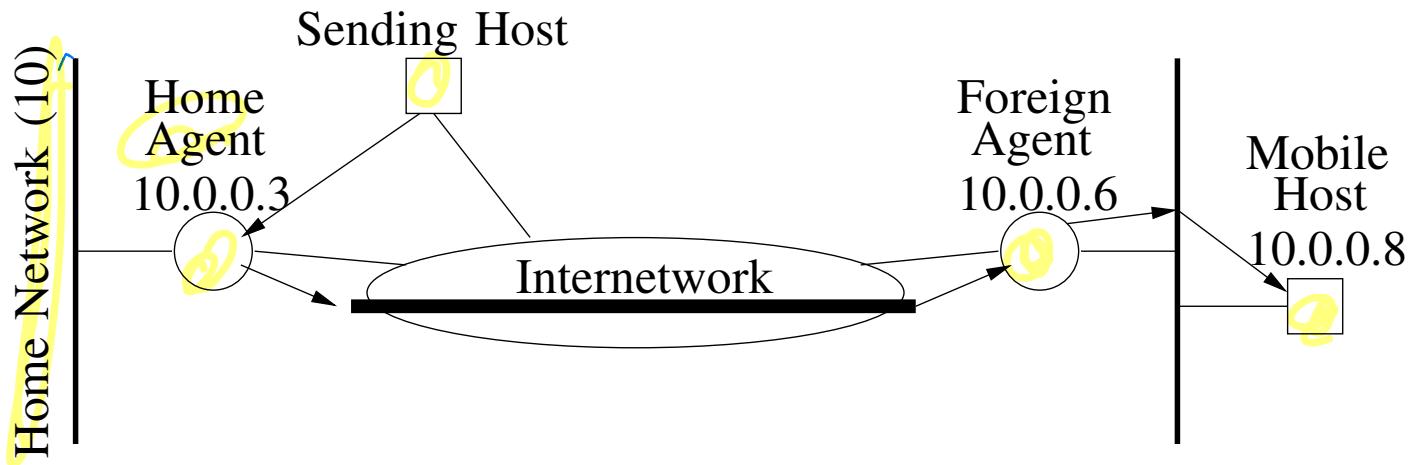
Miscellaneous

Measuring Performance of Routing

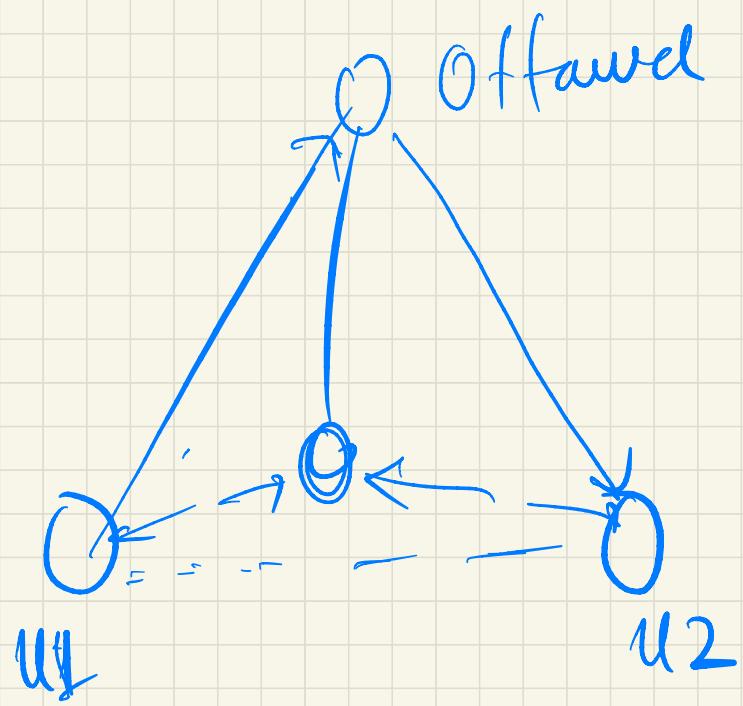
- Weights are placed on links according to either of several rules. ARPANET considered the following three metrics in historical order:A diagram showing two nodes, u and v, represented by blue circles. A horizontal line segment connects them, representing a link. Above the link, there is a small circle containing the letter 'w', with a blue arrow pointing from it to the link, indicating the weight of the link.
 - Queue Length Metric:** # of packets queued and waiting for transmission on each link.
 - Delay Metric:** Link bandwidth and latency (measured by timestamping packets).
$$\text{Delay} = (\text{DepartTime} - \text{ArrivalTime}) + \text{TransmissionTime} + \text{Latency.}$$
 - Utilization Metric:** Link Utilization averaged over last reported utilization together with limits on how much the measurement could change over time from previous value.
- SNMP (Simple Network Management Protocol) is a management tool for monitoring.

Routing for Mobile IP

- Mobile applications require smooth and transparent transition of usability from host-to-host and across platforms in the course of mobility.



- Routes provided may be suboptimal: Sending and Mobile node may be on the same network, but the home network of the Mobile nodes is far. This is known as the **triangle problem**. Solution is to let the sending node know the care-of address of the Mobile node.



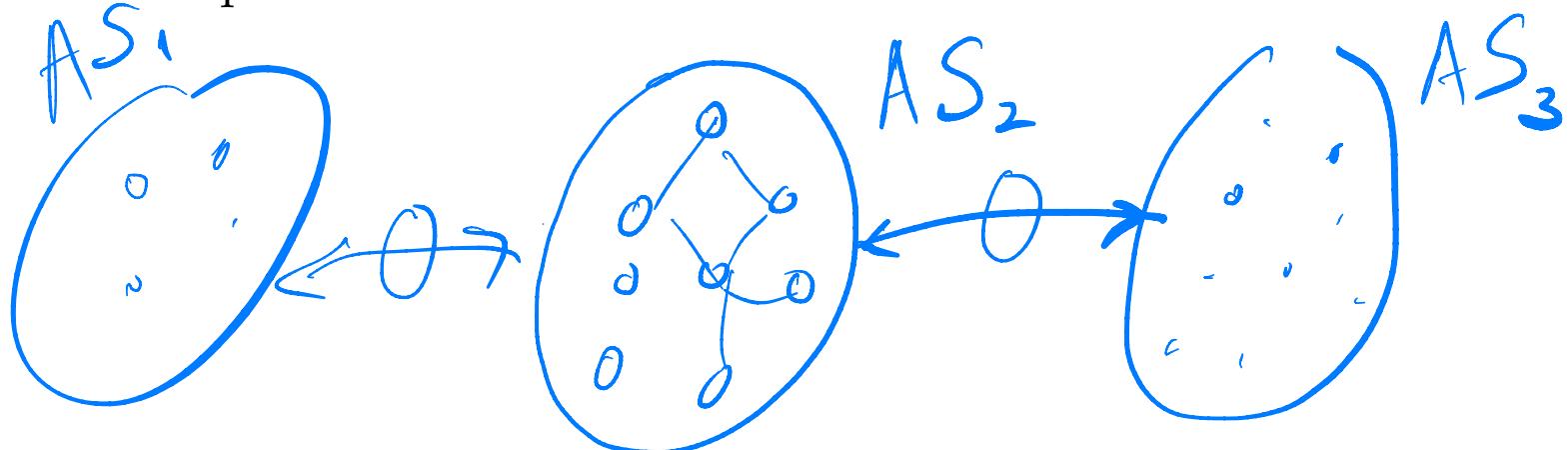
triangle

Routing for Mobile IP

1. FA and HA periodically announce their presence to the networks they are attached. This way the MH learns the address of the HA.
2. When the MH attaches to a foreign network it hears an advertisement from a FA and registers with it.
3. The FA contacts the HA providing a care-of address. This is the IP address of the FA.
4. A Sending host that wants to send a packet to a MH will send it with a destination address equal to home address of that node.
5. The HA tunnels the packet and sends to the FA.
6. The FA unwraps the packet and sends it to the Mobile node.

Autonomous Systems

- An autonomous system consists of a number of subnets exchanging packets via routers that are using the same routing protocol
- Routers are managed by a single or cooperating organizations
- Routing protocol used by an AS called *interior* routing protocol
- Since all routers are managed by one organization the protocol can be optimized to best serve the users of the AS



Internets

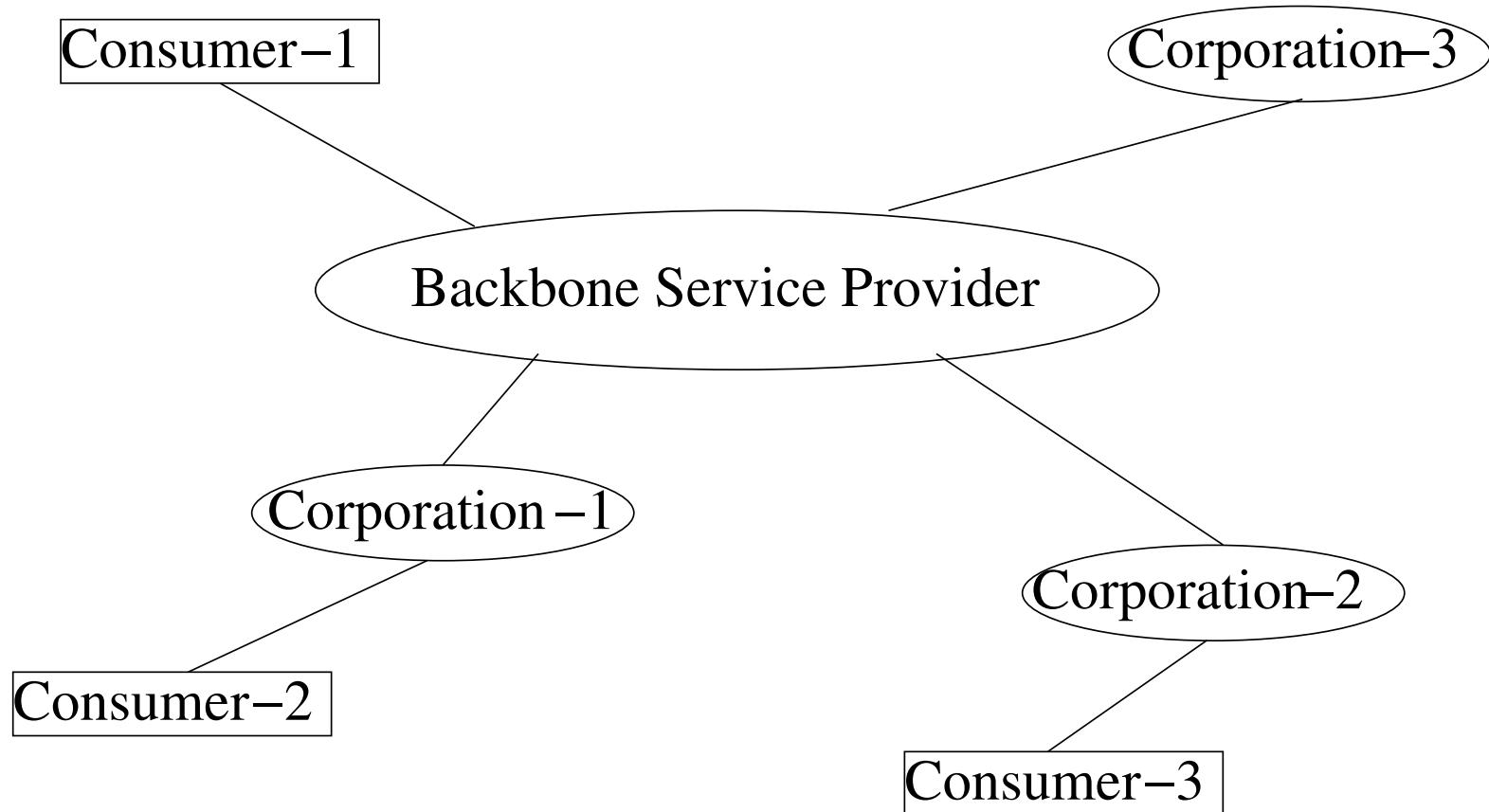
- An internet will connect a number of different autonomous systems run by many different organizations and running many different interior routing protocols
- Routers used to connect different ASs are often called *gateways*
- Protocols used by gateways are called *exterior* routing protocols

Bridged LANs

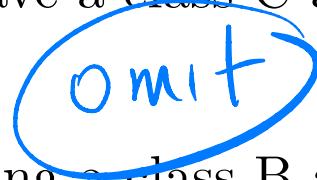
- Bridges connect LANs at the MAC sublayer
- The resulting network is non hierarchical
- Two standard methods of performing routing on bridged LANs are:
 - Spanning tree routing
 - Source routing
- Both schemes assume unique IDs for nodes and allow nodes to be turned off/on and to move from one location to another

remove

Inter Domain Routing in a Network of ASs



Classless Routing (CIDR)

- The IP-address A, B, C class system has inefficiencies. E.g., a two-host network with two hosts will have a class C address leading to a utilization of $\frac{2}{255}$. 
Omit
- Efficiency improves if instead of allocating a class B address (amounting to 64K addresses) we provide a sufficient amount of class C addresses (i.e., blocks of 256 addresses). For Autonomous systems with more than 256 hosts this gives a utilization $\geq 50\%$.
- This raises the following problem: if an autonomous system is allocated n blocks of class C addresses then every backbone router needs n entries for this AS. If we had assigned a single class C address then we would need only one entry.
- How do we optimize address utilization and routing table size?

Classless Routing (CIDR)

- CIDR addresses this problem by aggregating routes.
- Rather than assigning class C addresses at random they are assigned as a block.
- E.g., 20 class C addresses in an interval of numbers from 195.34.17 195.34.36 can be assigned to an AS.
- The BGP (version 4) protocol is defined to understand this.
- All that is needed, is for the router to remember the interval.
- This can be done by remembering one of the two addresses and the range (in this case 20).

Inter Domain Routing

- Routing usually partitioned into inter-domain and intra-domain.
- Inter-domain routing has its origins on ARPANET and its design was influenced by the internet.
 1. EGP (Exterior Gateway Protocol: early version of BGP)
 2. BGP (Border Gateway Protocol)
 3. IDRIP (Inter Domain Routing Protocol)
- Protocols operate in conjunction with IP.
- Information exchanged between hosts via a connectionless protocol.

These are quite complex protocols

BGP Considerations

Interdomain routing is difficult and affected by

1. Scale factors,
2. Different routing protocols on different Autonomous Systems,
3. Security.

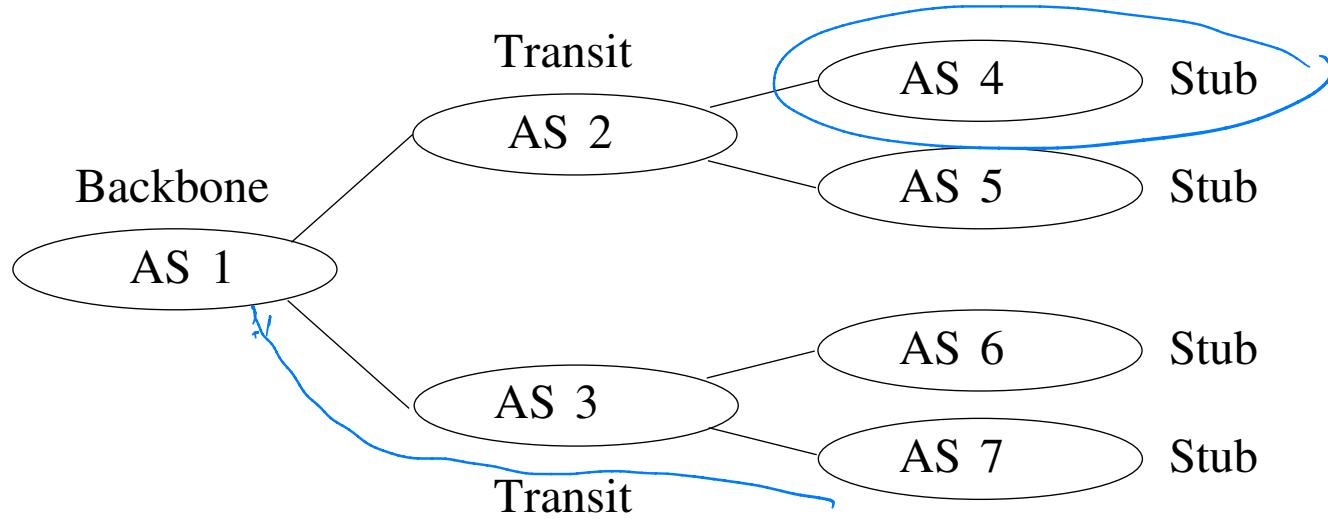
There are two types of traffic **local** and **transit**.

Interdomain routing assumes internet is a collection of Autonomous Systems. There are three types of ASs:

1. **stub AS**: has only a single connection to one other AS.
2. **multihomed AS**: has connections to more than one other AS but refuses to carry transit traffic.
3. **transit AS**: designed to carry transit traffic.

BGP

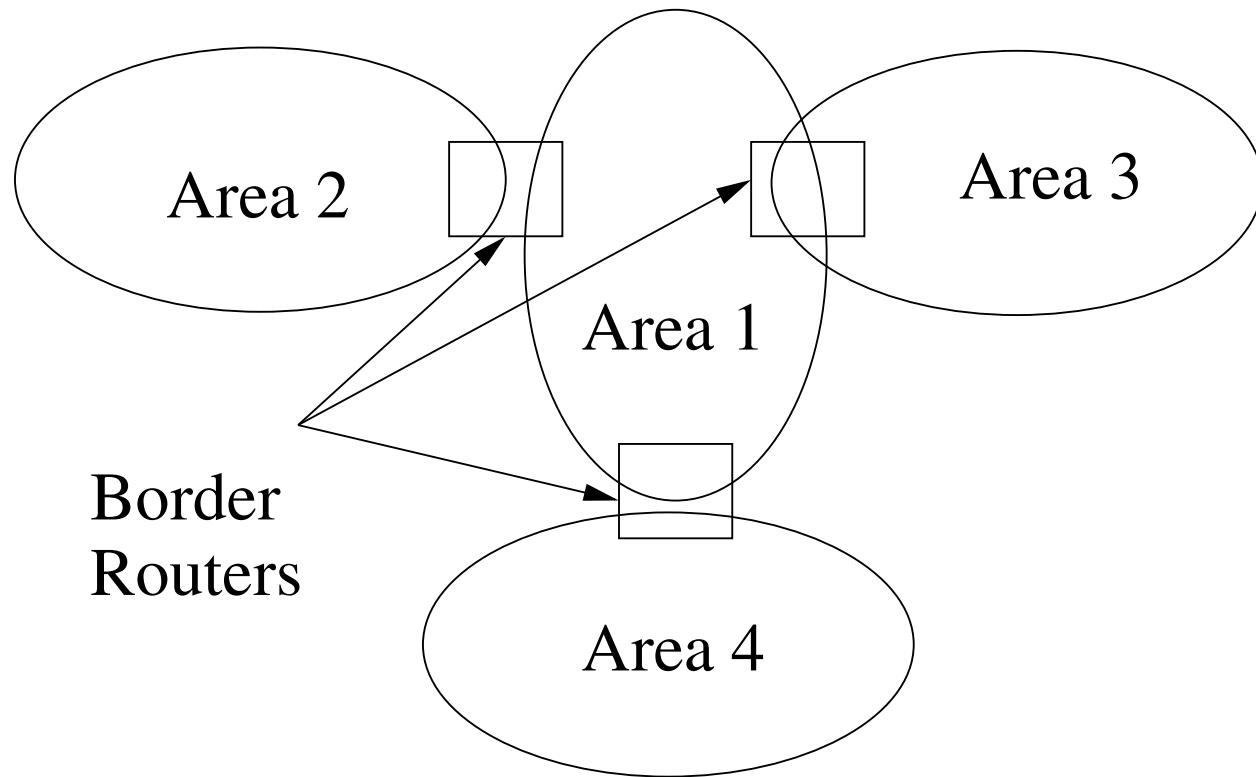
Unlike Distance-Vector or Link-State Protocols, BGP advertises complete paths as an enumerated list of ASs..



A₂ can advertise reachability of A₄ and A₅; A₃ can advertise reachability of A₆ and A₇. The backbone network on receiving this information, can advertise a path on how a node can be reached.

Routing Areas

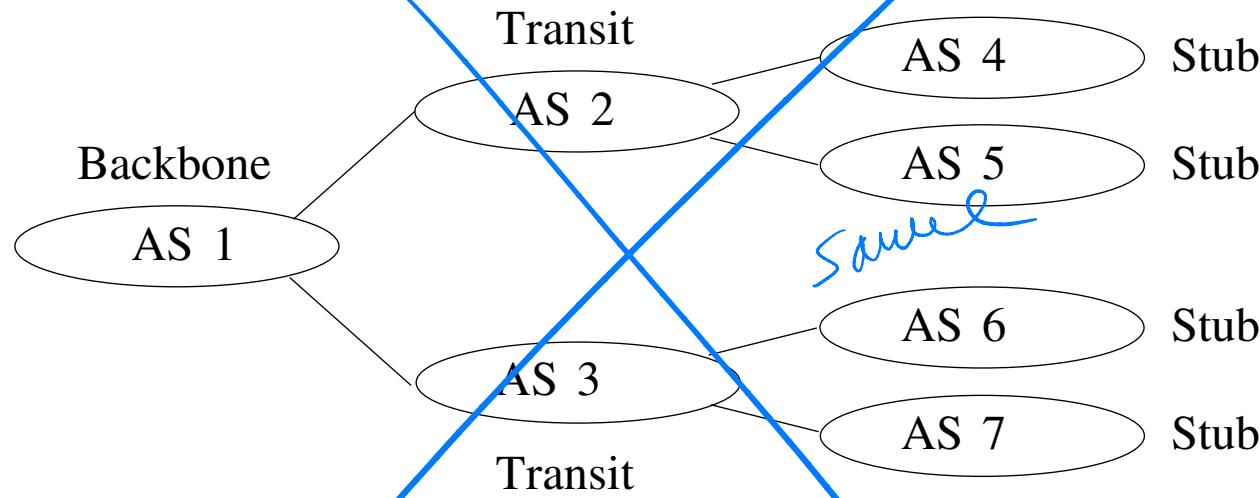
Administrators can divide a network into routing areas.



Here we try to create hierarchies at the cost of hindering optimal routing. But this is an essential tradeoff!

BGP

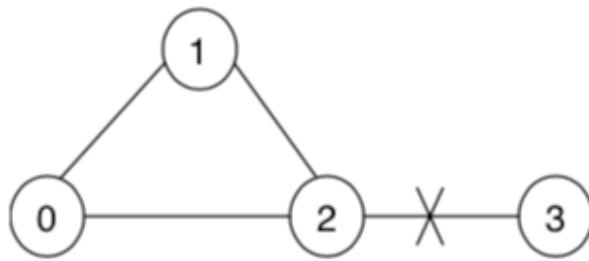
Unlike Distance-Vector or Link-State Protocols, BGP advertises complete paths as an enumerated list of ASes..



A₂ can advertise reachability of A₄ and A₅; A₃ can advertise reachability of A₆ and A₇. The backbone network on receiving this information, can advertise a path on how a node can be reached.

Exercises^a

1. Consider RIP in the network below:

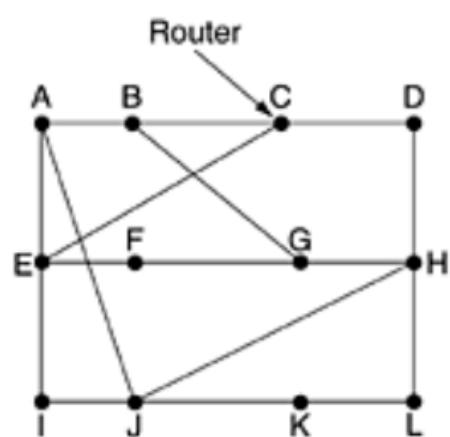


- (a) Compute, distances $d(1, 3)$, $d(2, 3)$ without errors on edges.
(b) What is the new value of $d(2, 3)$ when the link $(2, 3)$ fails?
(c) Node 1 does not know yet of the failure of edge $(2, 3)$. What information does it pass to nodes 0, 2?
(d) To what value does node 2 update $d(2, 3)$?
(e) What is node 1's subsequent update?
2. Consider a path consisting of n links connecting node s to node t . Let p_{ij} be the probability that the i -th link fails. Assuming

^aDo not submit.

that links fail independently at random what is the probability that the path succeeds to transmit a message?

3. (*) Consider a network and let p_e be the probability that the link e fails. Assume that link failures are independent. Use Dijkstra's algorithm and your solution of Exercise 2 to design an algorithm so that for any pair of nodes s, t it finds a path that maximizes the probability that a message from s to t will be transmitted successfully. **Hint:** Consider your answer in Exercise 2 and optimize the quantity arising when you take its logarithm.
4. Consider the network depicted below in which delay is being used as a metric and the router knows the delay to each of its neighbors. Once every T msec each router sends to each neighbor a list of its estimated delays to each destination. It also receives a similar list from each neighbor. Imagine that one of these tables has just come in from neighbor X , with X_i



New estimated delay from J

To	A	I	H	K	Line
A	0	24	20	21	8 A
B	12	36	31	28	20 A
C	25	18	19	36	28 I
D	40	27	8	24	20 H
E	14	7	30	22	17 I
F	23	20	19	40	30 I
G	18	31	6	31	18 H
H	17	20	0	19	12 H
I	21	0	14	22	10 I
J	9	11	7	10	0 -
K	24	22	22	0	6 K
L	29	33	9	9	15 K

JA delay is 8 JI delay is 10 JH delay is 12 JK delay is 6

Vectors received from J's four neighbors

New routing table for J

being X 's estimate of how long it takes to get to router i . If the router knows that the delay to X is m msec, it also knows that it can reach router i via X in $X_i + m$ msec. By performing this calculation for each neighbor, a router can find out which estimate seems the best and use that estimate and the corresponding line in its new routing table.

The updating process is illustrated in the table. The first four columns of part show the delay vectors received from the neighbors of router J . A claims to have a 12 msec delay to B , a 25 msec delay to C , a 40 msec delay to D , etc. Suppose that J has measured or estimated its delay to its neighbors, A, I, H, K as 8, 10, 12, 6 msec, respectively.

Appendix

Route Calculation in LSP: Dijkstra's Algorithm

- We describe Dijkstra's Algorithm as it generates shortest paths.
- A modification can be made which provides the LSPs (these are trees forming the view from a node)
 - N set of nodes in network.
 - $l(i, j)$ (non-negative) cost associated with edge $\{i, j\}$.
 - Let s be the node executing the algorithm in order to find shortest paths to all other nodes in the network.
 - M is the set of nodes incorporated so far by algorithm.
 - $C(n)$ is the cost of the path from s to node n .

Dijkstra's Algorithm

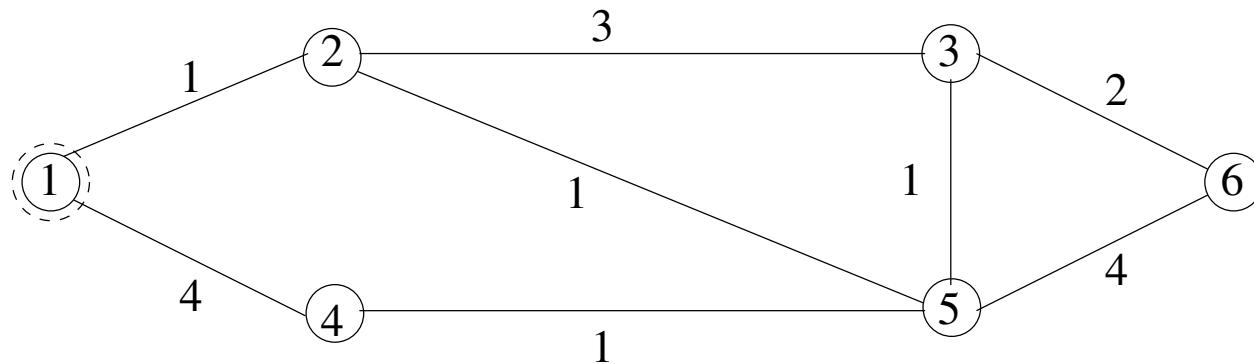
1. $M = \{s\}$
2. **for each** $n \in N \setminus \{s\}$
3. $C(n) = l(s, n)$
4. **while** $N \neq M$ **do**
5. $M = M \cup \{w\}$ such that $C(w)$ is min for all $w \in N \setminus M$
6. **for each** $n \in N \setminus M$
7. $C(n) = \min\{C(n), C(w) + l(w, n)\}$

Route Calculation in LSP: Dijkstra's Algorithm

- Also finds shortest paths from all nodes to some fixed destination (or source)
- Requires that all edge weights are nonnegative (not a restriction for most network applications)
- Shortest paths found in order of increasing path length.
- Crucial idea:
 - During the k th step the k th closest node to the destination is found by considering the distance of nodes not among the $k - 1$ closest to any node among the $k - 1$ closest

Example: Dijkstra's Algorithm

Start node is 1

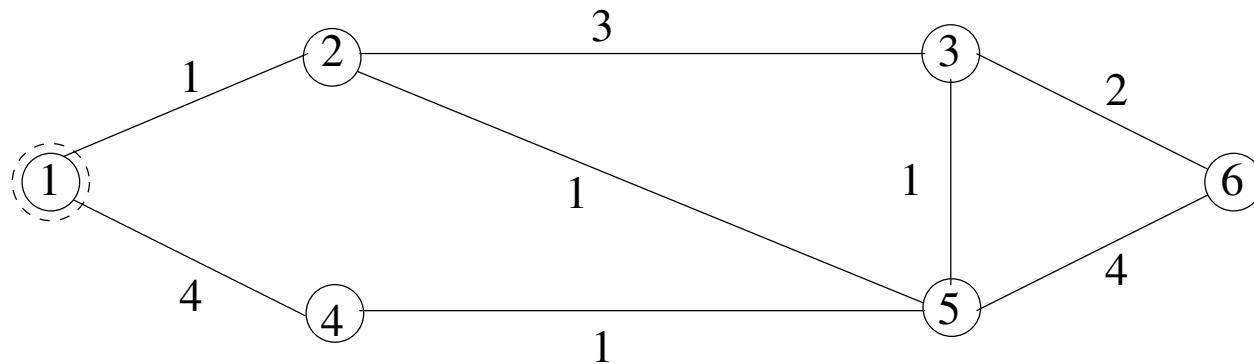


Iteration 1: Compute all costs to 1. Update min cost routes to 1.

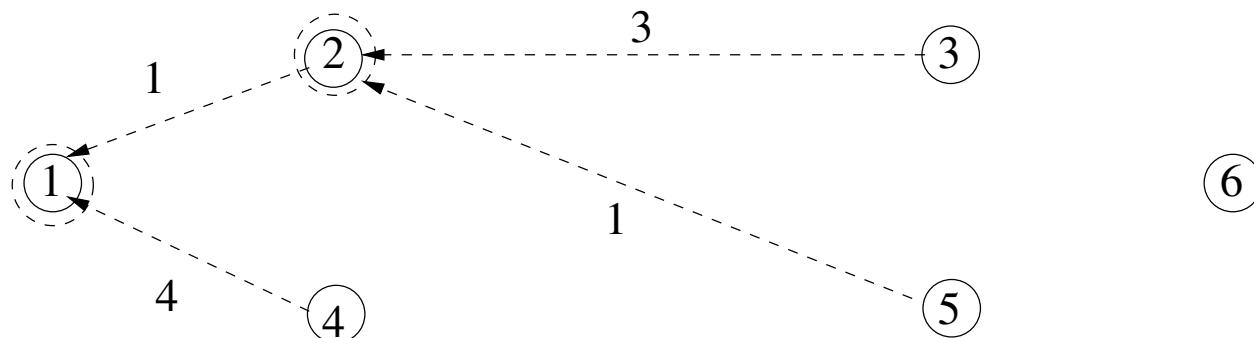


Example: Dijkstra's Algorithm

Start node is 1

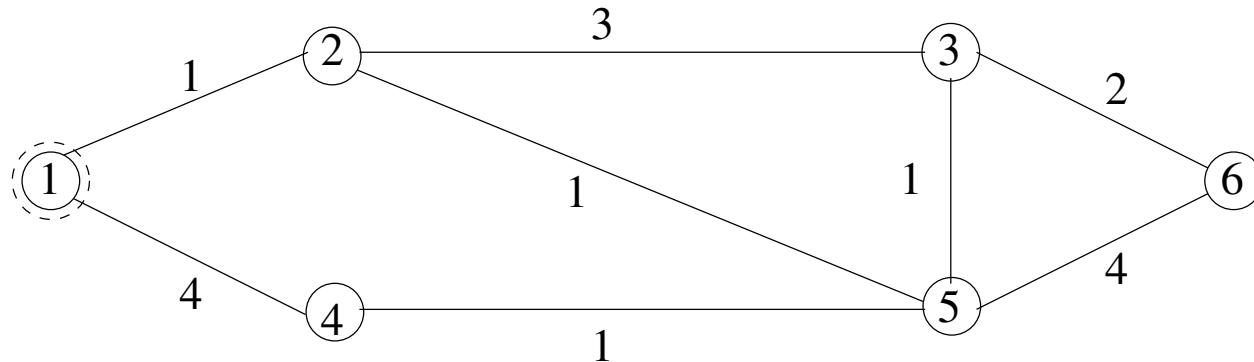


Iteration 2: Add min cost node to M (node 2). Update min cost routes to 1.

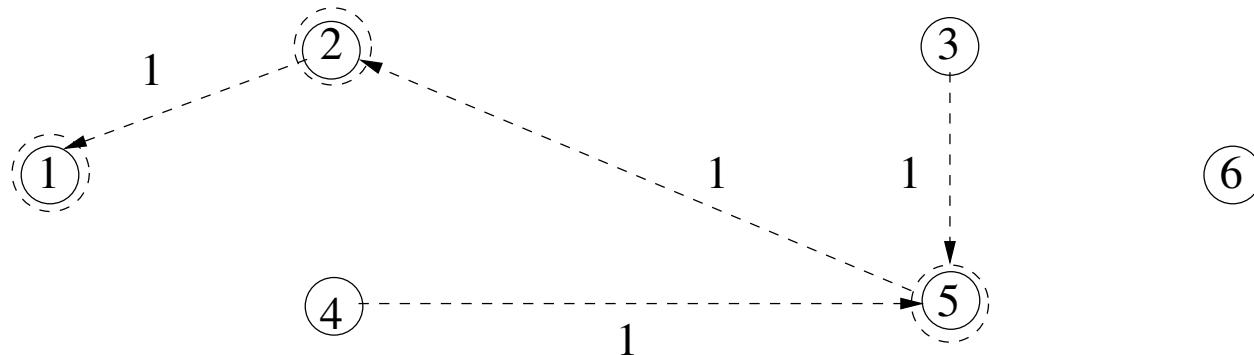


Example: Dijkstra's Algorithm

Every node executes Dijkstra's algorithm

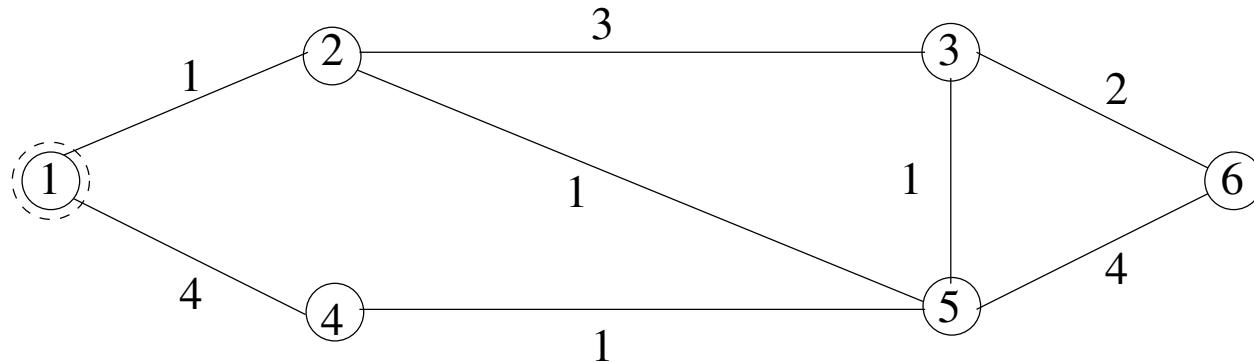


Iteration 3: Add min cost node to M (node 5). Update min cost routes to 1.

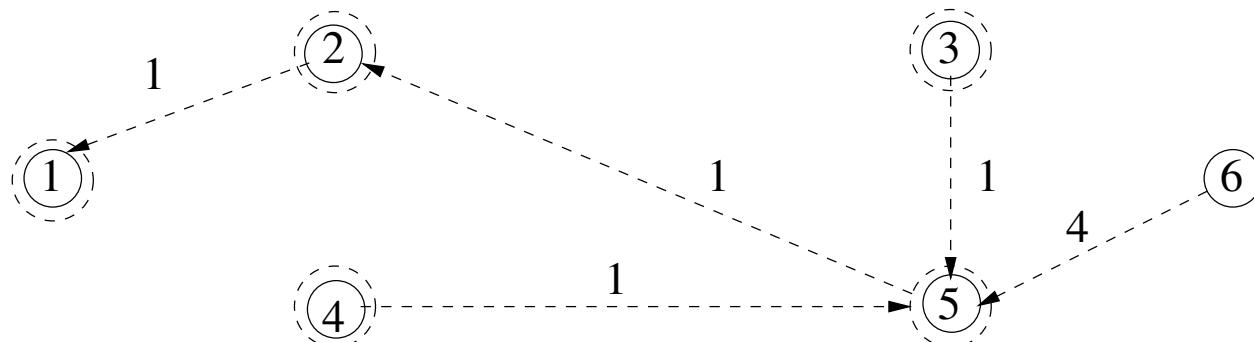


Example: Dijkstra's Algorithm

Every node executes Dijkstra's algorithm

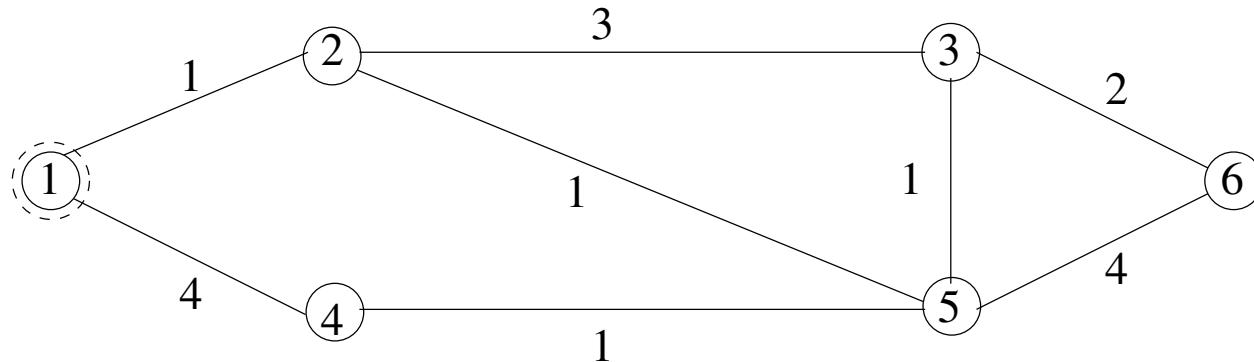


Iterations 4: Add min cost node to M (node 3). Update min cost routes to 1.

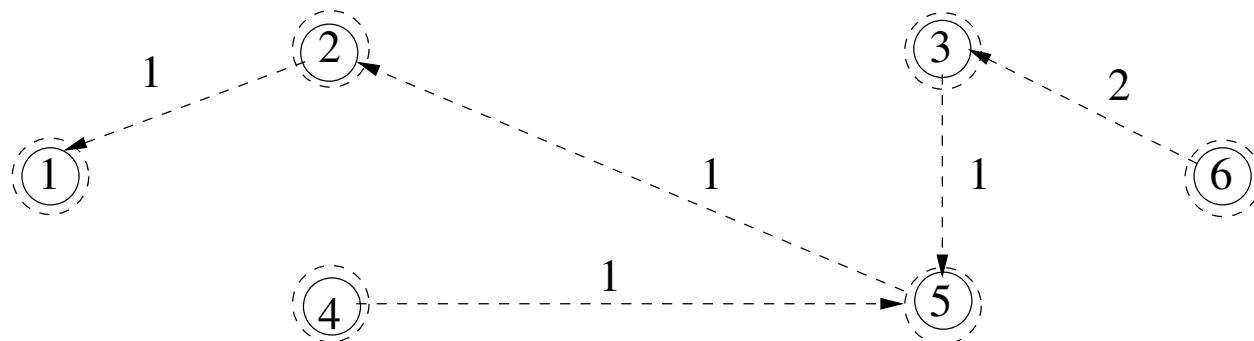


Example: Dijkstra's Algorithm

Every node executes Dijkstra's algorithm



Iteration 5: Add min cost node to M (node 6). Update min cost routes to 1.



Analysis

- Algorithm can be implemented so as to run in time $O(n^2)$, where n is the number of nodes.
- The algorithm computes weights of paths not the paths
- Can be easily modified to compute the paths:
- The last edge found in update step 3 is the first edge in a shortest path to destination and can be used to compute a shortest path tree and to compute routing tables

Spanning Trees

Spanning Trees

- A spanning tree of a network is a subnetwork that is a tree (i.e., contains no cycles) and includes all of the nodes of the network
- If the edges of the network are weighted (e.g., representing average delay expected on a given LAN) a minimum weight spanning tree is one with minimum sum of edge weights
- Two standard algorithms for computing MST are:
 - Prim's algorithm
 - Kruskal's algorithm

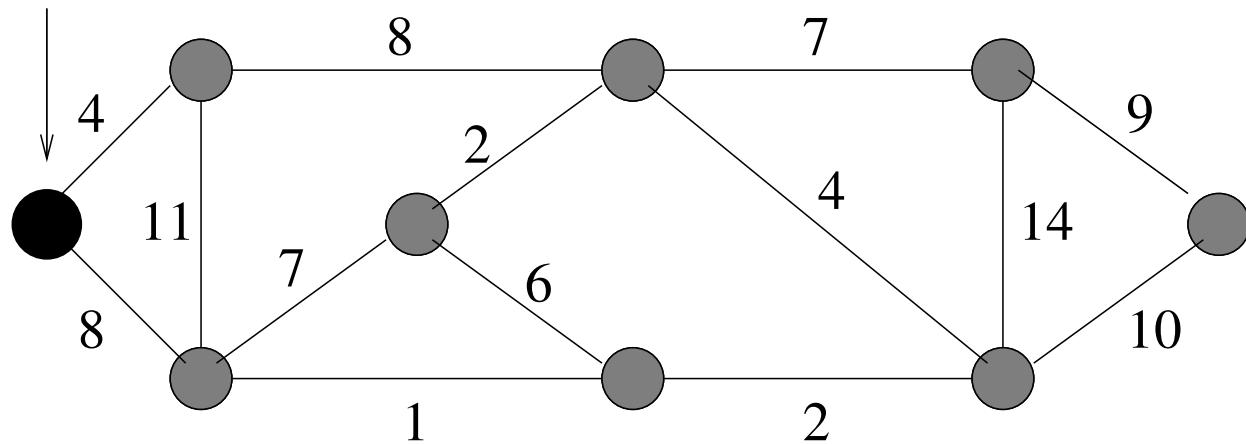
Spanning Tree Terminology

- $G = (V, E)$ is an undirected graph
- V is the set of nodes (vertices)
- E is the set of edges
- $w_{i,j}$ is the weight of the edge (i, j)
- A *spanning tree* is an acyclic subgraph containing all nodes
- Weight of a tree is the sum of its edge weights
- Minimum weight spanning tree (MST) is ST of minimum weight

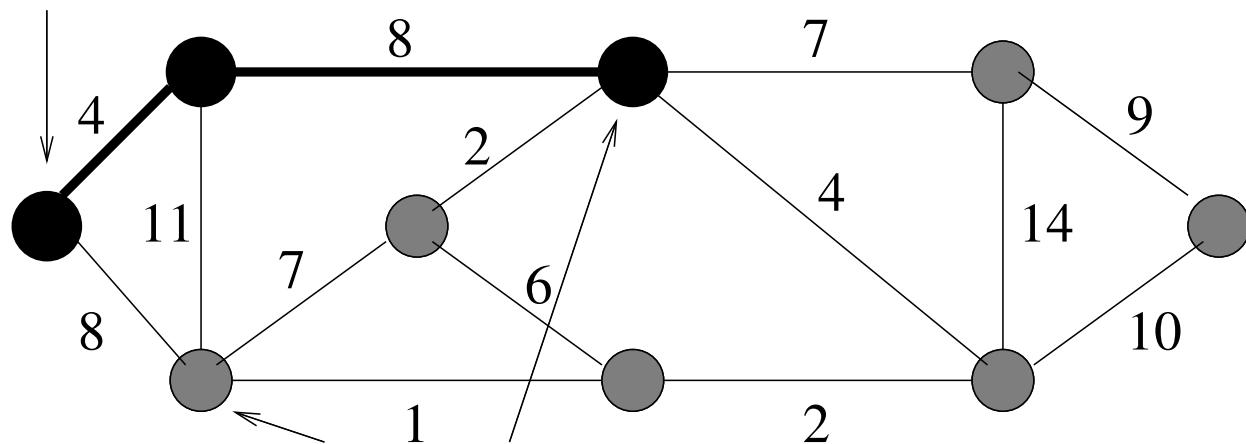
Prim's Algorithm (Jarnik, 1930)

- P is set of nodes in tree and D_i is min weight edge from node i to a node in P
- Initially $P = \{1\}$, and $D_i = w_{i,1}$ if $(i, 1)$ exists, ∞ otherwise
 1. Find $i \notin P$ such that D_i is minimum
 2. $P = P \cup \{i\}$
 3. For $j \notin P$, $D_j = \min\{D_j, w_{j,i}\}$
 4. Go back to 1
- Can be implemented in $O(|E| + |V| \log |V|)$ time

Root node.

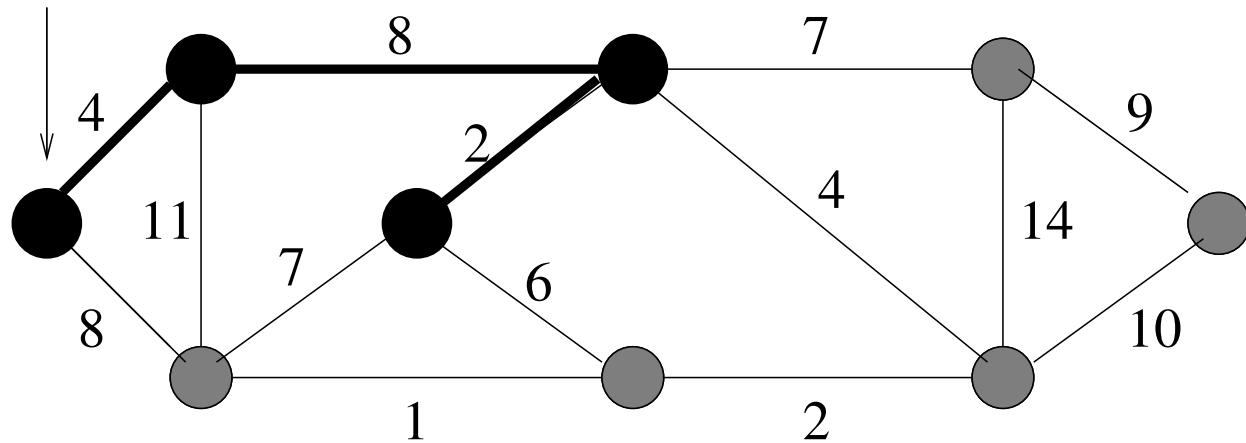


Root node.

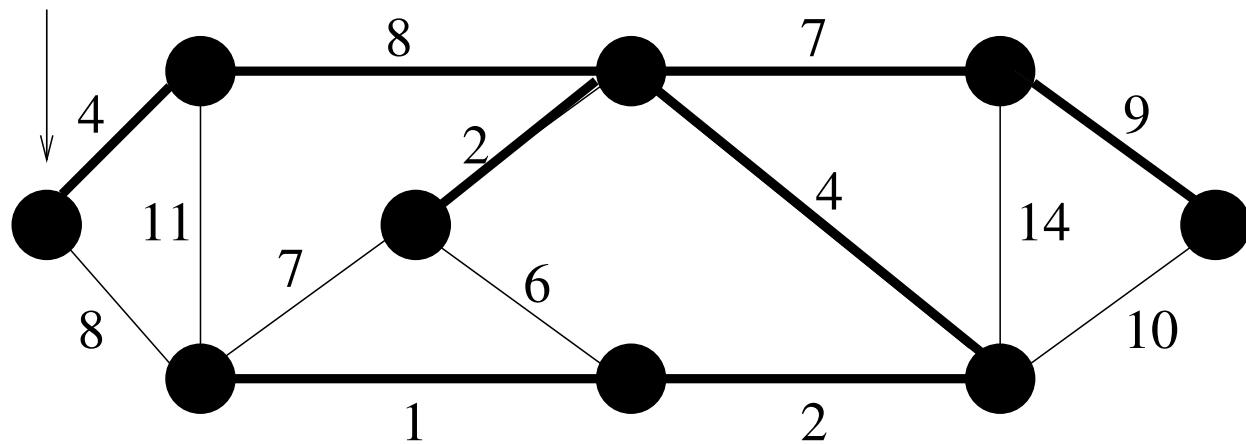


We have a choice: can add either of these two nodes.

Root node.



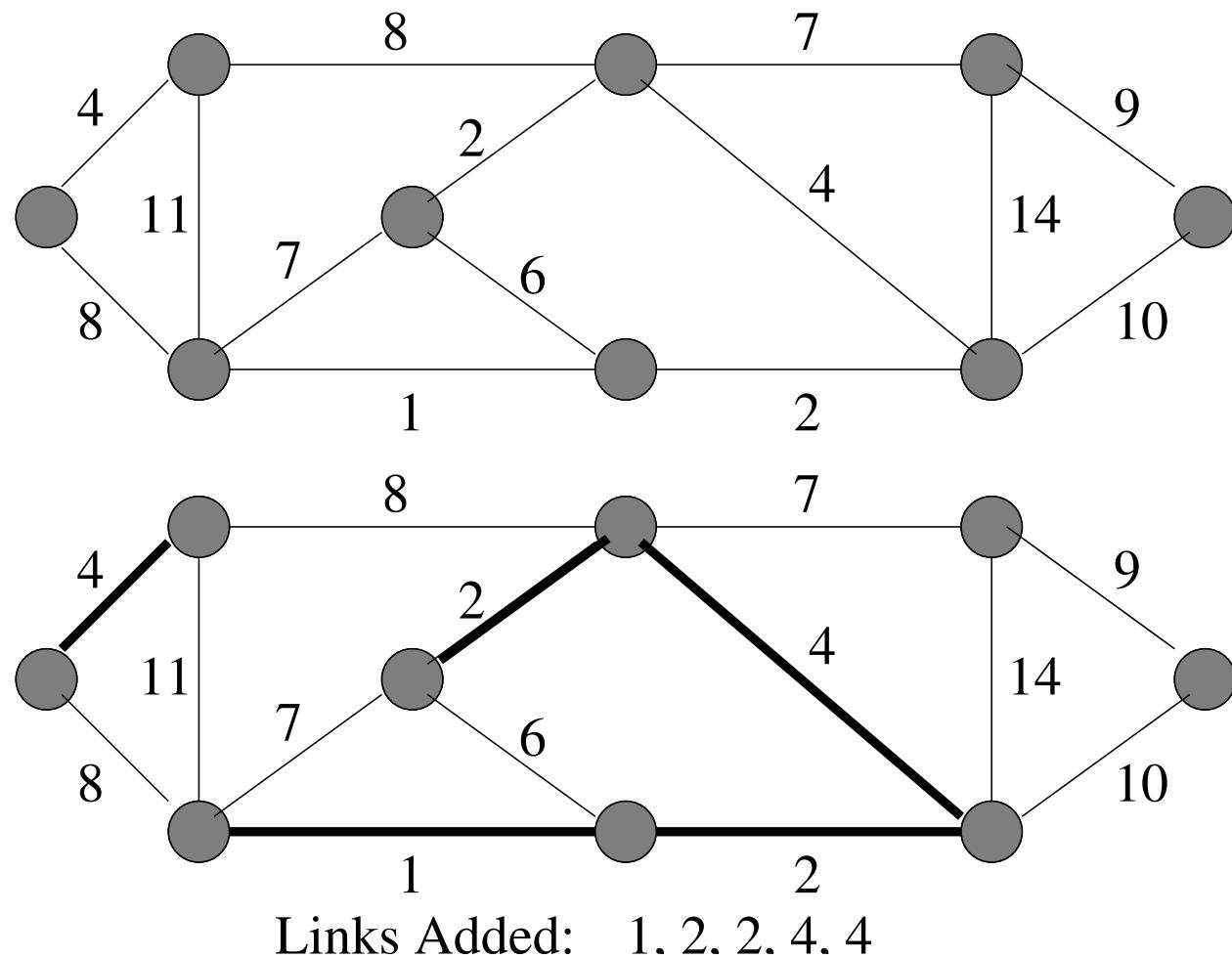
Root node.

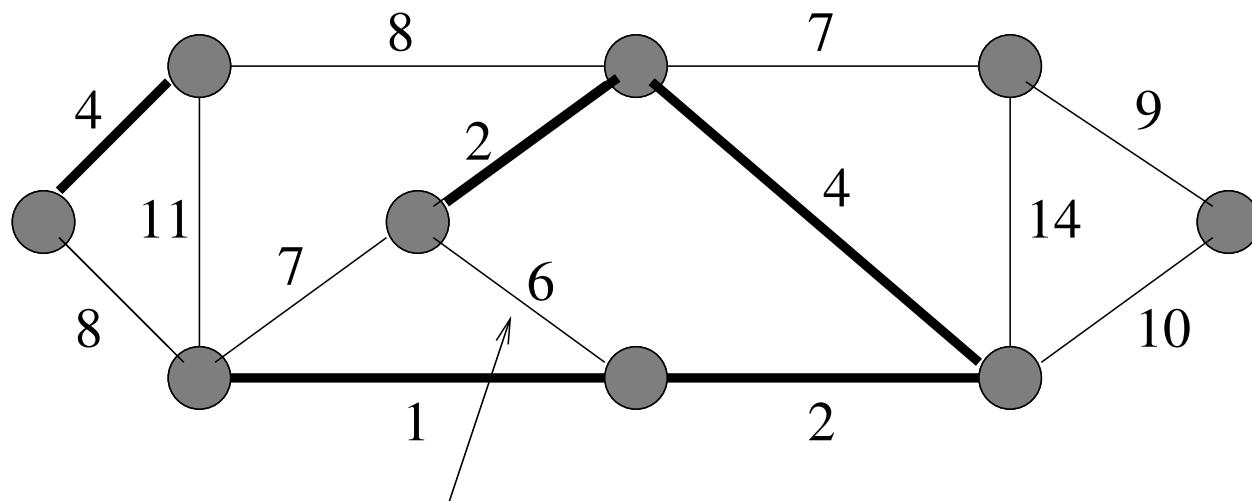


Then we add nodes adjacent to links
4, 2, 1, 7, 9 in this order

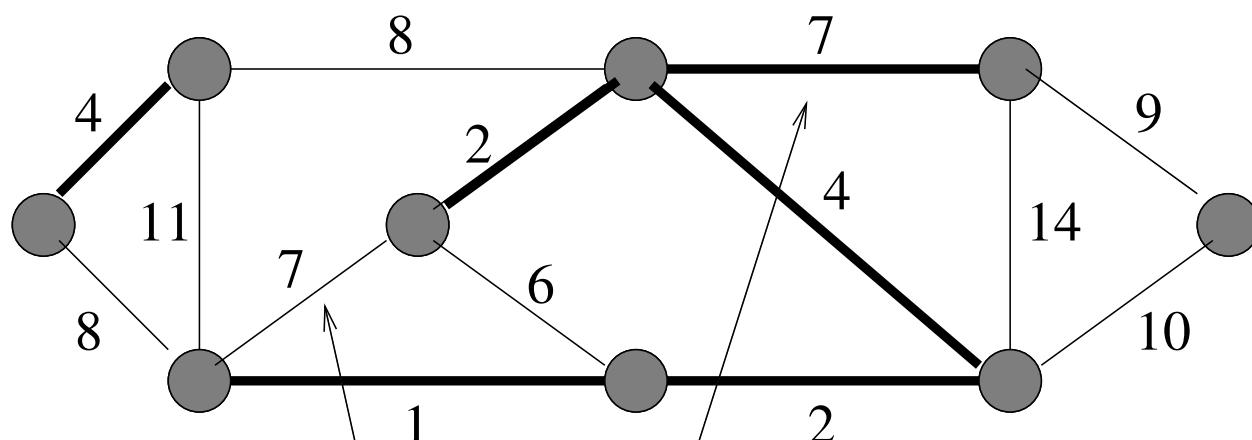
Kruskal's Algorithm (Boruvka, 1926)

- Algorithm
 1. Sort the edges of G in increasing order
 2. Consider edges in order and add edge to tree if the result does not form a cycle
- Time complexity is $O(|E| \log |E|)$
- Can be implemented in a distributed manner and used to elect a leader



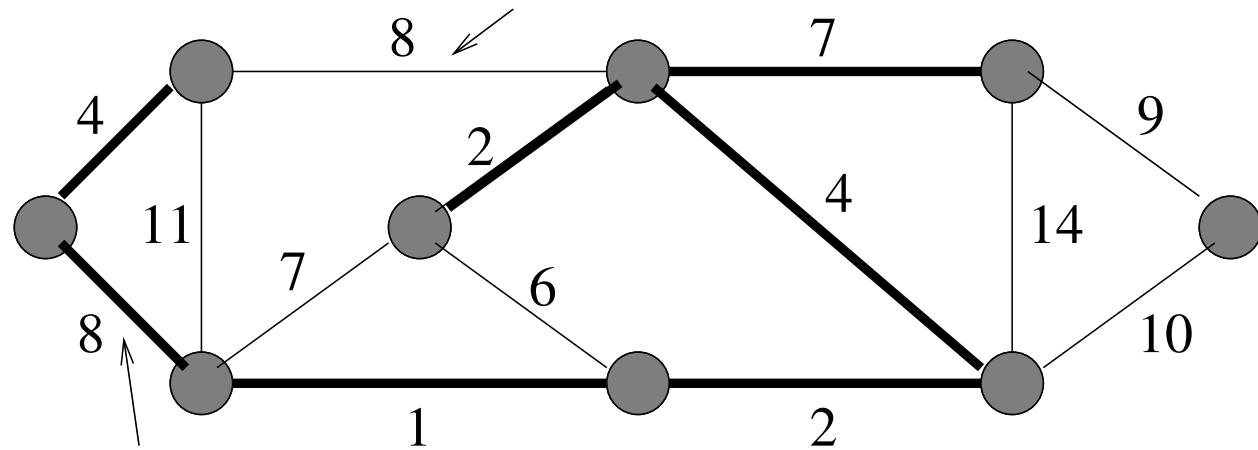


We cannot add this link: creates a cycle

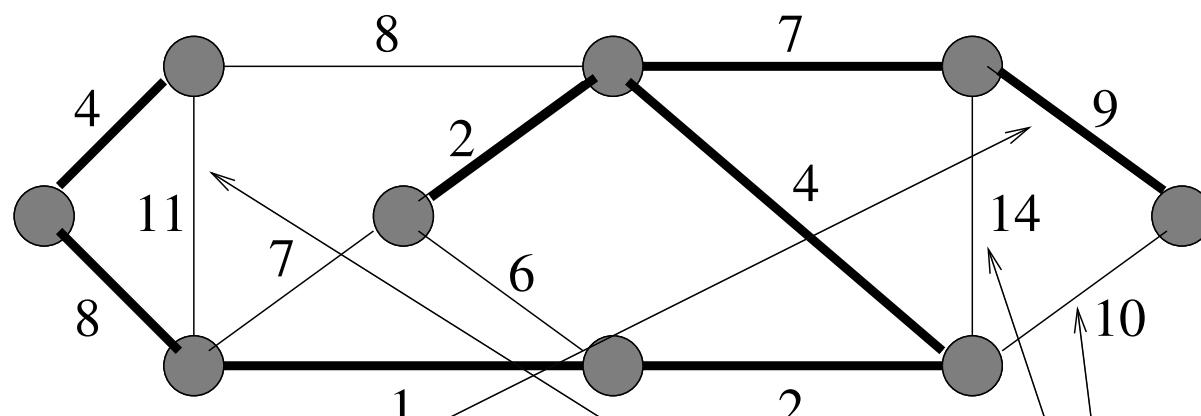


We can add the 7 that does not create a cycle

We cannot add this 8.



We can add this 8.

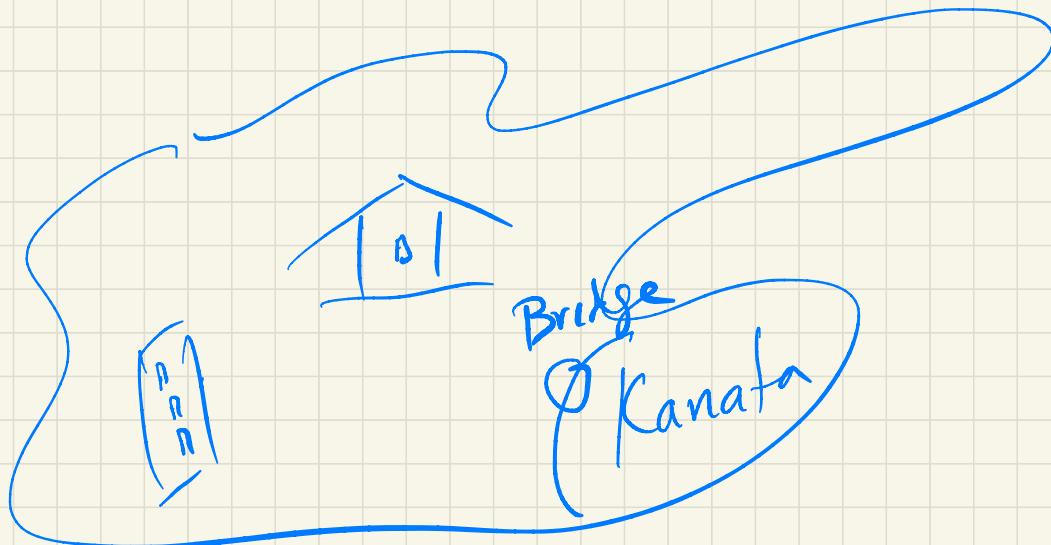


We add the 9. None of 10, 11, 14 can be added

Internetworking (IP)

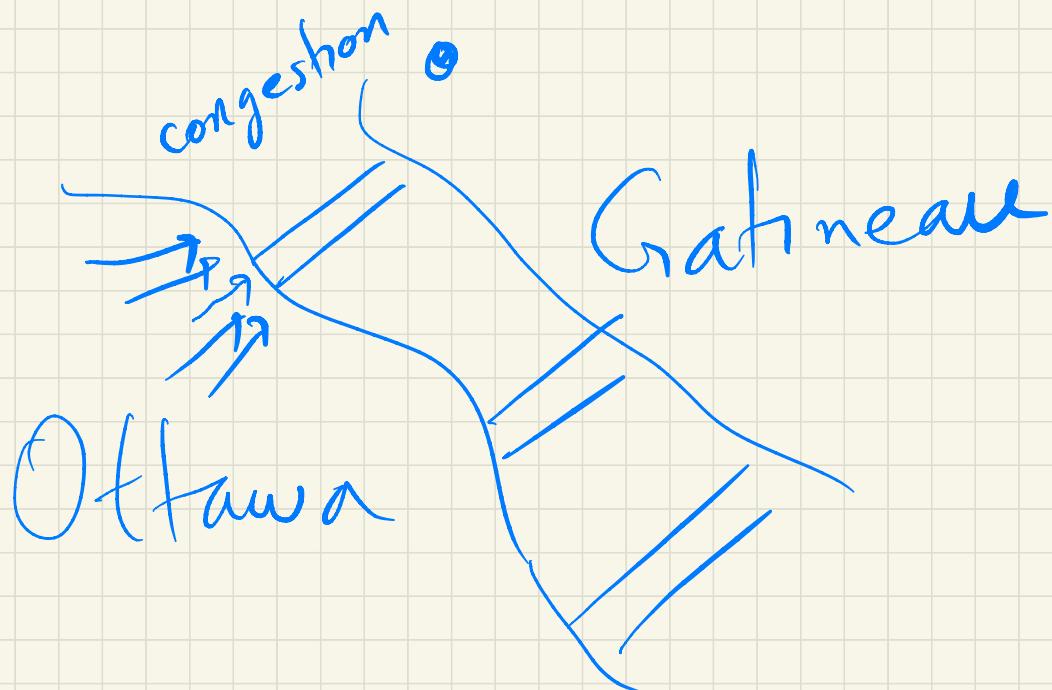
IP/TCP

Big City ; Post Office



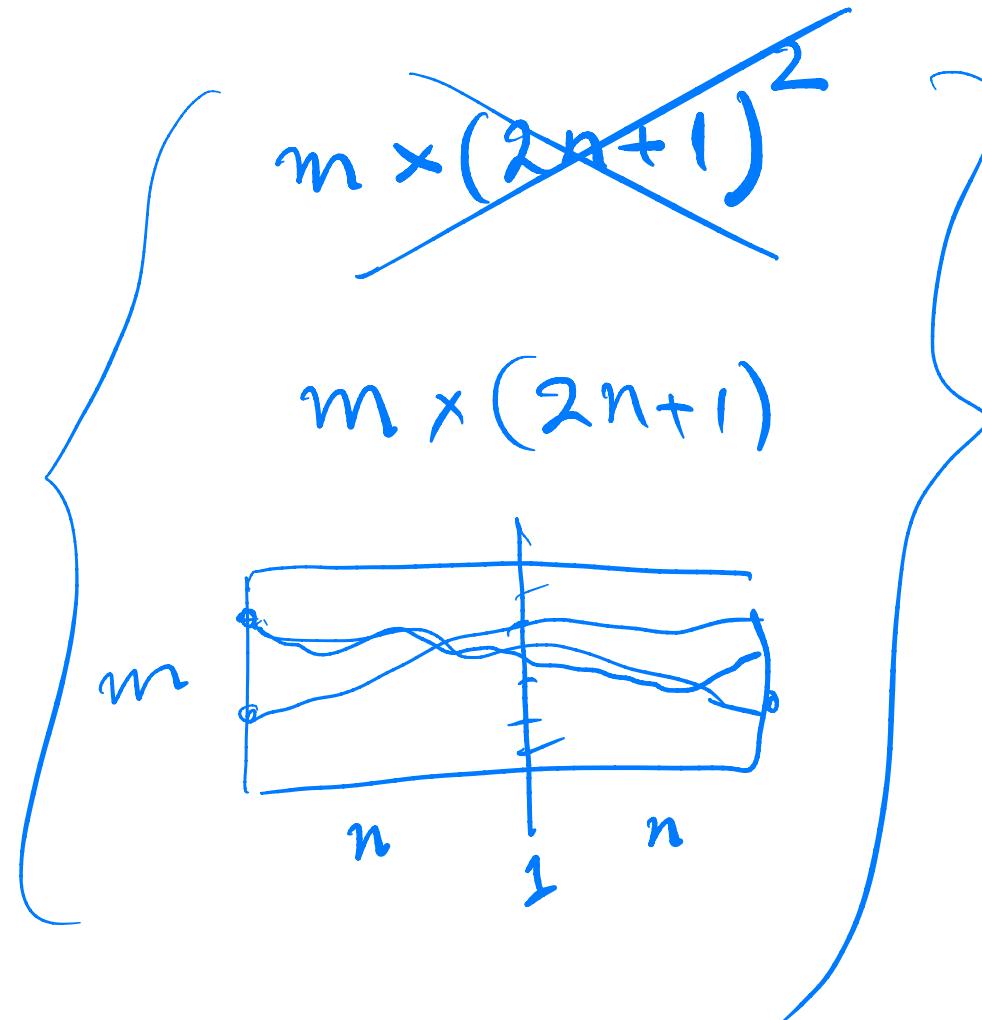
IP: Provides addresses

TCP: Controls traffic



Outline

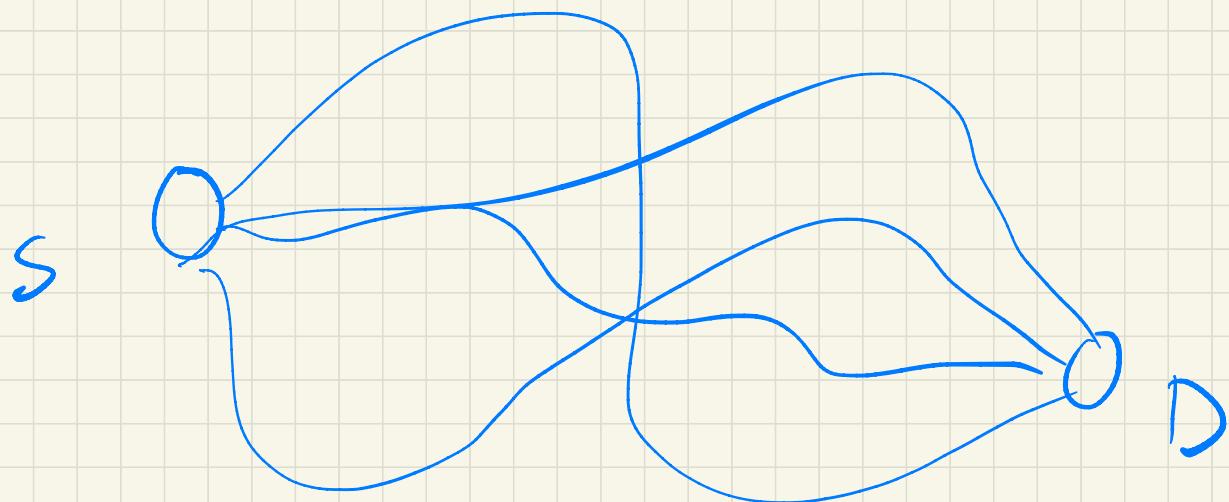
- Internetworks
 - IPv4
- Address Resolution
- IPv6



IP Networks

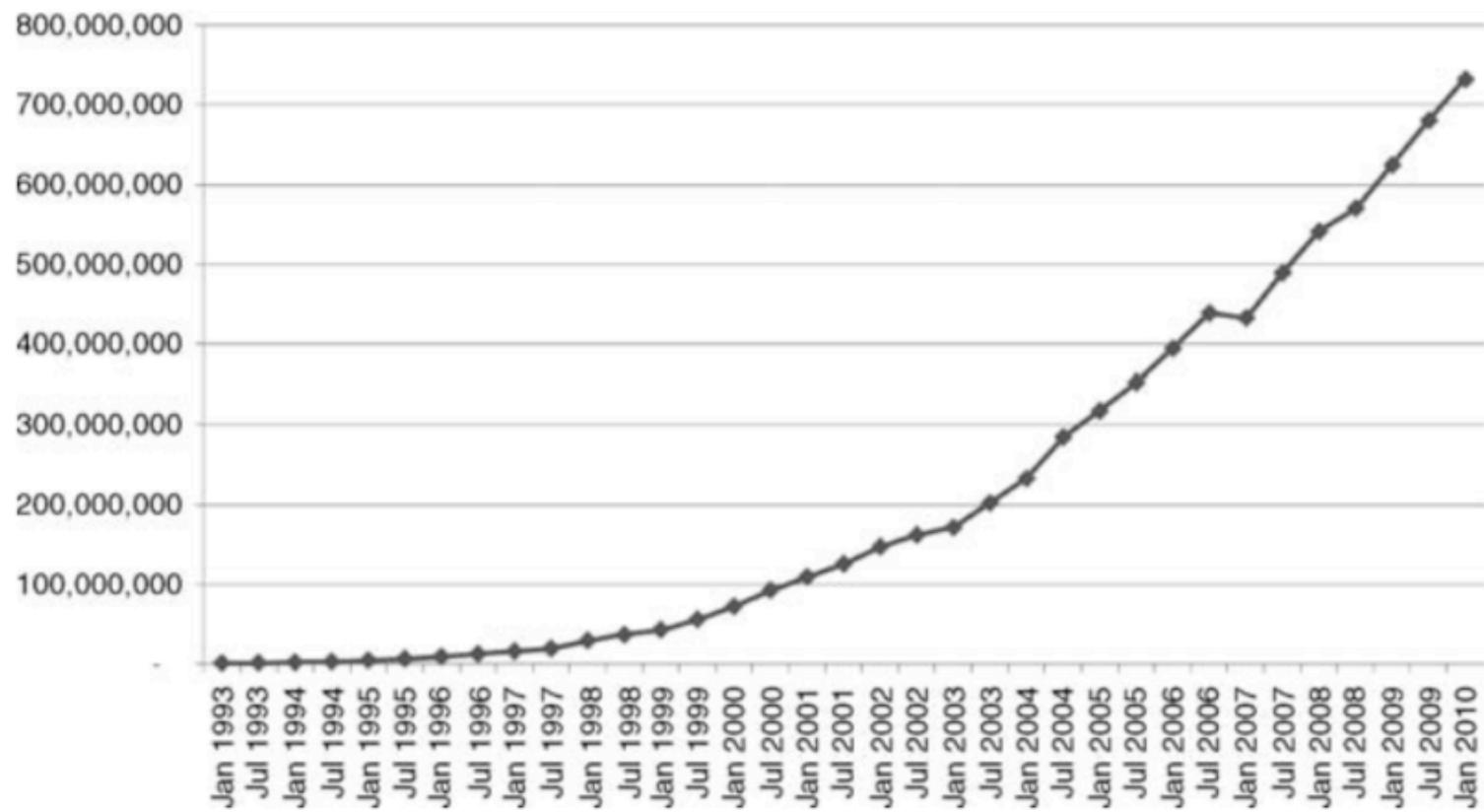
- IP is the most widely deployed network layer protocol worldwide.
- Emerging from a U.S. government sponsored networking project for the U.S. Department of Defense begun in the 1960s, the TCP/IP suite has evolved and scaled to support networks from hundreds of computers to hundreds of millions today.
- The number of devices or hosts on the Internet exceeded 730 million as of early 2010 with average annual additions of over 75 million hosts per year.
- The fact that the Internet has scaled rather seamlessly from a research project to a network of over 730 million computers is a testament to the vision of its developers and robustness of their underlying technology design.

We would use in the past the
"circuit switching" model



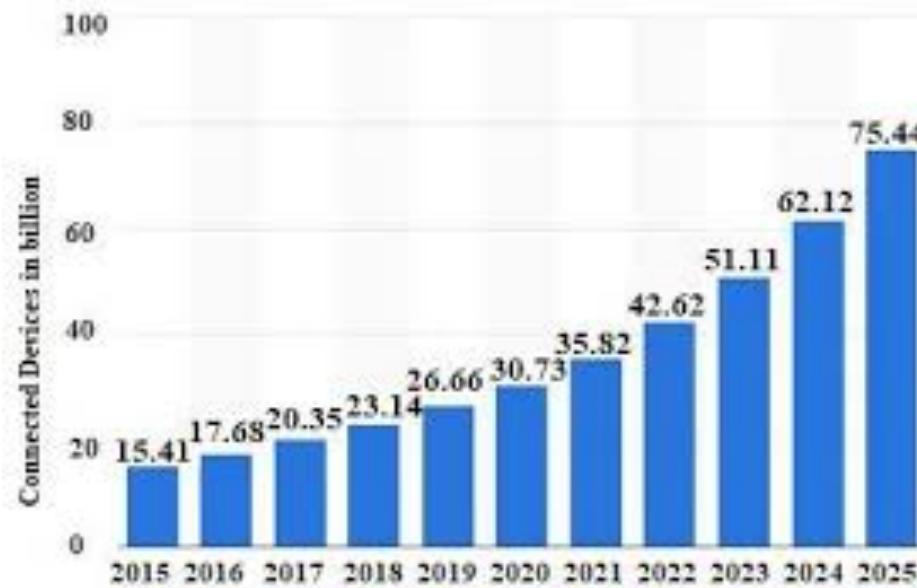
IP Networks

- Growth of Internet hosts during 1993-2010



IoT Devices

- Growth of IoT Devices 2015 to 2025



- The number of devices connected to the internet reached 22 billion worldwide at the end of 2018

IP Networks

- IP was “initially” defined in 1980 in Request for Comments (RFC) 760 and 791, edited by Jon Postel^a
- Postel pointed out in his preface, that RFC 791 is based on six earlier editions of the ARPA^b Internet Protocol, though it is referred to in the RFC as version 4 (IPv4). *v = version*
- RFC 791 states that the Internet Protocol performs two basic functions: addressing and fragmentation.
 - Addressing assures unique addressability of hosts
 - Fragmentation deals with splitting messages into a number of IP packets so that they can be transmitted over networks that have limited packet size constraints, and reassembly of packets at the destination in the proper order.

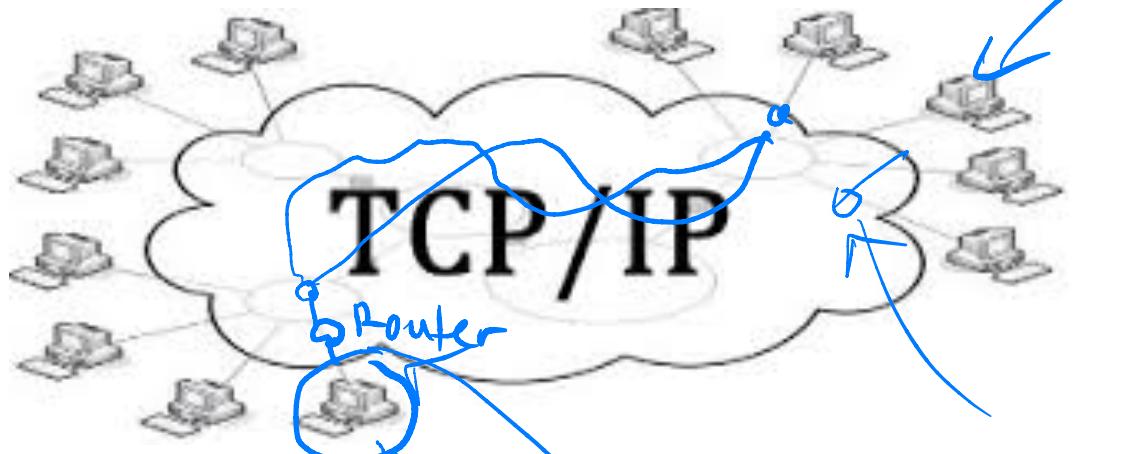
^a1943 to 1998 was also editor of RFC documents; his obituary was published as an RFC by Cerf.

^bAdvanced Research Projects Agency, a U.S. Department of Defense agency

Internetworks

Network

- Hosts are connected via a network cloud.



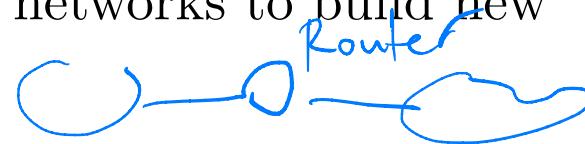
- Its “nerve veins” are based on TCP/IP

IPv4

- Internet Protocol version 4 (IPv4) is the fourth revision in the development of the Internet Protocol (IP) and the first version of the protocol to have been widely deployed.
- Described in IETF (Internet Engineering Task Force) publication RFC 791 (September 1981), replacing an earlier definition (RFC 760, January 1980).
- It is a connectionless protocol for use on packet-switched Link Layer networks
- It operates on a best effort delivery model, in that it does not guarantee delivery, nor does it assure proper sequencing or avoidance of duplicate delivery.
- These aspects, including data integrity, are addressed by an upper layer transport protocol, such as the Transmission Control Protocol (TCP)

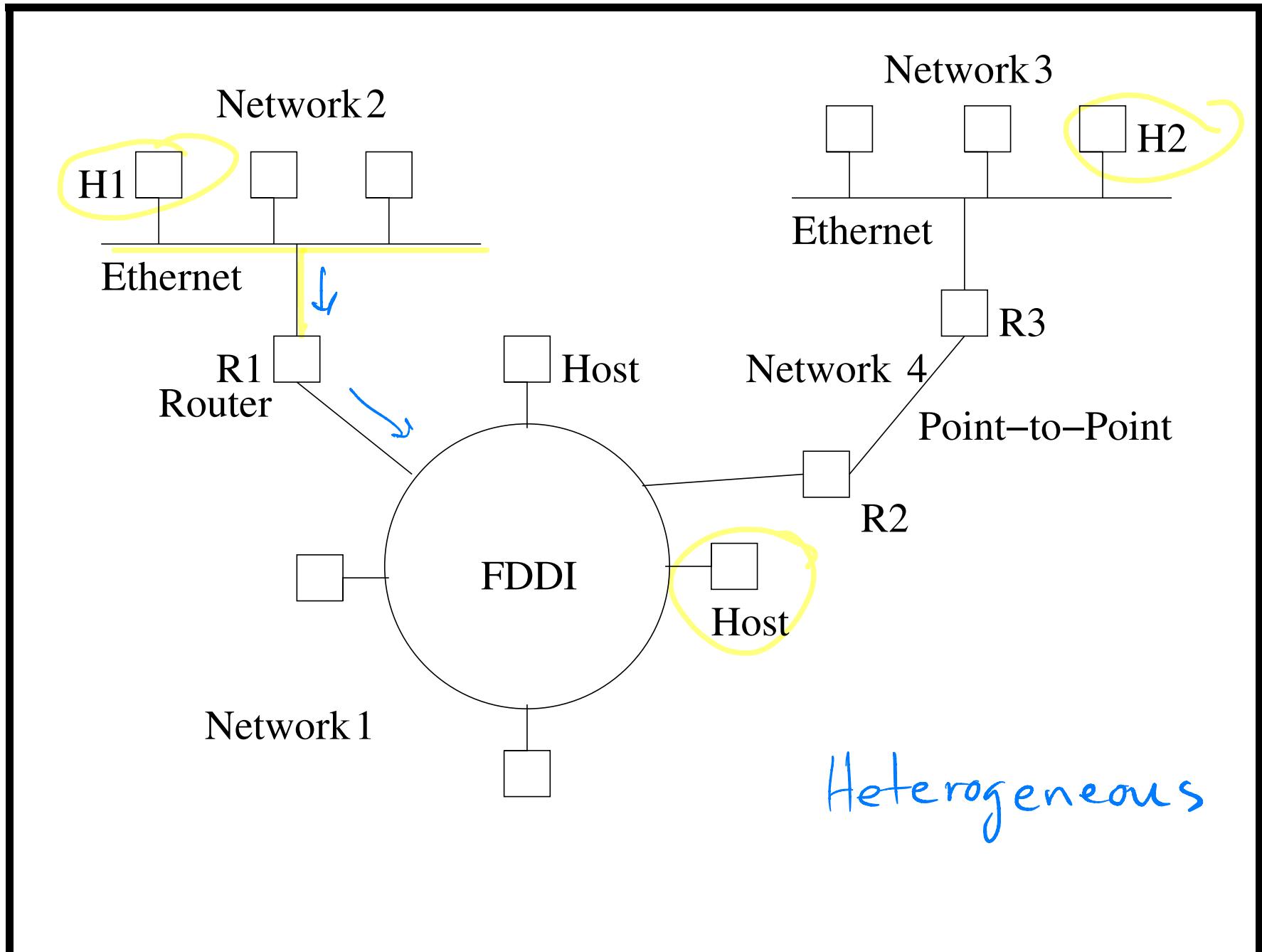
Internetworks

- Important to be able to interconnect networks to build new and larger ones.
- LAN approaches are limited: they do not scale well and they cannot handle heterogeneity.
- An internetwork is an arbitrary collection of networks interconnected to provide host-to-host packet delivery service.
- The networks are interconnected with special nodes called routers, and gateways.
- IP is the main protocol for interconnecting: invented by Kahn and Cerf^a

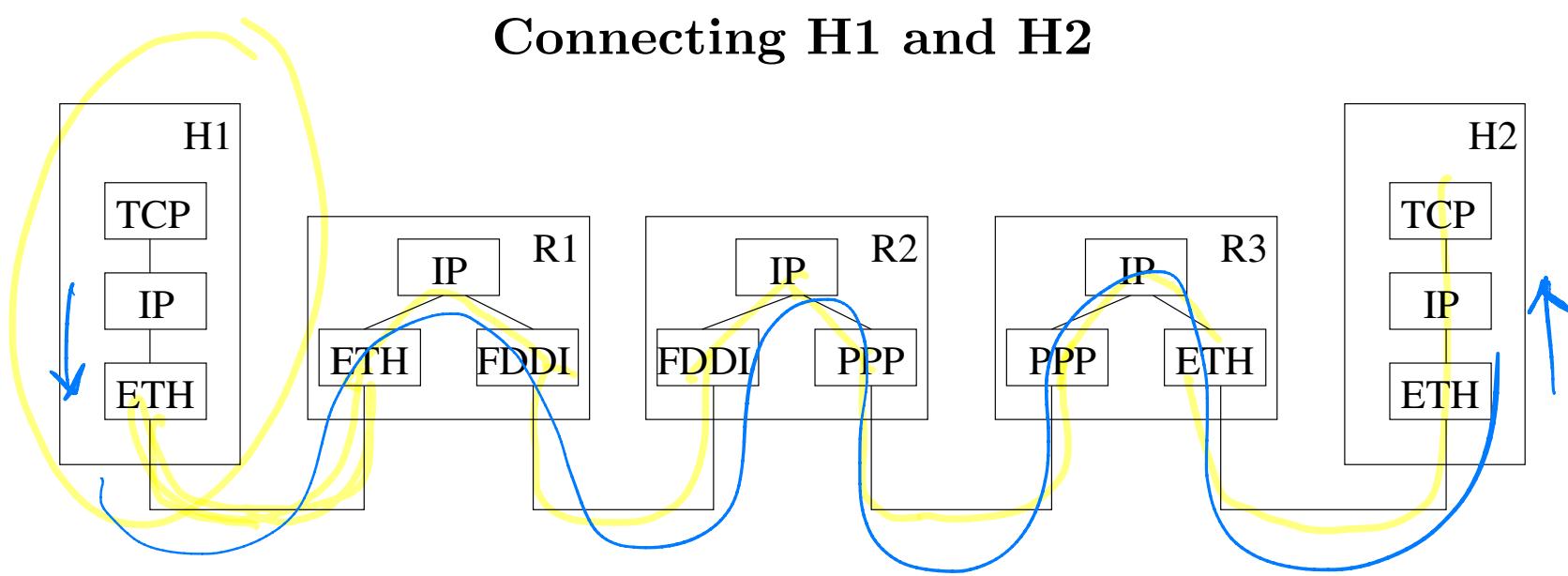


CU Learn

^aDecember 1997: presented with National Medal of Technology by Bill Clinton, “for creating and sustaining development of Internet Protocols and continuing to provide leadership in the emerging industry of internetworking”.



Connecting H1 and H2



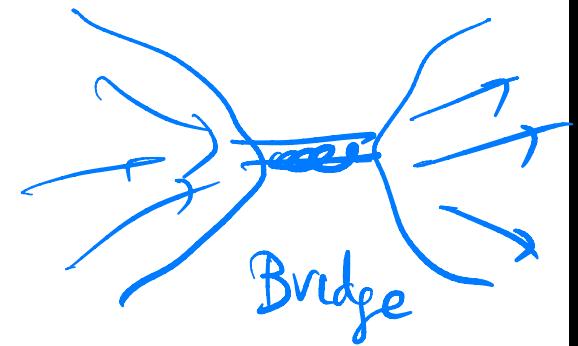
1. Leave H1 from Network 1 Ethernet via Router R1
2. From Router R1 to FDDI token ring.
3. From FDDI token ring to R2.
4. From router R2 to router R3 via Point-to-Point Network.
5. From router R3 to Network 3 Ethernet.

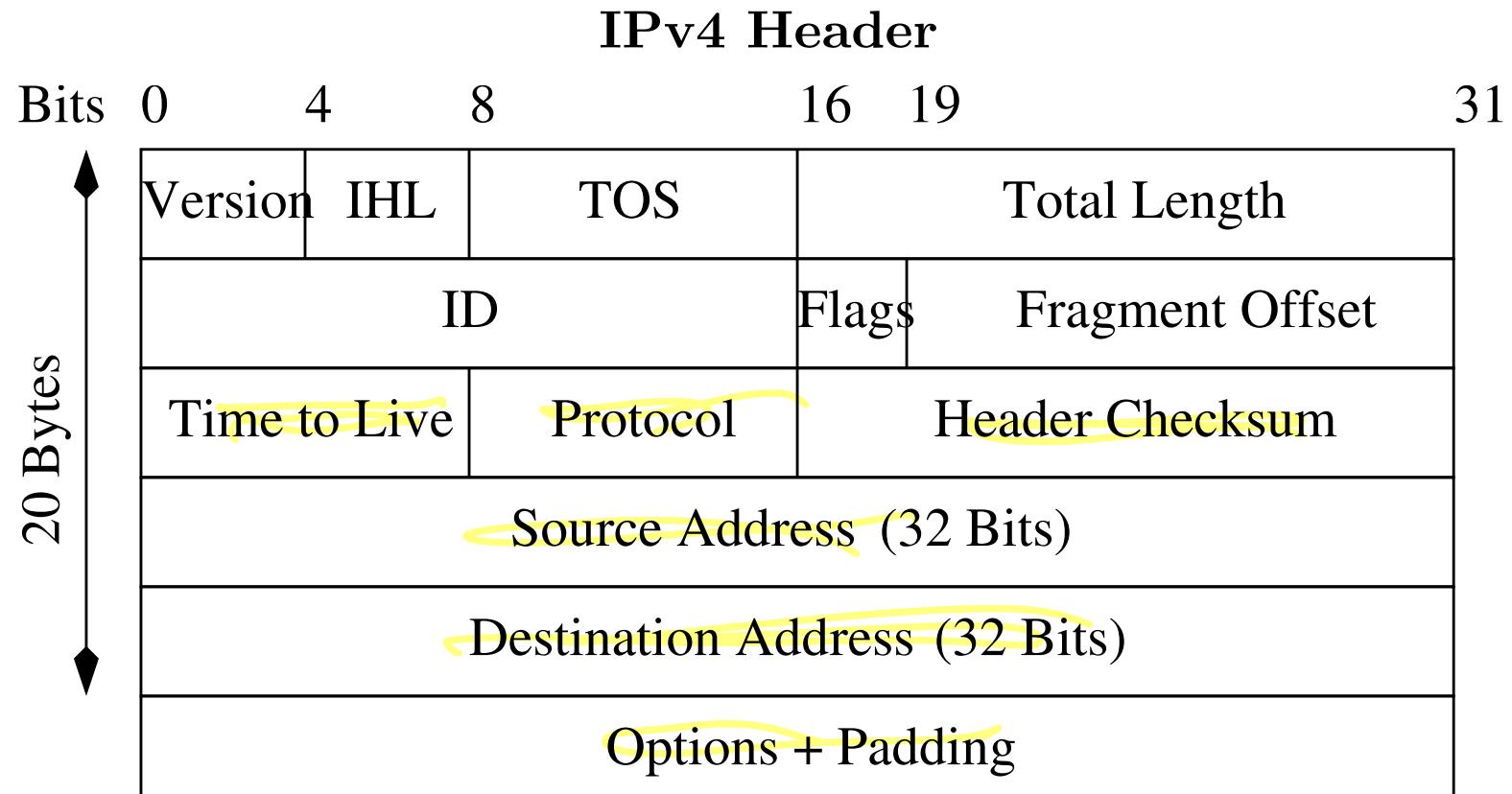
Service Model (Ability to Run over Anything!)

- The philosophy was to give a protocol that will be undemanding enough for just about any existing technology.
- The service model has two parts:
 1. Addressing Scheme.
 2. Datagram delivery connectionless model.
- The addressing scheme provides a way to identify all hosts.
- The datagram delivery scheme is called best effort because it makes no guarantees.
- Datagrams are frames that are sent in a connectionless manner. Of course, they must include sufficient information to enable delivery. **IPv4 Datagrams** consist of a header plus a number of data bytes.

IP

- IP is the vehicle for traffic management.
- IP based internets were designed to support delay insensitive applications.
QoS = Quality of Service
- Today they face the following design requirements.
 - Control congestion
 - Provide low delay
 - Provide high throughput
 - Support QoS
 - Provide fair service
- All these issues fall under the category of traffic management.





IPv4 HEADER

IPv4 Header

- IP with no options is 20 Bytes.
- IHL is header length in 32-bit words.
- TOS (Type Of Service): provides guidance on selecting next hop and relative allocation of router resources.
- TOS subfield: provides route selection, subnetwork service, queuing discipline. These are specified with “certain rules”.
- Precedence Subfield: indicates the degree of urgency from highest level of “Network Control” to lowest level “Routine”. These provide appropriate Queue Service and Congestion Control.
- IPv4 options: Security, Timestamping, Source routing, Route Recording.

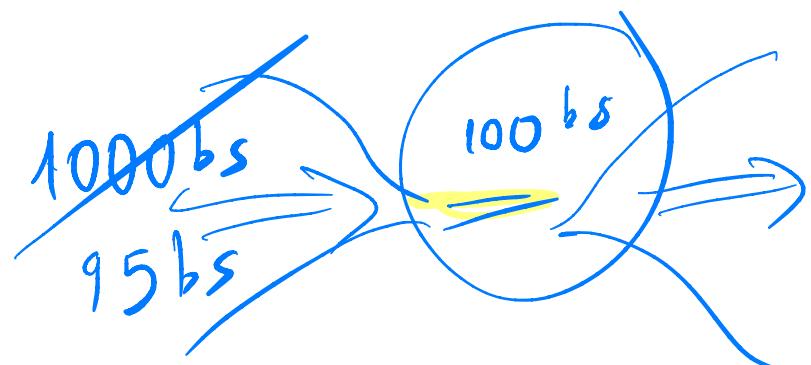
32
2

IPv4 TOS Field

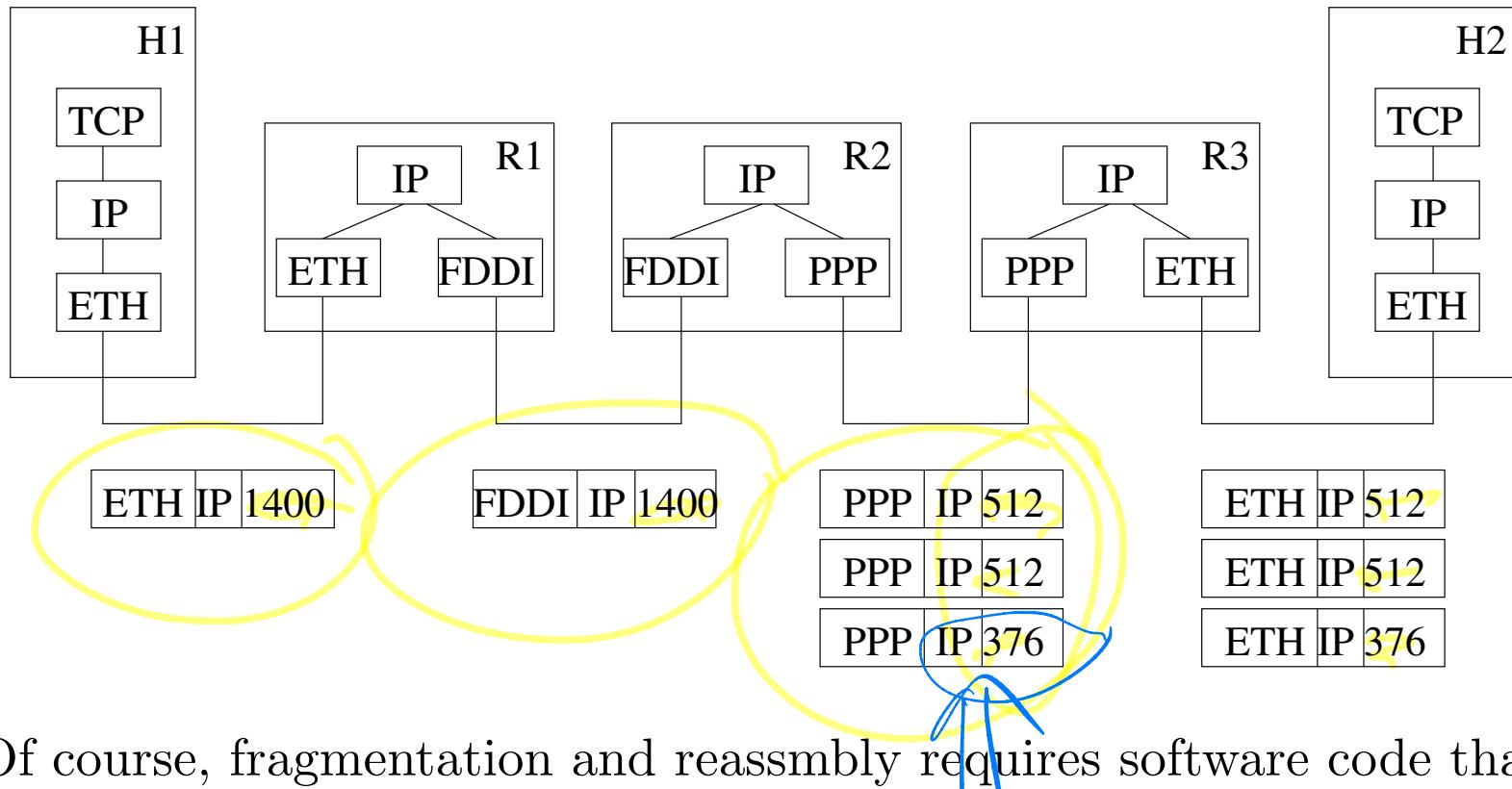
Bits	0	1	2	3	4	5	6	7
	Precedence			TOS			0	

IPv4 TOS-field

Precedence	TOS
111 Network Control	1000 Minimize Delay
110 Internetwork Control	0000 Maximize Throughput
101 Critical	0010 Maximize Reliability
100 Flash Override	0001 Minimize Monetary Cost
011 Flash	0000 Normal Service
010 Immediate	
001 Priority	
000 Routine	



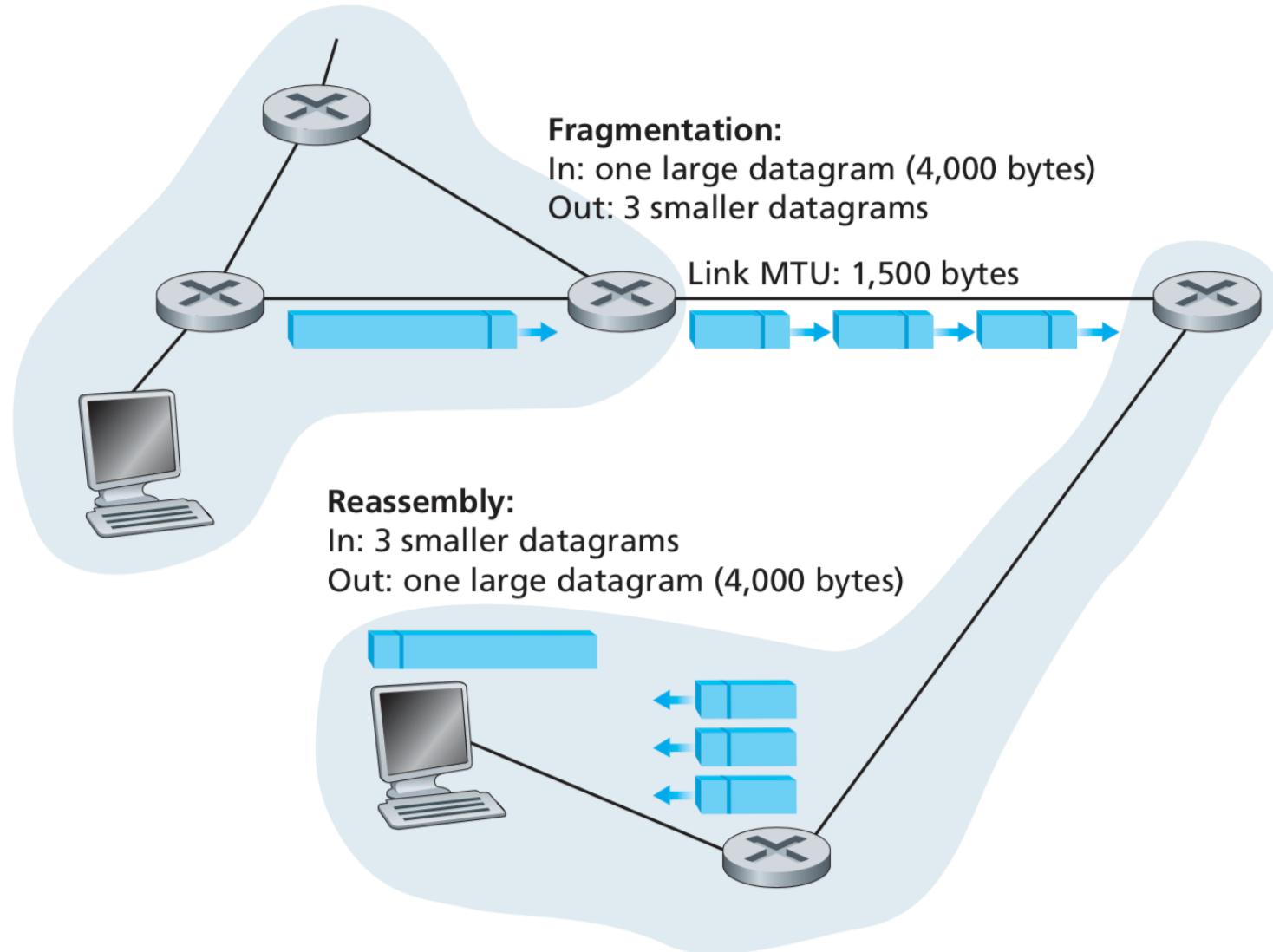
Fragmentation/Reassembly



Of course, fragmentation and reassembly requires software code that performs the required transformations from one packet format to another.

maximum transmission unit

MTU (Max Transmission Unit) Fragmentation/Reassembly



IP Addressing

- In addition to physical addresses (contained in NICs) nodes have 32 bit IP addresses.
- It is a two level hierarchy consisting of the net ID and the Host ID: net ID identifies the network the host is connected.
- All hosts connected to the same network have the same net ID.
- IP addresses look like:

Class	Net-ID	Host-ID
-------	--------	---------

- The lengths of Class, Net-ID, and Host-ID are variable, but the total length is 32 bits.

K2Q5R7 Street #

IP Addressing

- There are five classes of addresses: A, B, C, D, E.

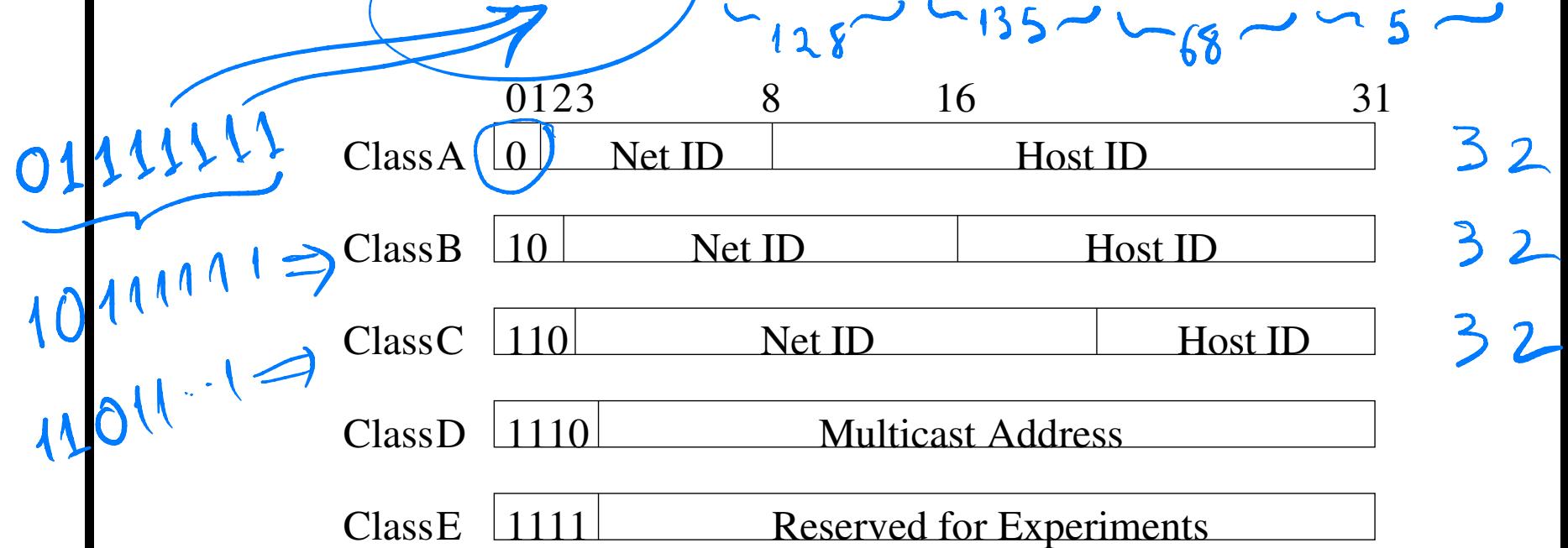
Class	Net ID	Host ID
A	7 bits	24 bits
B	14 bits	16 bits
C	21 bits	8 bits

$$\begin{aligned} & 2^{16} \\ & 1,000 \cdot 2^6 \\ & 64,000 \end{aligned}$$

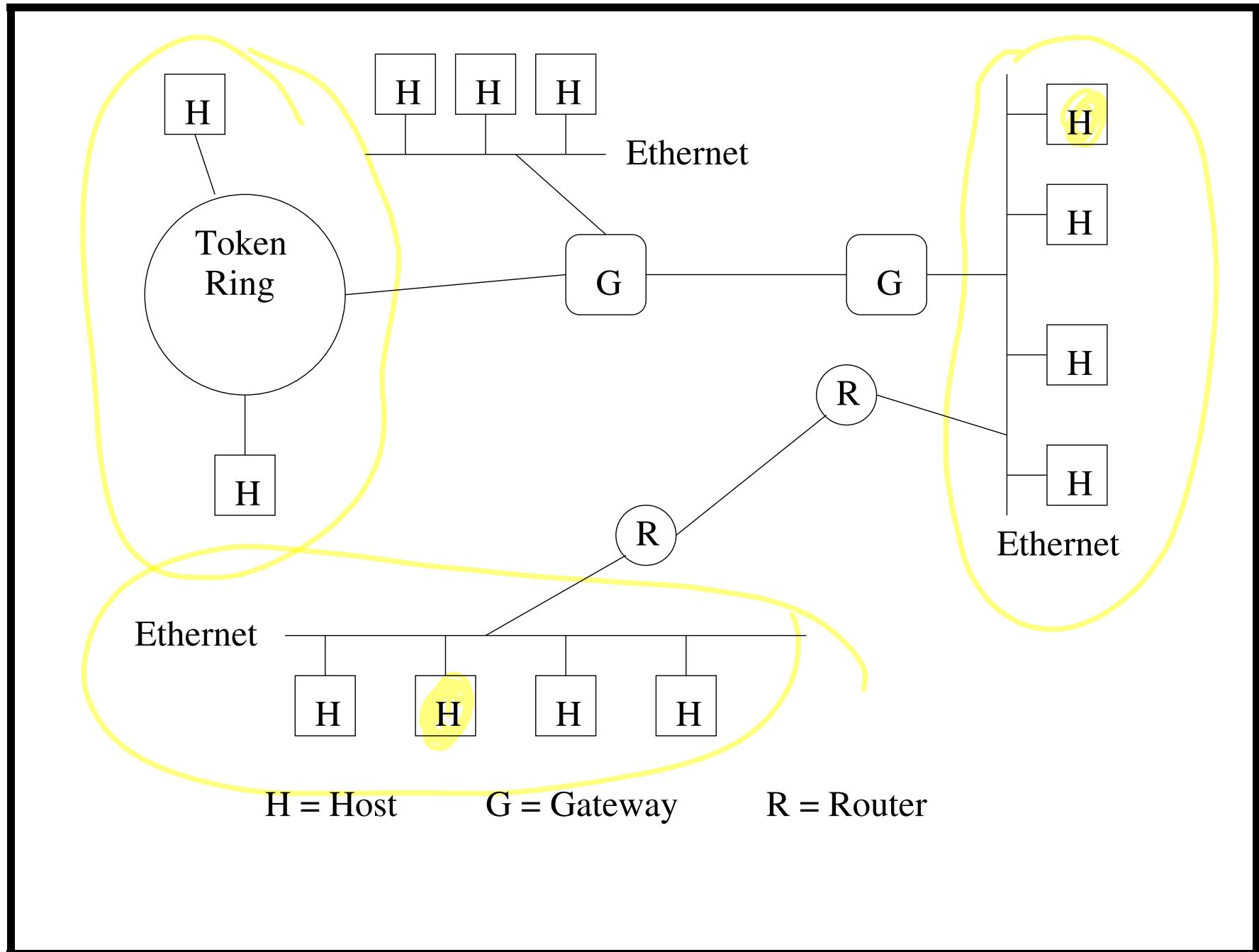
- D is used for multicasting and E for experiments.
- IDs with all 0s or all 1s are used for broadcasting:
 - immediately after booting up a host may not know its ID.
 - So host will transmit packets with all 0s while trying to find out correct ID.

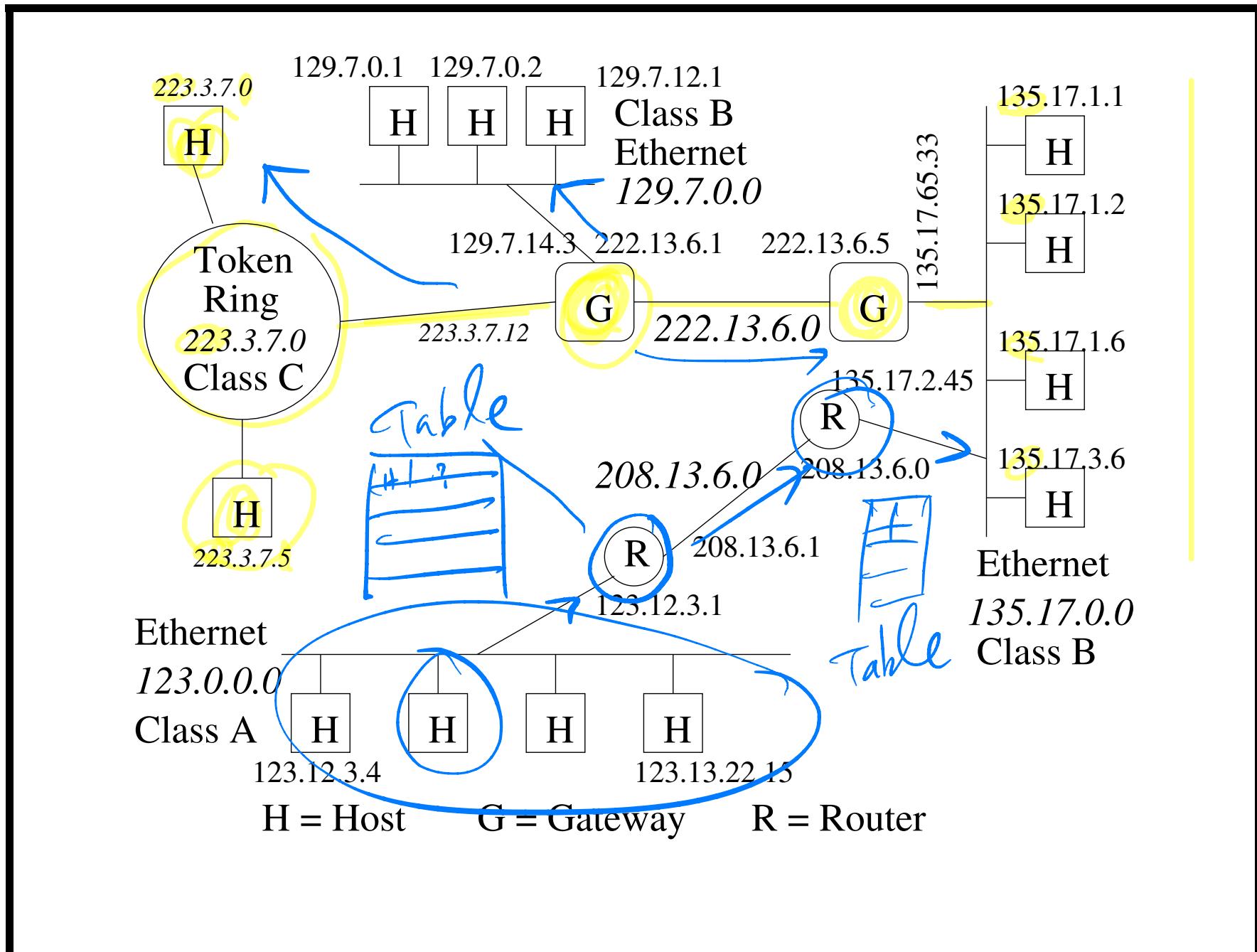
IP Addressing

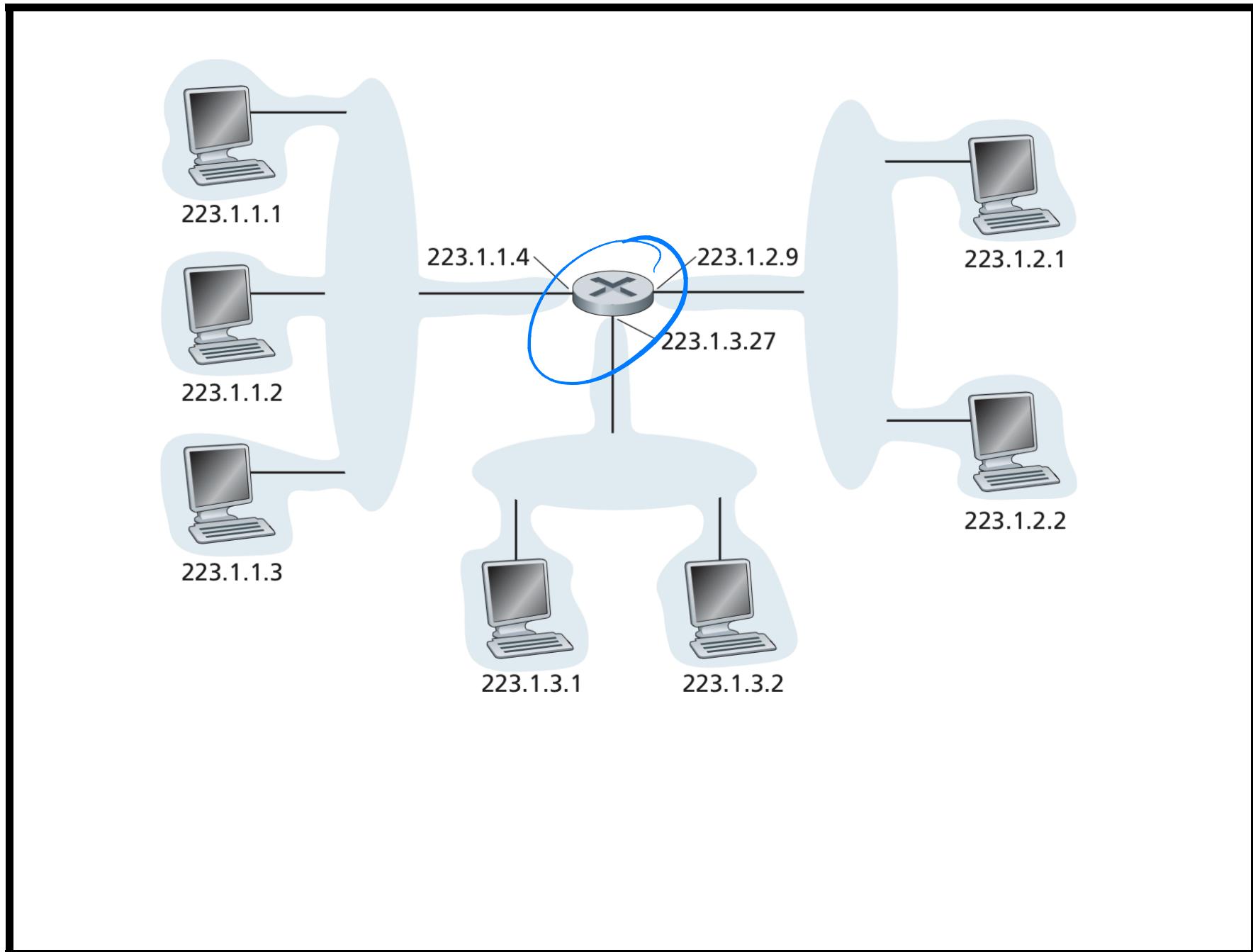
- Addresses broken into four bytes and written as: $X.Y.Z.W$ in decimal: $128.135.68.5 = 10000000.10000111.01000100.00000101$.

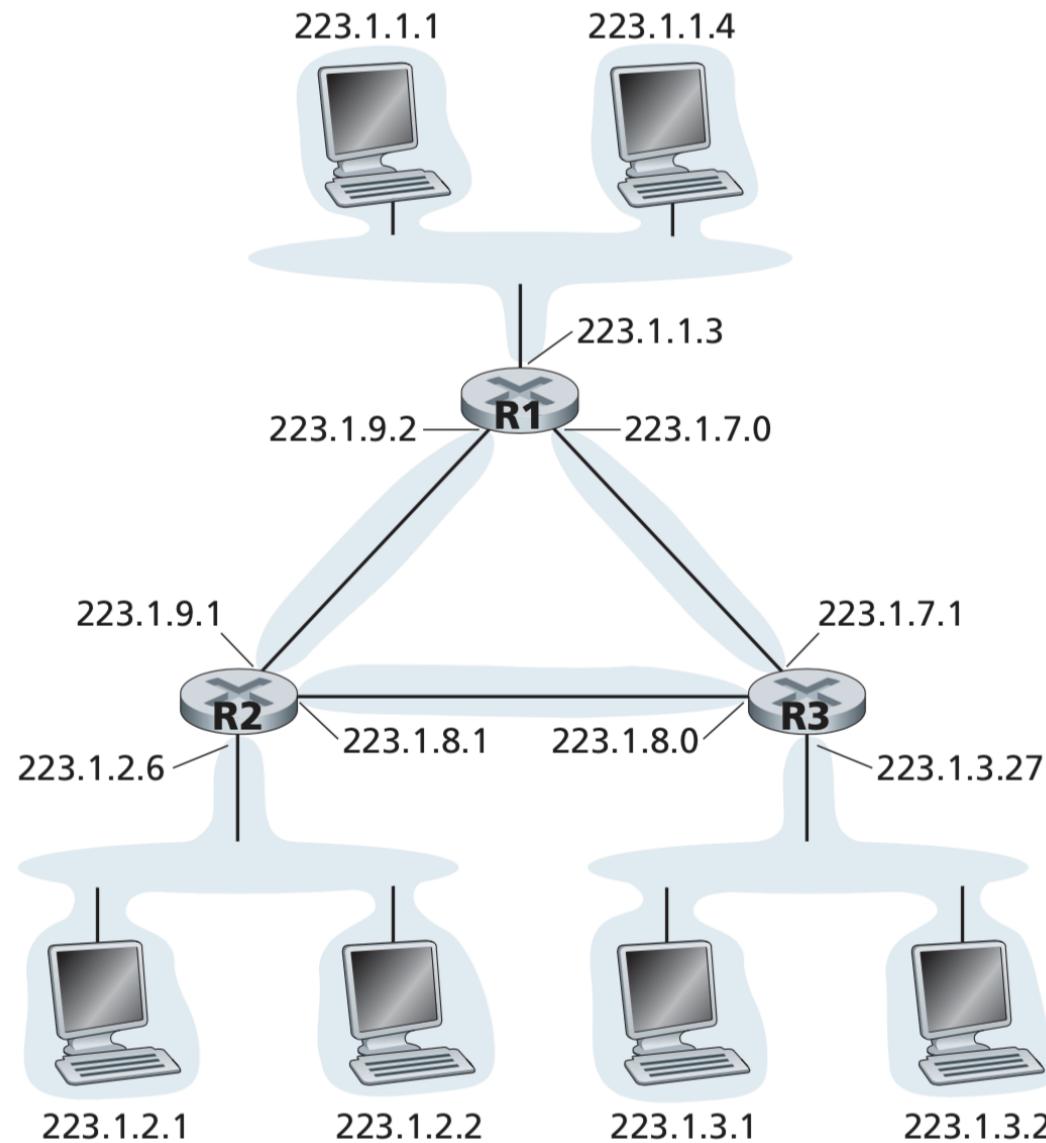


- Class 10, Net ID 00000010000111, Host ID 010001000000101.
- 127.Y.Z.W is a special “loopback” address: packet with this address returns back to host









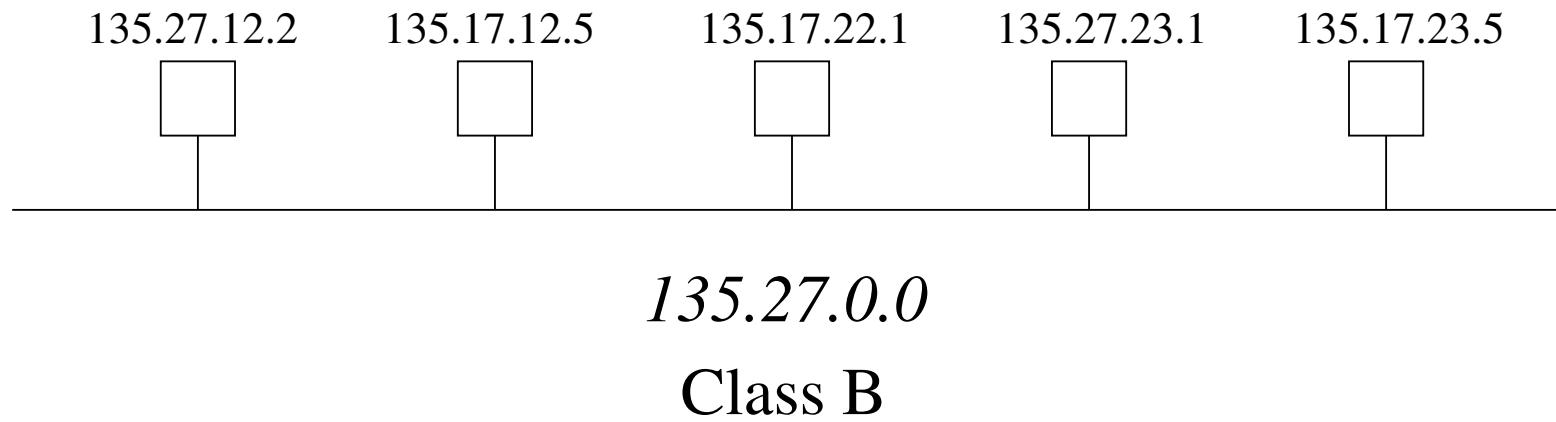
IP Addressing and the Network

- An IP address defines the host's connection to its network.
- A device connected to more than one network (e.g., router) will have more than one internet address.
- In fact, a different address for each network connected to it.
- To reach a host on the internet:
 - first we reach the network using the first portion of the address,
 - second we reach the host itself using the last portion of the address.
- Hence, Classes A, B, C have only two levels!

IP imposes a
hierarchical routing
scheme!

Two-level Hierarchy

A network may have a Class B address.



Because it is a two level hierarchy they cannot be grouped into a “less flat” scheme.

Solution: Subnetting!

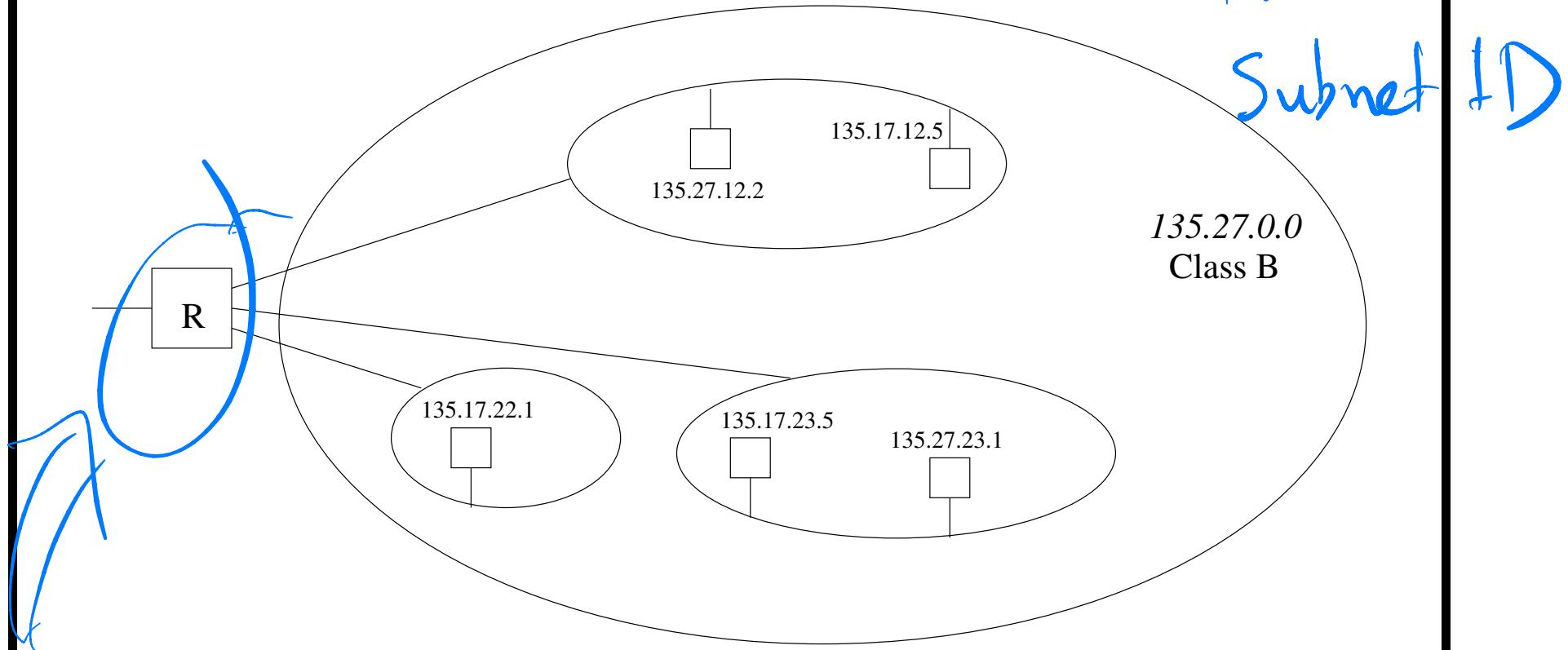
Subnetting

Subnetting

The rest of the internet does not need to be aware of the subnet division! The Router is aware of subnetting!

64,000

Subnet ID

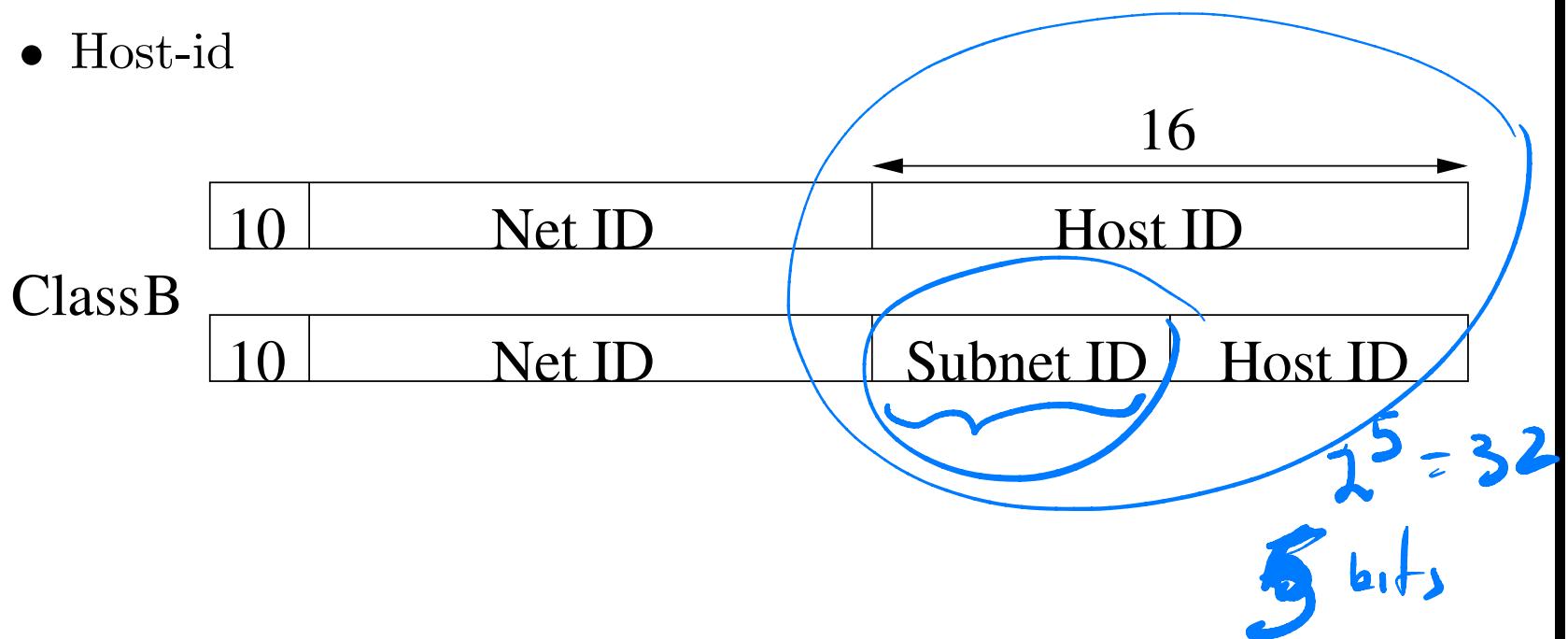


135.27 is the destination net-id, while 22.1 is a host-id.

More Levels

Now we have three levels in the hierarchy:

- Net-id (135.17)
- Subnet-id (12, 22, 23)
- Host-id



More Levels: Example

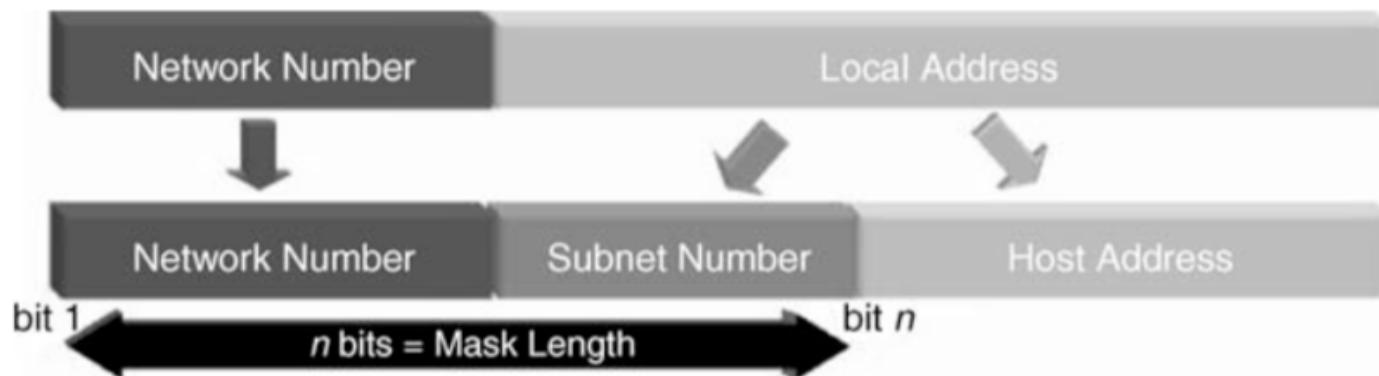
As an example consider an IP address with Net ID 150.100.

If an organization has many LANs each of less than 100 hosts then

- 7 bits suffice to represent hosts within each subnet, and
- remaining 9 bits are left to identify subnets within organization.

IP Subnetting

- A subnetwork or subnet is a logical subdivision of an IP network. The practice of dividing a network into two or more networks is called subnetting.
- Subnetting provides routing boundaries for communications and routing protocol updates.



- Subnetting is facilitated by specifying a network mask along with the network address.

IP Forwarding

IP Datagram Forwarding Algorithm: Source → Destination

To forward IP packets:

- { 1. Participating nodes compare network part of destination address to see if they are connected to same network as destination.
- { 2. If match occurs and destination is in same physical network the packet can be directly delivered.
- { 3. If node is not connected to same physical network all destination datagrams are sent to a router (in general there could be more than one router).
- { 4. The router does this by consulting a forwarding table.
5. There is also a default address used if none of the entries of the forwarding table match the destination address.

In conjunction with TCP

Address Resolution

Address Resolution

- How to re-translate an IP address to an address understood by a local host, e.g. an Ethernet address?
- Each host maintains a table of address pairs:
(IP-address, Physical-address)
Table can be managed either by an administrator or better yet dynamically by the host.
- Translation accomplished by ARP (Address Resolution Protocol).
- ARP enables hosts to build such tables. Moreover, ARP performs updating approximately every 15 minutes.
- ARP performs queries that take advantage of broadcasting capabilities of local networks, like Ethernet.

Mapping IP-Addresses \leftrightarrow Ethernet-Addresses

0	8	16	31		
Hardware Type = 1		ProtocolType = 0x0800			
HLen = 48	PLen = 32	Operation			
SourceHardwareAddress					
SourceHardwareAddress		SourceProtocolAddress			
SourceProtocolAddress		DestHardwareAddress			
DestHardwareAddress					
DestProtocolAddress					

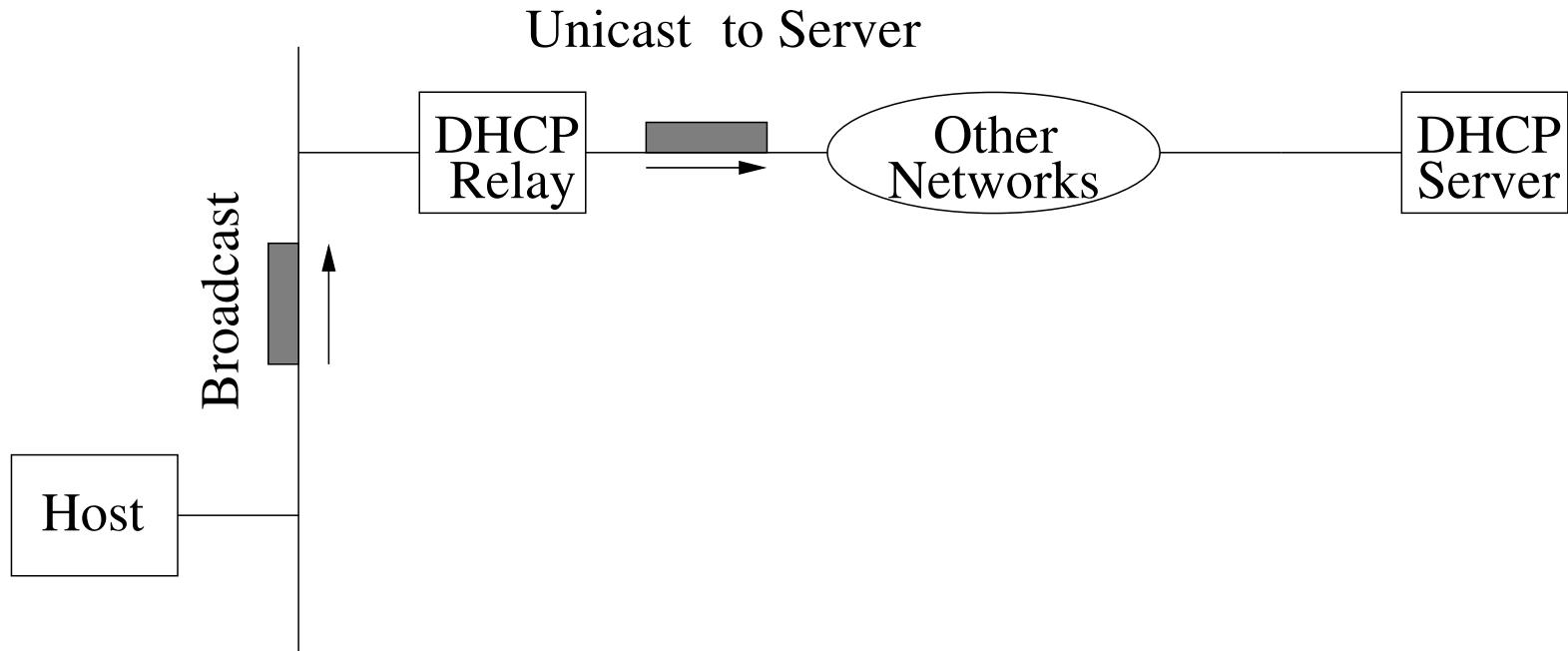
HardwareType is physical network (e.g., Ethernet), ProtocolType is higher level protocol (e.g., IP), HLen, PLen are the Hardware and Protocol address lengths, Operation specifies if this is a request for response.

IPv4: Not Enough Addresses

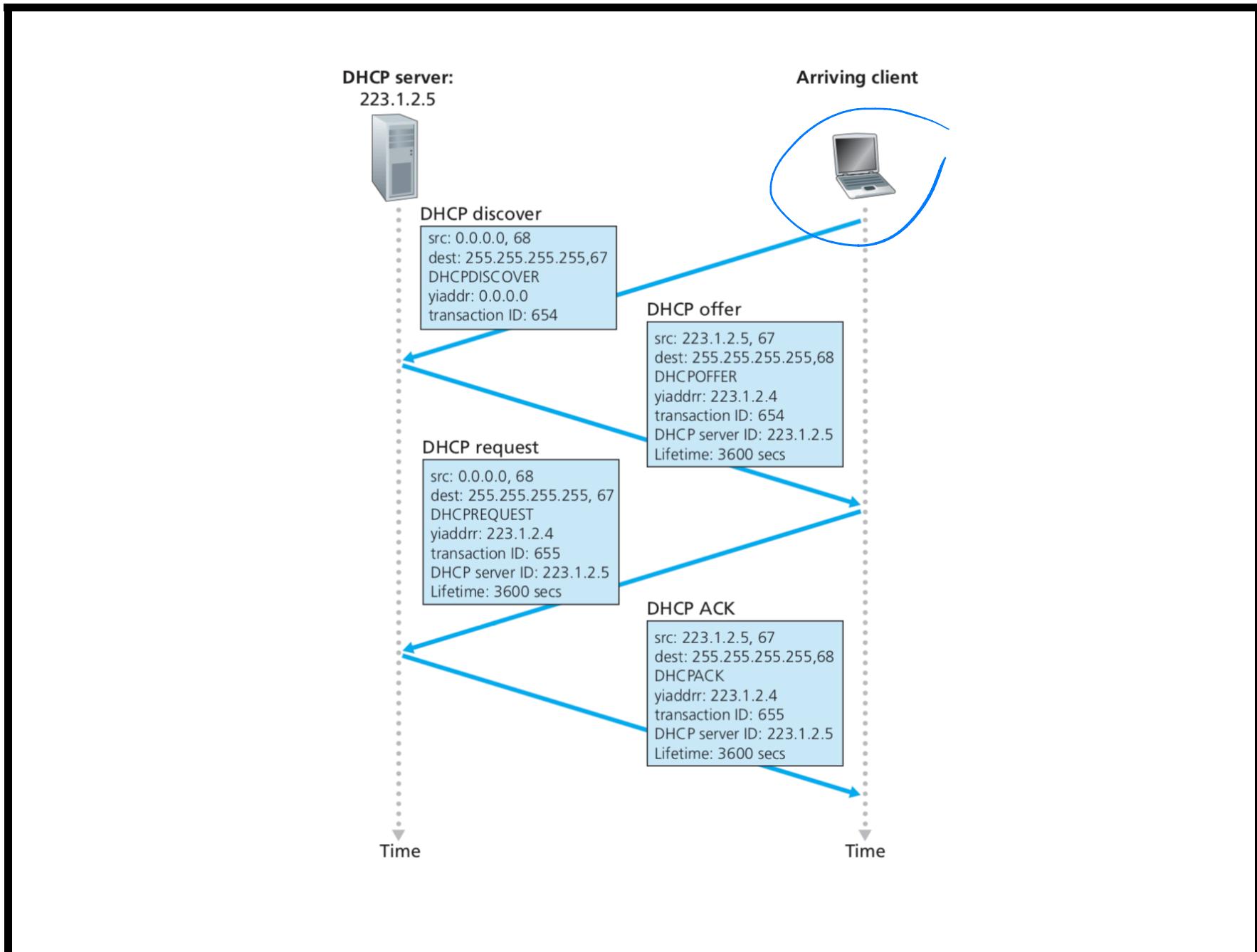
- Unlike Ethernet addresses, IP addresses cannot be configured once and for all into a host at manufacturing.
- Before it can start sending packets the host also needs to know the address of a default router.
- Operating systems enable IP address configuration. Configuring them directly and/or manually can be a lot of work.
- The protocol that automates this process is called DHCP (Dynamic Host Configuration Protocol).
- Every administrative domain (e.g., large company) has at least one DHCP server: DHCP saves the administrator from having to walk around each host.
- Information is stored in a table.

Dynamic Host Configuration Protocol (DHCP)

Newly booted hosts sends a DHCPDiscover message through a DHCP relay.



Used widely by Internet Providers because it maximizes usage of their limited number of IP addresses.



DHCP Packet Format

Operation	HType	HLen	Hops		
Xid					
Secs		Flags			
ciaddr (client IP address)					
yiaddr (your IP address)					
siaddr					
giaddr					
chaddr (client hardware address)					
sname (server name)					
file					
options (defaults, etc)					

Message sent using UDP (runs over IP): it provides a demultiplexing key that says **I am a DHCP packet.**

How DHCP Works

Host broadcasts a DHCP Discover message in its physical network.

Network server(s) respond with a DHCP Offer message.

Host selects one of the offers and broadcasts a DHCP Request message (that includes IP address of server).

Server acknowledges message with a DHCP ACK and assigns IP address for a period of time T with two thresholds T_1, T_2 (usually, $T_1 = T/2$ and $T_2 = 7T/8$).

When T_1 expires host attempts to extend lease by sending DHCP Request to same server. If accepted host also gets new values T', T'_1, T'_2 . If host does not receive DHCP ACK by time T_2 then it broadcasts to any server in the network.

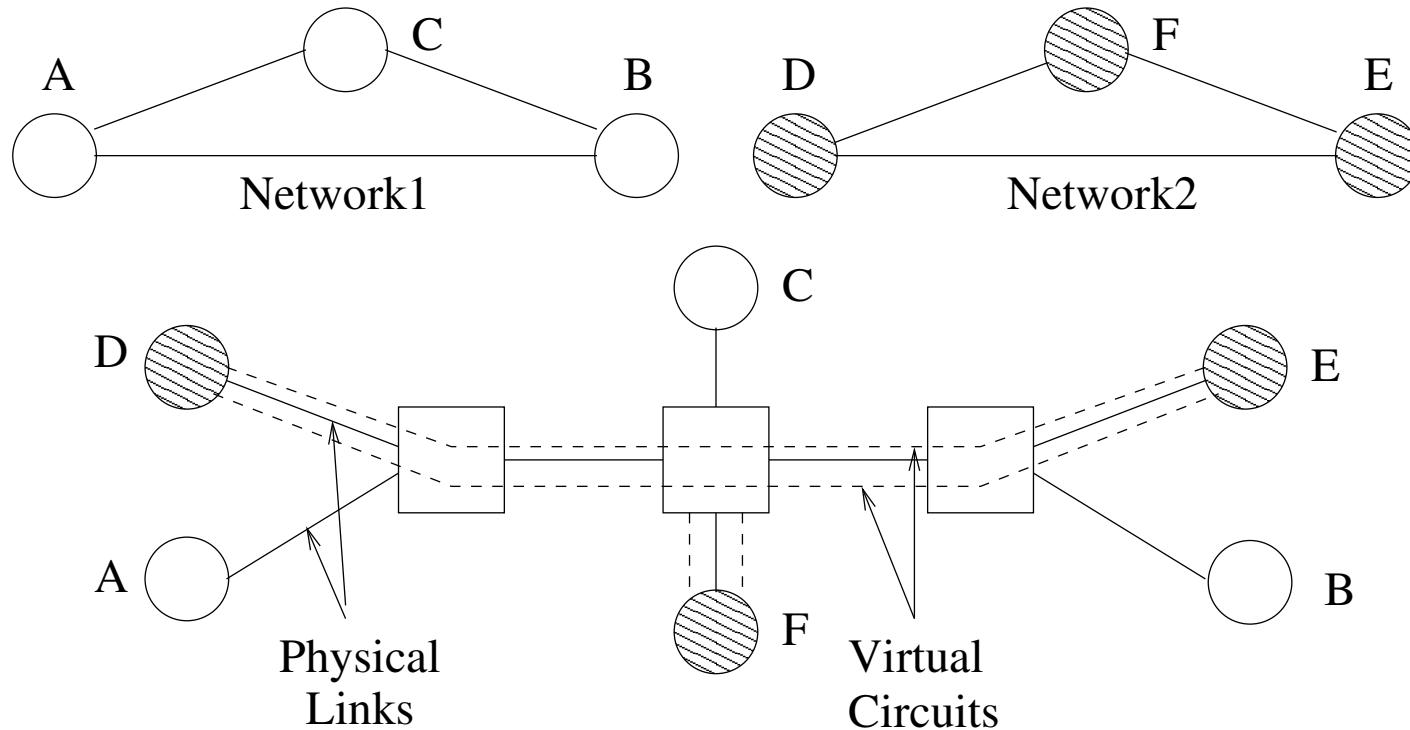
If no ACK is received by time T then host must relinquish old IP address and begin anew.

Reporting Errors

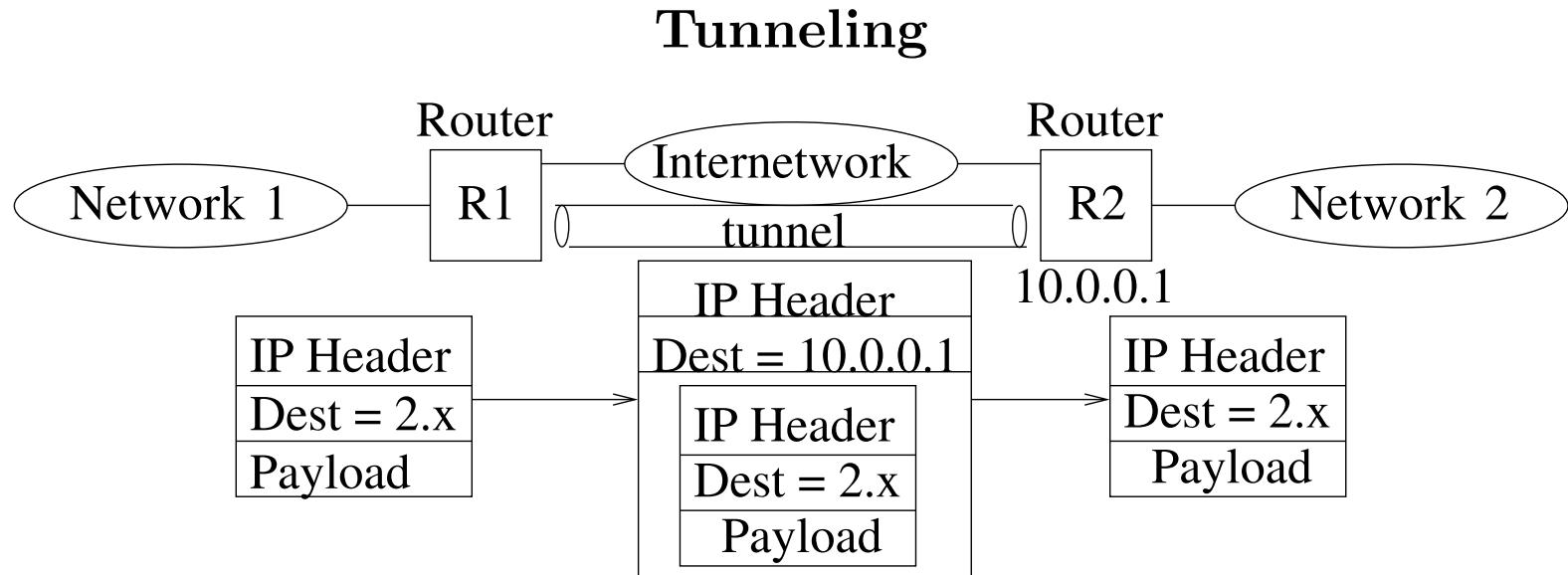
- Whenever a router or host is unable to deliver a message, IP reports errors.
- IP has a companion protocol, called Internet Control Message Protocol (ICMP) that defines error messages, including unreachable destination host, failure of datagram arrival, etc
- Check in your UNIX machine the command **ping!**
- ICMP control messages include: ICMP-Redirect, that tells the source host that there is a better route to destination.

Tunneling

Tunneling Old Virtual Private Networks



Two separate private networks, Network1 and Network2, can share common switches.



Original packet is encapsulated, receives a new IP header and the address of the destination router. It is then transported through the Internetwork. At the destination router it is decapsulated and forwarded to its “local” destination.

IPv4 to IPv6

IPv6

- Surge in demand for IP address space stimulated the IETF^a to define a new version of the Internet Protocol that would provide more addressing capacity to meet then and anticipated future address requirements.
- Every Regional Internet Registry (RIR) had issued notifications to the Internet community at large that IPv4 space availability is limited and will be exhausted within “a few years.”
- The primary objective for version 6 was essentially to redesign version 4 based on the prior 20 years of experience with IPv4.

^athe engineering and standards body of the Internet

IPv6 Wishlist

- IPv4 not sufficient!
 - Support for real time services.
 - Security support.
 - Autoconfiguration.
 - Enhanced routing functionality (e.g., mobility support).
- Also a transition plan IPv4 → IPv6 was necessary.

IPv4 Address Depletion

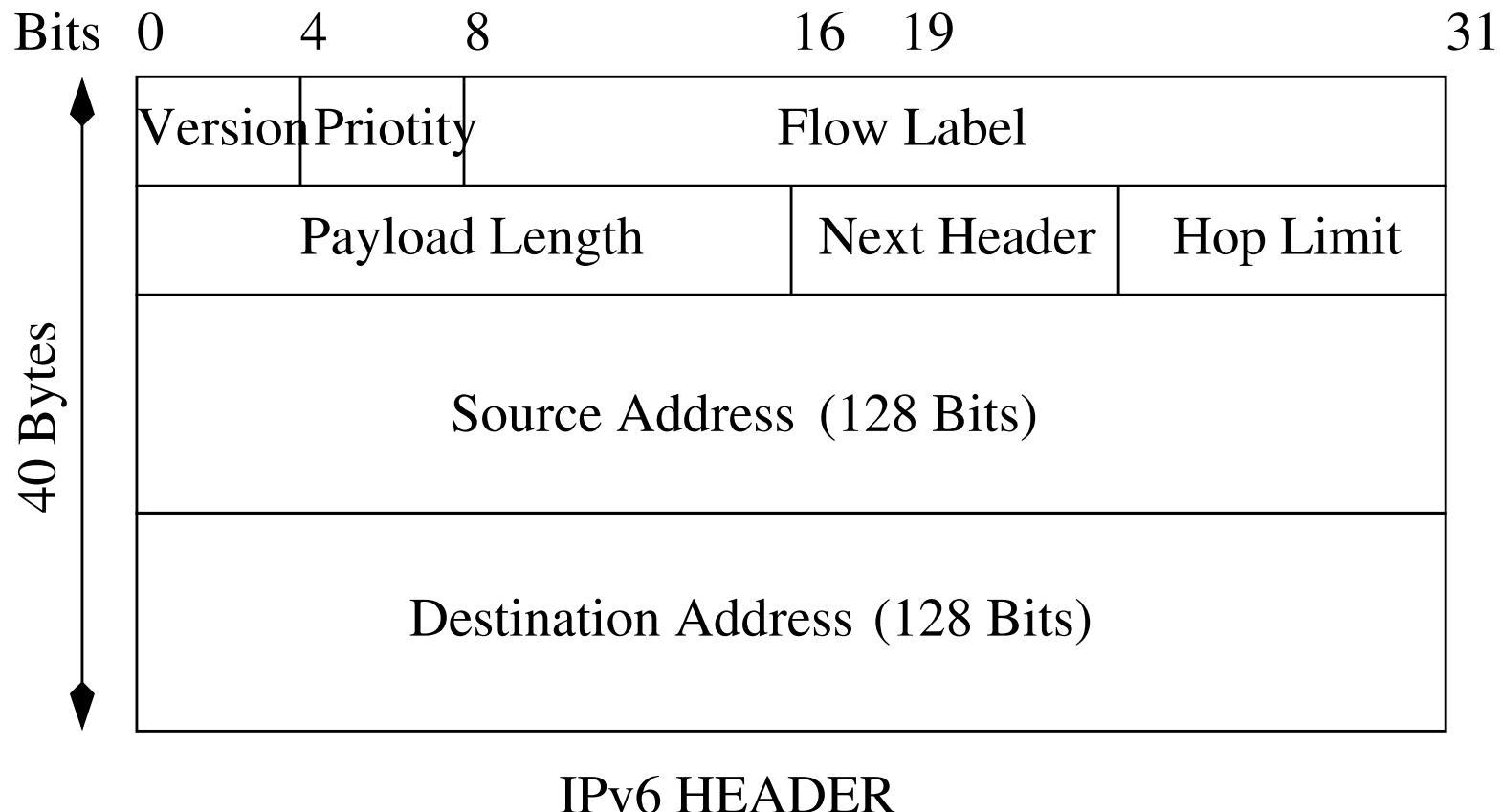
- February 3, 2011: in a ceremony in Miami, the Internet Assigned Numbers Authority (IANA) assigned the last batch of address blocks to the Regional Internet Registries (RIRs): over 80 million combined potential addresses (officially depleting the global pool of completely fresh blocks of addresses).
- APNIC was the first RIR to exhaust its regional pool on 15 April 2011.
- There remains a small amount of address space reserved for the transition to IPv6, which will be allocated in a much more restricted way.

IPv6 Header

- IPv6: designed to accommodate higher speeds. IPv4 uses 32 bit address space. IPv6 uses 128 bit address. IPv4 can address up to 2^{32} (= 4 billion) nodes. IPv6 can address up to $2^{128} = (2^{32})^4$ hosts.
- IPv6 Format: An IPv6 packet has the form: IPv6-header, extension field, . . . , extension header, format level PDU (Protocol Data Unit).
- IPv6 header:
- Priority Field: defines types of traffic.
- Flow labels: e.g. multimedia traffic consists of audio flow, video flow, data flow.

IPv6 Header

Has less fields than IPv4 (for less processing).



IPv6 Addressing (No Classes Being Used)

Prefix	Use	Prefix	Use
0000 0000	Reserved	0001	Unassigned
0000 0001	Unassigned	001	Global Unicast
0000 001	NSAP allocation	1111 1110 10	Link Local Use
0000 010	IPX allocation	1111 1110 11	Site Local Use
0000 011	Unassigned	1111 1111	Multicast
0000 1	Unassigned		

Large address chunks unassigned to allow for future growth.

NSAP used for ISO, **IPX** for Novell. **Link local** and **Site local** enable address construction without concern for global addresses (useful for autoconfigurations), **Multicast** is for multicast addresses, by zero extending with a byte of 0s one assigns IPv4-compatible and IPv4-mapped IPv6 addresses.

IPv6 Packet with Extension Headers

IPv6 treats options as extension headers appearing in certain order.
Thus routers can determine relevant options quickly.

	Size (Bytes)
IPv6 Header	40
Hop-by-Hop Options Header	Variable
Routing Header	Variable
Fragment Header	8
Authentication Header	Variable
Encapsulation Security Header	Variable
Destination Options Header	Variable
TCP Header	20
DATA	Variable

Assigning Addresses

- Three types of addresses: Unicast, Anycast (different interfaces), Multicast (different nodes).
- Hop-by-Hop Options Header: carries optional information that must be examined (like next header, header extension length, options).
- Fragment Header: fragmentation is done only by source nodes not routers. Nodes perform a path discovery algorithm to determine the smaller max transmission unit. With this knowledge source nodes fragments data. Fragment header has several flags and data itself.
- Routing Header: contains a list of one or more intermediate nodes to be visited along the way. Fields include: Next Header, Header Extension Length, Routing Type, etc Destinations
- Options Header: carries optional information.

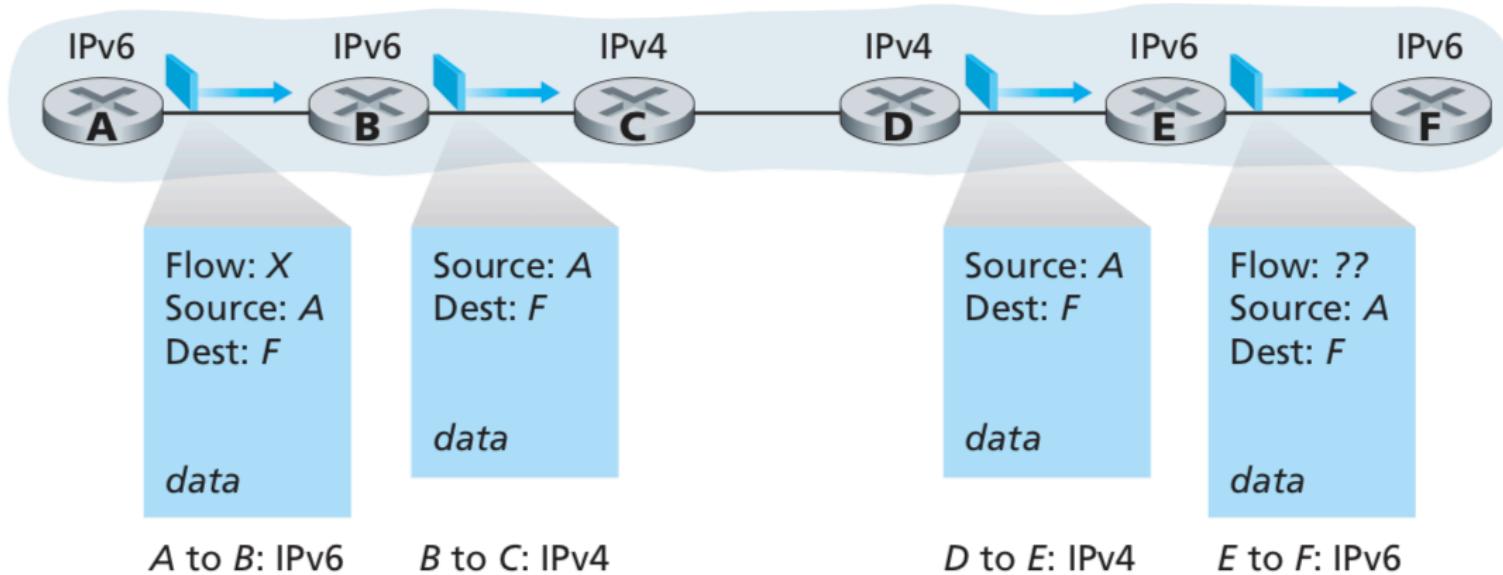
IPv6 Address Notation

- Hexadecimal digits are being used; represented in eight 16-bit blocks. block1:block2:block3:block4:block5:block6:block7:block8
- One set of contiguous 0s can be omitted: block1::block7:block8
- An IPv4-mapped address, like 128.33.87.51 is now written as :: 00FF : 128.33.87.51
- 001 prefix used for global unicast addressing.
- 010 prefix used for IPv6 provider based address. Here, registry IDs are provided as common identifiers, e.g., European, American, etc.

IPv6 Address Notation

- DHCP provides IPv4 autoconfiguration. So does IPv6.. This is done as follows:
 1. obtain correct subnet address prefix (through a router), and
 2. unique interface ID (like Ethernet address).
- IPv6 provides for anycast addresses: selects one of a set of any. Also multicast and security provided.

Transitioning from IPv4 to IPv6

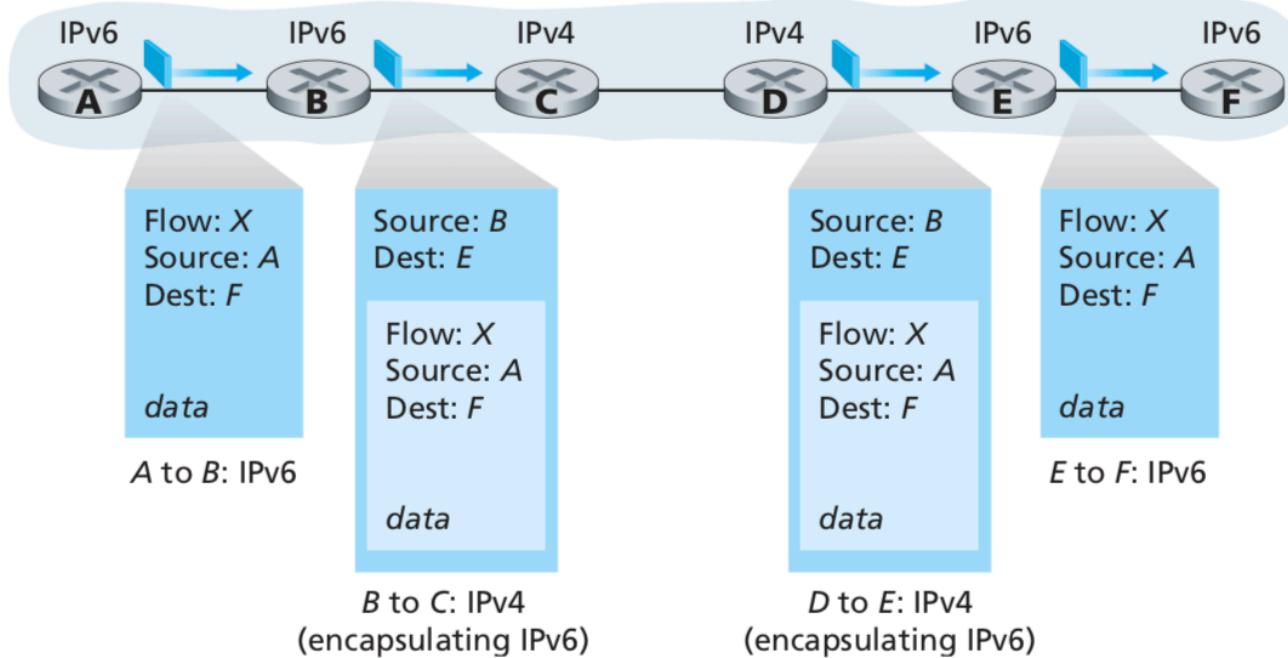


Tunneling in IPv6

Logical view



Physical view



IPv6 Neighbor Discovery

- Enables a node to discover the IPv6 subnet address on which it is connected.
- Enables IPv6 nodes to automatically identify routers on the subnet.
- The discovery process entails each router periodically sending advertisements on each of its configured subnets indicating its IP address, its ability to provide default gateway functionality, its link layer address, the network prefix(es) served on the link including corresponding prefix length and valid address lifetime, as well as other configuration parameters.

IPv6 Deployment (1/4)

- The introduction of Classless Inter-Domain Routing (CIDR) in the Internet routing and IP address allocation methods in 1993 and the extensive use of network address translation (NAT) delayed the inevitable IPv4 address exhaustion, but the final phase of exhaustion started on February 3, 2011.
- Despite a decade long development and implementation history as a Standards Track protocol, general worldwide deployment was still in its infancy: as of October 2011, about 3 % of domain names and 12 % of the networks on the internet have IPv6 protocol support.
- IPv6 has been implemented on all major operating systems in use in commercial, business, and home consumer environments.
- IoT (Internet of Things) is giving a significant boost to IPv6.

IPv6 Deployment (2/4)

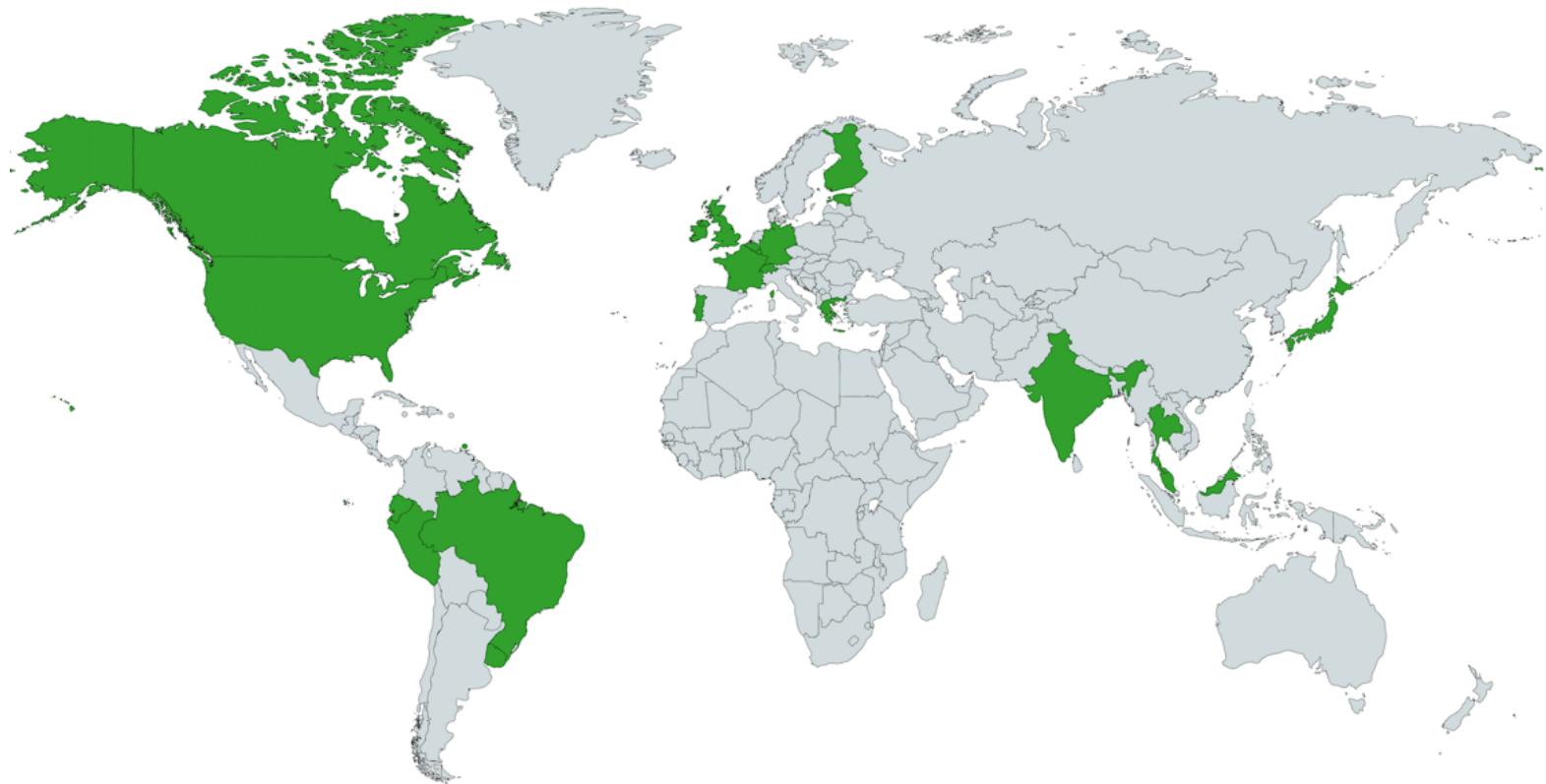
- Since 2008, the domain name system can be used in IPv6 at major web sites like Google, although sometimes with extra configuration.
- IPv6 was first used in a major world event during the 2008 Summer Olympic Games, the largest showcase of IPv6 technology since the inception of IPv6.
- Countries like China or the Federal U.S. Government are also starting to require support for IPv6 on their equipment.
- Finally, modern cellular telephone specifications mandate IPv6 operation and deprecate IPv4 as an optional capability.

IPv6 Deployment (3/4)

- IPv6 deployment continues to increase around the world.
- As of 2018
 - Over 25% of all Internet-connected networks advertise IPv6 connectivity.
 - 49 countries deliver more than 5% of traffic over IPv6, with new countries joining all the time.
 - In 24 countries IPv6 traffic exceeds 15%.

IPv6 Deployment (4/4)

- Countries with Pv6 Deployment greater than 15% as of 2018.^a



^aSource: Internet Society

References

- T. Rooney, Rooney-IP Address Management Principles and Practice, IEEE, 2011.
- RFC 760, DoD standard Internet Protocol, January 1980
- RFC 761, DoD standard Transmission Control Protocol, 1980.

Appendix

Masking

IP Masking

- It is called a subnet masking because it is used to identify network address of an IP address by performing a bitwise AND operation on the netmask.
- A Subnet mask is a 32-bit number that masks an IP address, and divides the IP address into network- and host-address.
- Class A network has a mask length 8, a class B 16, and C 24.
- By essentially extending the length of the network number that routers need to examine in each packet, a larger number of networks can be supported, and address space can be allocated more flexibly.
- For IPv4, a network may also be characterized by its subnet mask or netmask, which is the bitmask that when applied by a bitwise AND operation to any IP address in the network, yields the routing prefix.

IP Addressing and Subnets: Masking

- A subnet mask separates the IP address into the network and host addresses (*<network><host>*).
- Subnetting further divides the host part of an IP address into a subnet and host address (*<network><subnet><host>*) if additional subnetwork is needed.
- Masking extracts the address of the physical network from an IP address.
- If there is no subnet it merely extracts the network address from the IP address.
- If there is subnet division then it extracts the subnet address from the IP address.
- Masking is also used to “hide” addresses.

Global Connectivity and Scalability: Subnets

It is a mistake to assign one network number per physical network.

Address assignment inefficiencies arise:

1. A network with three nodes may be using an entire class C network address (thereby wasting 252 useful addresses).
2. A network with slightly more than 255 hosts is using a class address thereby wasting 64,000 addresses.
3. The more forwarding numbers exist the larger the forwarding tables.

Subnetting

- **Subnetting** takes a single IP network address and allocates it to several physical networks referred to as subnets, e.g., a University Campus, Company, etc.
- Subnets should be physically close to each other.
- Knowing the addresses of a few entry gateways should be enough.

Subnet Masks

1. The mechanism allowing a network number to be shared among multiple networks is called **subnet masking**.
2. This gives rise to the **subnet number**: all hosts on the same network have the same subnet number.
3. Hosts on different physical networks share single network #.
4. Subnet masks introduce another level of hierarchy into IP-addressing.
5. Subnet masks written down like IP addresses, e.g.
255.255.255.3

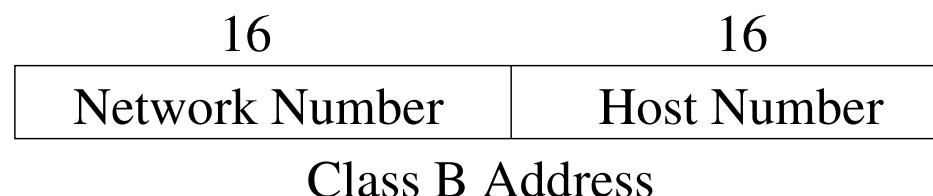
Subnet Masks

To share a single class B address use a subnet mask, say

$255.255.255.0 = 11111111.11111111.11111111.00000000$

I.e., lowest 8 bits are the host numbers.

Address now has three parts: network part, subnet part, and host part.



24 1s	8 0s
-------	------

Subnet Mask: 255.255.255.0



Subnet Masks (Example)

Suppose a router is given a destination address D and a pair (I, M) where I is an IP address and M a mask.

Router verifies the condition $I = D \& M$ (bitwise AND) to test whether or not D belongs to the same subnet.

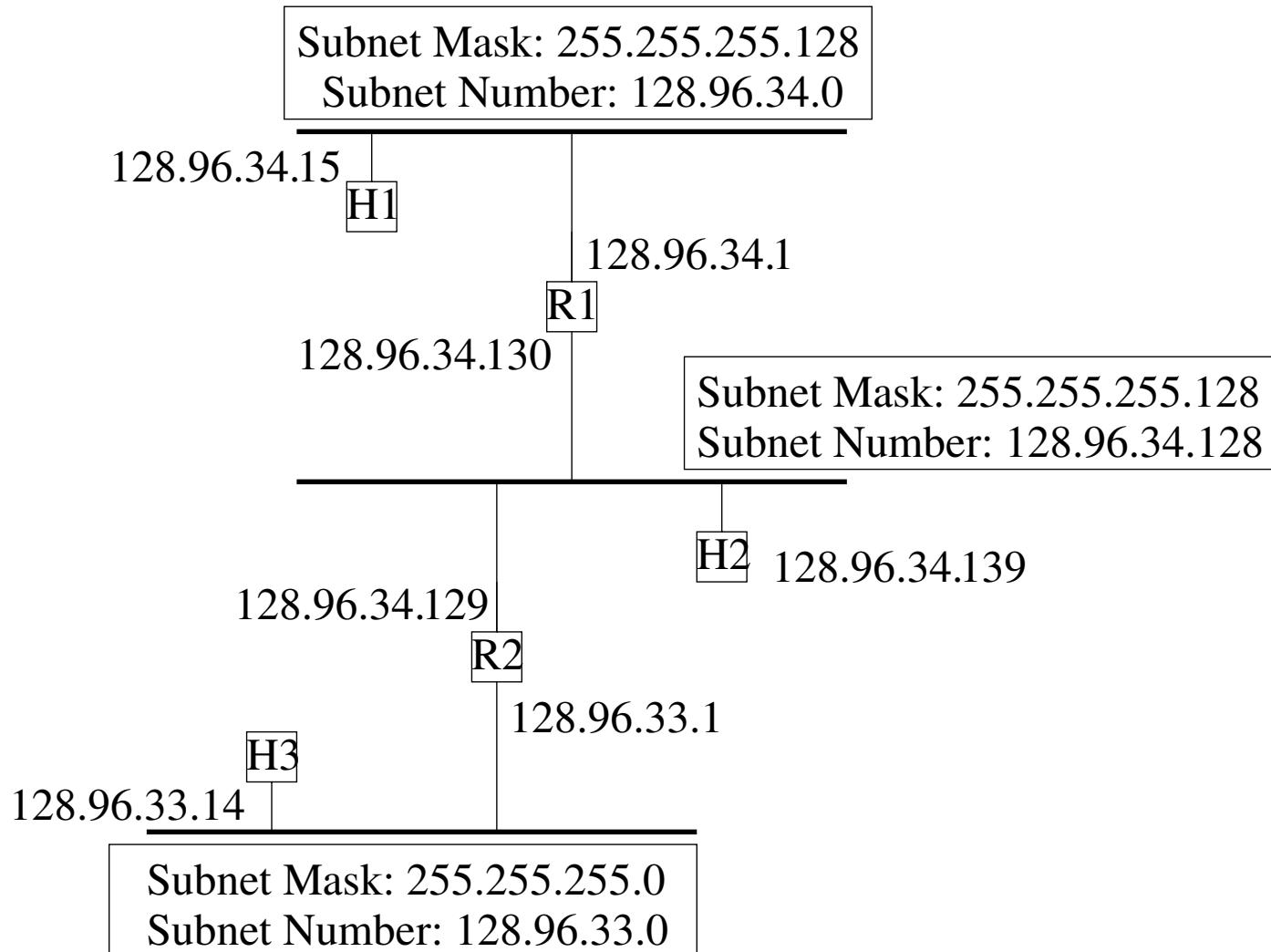
For example, in

$I = 128.10.0.0 = 10000000 \quad 00001010 \quad 00000000 \quad 00000000$

$M = 255.255.0.0 = 11111111 \quad 11111111 \quad 00000000 \quad 00000000$

$D = 128.10.2.3 = 10000000 \quad 00001010 \quad 00000010 \quad 00000011$

the mask M is used as follows: it “matches” the first sixteen bits of the IP address I (ignoring the rest).



Routers R1 and R2 have a different subnet number for each subnet

Subnet Masks and Subnet Number

1. Hosts are configured with an address and the subnet mask.
2. The bitwise AND of these two numbers defines subnet number of the given host as well as all the hosts in the same subnet.

Host	H1	H2
Subnet Number	128.96.34.15	128.96.34.139
Subnet Mask	255.255.255.128	255.255.255.128
BIT-WISE AND	128.96.34.0	128.96.34.128

H1 forwards to H2: H1 calculates AND of H2's subnet address (128.96.34.139) with subnet mask (255.255.255.128). If result is equal to H1's Subnet Number (128.96.34.128) then it is delivered to NextHop for H2 of its forwarding table. If it is not equal to H1's Subnet Number then packet is forwarded to H1's default router.

Masking: Example

If IP address 150.100.12.176 arrives from outside use a binary AND between IP address and mask 255.255.255.125 to determine subnet.

IP address 10010110.01100100.00001100.10110000

mask 11111111.11111111.11111111.10000000

AND 10010110.01100100.00001100.10000000

IP-address	Mask	Network/Subnetwork-address
141.14.3.22	255.255.0.0	141.14.0.0 (without subnetting)
141.14.3.22	255.255.255.0	141.14.3.0 (with subnetting)

END-TO-END CONNECTIVITY (TCP)

Outline

- UDP
- TCP
 - Headers
 - Connection Establishment
 - Transition Diagram
 - Data Transfer
 - Congestion Control
- Equilibria: A Model for TCP

Algorithm

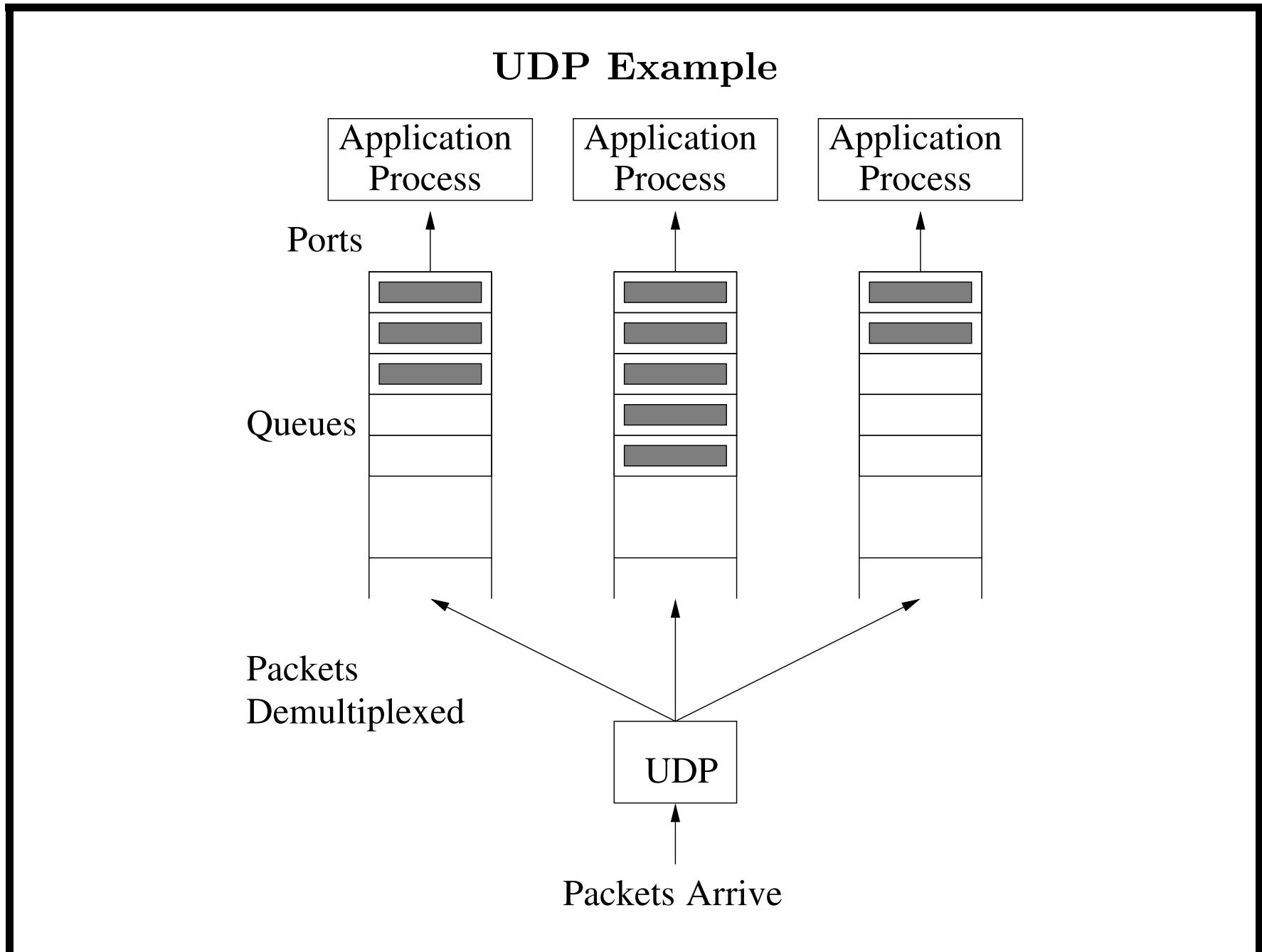
Flow : How many packets
should I send !

UDP

UDP

- UDP^a (User Datagram Protocol) is an alternative communications protocol to Transmission Control Protocol (TCP) 
- It is used primarily for establishing low-latency and loss-tolerating connections between applications on the internet.
- Computer applications can send messages, in this case referred to as datagrams, to other hosts on an Internet Protocol (IP) network.
- It has no handshaking dialogues, and thus exposes the user's program to any unreliability of the underlying network; there is no guarantee of delivery, ordering, or duplicate protection.

^aUDP was designed by David P. Reed in 1980 and formally defined in RFC 768.



UDP (A Simple Demultiplexer)

- Many processes are running on a given host.
- UDP extends the host-to-host delivery service into a process-to-process communication service and allows multiple application processes on each host to share the network.
- In UDP processes are identified by an abstract locator (also called port or mailbox).
- TCP is a connection-oriented protocol and UDP is a connection-less protocol. TCP establishes a connection between a sender and receiver before data can be sent. UDP does not establish a connection before sending data.

Ports

- Port numbers-mapping is published periodically in an RFC, and in UNIX they are available in file: /etc/services.
- Ports are implemented by message queues. Arriving messages are appended in the queue. No flow control is available.
- An application process removes messages from the queue, if there are any, otherwise the process blocks.
- How does one process learn the port number of another process?

UDP and Processes

- Servers accept messages in a well-known port:
 - Client process initiates message exchange with a server process: client's message has its port number in the message header and server can reply to it.
- Source process sends message to port, and application process receives message from port.

0	16	31
SourcePort	DestPort	
Checksum	Length	
Data		

- At most $2^{16} \approx 64K$ ports are possible. Not enough, but they need to be interpreted only in a single host and not across entire internet.

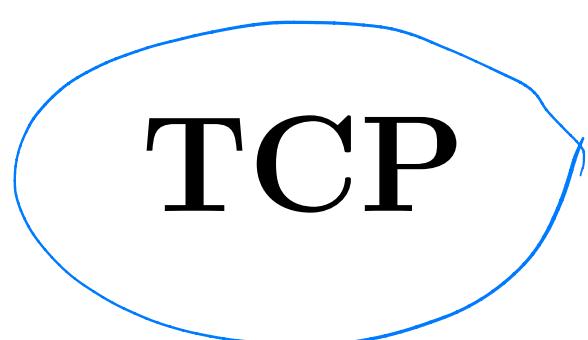
UDP and Processes

- UDP is suitable for purposes where error checking and correction are either not necessary or are performed in the application;
- UDP avoids the overhead of such processing in the protocol stack.
- Time-sensitive applications often use UDP because dropping packets is preferable to waiting for packets delayed due to retransmission, which may not be an option in a real-time system.

UDP Usage

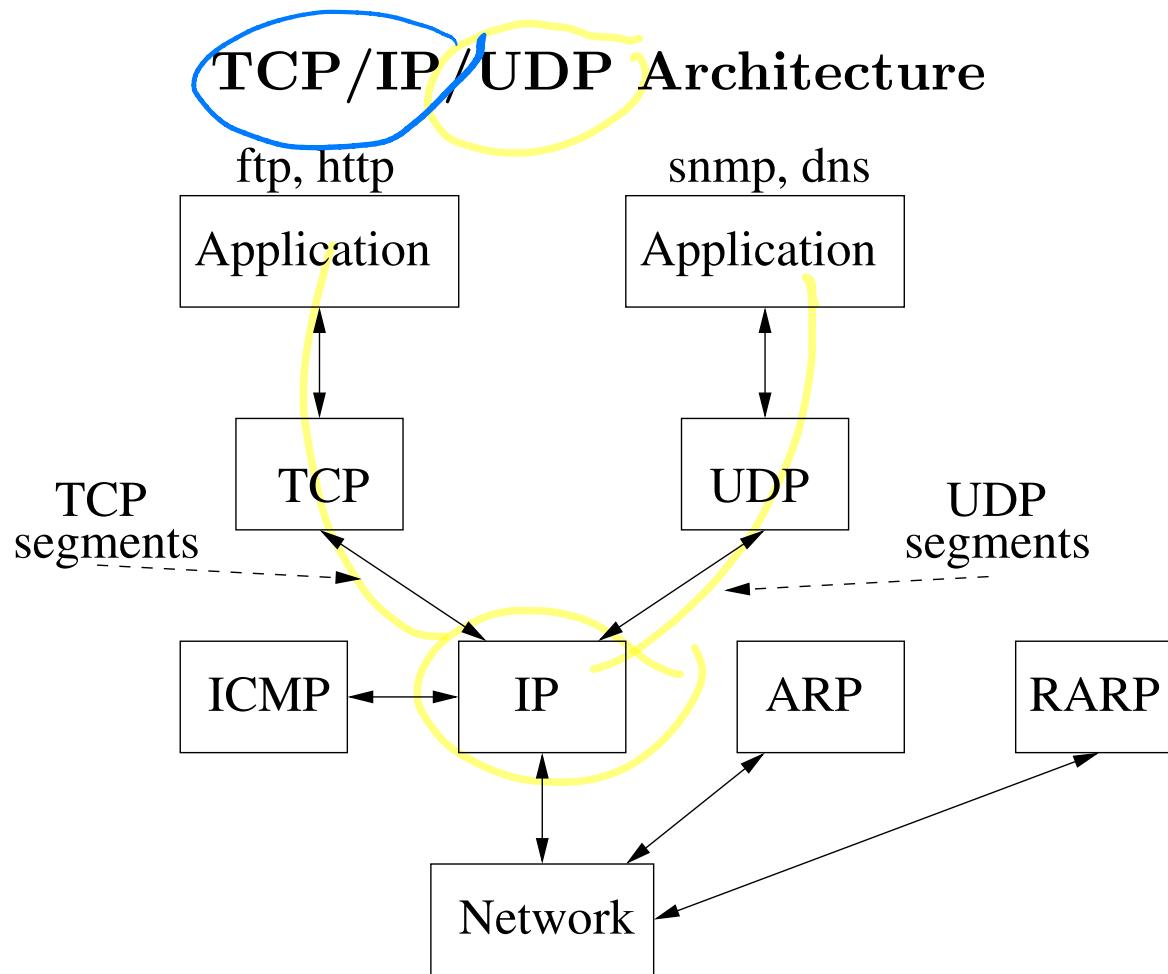
- Numerous key Internet applications use UDP, including:
 - the Domain Name System (DNS), where queries must be fast and only consist of a single request followed by a single reply packet,
 - the Simple Network Management Protocol (SNMP),
 - the Routing Information Protocol (RIP), and
 - the Dynamic Host Configuration Protocol (DHCP).
- UDP is best suited for applications that require speed and efficiency.
 - VPN tunneling, Streaming videos, Online games, Live broadcasts, Domain Name System (DNS), Voice over Internet Protocol (VoIP), Trivial File Transfer Protocol (TFTP)

{ UDP does not
require any complicated
structure }



TCP

- TCP is based on the **End-to-End** connectivity paradigm:
functions should not be provided at lower system levels unless they can be correctly implemented at that level.
- TCP is a **reliable** byte stream protocol. Its main features are
 1. A TCP sliding window protocol. (*Sequence numbers*)
 2. TCP connections have very variable round trip times.
 3. TCP Packets may arrive out of order.
 - { 4. TCP includes mechanism so that connections learn of each other's resources.
 - { 5. TCP includes mechanisms to monitor congestion and control resource allocation.

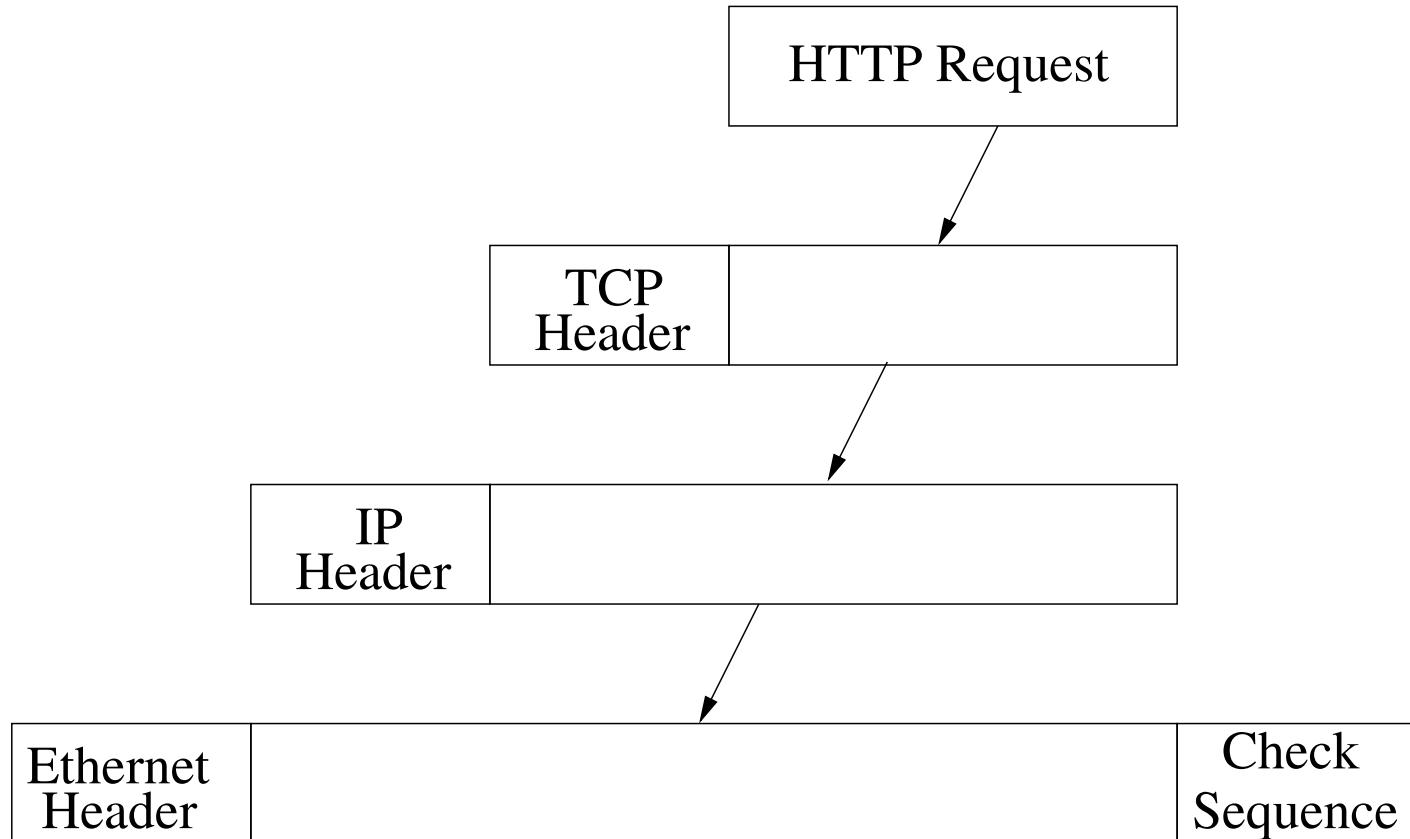


ARP = Address Resolution Protocol

RARP = Reverse Address Resolution Protocol

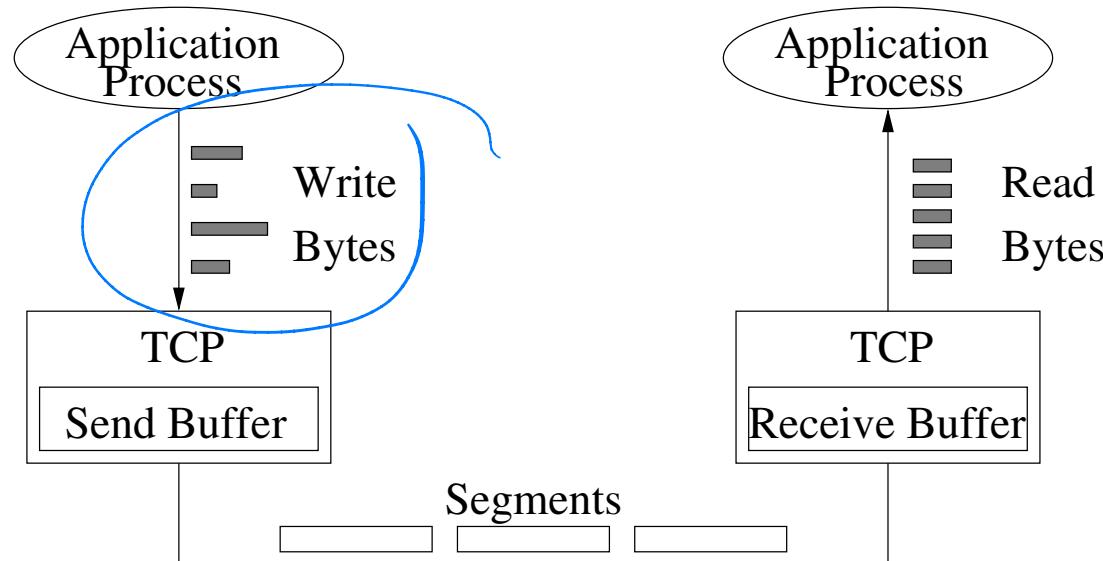
ICMP = Internet Control Message Protocol

TCP/IP Packet Encapsulation

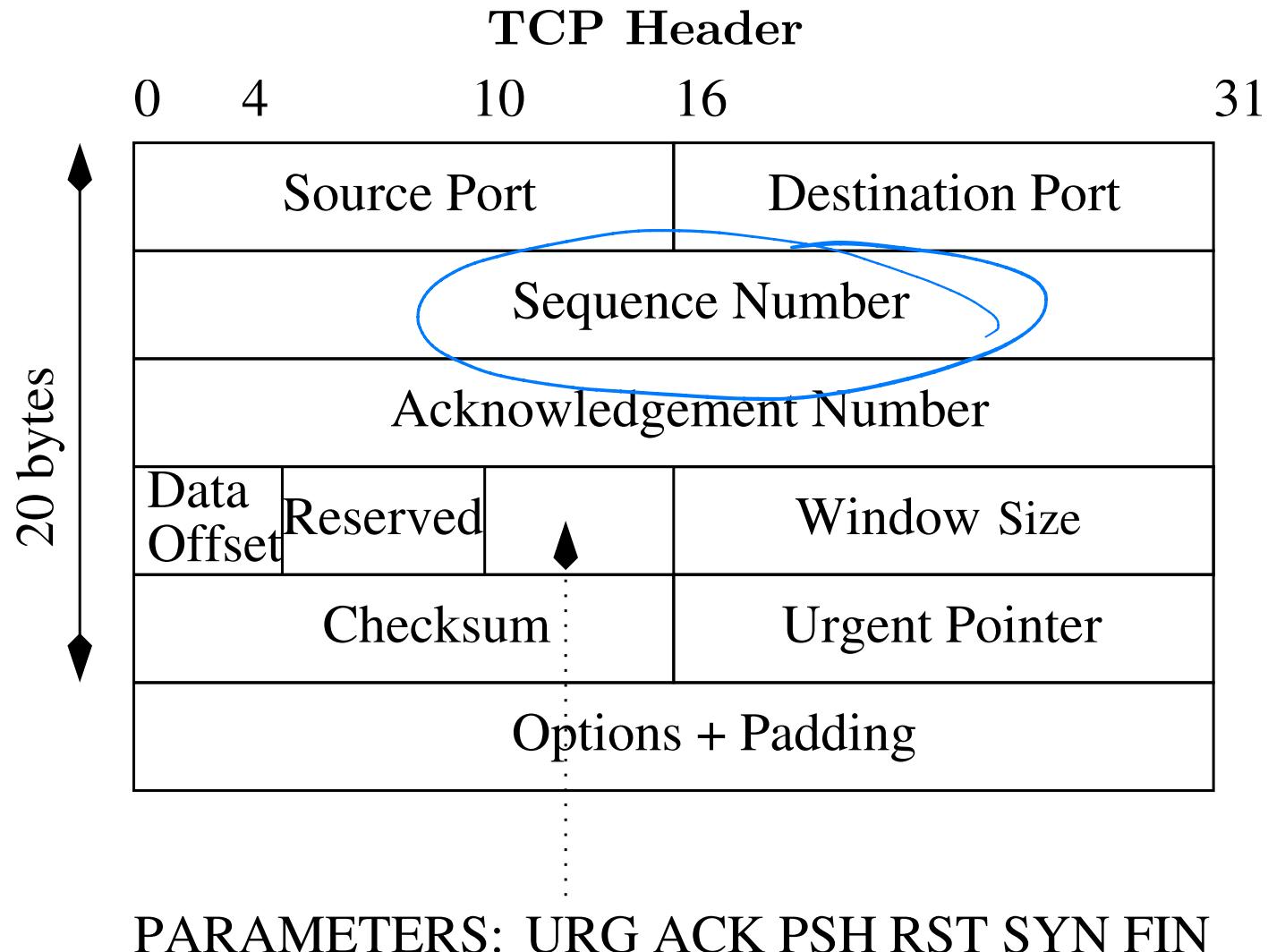


TCP Segment Transmission

- TCP is a byte oriented protocol: sender writes bytes into a TCP connection and receiver reads them out of the TCP connection.

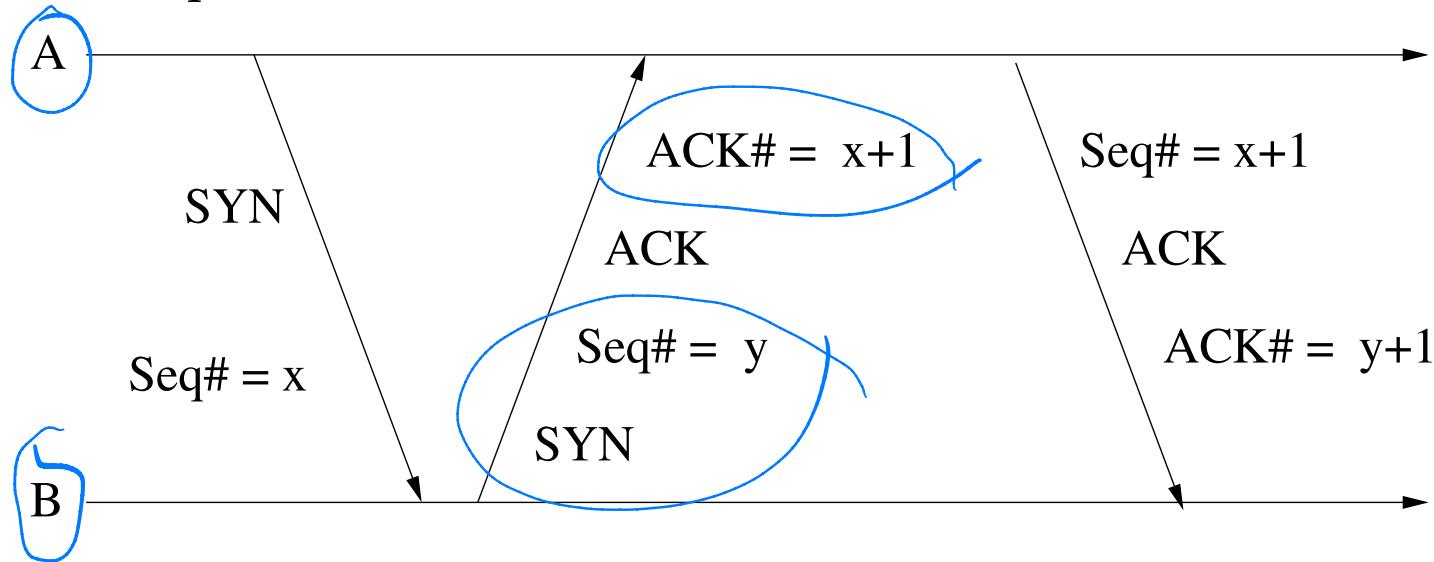


- TCP maintains a variable MSS (Max Segment Size). It decides to send when it has collected “enough” (= MSS) bytes and/or sending process explicitly requests packets. A timer can also trigger transmissions.



TCP Connection Establishment

- Three-way handshake:
 1. A sends request by setting its SYN bit and registers initial sequence number.



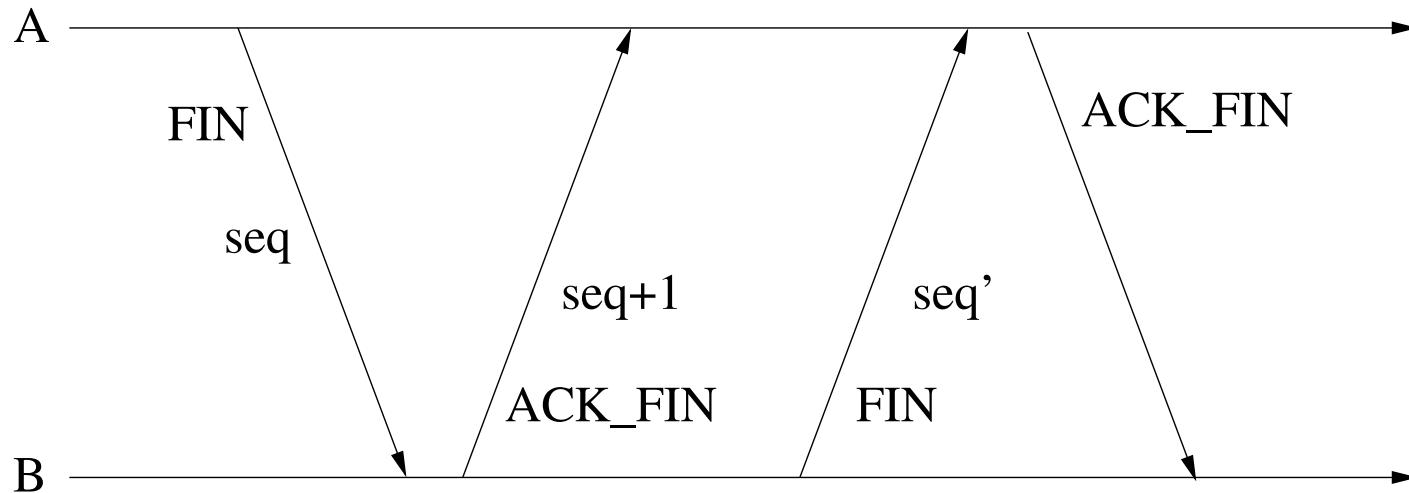
2. B responds by setting its own SYN bit and sends its own sequence number.
3. A acknowledges.

Parameters

- A and B want to agree on the connection parameters.
 1. A sends to B the starting segment number it plans to use.
 2. B acknowledges and sends to A its own starting segment number.
 3. A responds by acknowledging B's segment number.
- In addition a timer is scheduled and if expected response is not received the segment is retransmitted.

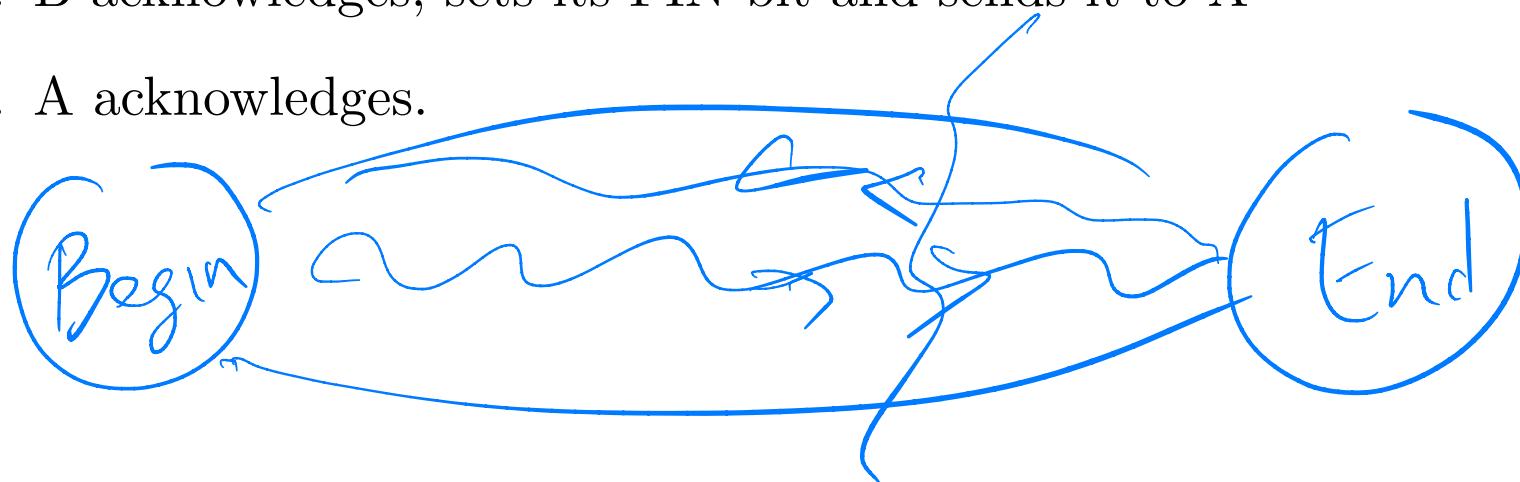
Closing a TCP Connection

1. A sets and sends its FIN bit with a sequence number.



2. B acknowledges, sets its FIN bit and sends it to A

3. A acknowledges.



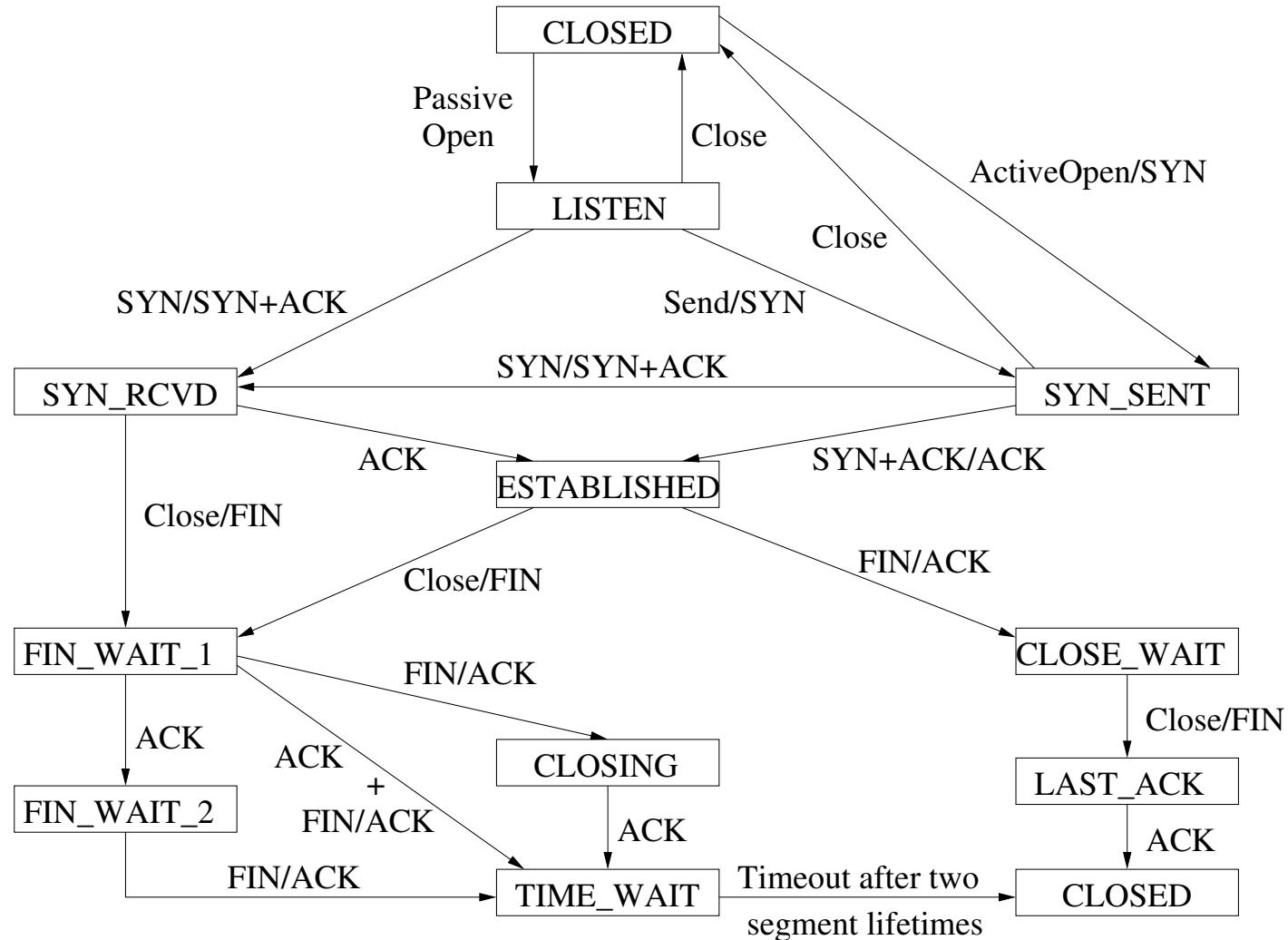
TCP Transition Diagram

- The TCP Transition Diagram defines the semantics of peer-to-peer and service interface.
- The Syntax is given by the TCP headers.
 - All connections start in the CLOSED state.
 - Connections move from state to state through the arcs.
 - Tags on arcs are of the form Event/Action. E.g., SYN/SYN+ACK means when SYN arrives transition is made and a reply ACK+SYN given.
 - Passive open causes TCP to move to LISTEN, followed by an active open and eventually to ESTABLISHED state.

TCP Transition Diagram

- Transitions are triggered when
 1. a segment arrives from a peer.
 2. an operation (e.g. Active Open) on TCP is invoked by local application process.
- Passive open causes TCP to move to LISTEN, followed by an active open and eventually to ESTABLISHED state.

TCP Transition Diagram



TCP Sequence Numbers

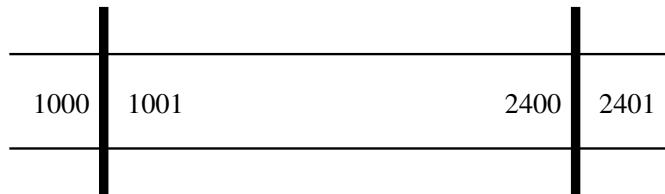
- In a TCP session, the client (on either side) maintains a 32-bit sequence number which it uses to keep track of how much data it has sent.
- This sequence number is included on each transmitted packet, and acknowledged by the opposite host as an acknowledgement number to inform the sending host that the transmitted data was received successfully.
- When a host initiates a TCP session, its initial sequence number may be any value $0 \dots 2^{32} - 1$. Sequence numbers are relative to this initial sequence number of that stream.

TCP Data Transfer (Sliding Window)

- Each byte transmitted has a sequence number.
- When sending a segment sender includes a sequence number of the first byte in segment data field. (i, j) $(533, 100)$
- Receiver acknowledges an incoming segment with a message of the form $(A = i, W = j)$, meaning
 - all bytes through sequence number $i - 1$ are acknowledged, next expected byte has sequence number i .
 - permission is granted to send an additional window W of bytes of data, i.e. the j bytes corresponding to sequence numbers $i..i + j - 1$.
- Protocol known as Credit Allocation Scheme

TCP Credit Allocation: Sending

USER A



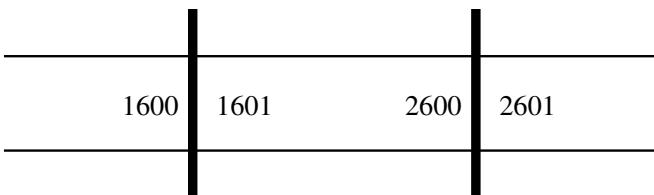
$SN = 1001$

$SN = 1201$

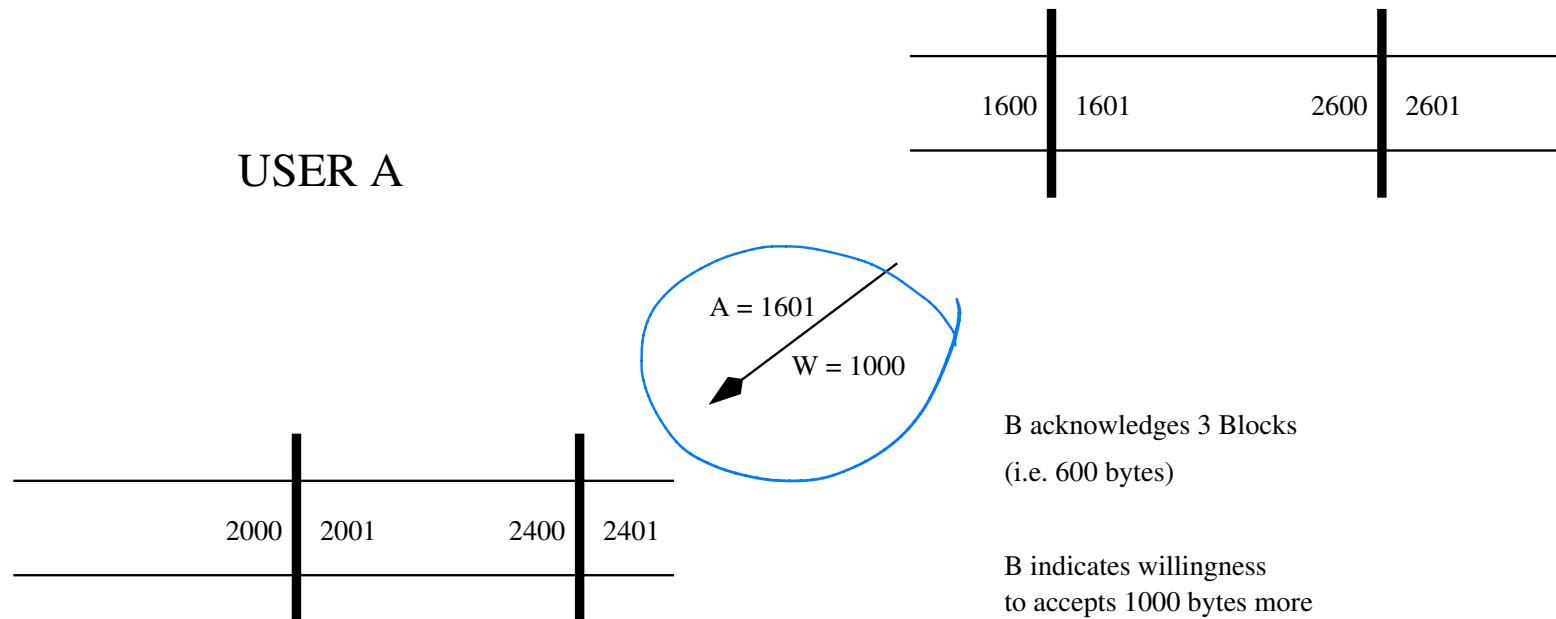
$SN = 1401$

Packets Sent
in Blocks of
200 Bytes

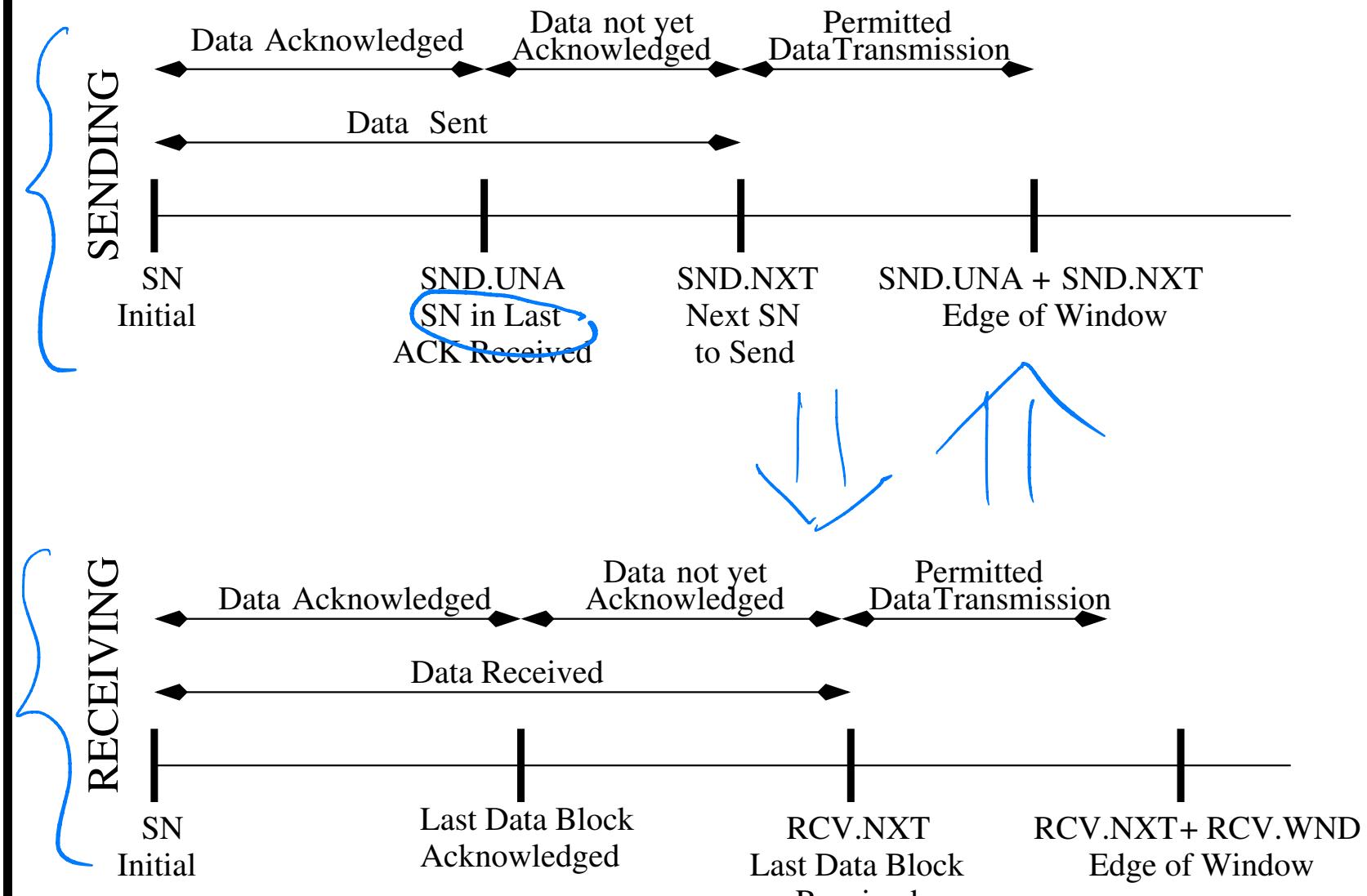
USER B



TCP Credit Allocation: Acknowledging

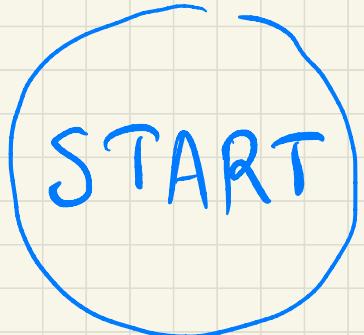


TCP Credit Allocation: Acknowledging

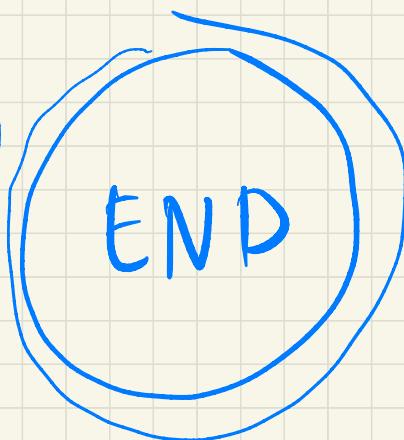
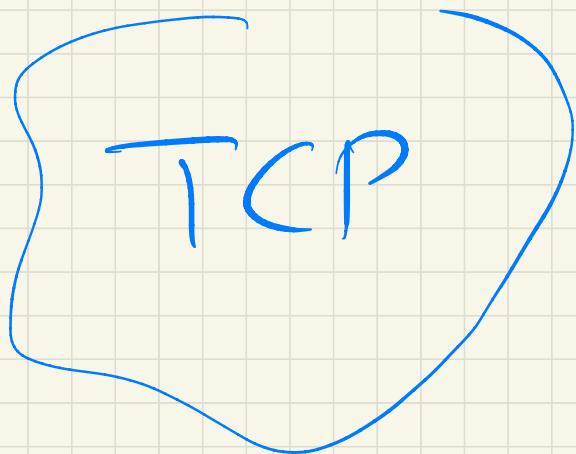


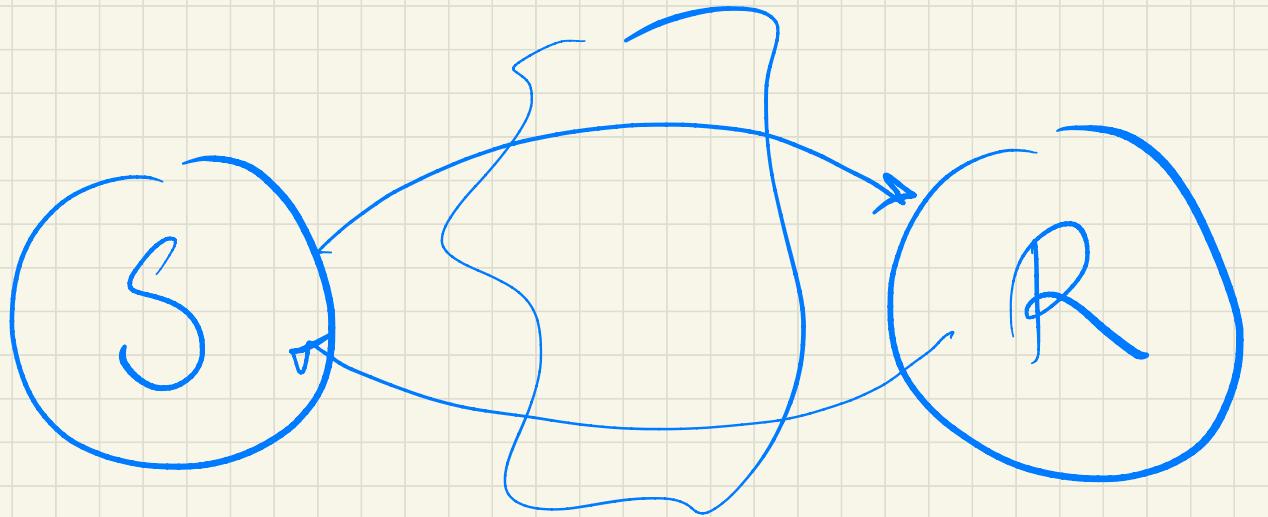
TCP.

Key to everything are TCP
sequence numbers



Seq.#





Clock
Delay

Cloud

Clock
Delay

Flexibility of Credit Allocation Scheme

- Assume last message issued by B was $(A = i, W = j)$ and the last byte of data reserved by B was the byte numbered $i - 1$.
 - To increase credit to an amount k ($k > j$) when no additional data have arrived B issues $(A = i, W = k)$.
 - To acknowledge an incoming segment containing m bytes of data ($m < j$) without granting additional credit B issues $(A = i + m, W = j - m)$.
 - Receiver is not required to acknowledge data received immediately but can issue cumulative acknowledgement.

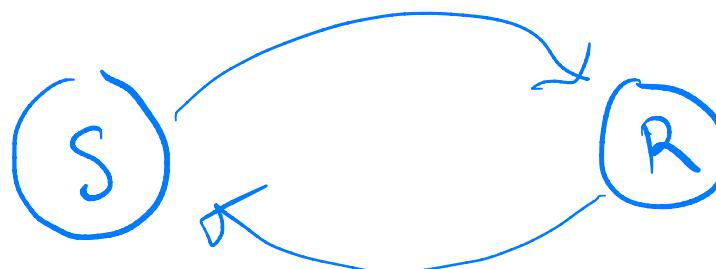
$$(A = i, W = j)$$

$$(A = i + j - 1, W = 2j) \text{ well}$$

$$(A = i + 2, W = j/3)$$

TCP Retransmission Strategies

- TCP has no explicit negative acknowledgement. It relies instead on positive ACK.
 - A timer is associated with each segment as it is sent. If timer expires before acknowledgement then it must retransmit.
 - What is the appropriate value of the timer?
 - * Too small? We have unnecessary retransmissions!
 - * Too large? We have sluggishness!
 - * Right value? Slightly above round-trip delay!
- Possible strategies: Adaptive and Nonadaptive.

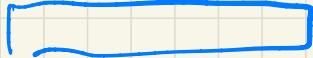


TCP Average Retransmission Timers

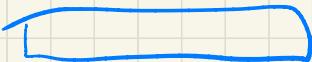
RTT

ping

- Most TCP implementations attempt to estimate the current round trip delay by observing the pattern of delay for recently transmitted segments.
- How does TCP determines timers?
- It runs a complex algorithm which updates periodically values. For example, define the parameters
 - $RTT(i)$ = round trip time observed for i th transmitted segment.
 - $ARTT(k)$ = average round trip time for first k segments, where
$$ARTT(k) = \frac{1}{k} \sum_{i=1}^k RTT(i)$$
- Segment is the unit of transmission used by TCP.



1

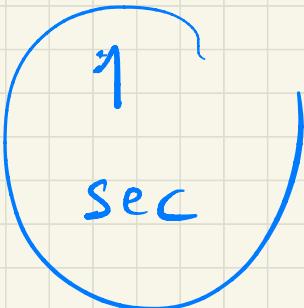
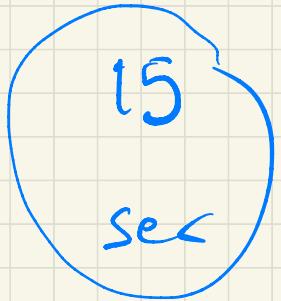


2

... . .



l

7
sec

$$\frac{1}{k} (15 + 1 + \dots + 7)$$

TCP Average Retransmission Timers

- Therefore we have the formula

$$ARTT(k+1) = \frac{1}{k+1} \sum_{i=1}^{k+1} RTT(i) \quad (1)$$

- It might be expensive to compute each time for k , the following recursive formula is useful

$$\begin{aligned} ARTT(k+1) &= \frac{1}{k+1} \sum_{i=1}^k RTT(i) + \frac{1}{k+1} RTT(k+1) \\ &= \underbrace{\frac{k}{k+1} ARTT(k)}_{\text{ARTT}(k)} + \frac{1}{k+1} RTT(k+1) \end{aligned}$$

- Therefore the computation of $ARTT(k+1)$ involves only $ARTT(k)$ and $RTT(k+1)$.

$$ARTT(k) \xrightarrow{k} \frac{1}{k}RTT(1) + \frac{1}{k}RTT(2) + \dots + \frac{1}{k}RTT(k)$$

$\frac{1}{k}$ $\frac{1}{k}$ $\frac{1}{2}$

Dec for $k+1$ st segm

should be weighted more
heavily on the most recent

Exponential Averaging

- We can smooth the previous parameter $ARTT(k)$ by using

$$SRTT(k+1) = (1 - \alpha) \sum_{i=0}^k \alpha^i RTT(k+1-i) + \alpha^{k+1} RTT(0)$$


- From this we get

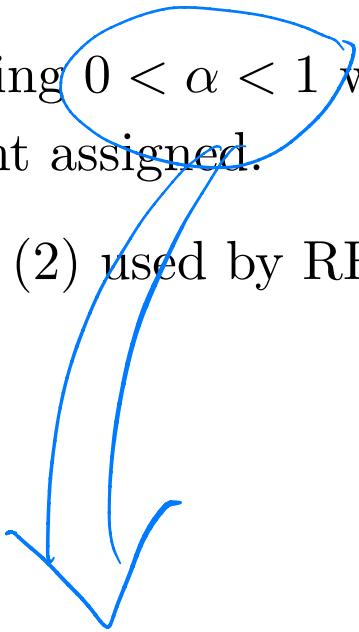
$$\begin{aligned}
 SRTT(k+1) & \quad \stackrel{i=0}{\overbrace{(1-\alpha), (1-\alpha)\cdot\alpha, (1-\alpha)\cdot\alpha^2, \dots}} \\
 &= (1 - \alpha) \sum_{i=0}^k \alpha^i RTT(k+1-i) + \alpha^{k+1} RTT(0) \\
 &= \alpha \left((1 - \alpha) \sum_{i=1}^k \alpha^{i-1} RTT(k - (i-1)) + \alpha^k RTT(0) \right) \\
 &\quad + (1 - \alpha) RTT(k+1)
 \end{aligned}$$

Exponential Averaging

- $SRTT(k)$ is called smoothed round trip time estimate.
- As shown above it is defined by the recursion

$$SRTT(k + 1) := \alpha SRTT(k) + (1 - \alpha)RTT(k + 1) \quad (2)$$

- By choosing $0 < \alpha < 1$ we observe that the further from k the less weight assigned.
- Equation (2) used by RFC 793 to make TCP estimates.



TCP Congestion Control

- IP is stateless and connectionless and has no provision for detecting congestion.
- TCP provides only end-to-end flow control and can deduce presence of congestion by indirect means.
- In general its knowledge of the network conditions is unreliable.
- There is no distributed control to bind together the various TCP entities.
- The only tool used for control is the sliding window mechanism.

Determining TCP Future Flow Rate

- TCP determines future flow rate by the rate of arrival of incoming ACKs which in turn depends on bottlenecks. Such bottlenecks may be due to
 - the internet
 - receiver
 - sender
- This gives rise to the need for retransmission timer management.
- In the sequel we consider three proposals (Algorithms).

TCP Traffic Variance

- Equation (2) enables TCP to adapt to round trip time changes but it is not always effective:
- TCP-peers vary in performance, traffic conditions may change abruptly, etc.
- To make estimates we try to use the probabilistic methodology of expectation, variance and standard deviation, i.e., we can “pretend”

$$X \leftarrow RTT \text{ and } E[X] \leftarrow ARTT$$

- and make use of the formula for mean deviation^a

$$MDEV(X) = \sqrt{E [|X - E[X]|^2]}$$

^aThis comes from the well-known formula for the variance.

TCP Traffic Variance

- Then we estimate the sample standard deviation of RTT as follows

$$\begin{aligned} AERR(k+1) &= RTT(k+1) - ARTT(k) \\ ADEV(k+1) &= \frac{1}{k+1} \sum_{i=1}^{k+1} |AERR(i)|^2 \\ &= \frac{1}{k+1} \sum_{i=1}^{k+1} |RTT(i) - ARTT(i-1)|^2 \end{aligned}$$

with $ARTT(0) = 0$

EXPONENTIAL RTO (Retransmission Timer) **Backoff**

- When a TCP sender timesout on a segment it must retransmit.
- Since timeout is probably due to network congestion it is best to reuse the RTO value.
- This may cause problems when many TCP connections are active.

algorithm

EXPONENTIAL RTO (Retransmission Timer) Backoff

- The following strategy is recommended.
 - TCP source increases its RTO value each time the same segment is retransmitted, this is referred to as the backoff process.
 - TCP source will have to wait longer time for a second retransmission.
 - Updating is done according to formula

$$RTO = qRTO$$

Usually $q = 2$.

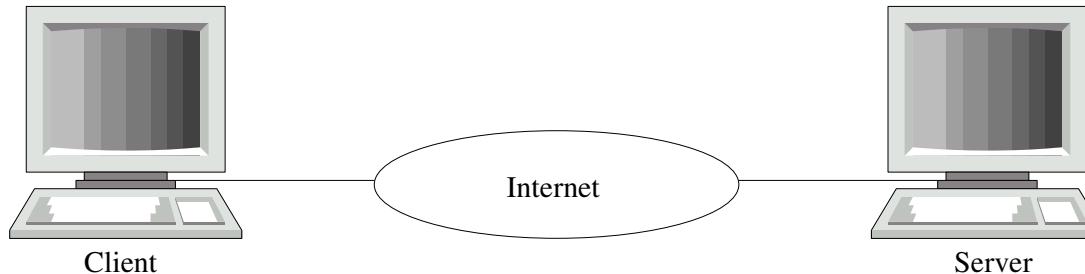
- Same technique used in Ethernet Carrier Sensing protocols.

Karn's Algorithm

- If no segments are retransmitted Jacobson's algorithm works well.
- When segments timeout and need to be retransmitted the RTT values in Jacobson's algorithm are discarded.
- A new suggestion is
 - Do not use the measured RTT for a retransmitted segment to update SRTT and SDEV.
 - Calculate the backoff RTO when a retransmission occurs.
 - Use the backoff RTO value for succeeding segments until an ACK arrives for a segment that has not been retransmitted.
 - When such an acknowledgement is received Jacobson's algorithm is used to compute future RTO values.

TCP and Client/Server

- It is not computers/users that communicate with each other, but rather the end-to-end application programs.

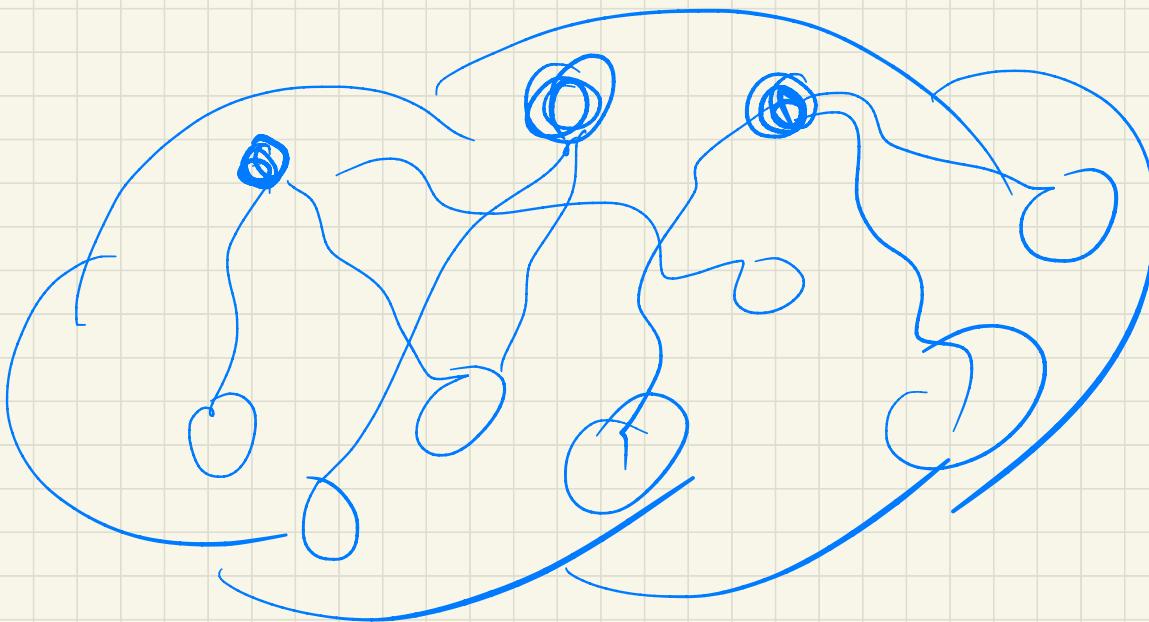


Each computer must know

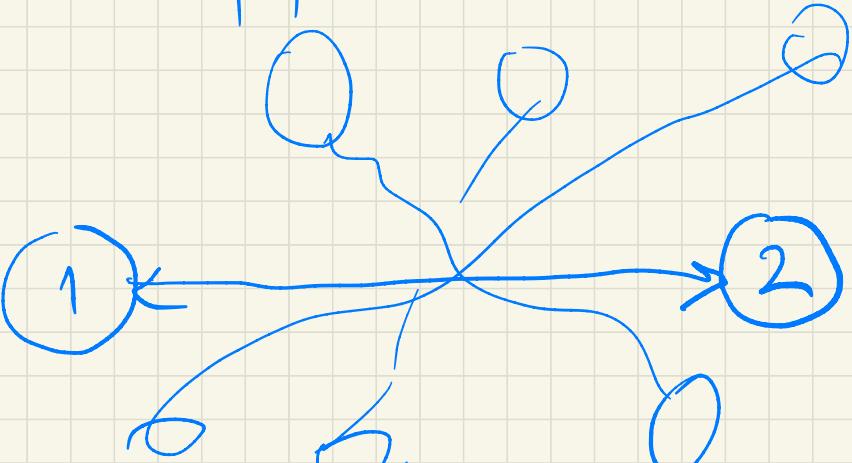
- Its own IP address.
- Its subnet mask.
- The IP address of a router.
- The IP address of a server.
- Stored in a config file, accessed by the bootstrap process (BOOTP).

Equilibria: A Model for TCP

Let's take a global view of
TCP traffic.



What happens when two



users exchange packets?

Equilibrium Conditions!

A TCP Model

- What's going on with TCP?
- Two things must happen:
 - You measure RTT to learn the behaviour of the traffic.
 - You react appropriately with RTO (timeouts) to improve performance. *Crucial is the TCP-number*
- Let's take a high level approach.
- We'll try to build a clear model that illuminates the behaviour of TCP.

Transferring Packets

- When a file is requested over the Internet, the computer that hosts that file breaks it into small packets of data that are then transferred across the network by the transmission control protocol of the Internet, known as TCP.



- The rate at which packets enter the network is controlled by TCP, which is implemented as software on the two computers that are the source and destination of the data.

General Algorithm: “Increase/Decrease”

- The general approach is as follows (Jacobson 1988).
 - When a link within the network becomes overloaded, one or more packets are lost; loss of a packet is taken as an indication of congestion, the destination informs the source, and the source slows down.
 - TCP then gradually increases its sending rate until it again receives an indication of congestion.
- This cycle of increase and decrease enables the source computers to discover and use the available capacity, and to share it between different flows of packets.

Works Very Well!

- TCP has been very successful as the Internet has evolved from a small-scale research network to today's interconnection of hundreds of millions of endpoints and links.
- Each of a large but indeterminate number of flows is controlled by a feedback loop that can know only of that flow's experience of congestion.
- A flow does not know how many other flows are sharing a link on its route, or even how many links are on its route.
- The links vary in capacity by many orders of magnitude, as do the numbers of flows sharing different links.
- It is remarkable that so much has been achieved in such a rapidly growing and heterogeneous network with congestion controlled just at the endpoints.

Why does “Increase/Decrease” Algorithm work so well?

- In recent years researchers have shed some light on TCP’s success, by interpreting the protocol as a ~~decentralized distributed algorithm~~ that solves an optimization problem, just as the decentralized choices of drivers in a road network solve an optimization problem.

deliver my
packet as fast as
possible please!

How TCP Works: ACKs (1/2)

- Packets transferred by TCP across the Internet contain sequence numbers indicating their order, and they should arrive at their destination in that order.
- When a packet is received at the destination, it is acknowledged: an acknowledgment is a short packet sent by the destination back to the source. $10,000 \quad 20$
- If a packet has been lost in the transfer, the source can tell this from the sequence numbers contained in the acknowledgments.
- The source keeps a copy of each packet sent until it has been positively acknowledged; these copies form what is called a sliding window, and allow packets lost in transfer to be sent again by the source.

$$\text{p} = \frac{1}{100} \quad \frac{1}{4} \quad \frac{20}{1,000} = \frac{1}{500}$$

How TCP Works: Increase/Decrease (2/2)

- Stored in the source computer there is a numerical variable known as the *congestion window* and denoted by w . 
- The congestion window directs the size of the sliding window in the following sense:
 1. if the size of the sliding window is less than w , then the computer increases it by sending out a packet;
 2. if the size of the sliding window is greater than or equal to w , then it waits for positive acknowledgments to come in, which have the effect of reducing the size of the sliding window and, as we shall see, increasing w as well.
- The size of the sliding window continually changes, moving in the direction of a target size that is given by the congestion window.

How Much to Increase/Decrease!

- How much should you increase the congestion window w ?
 - By a function $I(w)$ of w , i.e.,

$$w \leftarrow w + I(w)$$

$$\begin{aligned} I(w) &= 10w \\ I(w) &= 1,000w \\ I(w) &= 1 \end{aligned}$$

- How much should you decrease the congestion window w ?
 - By a function $D(w)$ of w , i.e.,

$$w \leftarrow w - D(w)$$

- The choice of $I(w)$ and $D(w)$ should reflect how aggressive/cautious you want to be!

Basic Rules: Algorithm

- The congestion window itself is not a fixed number:
 - rather, it is constantly being updated, and the precise rules for how this is done are critical for TCP's sharing of capacity.
- The rules currently used are as follows:
 1. Every time a positive acknowledgment comes in, w is increased by $I(w)$, and
 2. Every time a lost packet is detected, w is decreased by $D(w)$.
- Thus, if the source computer detects a lost packet, it realizes that there has been some congestion and backs off for a while, but if all its packets are getting through then it allows the rate at which it sends packets to inch up again.

Using Statistics and Probability

- The way one applies this methodology is by monitoring traffic parameters, like packet loss, and congestion at the nodes.
- For example, we can measure the probability that a packet is lost.
 - We can do this by measuring traffic at the nodes over longer periods of time.
 - Of course, the values we obtain will depend on time.

Algorithm

- Let p be the probability that a packet is lost.
- We execute the following algorithm:
- If the current window size is w then with probability

 1. $1 - p$ increase the congestion window by $I(w)$, and
 2. p decrease the congestion window by $D(w)$.
- $I(w), D(w)$ are as yet undefined!

Equilibrium

- The expected change in the congestion window w per update step is therefore

$$E[Change] = I(w)(1 - p) + (-D(w))p$$

- The expected change will be positive for small values of w , but will become negative if w is big enough.

- Equilibrium for w might arise when

$$E[Change] = 0$$

equilibrium

- Hence, this expression is zero when

$$I(w)(1 - p) = D(w)p$$

P

(3)

- If $I(w), D(w)$ are interesting functions of w we get information on “how window size relates to the packet loss probability p .

Example

- Assume that the packet error rate is p : this can be measured in the TCP protocol. Let

$$I(w) = \frac{1}{w} \text{ and } D(w) = \frac{w}{2}$$

$$\begin{aligned} w + \frac{1}{w} \\ w - \frac{w}{2} \end{aligned}$$

- At equilibrium (see Equation (3)) $I(w)(1 - p) = D(w)p$, i.e.,

$$\frac{1}{w}(1 - p) = \frac{w}{2}p$$

- So solving for w ,

$$w^2 = \frac{2(1 - p)}{p}$$

- Therefore

$$w = \left(\frac{2(1 - p)}{p} \right)^{1/2} = \sqrt{2} \left(\frac{1}{p} - 1 \right)$$

So w is inversely proportional to p , which is what you expect!

Exercises^a

1. Would it be reasonable for TCP to set the retransmission strategies between hosts by negotiation prior to a session of transmission of packets?
2. Discuss how TCP makes measurements using Retransmission Strategies so as to determine optimal congestion strategies.
3. What are similarities and differences of exponential backoff algorithms as used in TCP and in Ethernet?
4. Compare the effectiveness of Jacobson's and Karn's algorithms for TCP retransmission control.
5. Here are some choices for $I(w)$, $D(w)$:
 - (a) [Useless:] $I(w) = Aw$, and $D(w) = Bw$

^aDo not submit!

- (b) [Aggressive:] $I(w) = Aw^k$, $k \geq 2$, and $D(w) = Bw$
- (c) [Conservative:] $I(w) = Aw$, and $D(w) = Bw^k$, $k \geq 2$
- (d) [About Right:] $I(w) = 1/w$, and $D(w) = w/2$

What equilibrium conditions do they give rise to?

6. Why does TCP use the last algorithm?