Network Layer

Chapter 5

- Design Issues
- Routing Algorithms
- Congestion Control
- Quality of Service
- Internetworking
- Network Layer of the Internet

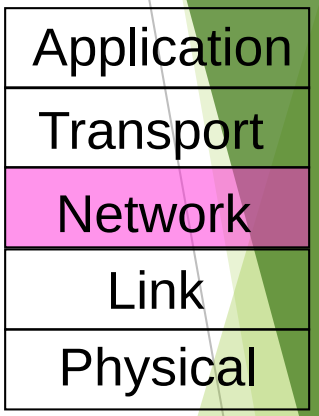
CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Revised: August 2011

The Network Layer

Il data link layer si occupa di spostare frame da un capo all'altro della linea di trasmissione.Il Network layer si occupa della trasmissione end-to-end.Per fare ció deve

Responsible for delivering packets between endpoints over multiple links



Questo layer deve provvedere a fornire servizi al transport layer. Tali servizi devono rispettare i seguenti vincoli: 1) indipender

Design Issues

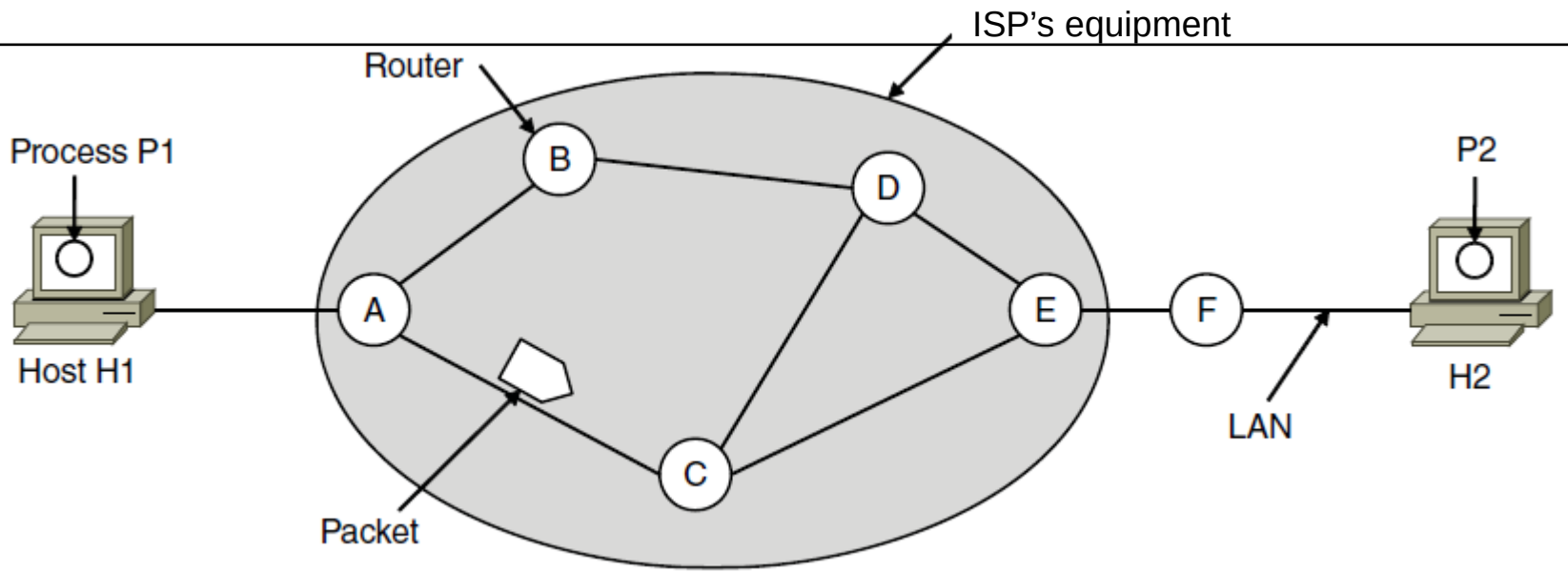
- Store-and-forward packet switching »
- Connectionless service – datagrams »
- Connection-oriented service – virtual circuits »
- Comparison of virtual-circuits and datagrams »

Attraverso i servizi connectionless, i pacchetti sono inoltrati nella reta individualmente e indirizzati indipendentemente. Non é richiesto nessun setup iniziale.I pac

Store-and-Forward Packet Switching

Hosts send packets into the network; packets are forwarded by routers

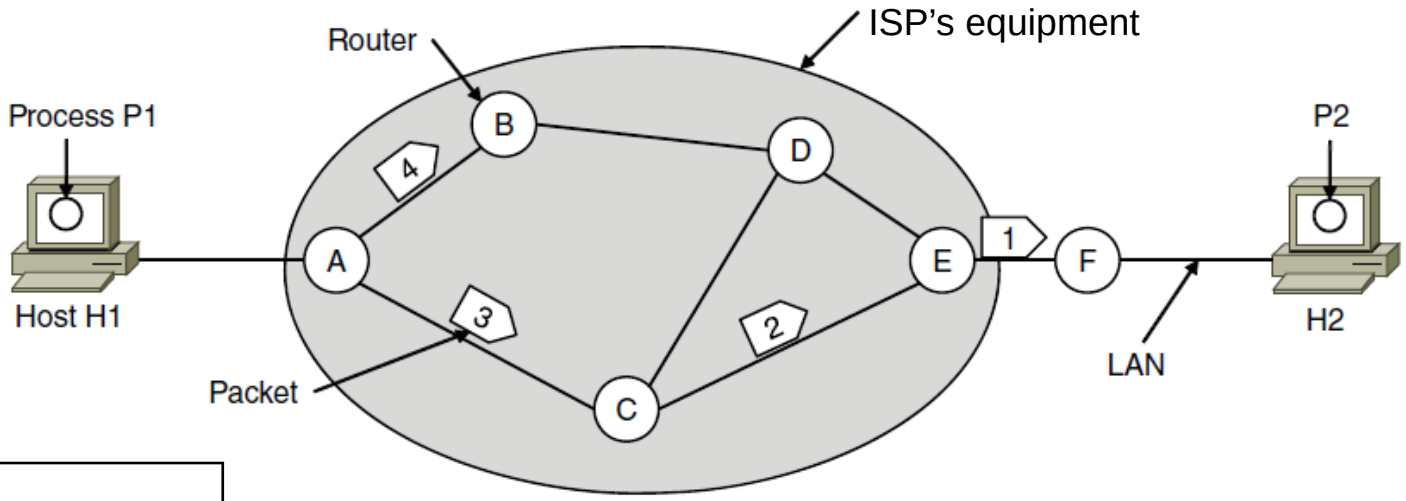
Un host inoltra i pacchetti al router piú vicino (della LAN o del ISP), il quale mantiene nel buffer ció che ha ricevuto fin quando l'host non ha finito di inotrare tutti i pacchetti. Sol



Connectionless Service – Datagrams

Il sistema operativo assegna un header al frame e lo invia alla rete. A questo pu

- Packet is forwarded using destination address inside it
  - Different packets may take different paths



L'algoritmo che gestisce la tabelle e prende

A's table (initially)

A	
B	B
C	C
D	B
E	C
F	C

A's table (later)

A	
B	B
C	C
D	B
E	D
F	D

C's Table

A	A
B	A
C	
D	E
E	E
F	E

E's Table

A	C
B	D
C	C
D	D
E	
F	F

Dest. Line

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Connection-Oriented – Virtual Circuits

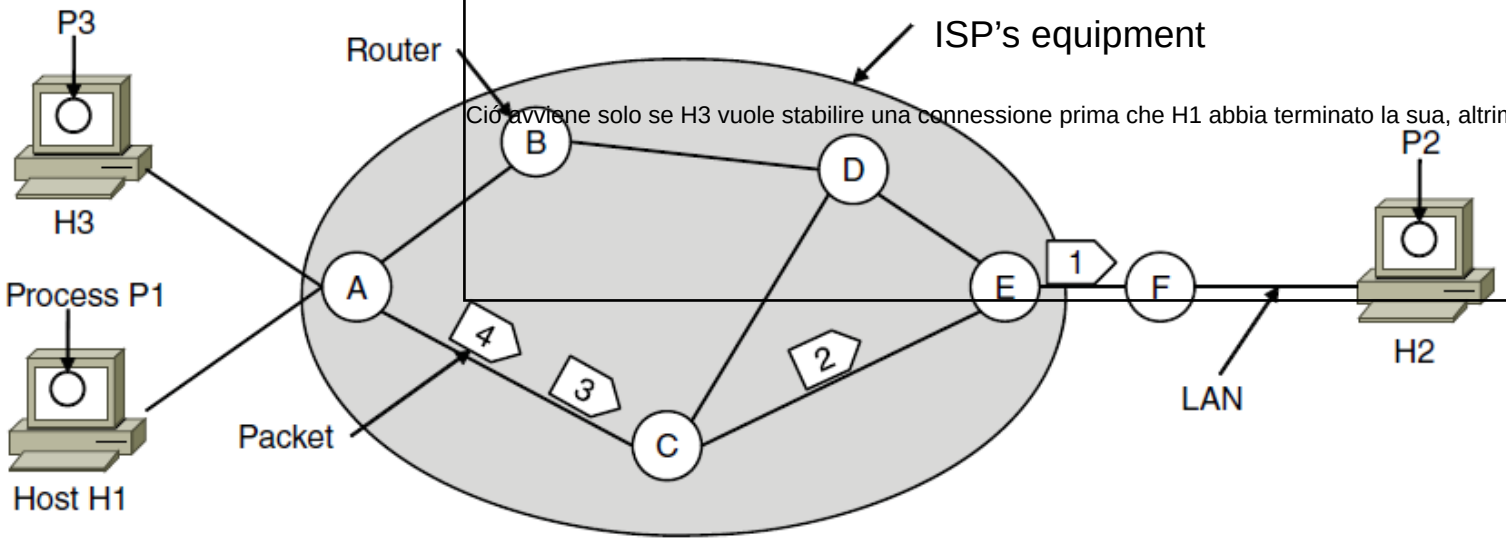
L'idea é di non dover prendere decisioni per ogni pacchetto

Quando una connessione é stabilita, viene scelto un percorso come parte del setup e viene salvata nelle tabelle dei router. Quando la co

- Packet is forwarded along a virtual circuit using tag inside it
  - Virtual circuit (VC) is set up ahead of time

Come mostrato nella figura, se un altro host H3 vuole intraprendere una comu

Ciò avviene solo se H3 vuole stabilire una connessione prima che H1 abbia terminato la sua, altrimenti la connession



connection identifier

A's table

In: Line	Tag	Line	Tag: Out
H1	1	C	1
H3	1	C	2

C's Table

A	1	E	1
A	2	E	2

E's Table

C	1	F	1
C	2	F	2

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Comparison of Virtual-Circuits & Datagrams

La principare differenza é che attraverso il virtual circuit si garantisce una certa qualitá del servizio e si evitano le congestioni piú f

Issue	Datagram network	Virtual-circuit network
Circuit setup	Not needed	Required
Addressing	Each packet contains the full source and destination address	Each packet contains a short VC number
State information	Routers do not hold state information about connections	Each VC requires router table space per connection
Routing	Each packet is routed independently	Route chosen when VC is set up; all packets follow it
Effect of router failures	None, except for packets lost during the crash	All VCs that passed through the failed router are terminated
Quality of service	Difficult	Easy if enough resources can be allocated in advance for each VC
Congestion control	Difficult	Easy if enough resources can be allocated in advance for each VC

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Routing Algorithms (1)

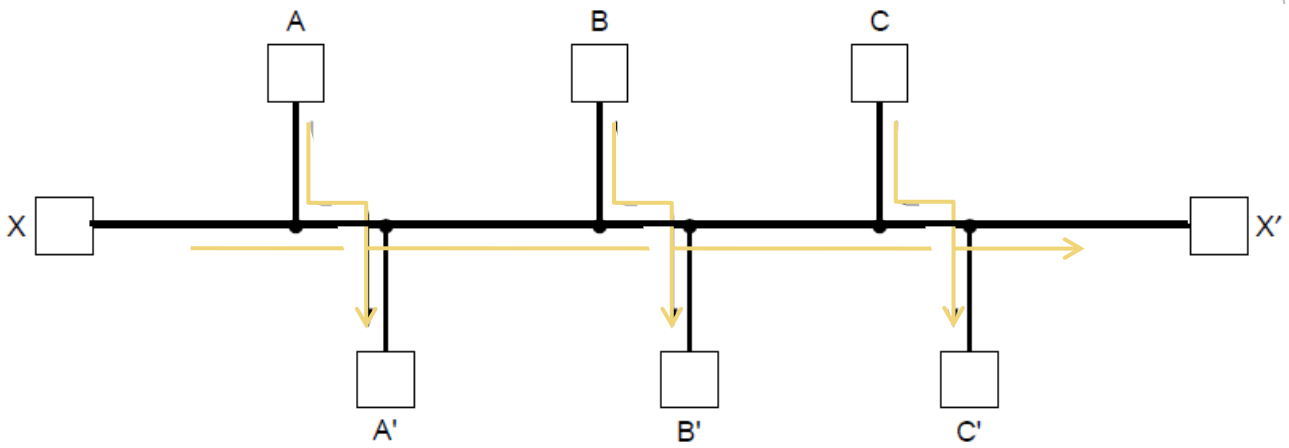
- Optimality principle »
- Shortest path algorithm »
- Flooding »
- Distance vector routing »
- Link state routing »
- Hierarchical routing »
- Broadcast routing »
- Multicast routing »
- Anycast routing »
- Routing for mobile hosts »
- Routing in ad hoc networks »

Il routing algorithm é quella parte del software che decide a quale

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Routing Algorithms (2)

- Routing is the process of discovering network paths
  - Model the network as a graph of nodes and links
  - Decide what to optimize (e.g., fairness vs efficiency)
  - Update routes for changes in topology (e.g., failures)



- For

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

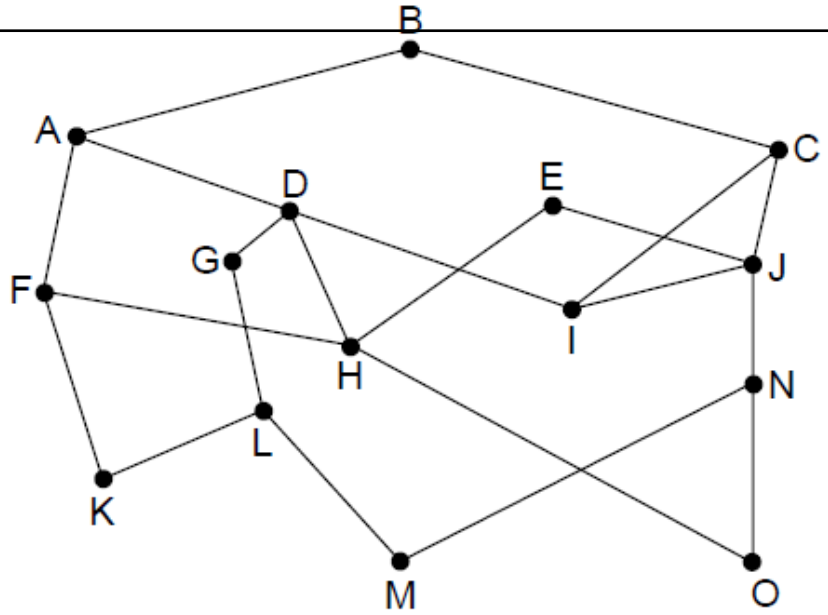
Pricipio di ottimalitá: se J é un nodo tra la via ottima che va da I a K, allora é un percorso ottimo anche tra J e K, e J e I.Questo principio serve come riferir

The Optimality Principle

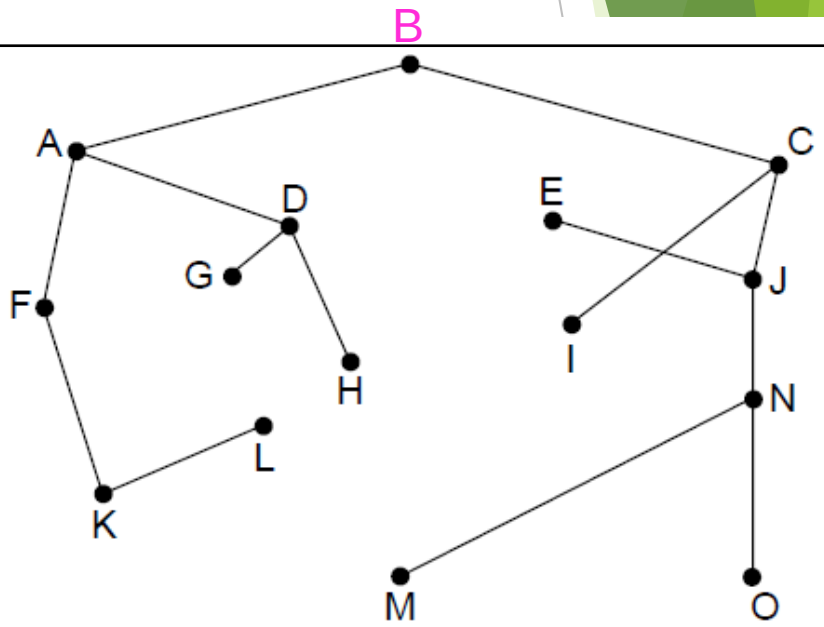
Each portion of a best path is also a best path; the union of them to a router is a tree called the sink tree

- Best means fewest hops in the example

Dato che il sink tree é un albero, anch'esso non avrá cicli, il che assicura che il pacchetto arrivi a destinazione entro un numero finito di salti. Nel mondo reale, p



Network



Sink tree of best paths to router B

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Shortest Path Algorithm (1)

Dijkstra's algorithm computes a sink tree on the graph:

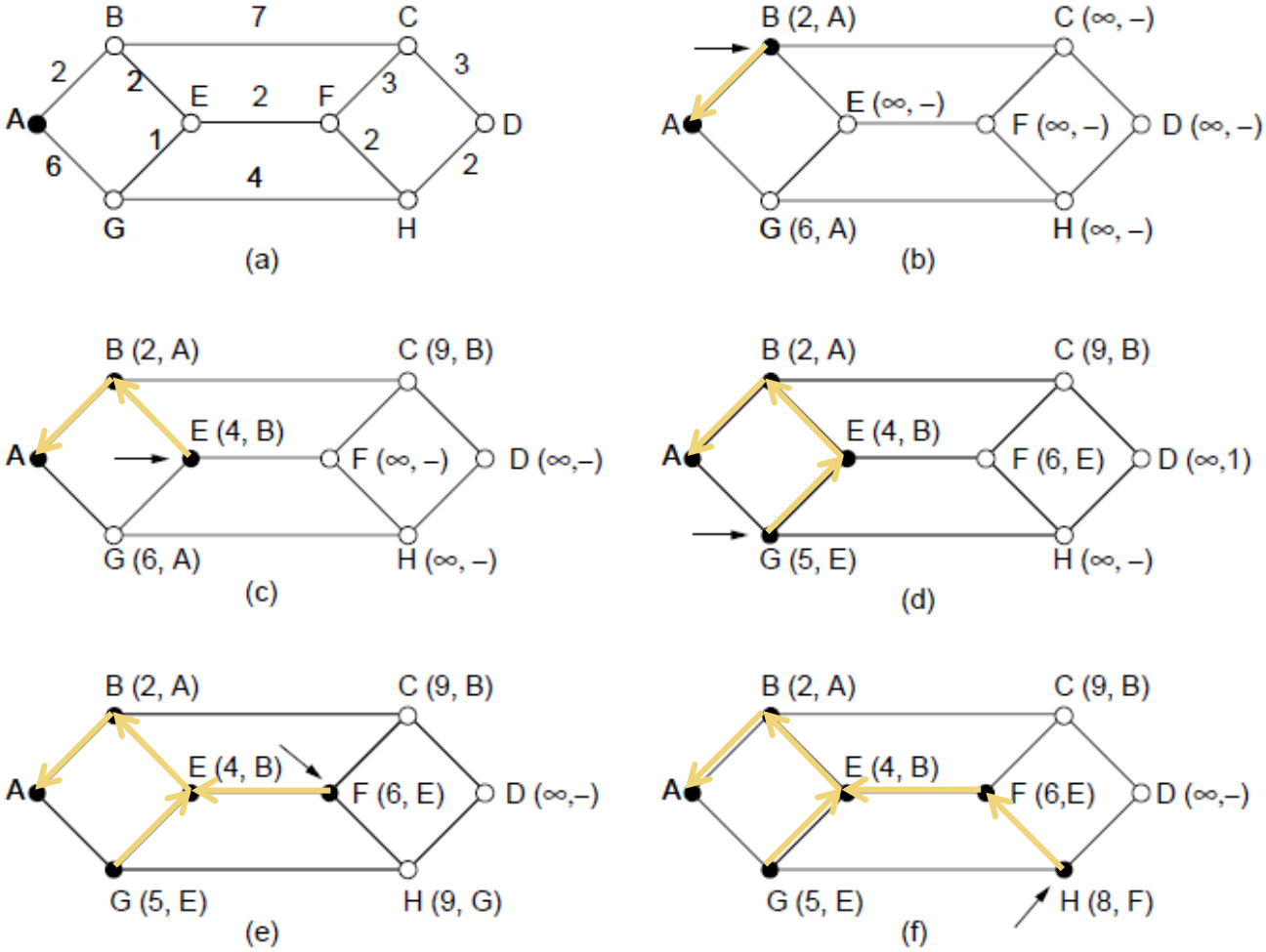
- Each link is assigned a non-negative weight/distance
- Shortest path is the one with lowest total weight
- Using weights of 1 gives paths with fewest hops

Algorithm:

- Start with sink, set distance at other nodes to infinity
- Relax distance to other nodes
- Pick the lowest distance node, add it to sink tree
- Repeat until all nodes are in the sink tree

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Shortest Path Algorithm (2)



A network and first five steps in computing the shortest paths from A to D. Pink arrows show the sink tree so far.

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Shortest Path Algorithm (3)

. . .

```
for (p = &state[0]; p < &state[n]; p++) {
      p->predecessor = –1;
      p->length = INFINITY;
      p->label = tentative;
}
state[t].length = 0;  state[t].label = permanent;
k = t;
do {
      for (i = 0; i < n; i++)
            if (dist[k][i] != 0 && state[i].label == tentative) {
                  if (state[k].length + dist[k][i] < state[i].length) {
                        state[i].predecessor = k;
                        state[i].length = state[k].length + dist[k][i];
                  }
            }
```

. . .

Start with the sink, all other nodes are unreachable

Relaxation step. Lower distance to nodes linked to newest member of the sink tree

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Shortest Path Algorithm (4)

. . .

```
    k = 0; min = INFINITY;
    for (i = 0; i < n; i++)
        if (state[i].label == tentative && state[i].length < min) {
            min = state[i].length;
            k = i;
        }
    state[k].label = permanent;
} while (k != s);
```

Find the lowest distance, add it to the sink tree, and repeat until done

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Flooding

- A simple method to send a packet to all network nodes
- Each node floods a new packet received on an incoming link by sending it out all of the other links
- Nodes need to keep track of flooded packets to stop the flood; even using a hop limit can blow up exponentially

Per evitare che pacchetti, con informazioni datate sulla topologia, continuino a vagare per la rete all'infinito, si aggiunge un contatore (inizializzato alla distanza tra mitt

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Distance Vector Routing (1)

Distance vector is a distributed routing algorithm

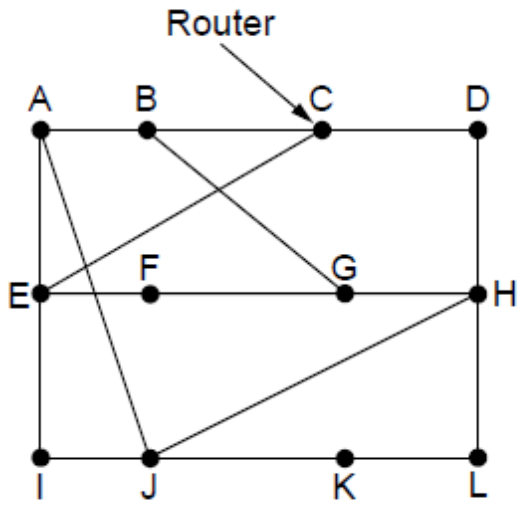
- Shortest path computation is split across nodes

Algorithm:

- Each node knows distance of links to its neighbors
- Each node advertises vector of lowest known distances to all neighbors
- Each node uses received vectors to update its own
- Repeat periodically

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Distance Vector Routing (2)



Network

Ogni router mantiene una copia della routing table contenente ogni nodo della re

To	A	I	H	K	New estimated delay from J	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		

New vector for J

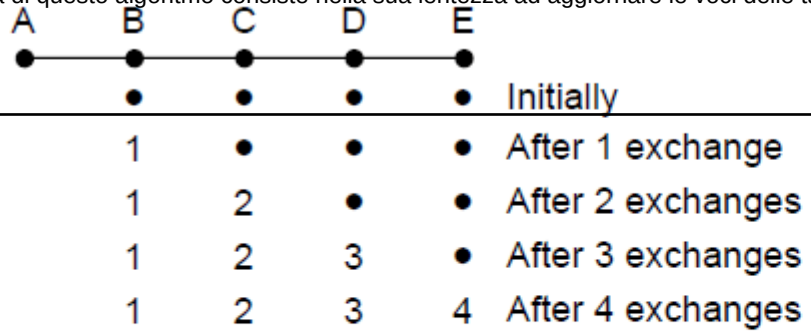
Vectors received at J from Neighbors A, I, H and K

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

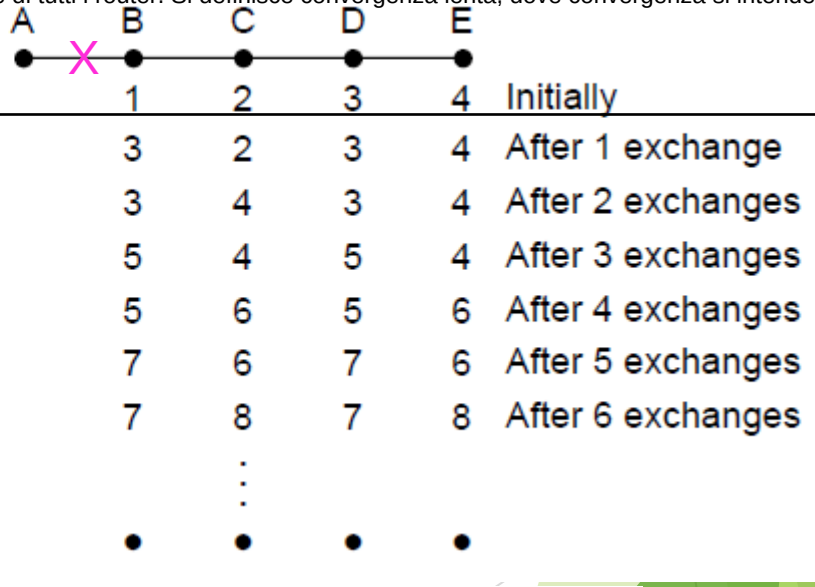
The Count-to-Infinity Problem

Failures can cause DV to "count to infinity" while seeking a path to an unreachable node

Il problema di questo algoritmo consiste nella sua lentezza ad aggiornare le voci delle tabelle di tutti i router. Si definisce convergenza lenta, dove convergenza si intende la c



Good news of a path to *A* spreads quickly



Bad news of no path to *A* is learned slowly

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Link State Routing (1)

Link state is an alternative to distance vector

- More computation but simpler dynamics
- Widely used in the Internet (OSPF, ISIS)

Algorithm:

- Each node floods information about its neighbors in LSPs (Link State Packets); all nodes learn the full network graph
- Each node runs Dijkstra's algorithm to compute the path to take for each destination

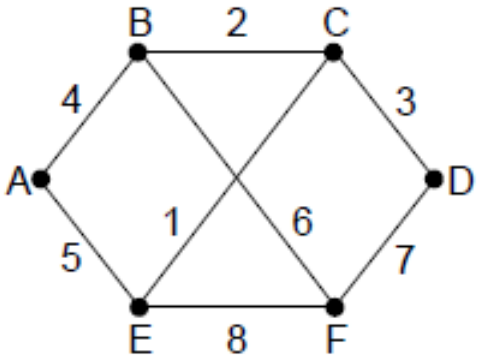
É composto da 5 step:1) scoprire i nodi vicini e sapere i loro indirizzi;2) stabilire la distanza o la metrica del costo;3) costruire il pacchetto con tutto quello che si sa;4) inv

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Link State Routing (2) – LSPs

LSP (Link State Packet) for a node lists neighbors and weights of links to reach them

Per il passaggio 3, il pacchetto inizia con l'identificatore del router che lo sta costruendo, seguito dal sequence number e dal timestamp.Rimane ora da decidere quando



Network

A	
Seq.	
Age	
B	4
E	5

B	
Seq.	
Age	
A	4
C	2
F	6

C	
Seq.	
Age	
B	2
D	3
E	1

D	
Seq.	
Age	
C	3
F	7

E	
Seq.	
Age	
A	5
C	1
F	8

F	
Seq.	
Age	
B	6
D	7
E	8

LSP for each node

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Per rendere efficace questo algoritmo, ogni router facente parte della stessa rete, deve avere in memoria la stessa topologia, pena inconsistenze come cicli, macchine irrag

Link State Routing (3) – Reliable Flooding

Seq. number and age are used for reliable flooding

- New LSPs are acknowledged on the lines they are received and sent on all other lines
- Example shows the LSP database at router B

I pacchetti che descrivono la topoplogia vengono sostanzimente distribuiti tramite flooding con sequence diagram e tempo di vita. I router tengono traccia di ogni pacchetto ch

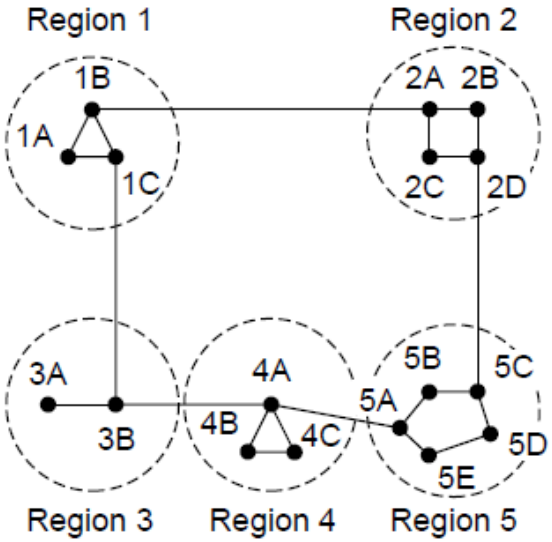
			Send flags			ACK flags			
Source	Seq.	Age	A	C	F	A	C	F	Data
A	21	60	0	1	1	1	0	0	
F	21	60	1	1	0	0	0	1	
E	21	59	0	1	0	1	0	1	
C	20	60	1	0	1	0	1	0	
D	21	59	1	0	0	0	1	1	

É possibile irrobustire questo algoritmo, ad esempio, non facendo inoltrare immediatamente un pacchetto che di aggiornamento della topologia immediatamente, invece il router

Hierarchical Routing

Per ridurre il numero di voci e di calcolo della topologia si puó dividere la rete in reg

- Hierarchical routing reduces the work of route computation but may result in slightly longer paths than flat routing



Full table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2A	1B	2
2B	1B	3
2C	1B	3
2D	1B	4
3A	1C	3
3B	1C	2
4A	1C	3
4B	1C	4
4C	1C	4
5A	1C	4
5B	1C	5
5C	1B	5
5D	1C	6
5E	1C	5

Hierarchical table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2	1B	2
3	1C	2
4	1C	3
5	1C	4

Best choice to reach nodes in 5 except for 5C

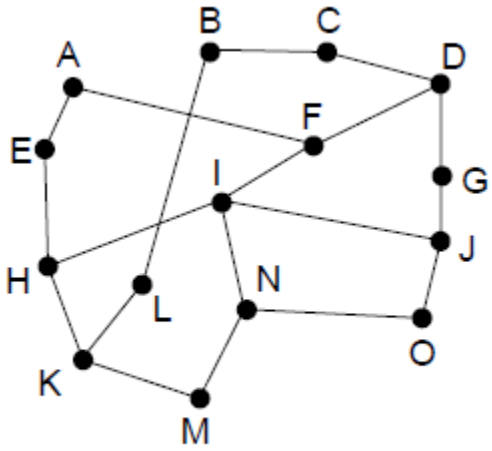
CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Broadcast Routing

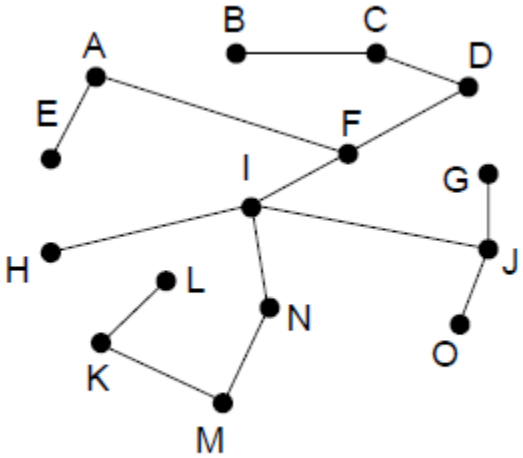
Banalmente basterebbe replicare il messaggio e inoltrarlo singolarmente ad ogni nodo. Ques

- Broadcast sends a packet to all nodes
  - RPF (Reverse Path Forwarding): send broadcast received on the link to the source out all remaining links
  - Alternatively, can build and use sink trees at all nodes

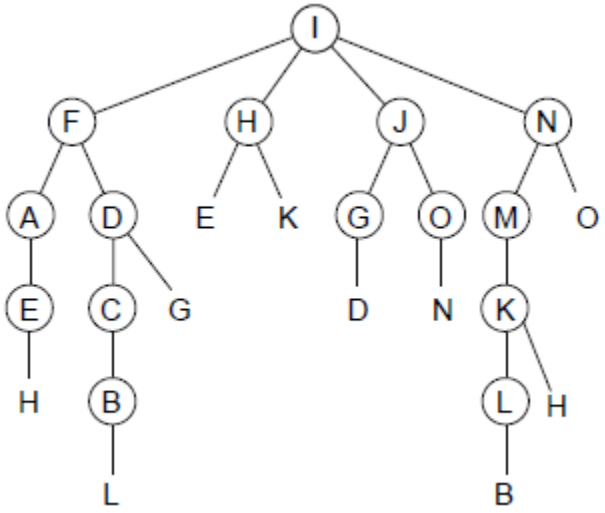
Un metodo molto efficace ed efficiente é il RPF. Se il pacchetto é stato ricevuto dalla linea verso il mittente, é abbastanza probabile che quella sia il percorso piú breve e quinc



Network



Sink tree for *I* is efficient broadcast



RPF from *I* is larger than sink tree

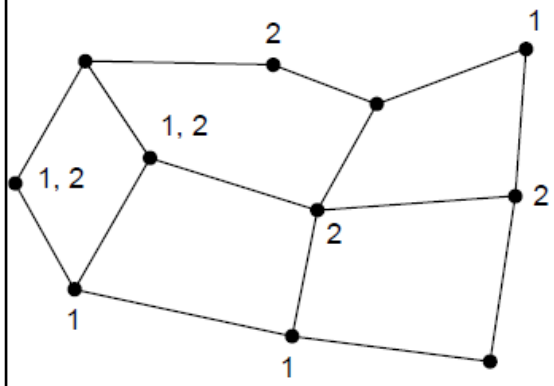
CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Multicast Routing (1) – Dense Case

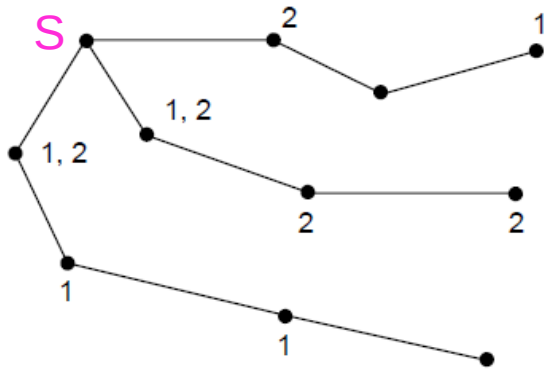
Tralasciando come viene implementato il gruppo (che sará argor

- Multicast sends to a subset of the nodes called a group
  - Uses a different tree for each group and source

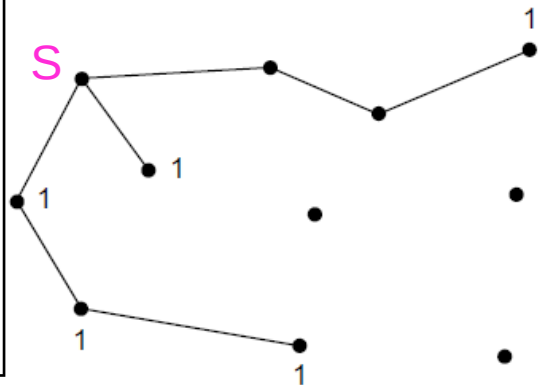
Il metodo piú facile per sfo



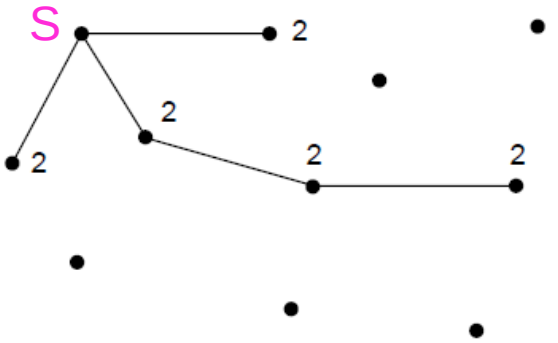
Network with groups 1 & 2



Spanning tree from source S



Multicast tree from S to group 1



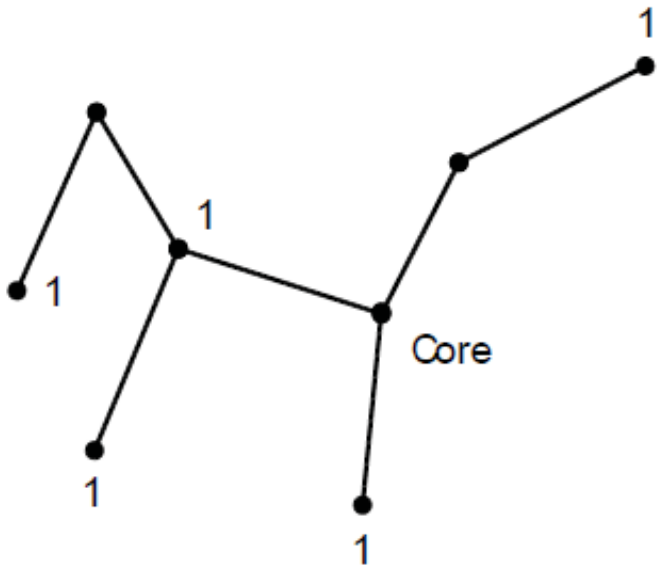
Multicast tree from S to group 2

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

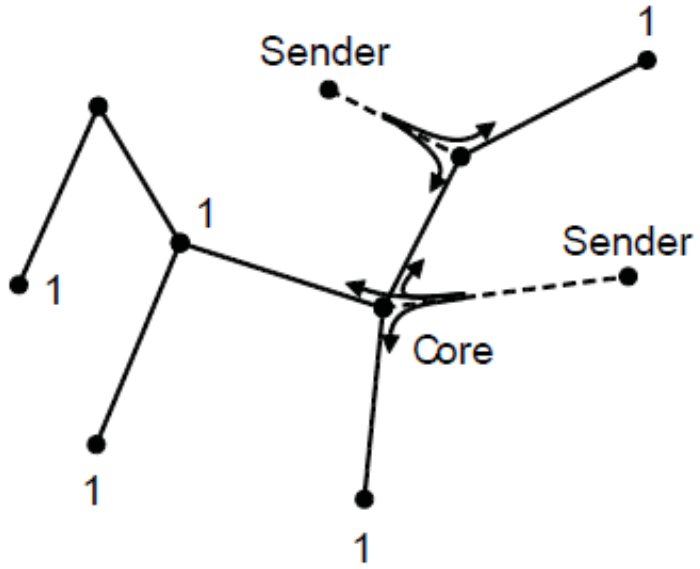
Multicast Routing (2) – Sparse Case

- CBT (Core-Based Tree) uses a single tree to multicast
  - Tree is the sink tree from core node to group members
  - Multicast heads to the core until it reaches the CBT
- p 1.

Un alternativa puó essere tramite code-based tree. C



Sink tree from core to group 1



Multicast is send to the core then down when it reaches the sink tree

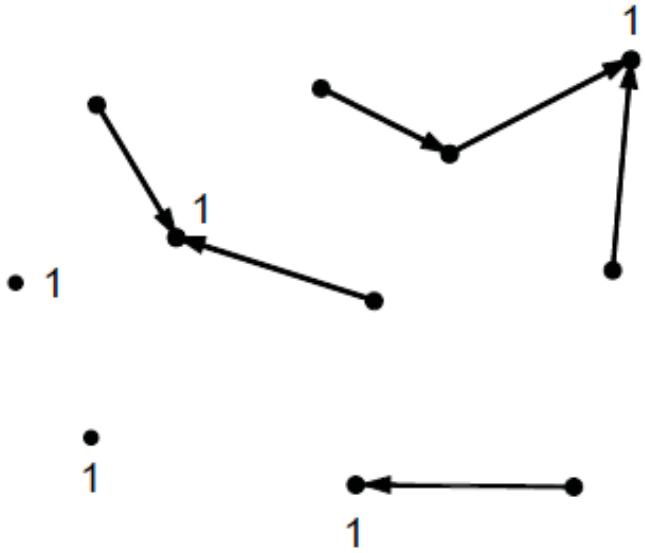
CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Anycasting consiste nel inviare al membro piú vicino appartenente al gruppo. L'unico vincolo é

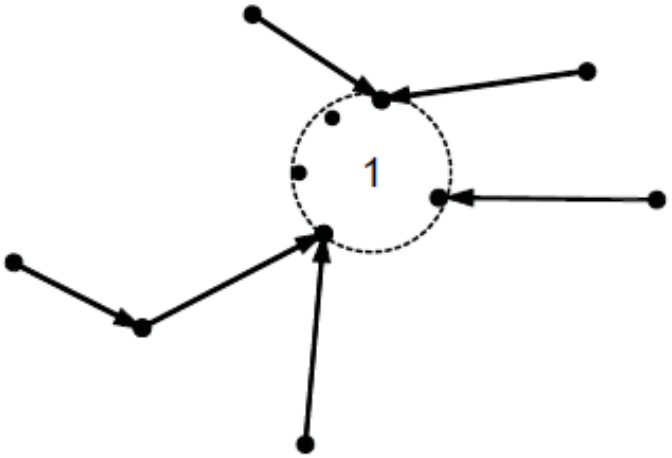
Anycast Routing

Anycast sends a packet to one (nearest) group member

- Falls out of regular routing with a node in many places



Anycast routes to group 1



Apparent topology of sink tree to "node" 1

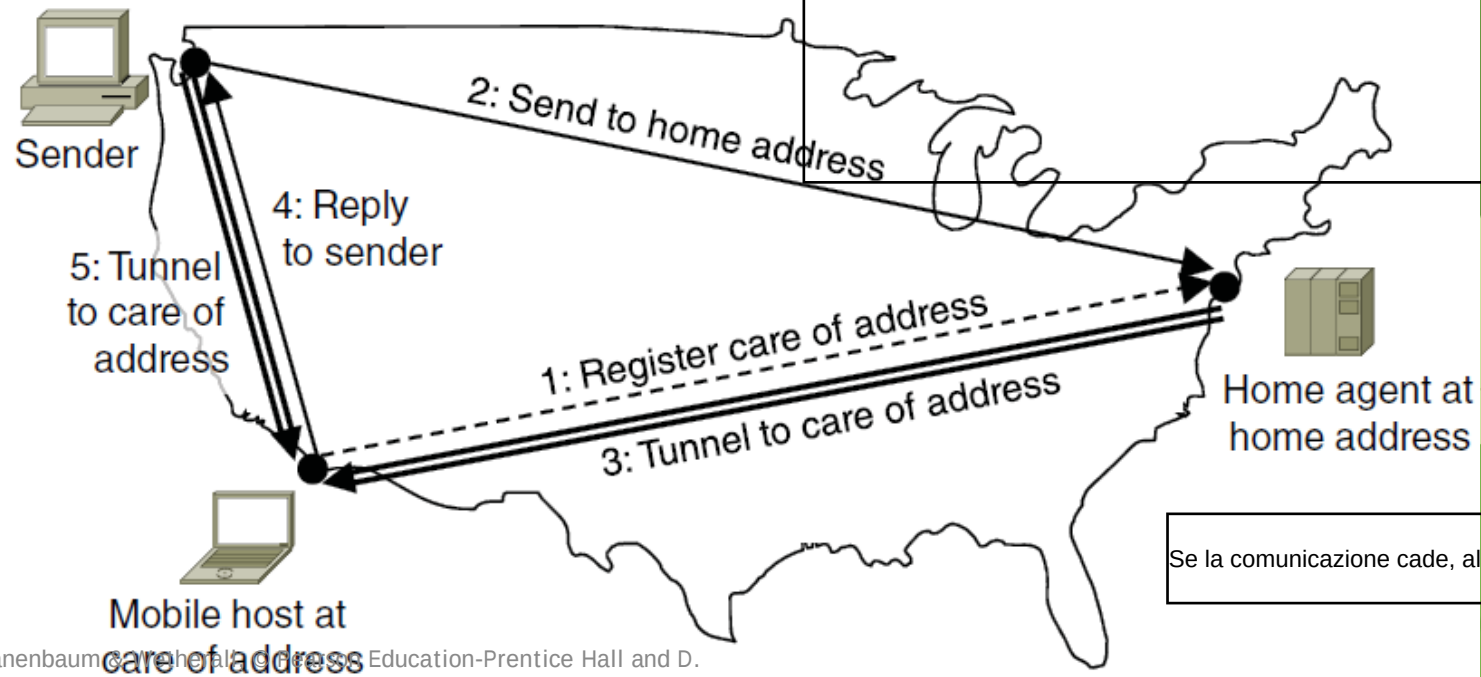
CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Routing for Mobile Hosts

Mobile Hosts sono dispositivi che possono cambiare posizion

- Mobile hosts can be reached via a home agent
  - Fixed home agent tunnels packets to reach the mobile host; reply can optimize path for subsequent packets
  - No changes to routers or fixed hosts

La comunicazione si stabilisce tramite una richesta di connessione alla rete, da pa

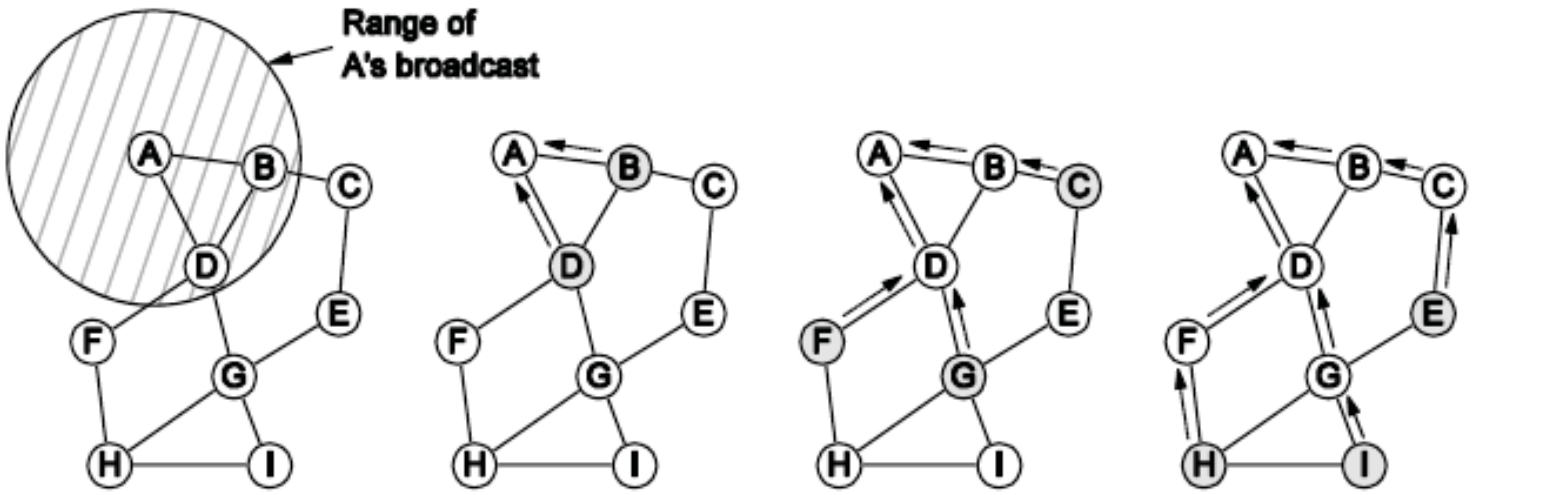


Se la comunicazione cade, allora puó semp

Routing in Ad Hoc Networks

Sono reti con nodi che possono sparire e aggiungersi e quindi con topologie molto variabili.

- The network topology changes as wireless nodes move
  - Routes are often made on demand, e.g., AODV (below)



A's starts to find route to *I*

A's broadcast reaches *B* & *D*

B's and *D*'s broadcast reach *C, F* & *G*

C's, *F*'s and *G*'s broadcast reach *H* & *I*

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Congestion Control (1)

Handling congestion is the responsibility of the Network and Transport layers working together

- We look at the Network portion here

- Traffic-aware routing »
- Admission control »
- Traffic throttling »
- Load shedding »

Si definisce congestione la situazione in cui la presenza di troppo pacchetti all'interno della

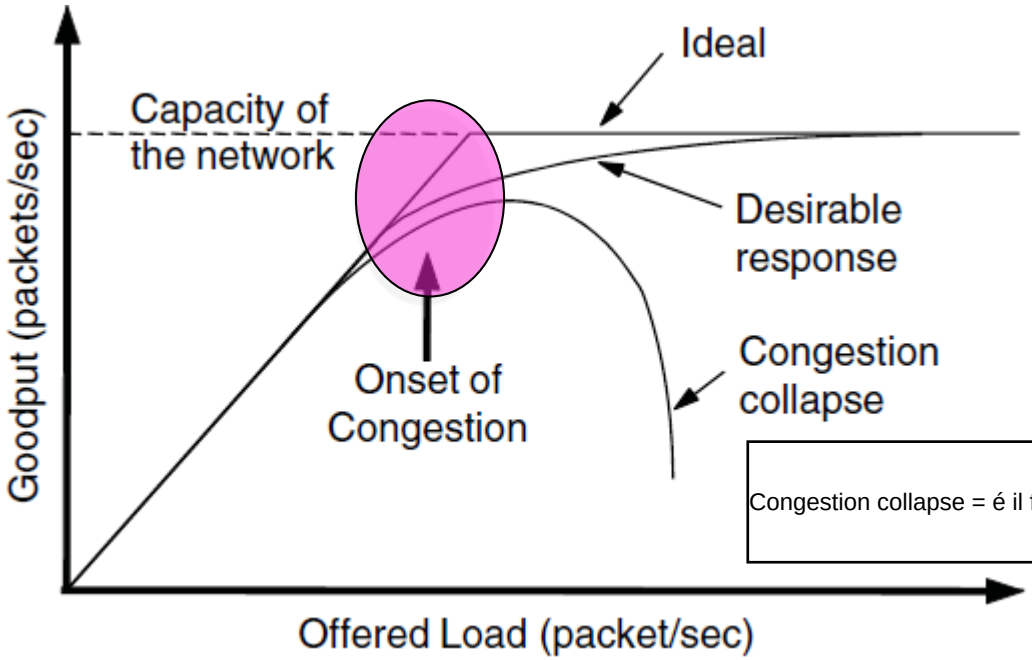
Congestion Control (2)

Congestion results when too much traffic is offered; performance degrades due to loss/retransmissions

- Goodput (=useful packets) trails offered load

Offered load = numero di pacchetti consegnati alla rete per

L'obbiettivo é creare una rete che eviti



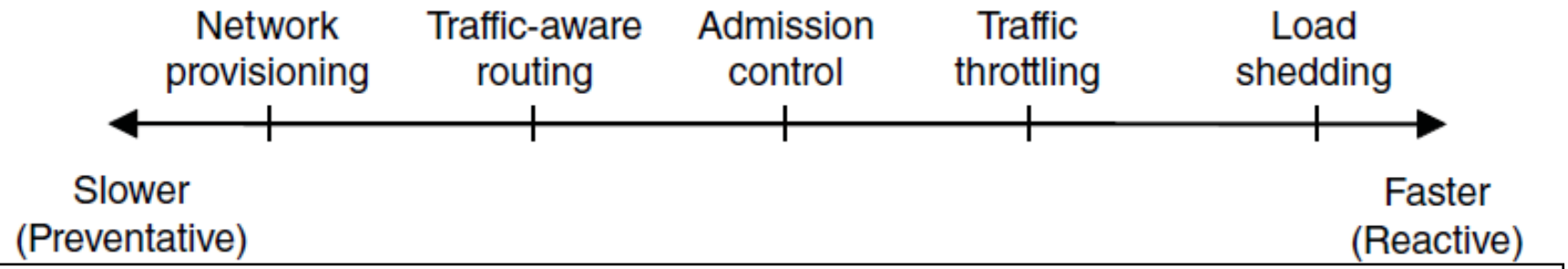
Congestion collapse = é il fenomeno di degradazione delle perforn

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Congestion Control (3) – Approaches

Network must do its best with the offered load

- Different approaches at different timescales
- Nodes should also reduce offered load (Transport)



Network rovvisioning: Aggiunta dinamica di (spare) router o di linee di backup o comprare bandwidth nell'open market.Traffic-Aware routing: Ripartire il traffico in piú ret

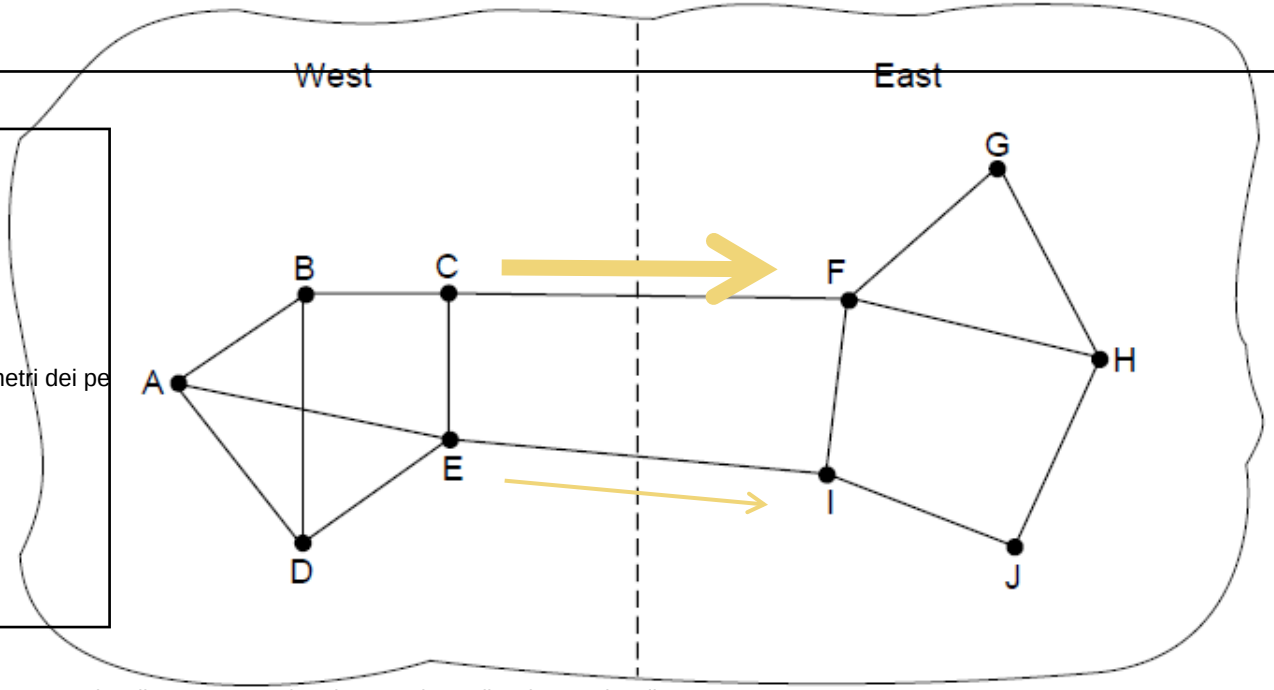
Traffic-Aware Routing

Choose routes depending on traffic, not just topology

- E.g., use *EI* for West-to-East traffic if *CF* is loaded
- But take care to avoid oscillations

La via piú semplice é assegnare dei pesi in funzione di capacitá di bandwidth, delay di propagazione, carico o media dei tempo di permanenza in queue (nel buffer). Il pe

Se si considera, come parametri dei pe



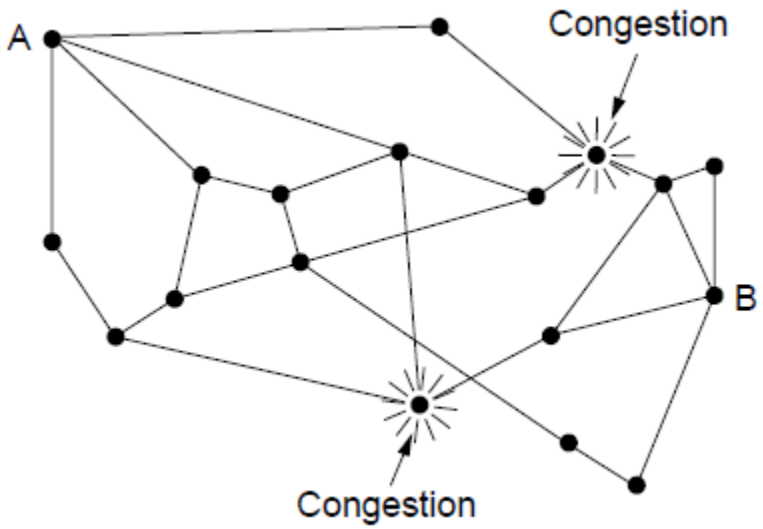
CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Admission Control

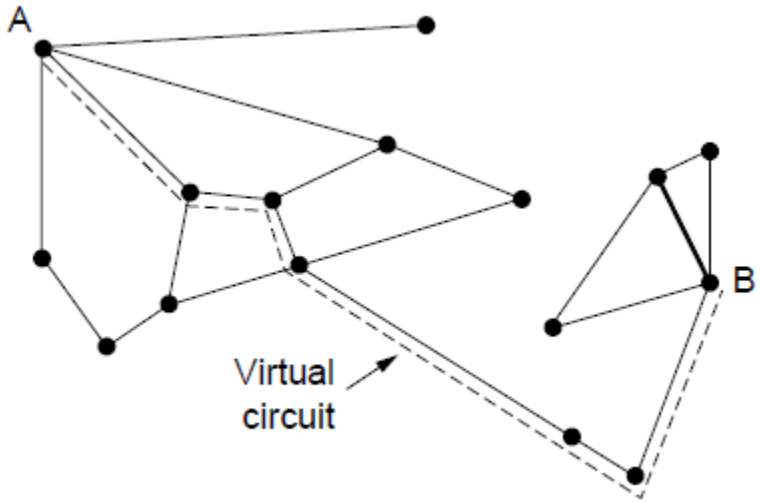
Admission control: ammettere nuove pacchetti se e solo se, con l'aggiunta di questi, non si incorre in congestione; utilizzata soprattutto per VC.Il traffico é tipicamente bursty (ad

- Admission control allows a new traffic load only if the network has sufficient capacity, e.g., with virtual circuits
  - Can combine with looking for an uncongested route

Armato di descrittore di traffico, la rete pió decidere se accettare traffico. Puó riservare abbastanza banda da evitare congestione. Il problema ora é calcolare quante connessio



Network with some congested nodes



Uncongested portion and route *AB* around congestion

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

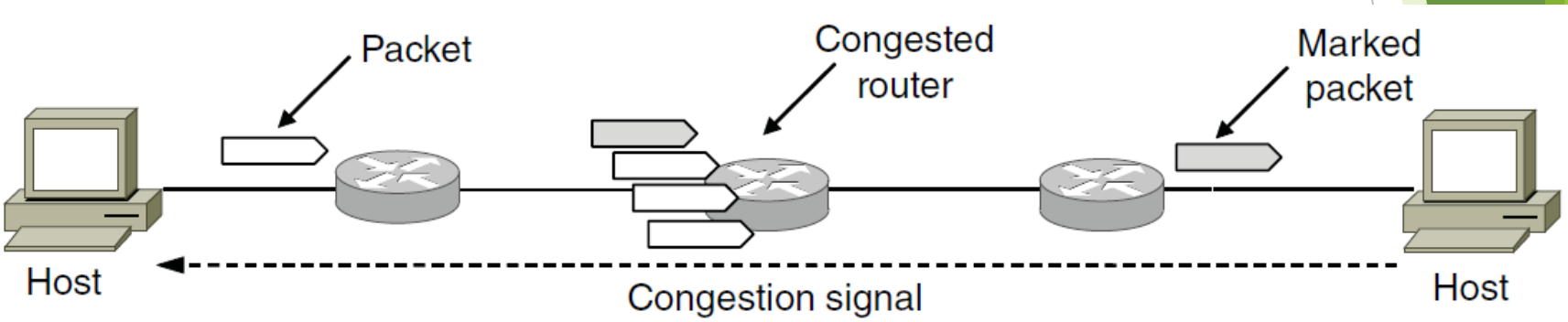
Puó essere utilizzato sia in reti di datagrammi che VC. Prima di tutto i router devono essere in grado di rilevare le congestioni, idealmente prima che si manifestino. Ci s

Traffic Throttling

I router devono comunicare messaggi di feedback per tempo ai mittenti che generano congesione. Questi messaggi devono essere minimali, altrimenti peggiorano la c

Congested routers signal hosts to slow down traffic

- ECN (Explicit Congestion Notification) marks packets and receiver returns signal to sender



Il metodo piú diretta é utilizzare dei "choke packets". In reti a datagrammi, i router congestionati, accettano randomicamente i pacchetti, causando choke packets che verranno in

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

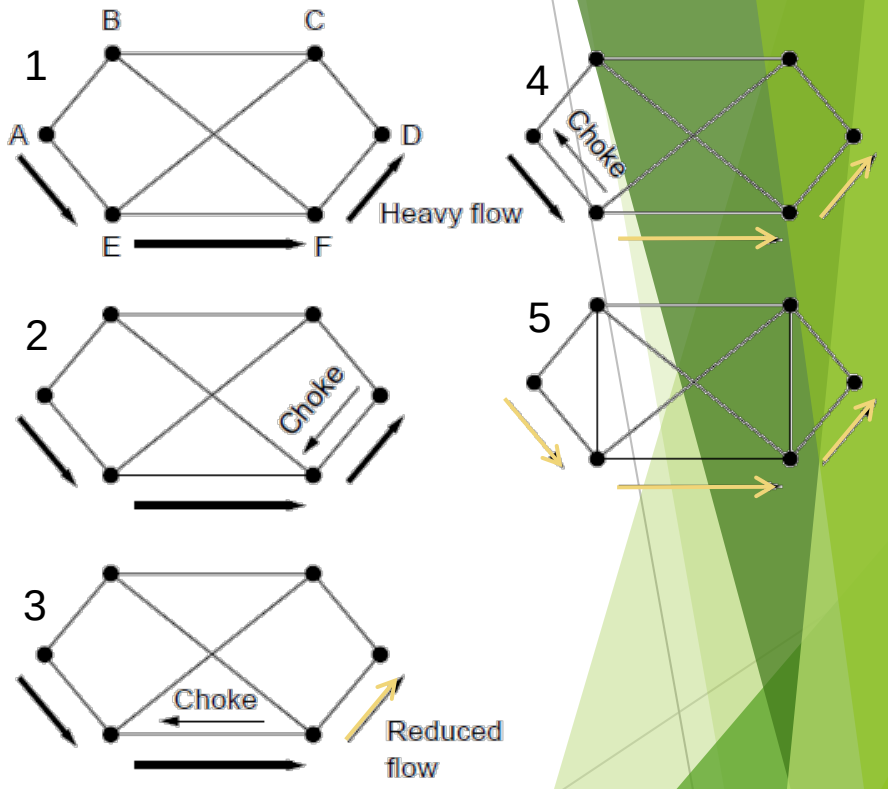
Un altro metodo é Hop-by-Hop Backpressure. Ogni qual volta un router rileva congestione, lo comunica al router precedente. Quest'ultimo limiterá immediatamente il th

Load Shedding (1)

When all else fails, network will drop packets (shed load)

Can be done end-to-end or link-by-link

Link-by-link (right) produces rapid relief

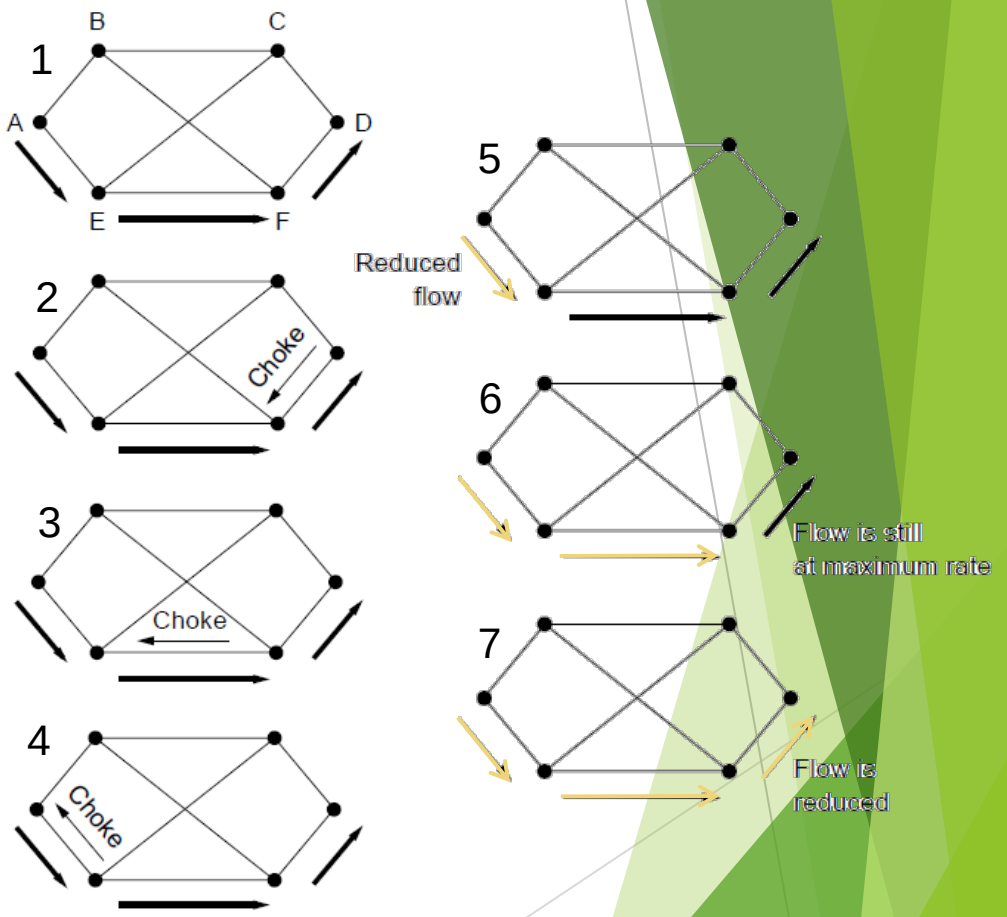


Metodo da utilizzare quando tutti gli altri hanno fallito. Una politica di scarto é segnare i

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Load Shedding (2)

End-to-end (right) takes longer to have an effect, but can better target the cause of congestion



CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Quality of Service

- Application requirements »
- Traffic shaping »
- Packet scheduling »
- Admission control »
- Integrated services »
- Differentiated services »

Fino ad adesso si é discusso di come la rete possa comportarsi sotto certe circostanze. Alcune applicazioni necessitano dei requisiti minimi entro i

Application Requirements (1)

Different applications care about different properties

- We want all applications to get what they need

- I parametri principali sono quelli presenti in tabella.Jitter: variabilitá nel delay o nel te

Application	Bandwidth	Delay	Jitter	Loss
Email	Low	Low	Low	Medium
File sharing	High	Low	Low	Medium
Web access	Medium	Medium	Low	Medium
Remote login	Low	Medium	Medium	Medium
Audio on demand	Low	Low	High	Low
Video on demand	High	Low	High	Low
Telephony	Low	High	High	Low
Videoconferencing	High	High	High	Low

"High" means a demanding requirement, e.g., low delay

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Application Requirements (2)

Network provides service with different kinds of QoS (Quality of Service) to meet application requirements

Network Service	Application
Constant bit rate	Telephony
Real-time variable bit rate	Videoconferencing
Non-real-time variable bit rate	Streaming a movie
Available bit rate	File transfer

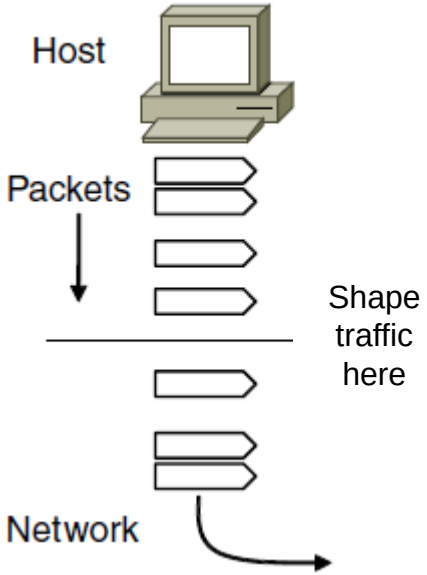
Example of QoS categories from ATM networks

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Traffic Shaping (1)

Traffic shaping regulates the average rate and burstiness of data entering the network

- Lets us make guarantees



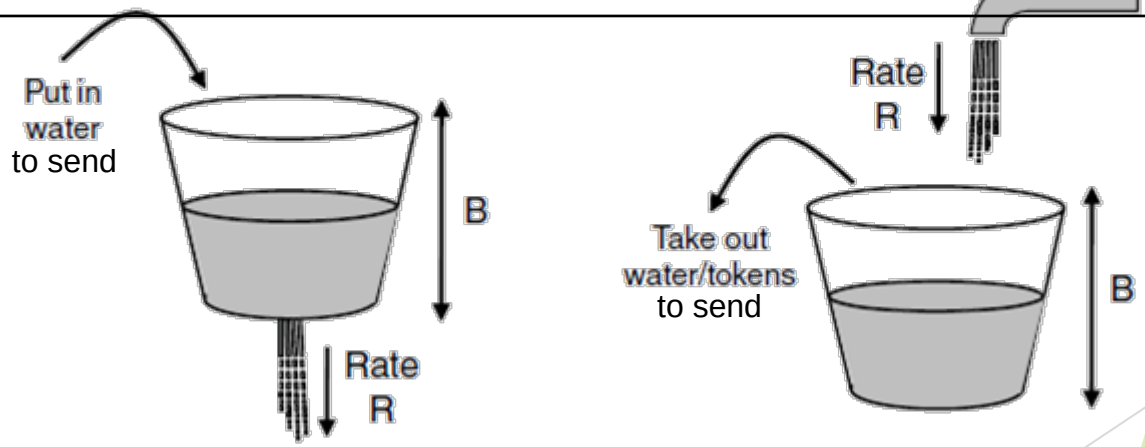
L'obiettivo é permettere alle applicazioni di trasmettere una larga scelta di traffici, che vada bene p

Traffic Shaping (2)

Token/Leaky bucket limits both the average rate (R) and short-term burst (B) of traffic

- For token, bucket size is B, water enters at rate R and is removed to send; opposite for leaky.

Leaky bucket algorithm: a prescindere da quanta acqua é presente nel secchio, essa esce a flusso costante R. Se il secchio di riempe, l'acqua esce dai bordi (ed é qundi pers



Leaky bucket (need not full to send)

Token bucket (need some water to send)

Traffic Shaping (3)

Host traffic
R=200 Mbps
B=16000 KB

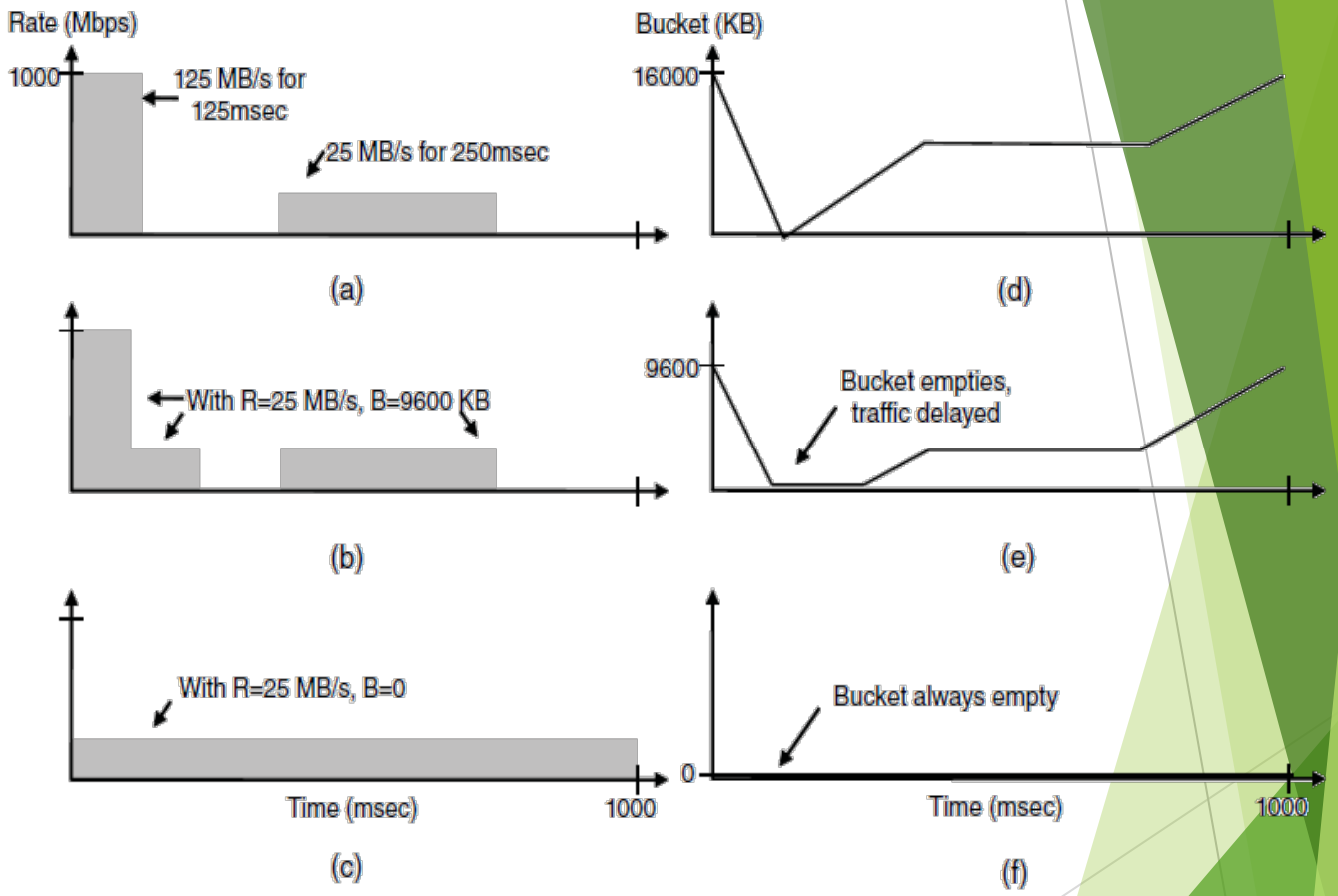
↓

Shaped by
R=200 Mbps
B=9600 KB

↓

Shaped by
R=200 Mbps
B=0 KB

Esempi con token bucket



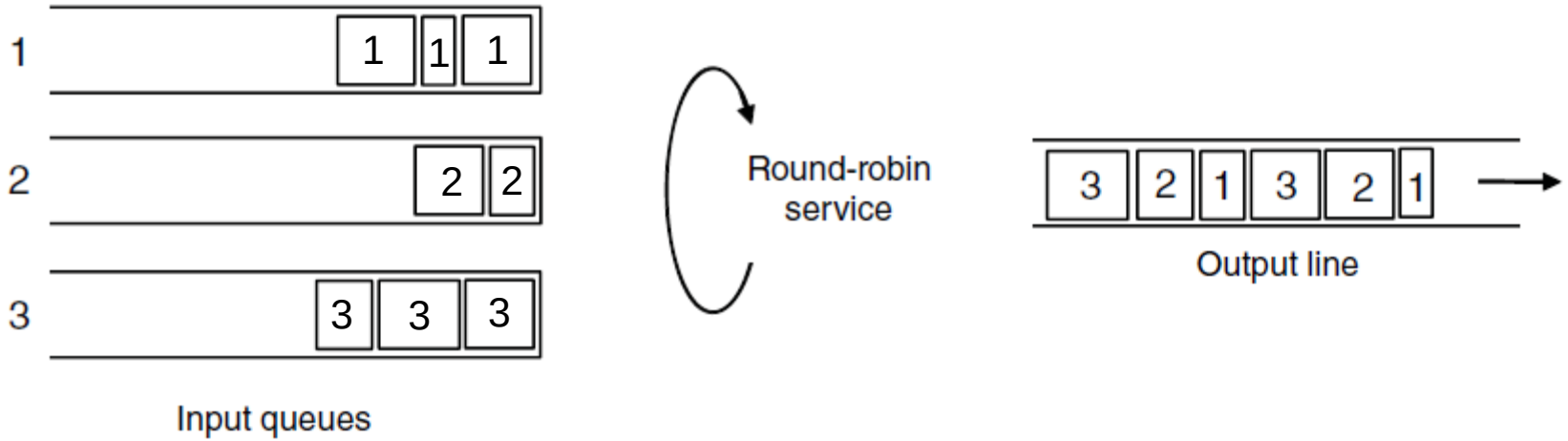
Smaller bucket size delays traffic and reduces burstiness

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Packet Scheduling (1)

Packet scheduling divides router/link resources among traffic flows with alternatives to FIFO (First In First Out)

Per fornire una garanzia di prestazioni, bisogna allocare risorse sufficienti su tutto il percorso fino a destinazione (il libro assume che tutti i pacchetti di un flusso, seguono lo s



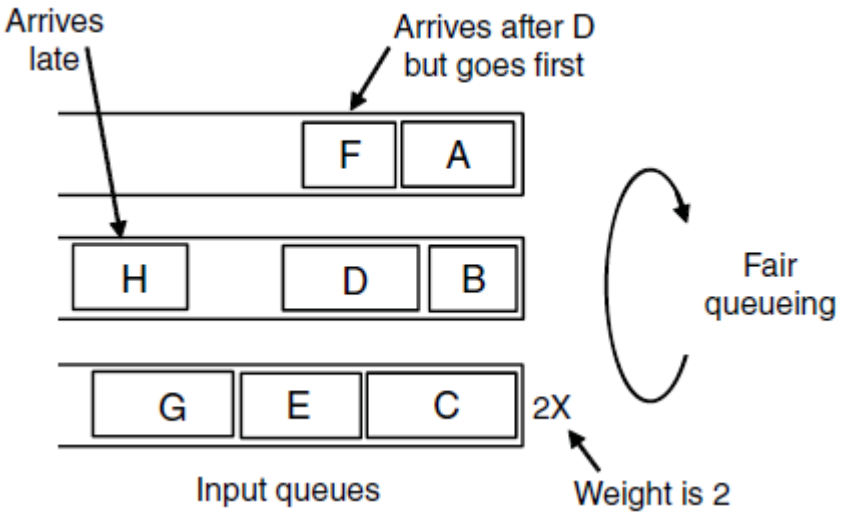
Example of round-robin queuing

Un algoritmo puó essere il fifo (sconsigliato poiché genera non rispetta i parametri di QoS) oppure il Fair Queueing (Round Robin) che puó essere eseguito su pacchetti o su b

Packet Scheduling (2)

Fair Queueing approximates bit-level fairness with different packet sizes; weights change target levels

- Result is WFQ (Weighted Fair Queueing)



Packet	Arrival time	Length	Finish time	Output order
A	0	8	8	1
B	5	6	11	3
C	5	10	10	2
D	8	9	20	7
E	8	8	14	4
F	10	6	16	5
G	11	10	19	6
H	20	8	28	8

$$F_i = \max(A_i, F_{i-1}) + L_i/W$$

Packets may be sent out of arrival order

Finish virtual times determine transmission order

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Admission Control (1)

Quelli fin ora sono i meccanismi base, ora si mettono in atto. Il client esegue

Admission control takes a traffic flow specification and decides whether the network can carry it

- Sets up packet scheduling to meet QoS

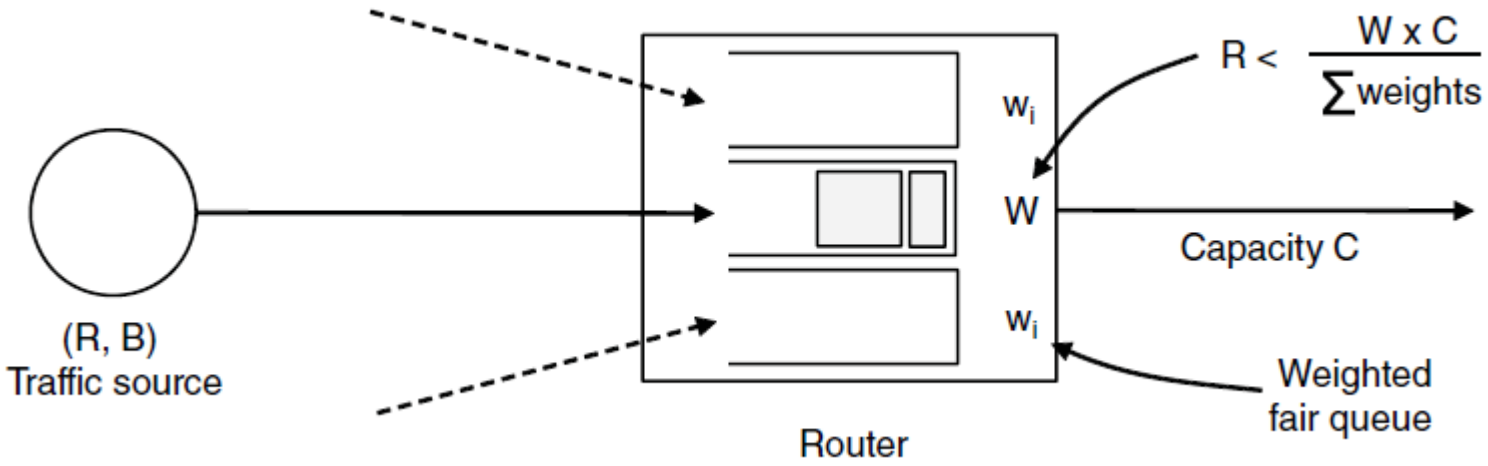
Parameter	Unit
Token bucket rate	Bytes/sec
Token bucket size	Bytes
Peak data rate	Bytes/sec
Minimum packet size	Bytes
Maximum packet size	Bytes

Example flow specification

Admission Control (2)

Construction to guarantee bandwidth B and delay D:

- Shape traffic source to a (R, B) token bucket
- Run WFQ with weight W / all weights > R/capacity
- Holds for all traffic patterns, all topologies



CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Integrated Services (1)

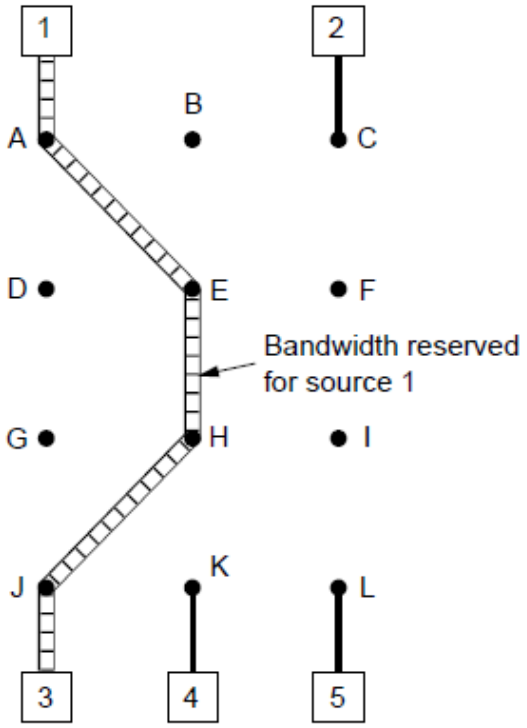
Design with QoS for each flow; handles multicast traffic.

Admission with RSVP (Resource reSerVation Protocol):

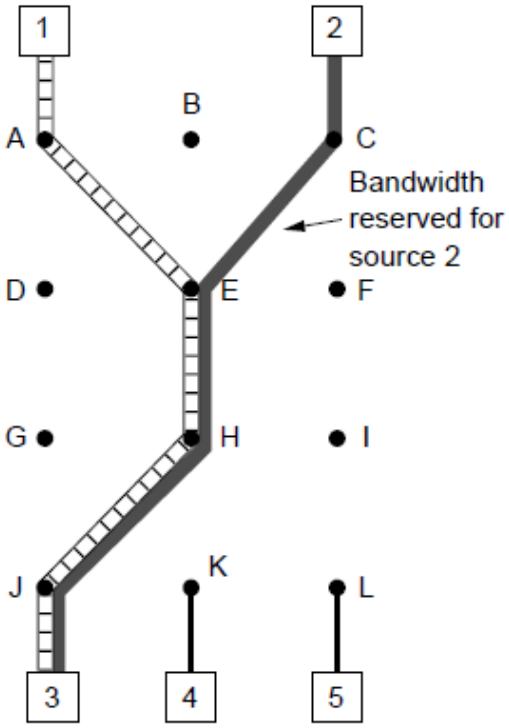
- Receiver sends a request back to the sender
- Each router along the way reserves resources
- Routers merge multiple requests for same flow
- Entire path is set up, or reservation not made

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

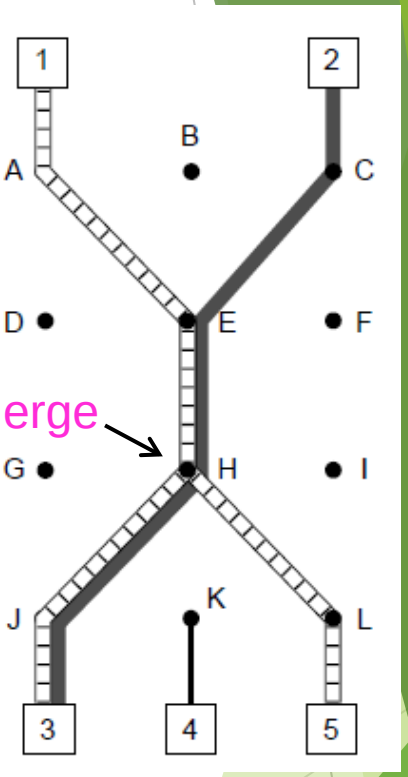
Integrated Services (2)



R3 reserves flow from S1



R3 reserves flow from S2

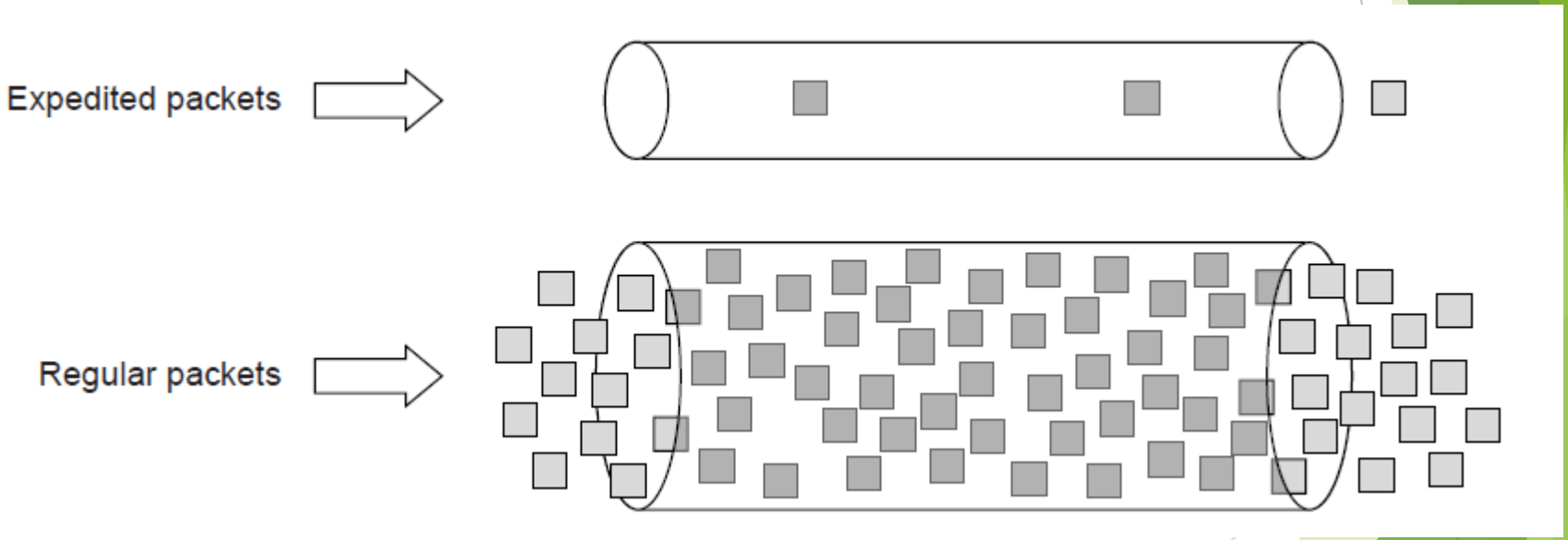


R5 reserves flow from S1; merged with R3 at H

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Differentiated Services (1)

- Design with classes of QoS; customers buy what they want
  - Expedited class is sent in preference to regular class
  - Less expedited traffic but better quality for applications

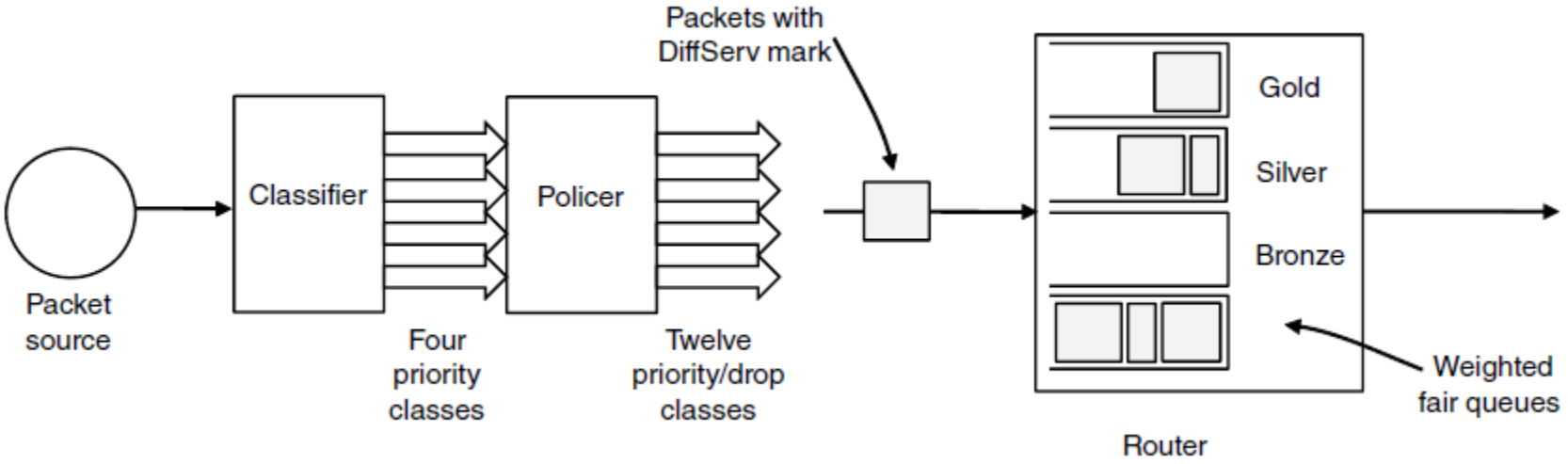


CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Differentiated Services (2)

Implementation of DiffServ:

- Customers mark desired class on packet
- ISP shapes traffic to ensure markings are paid for
- Routers use WFQ to give different service levels



CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Internetworking

Internetworking joins multiple, different networks into a single larger network

- How networks differ »
- How networks can be connected »
- Tunneling »
- Internetwork routing »
- Packet fragmentation »

How Networks Differ

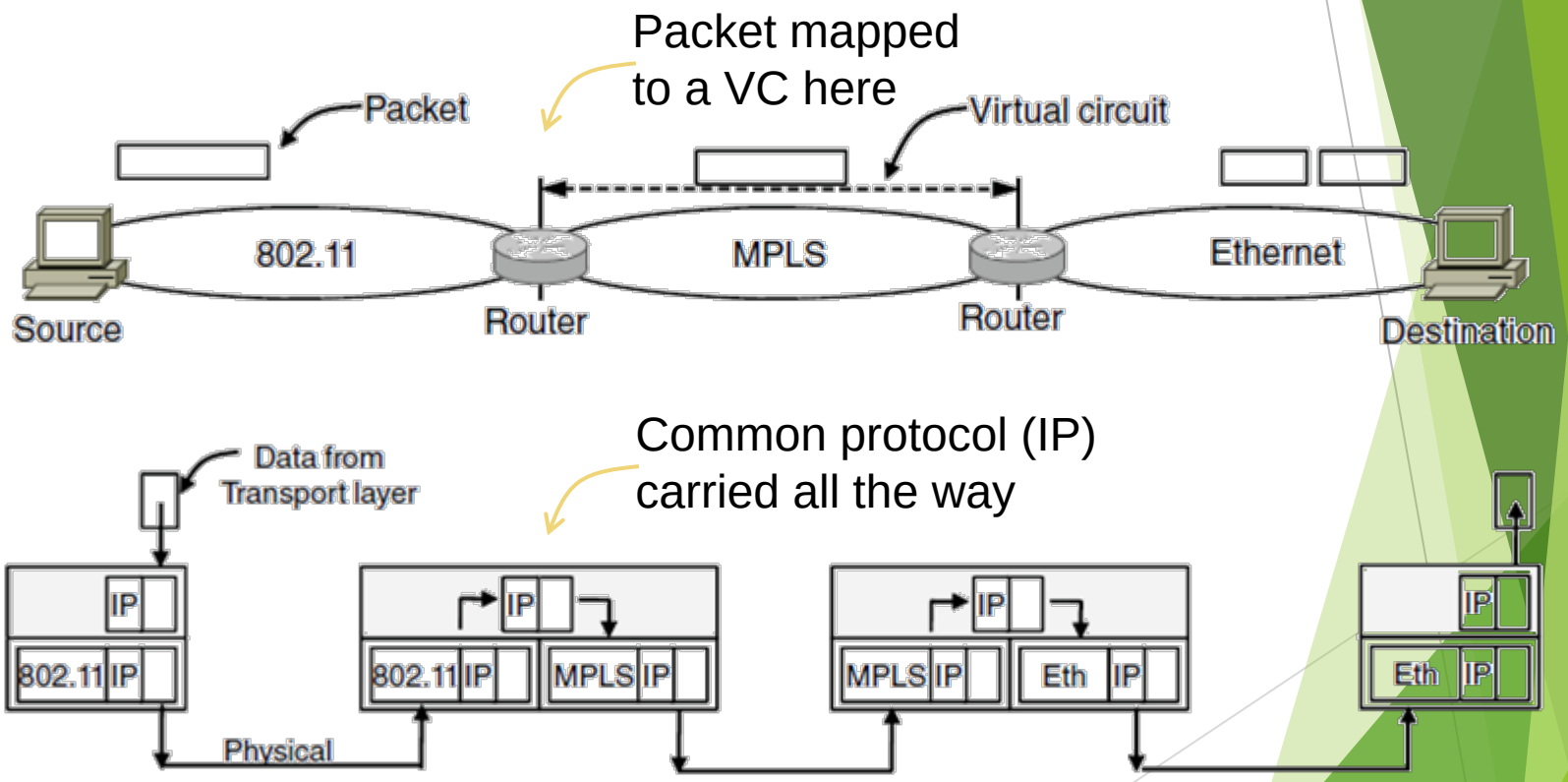
Differences can be large; complicates internetworking

Item	Some Possibilities
Service offered	Connectionless versus connection oriented
Addressing	Different sizes, flat or hierarchical
Broadcasting	Present or absent (also multicast)
Packet size	Every network has its own maximum
Ordering	Ordered and unordered delivery
Quality of service	Present or absent; many different kinds
Reliability	Different levels of loss
Security	Privacy rules, encryption, etc.
Parameters	Different timeouts, flow specifications, etc.
Accounting	By connect time, packet, byte, or not at all

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

How Networks Can Be Connected

- Internetworking based on a common network layer – IP

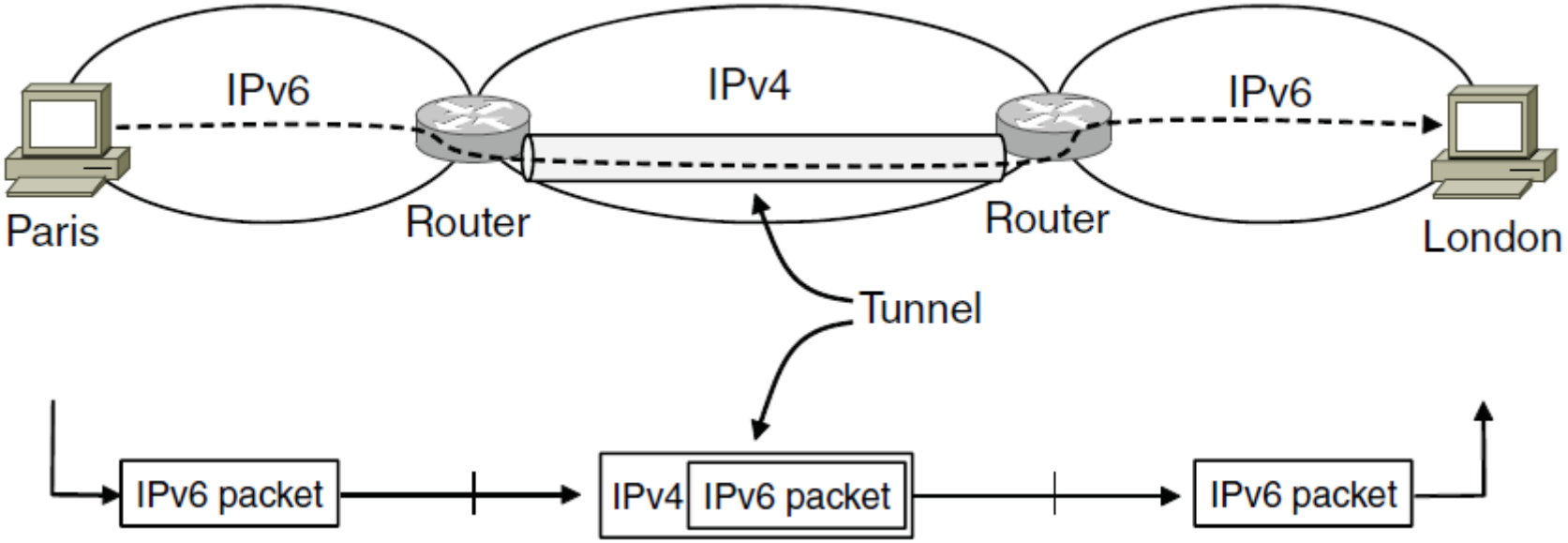


CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Tunneling (1)

Connects two networks through a middle one

- Packets are encapsulates over the middle

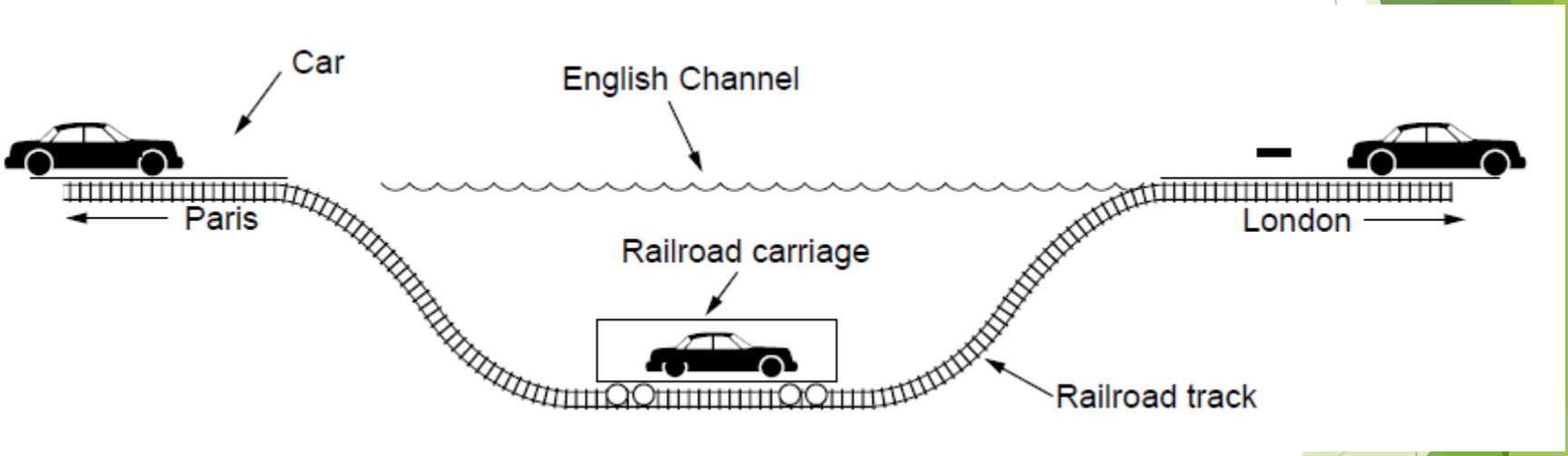


CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Tunneling (2)

Tunneling analogy:

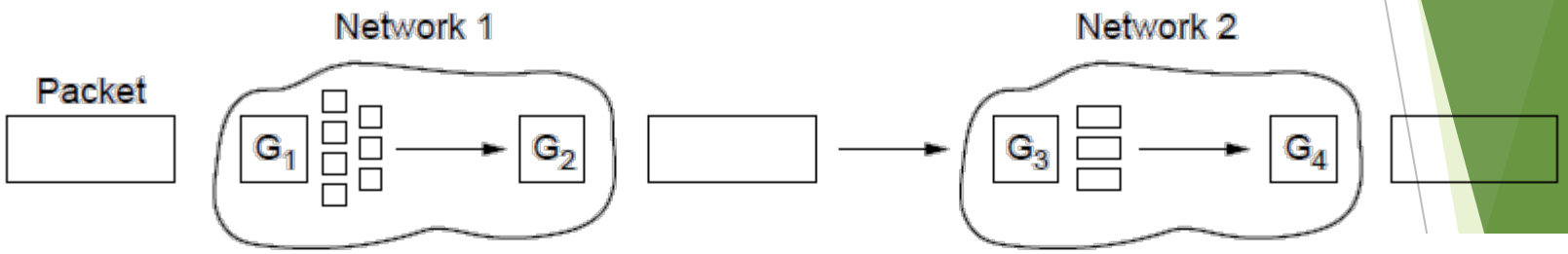
- tunnel is a link; packet can only enter/exit at ends



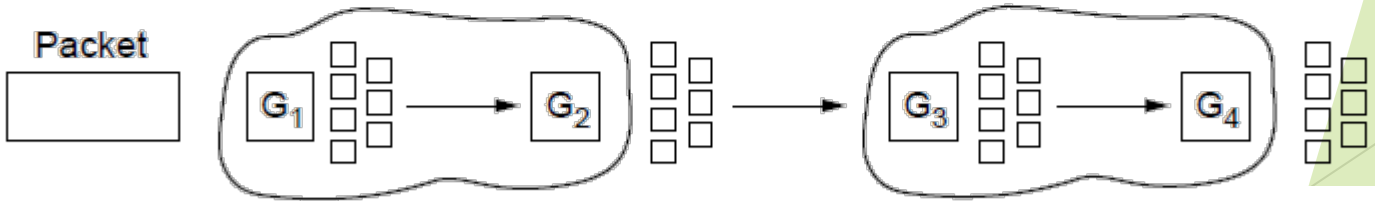
CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Packet Fragmentation (1)

- Networks have different packet size limits for many reasons
  - Large packets sent with fragmentation & reassembly



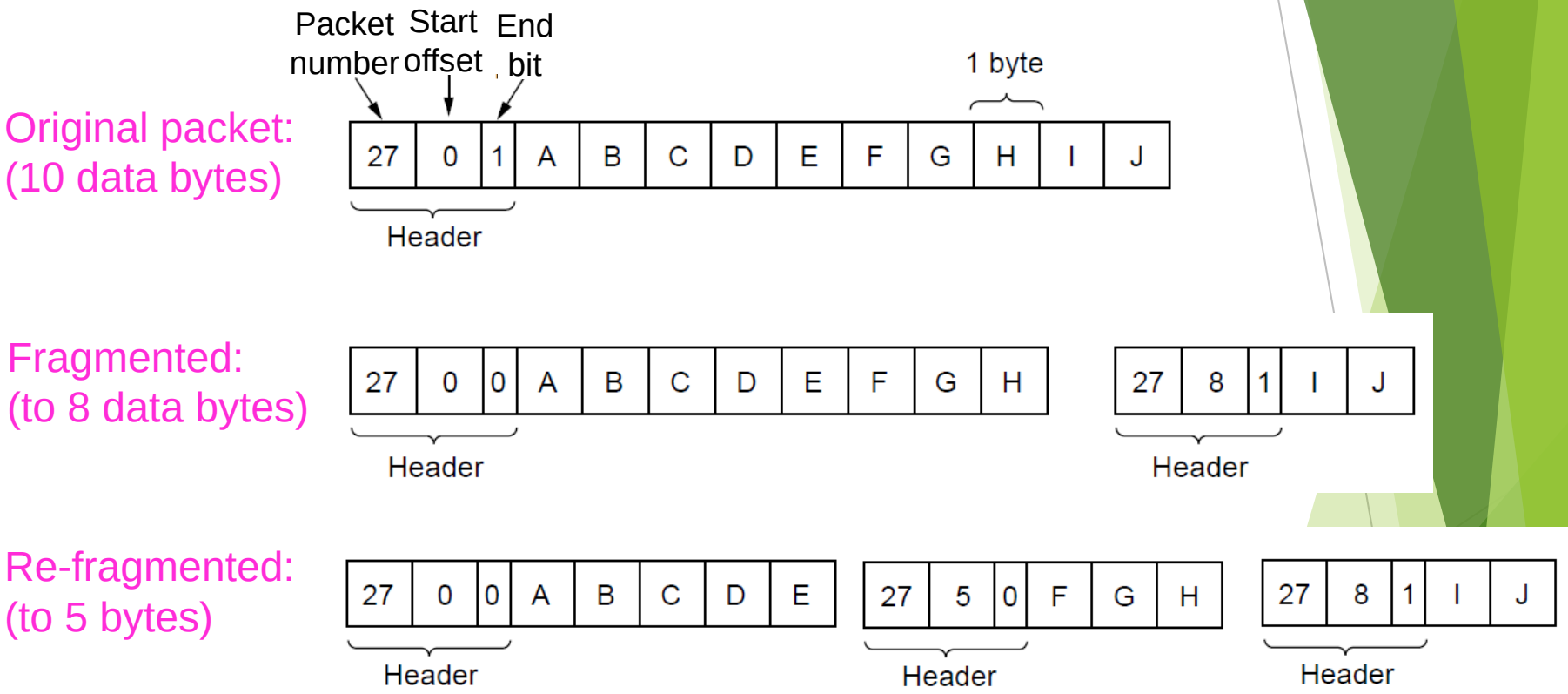
Transparent – packets fragmented / reassembled in each network



Non-transparent – fragments are reassembled at destination

Packet Fragmentation (2)

- Example of IP-style fragmentation:

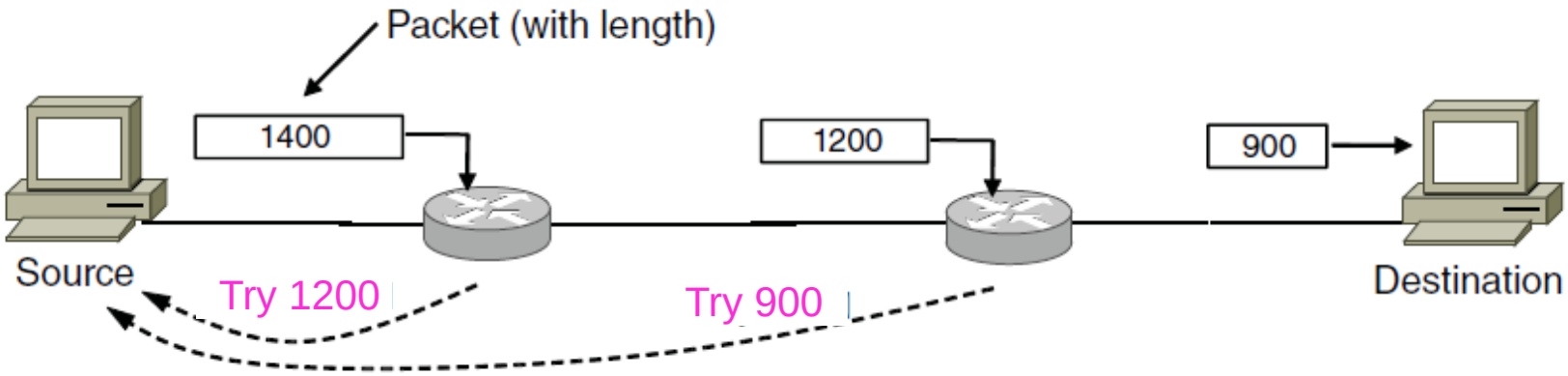


CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Packet Fragmentation (3)

Path MTU Discovery avoids network fragmentation

- Routers return MTU (Max. Transmission Unit) to source and discard large packets



CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Network Layer in the Internet (1)

- IP Version 4 »
- IP Addresses »
- IP Version 6 »
- Internet Control Protocols »
- Label Switching and MPLS »
- OSPF—An Interior Gateway Routing Protocol »
- BGP—The Exterior Gateway Routing Protocol »
- Internet Multicasting »
- Mobile IP »

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Network Layer in the Internet (2)

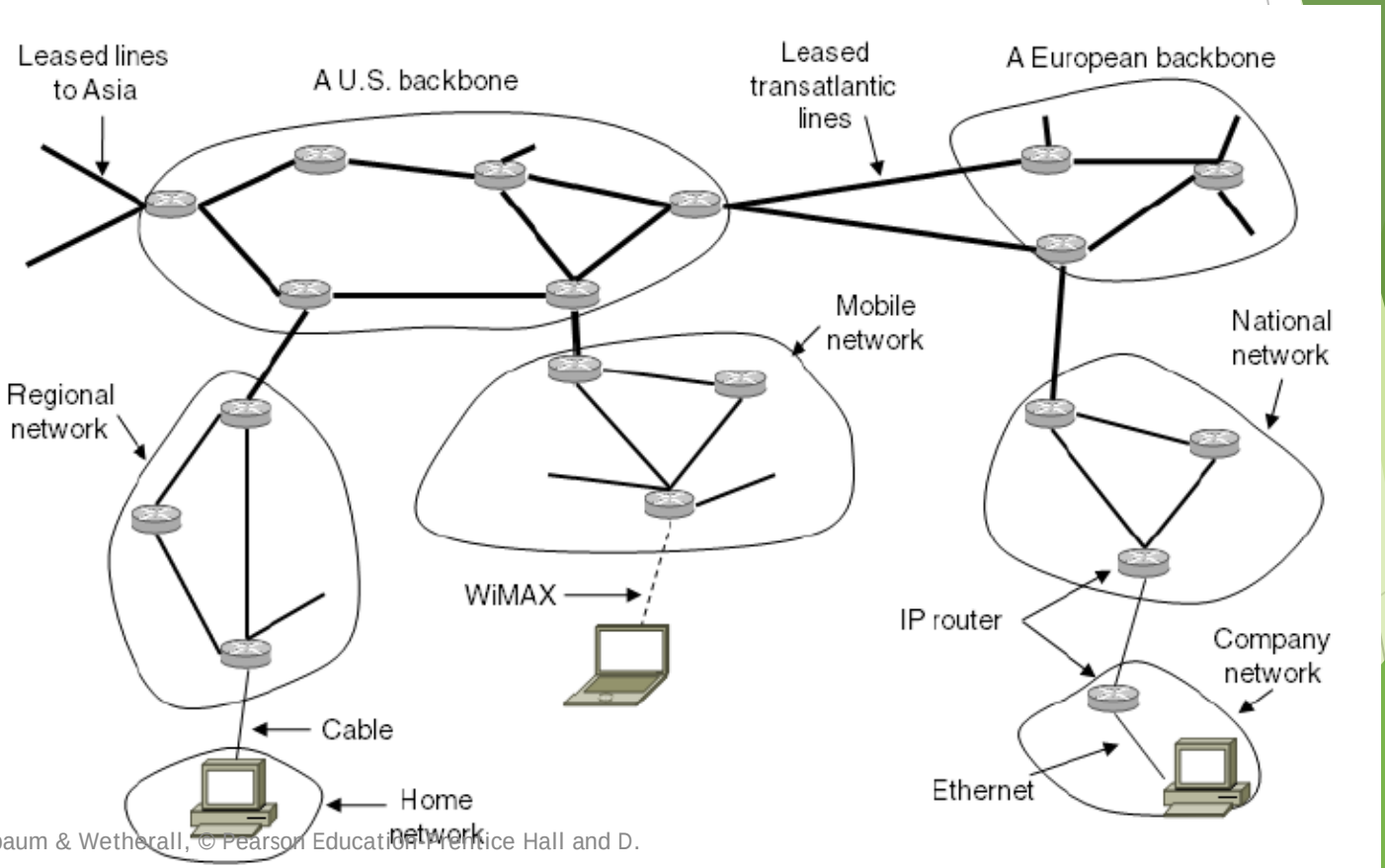
IP has been shaped by guiding principles:

- Make sure it works
- Keep it simple
- Make clear choices
- Exploit modularity
- Expect heterogeneity
- Avoid static options and parameters
- Look for good design (not perfect)
- Strict sending, tolerant receiving
- Think about scalability
- Consider performance and cost

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Network Layer in the Internet (3)

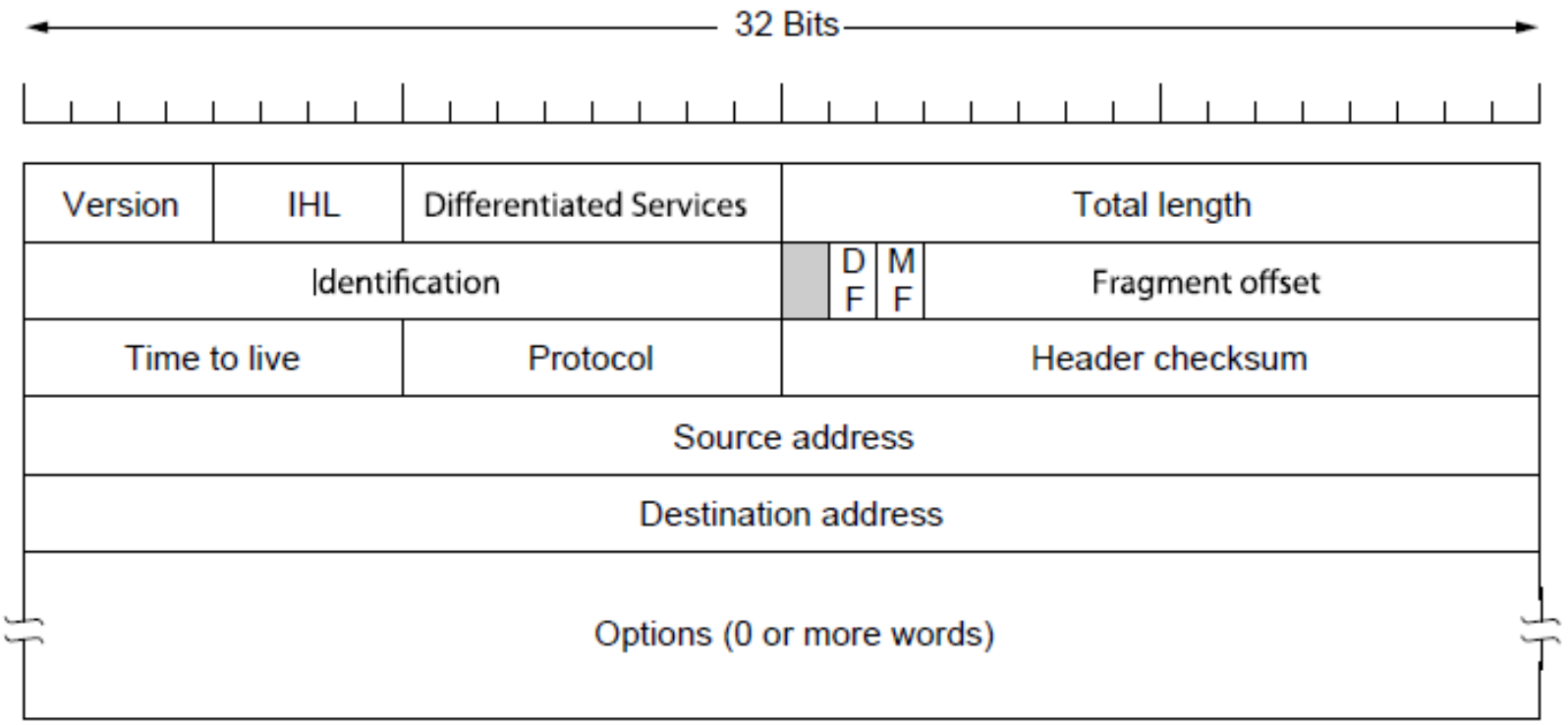
- Internet is an interconnected collection of many networks that is held together by the IP protocol



CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

IP Version 4 Protocol (1)

IPv4 (Internet Protocol) header is carried on all packets and has fields for the key parts of the protocol:

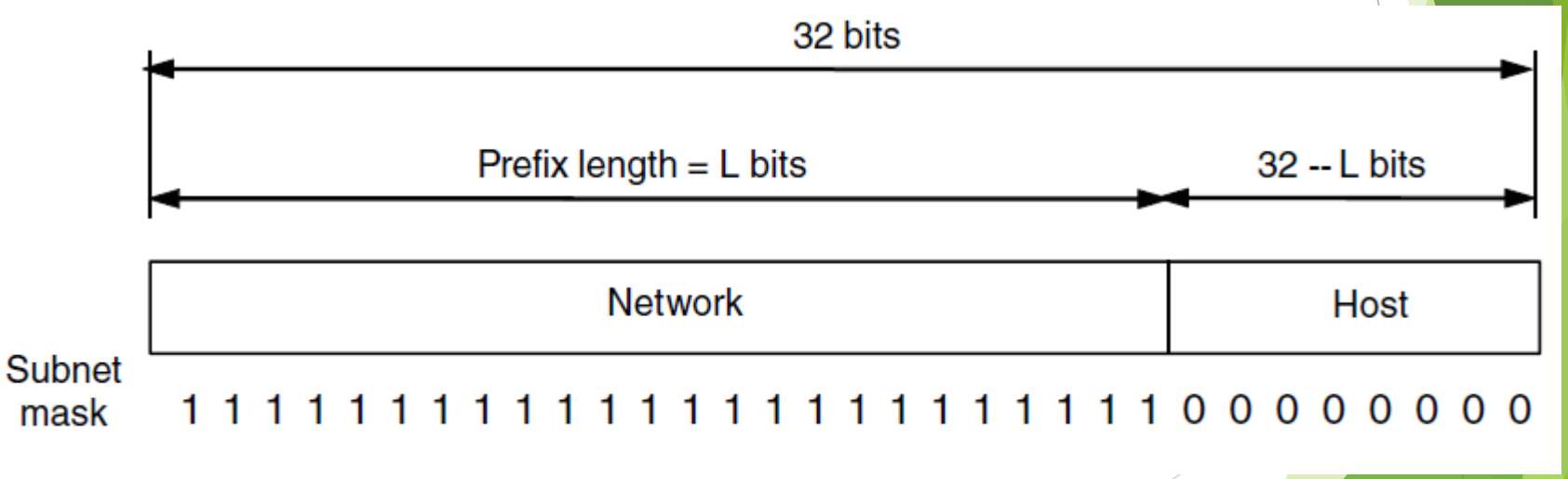


CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

IP Addresses (1) – Prefixes

Addresses are allocated in blocks called prefixes

- Prefix is determined by the network portion
- Has 2^L addresses aligned on 2^L boundary
- Written address/length, e.g., 18.0.31.0/24

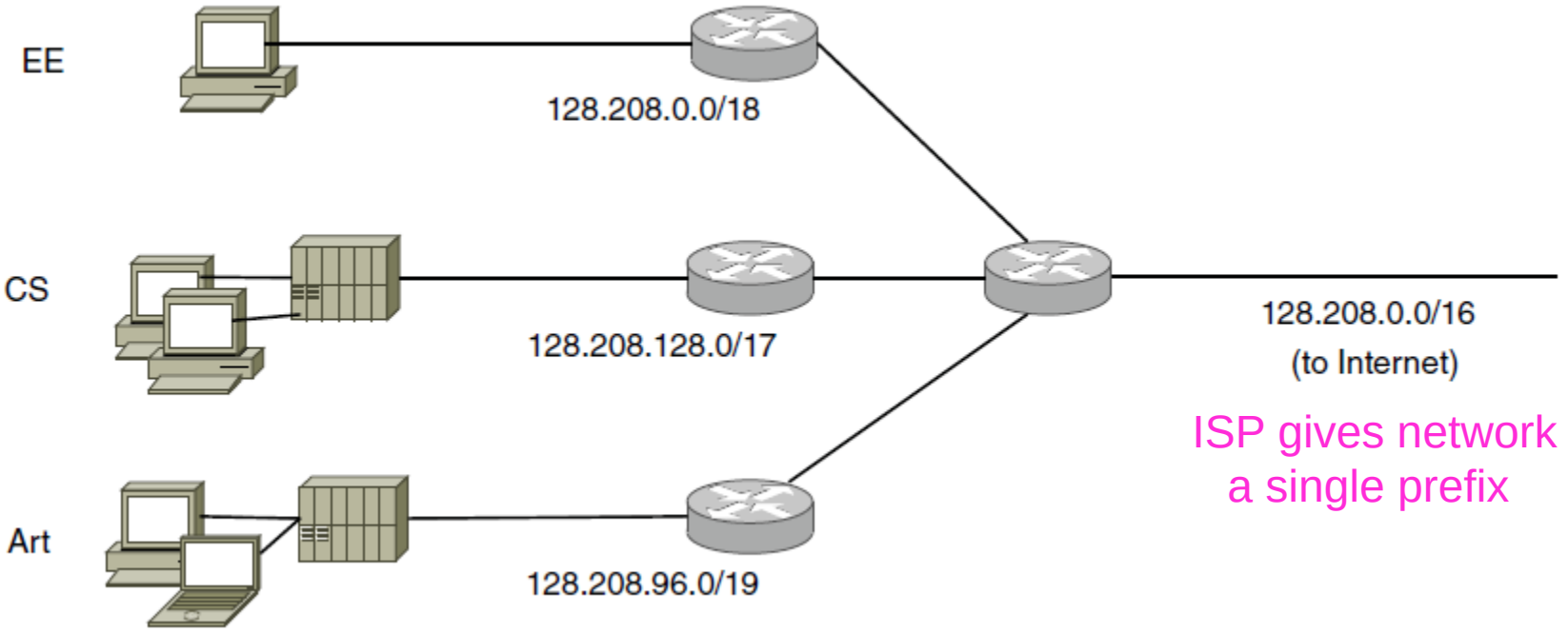


CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

IP Addresses (2) – Subnets

Subnetting splits up IP prefix to help with management

- Looks like a single prefix outside the network



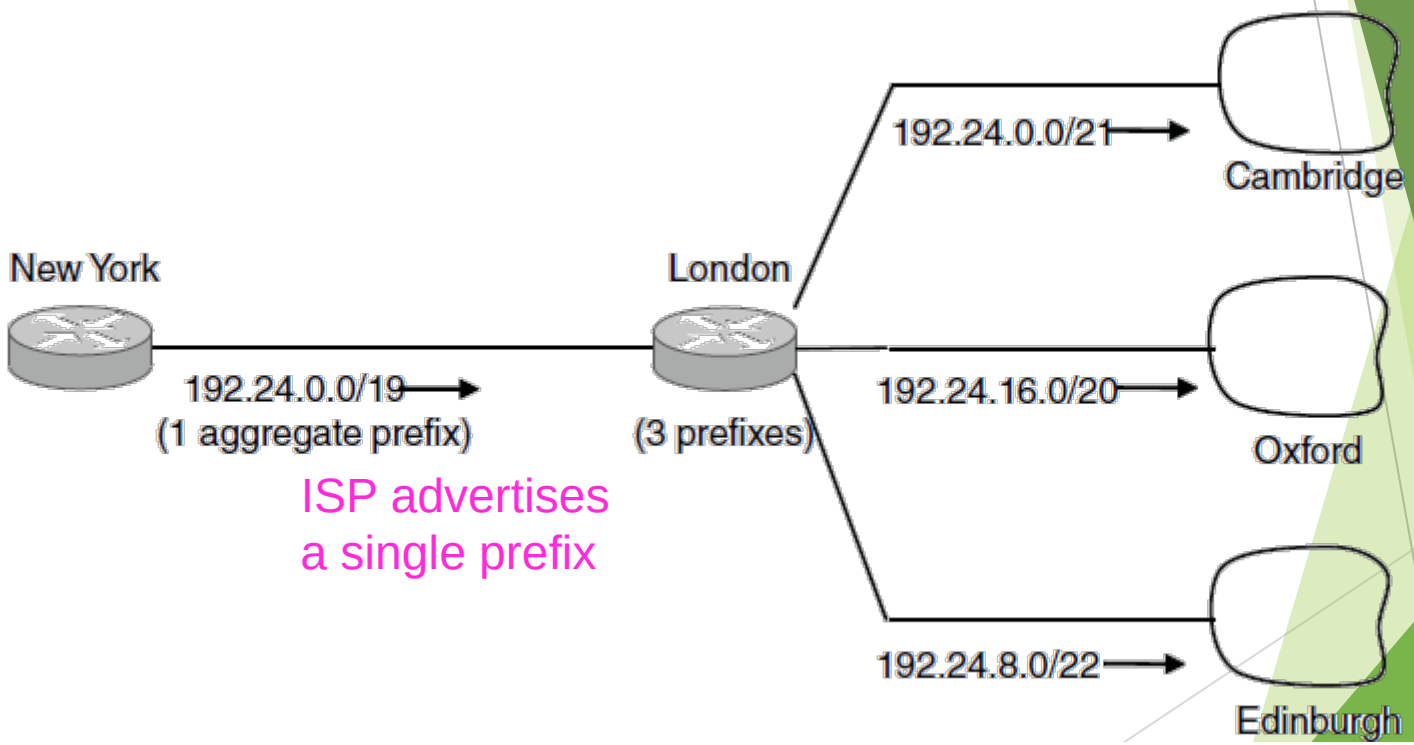
ISP gives network a single prefix

Network divides it into subnets internally

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

IP Addresses (3) – Aggregation

Aggregation joins multiple IP prefixes into a single larger prefix to reduce routing table size



ISP advertises a single prefix

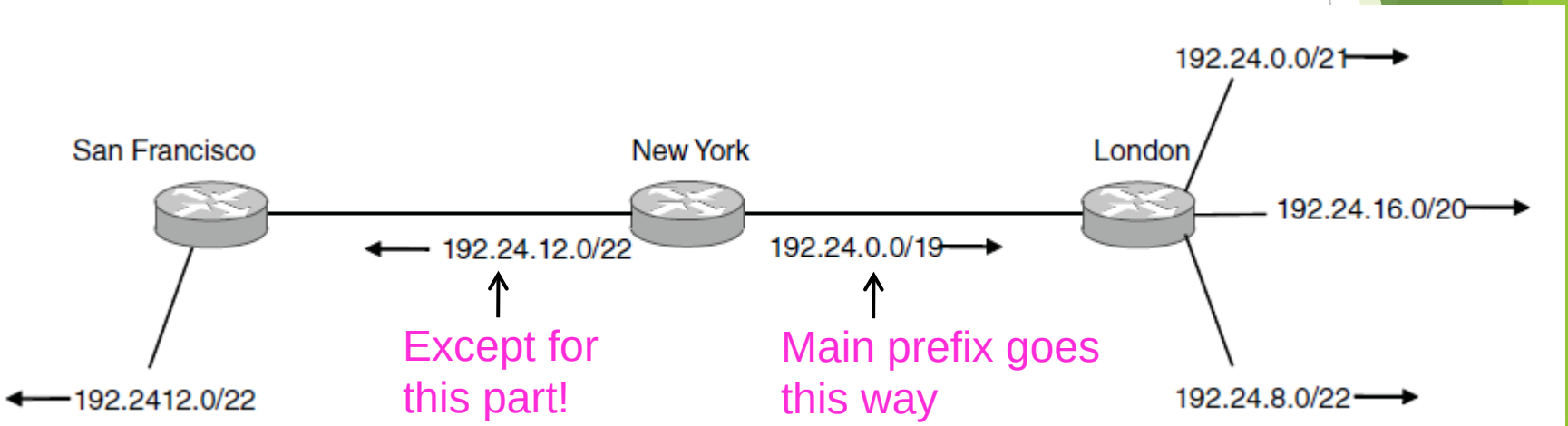
ISP customers have different prefixes

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

IP Addresses (4) – Longest Matching Prefix

Packets are forwarded to the entry with the longest matching prefix or smallest address block

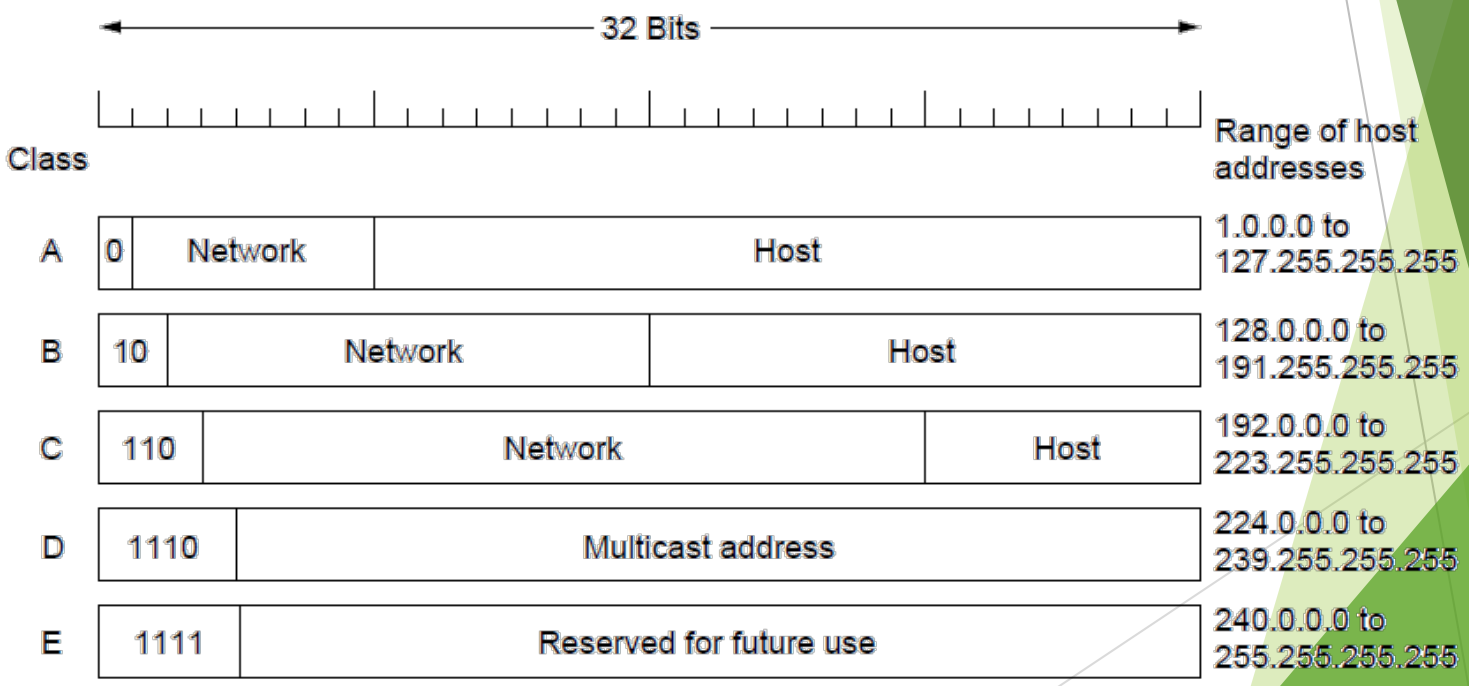
- Complicates forwarding but adds flexibility



CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

IP Addresses (5) – Classful Addresing

- Old addresses came in blocks of fixed size (A, B, C)
  - Carries size as part of address, but lacks flexibility
  - Called classful (vs. classless) addressing

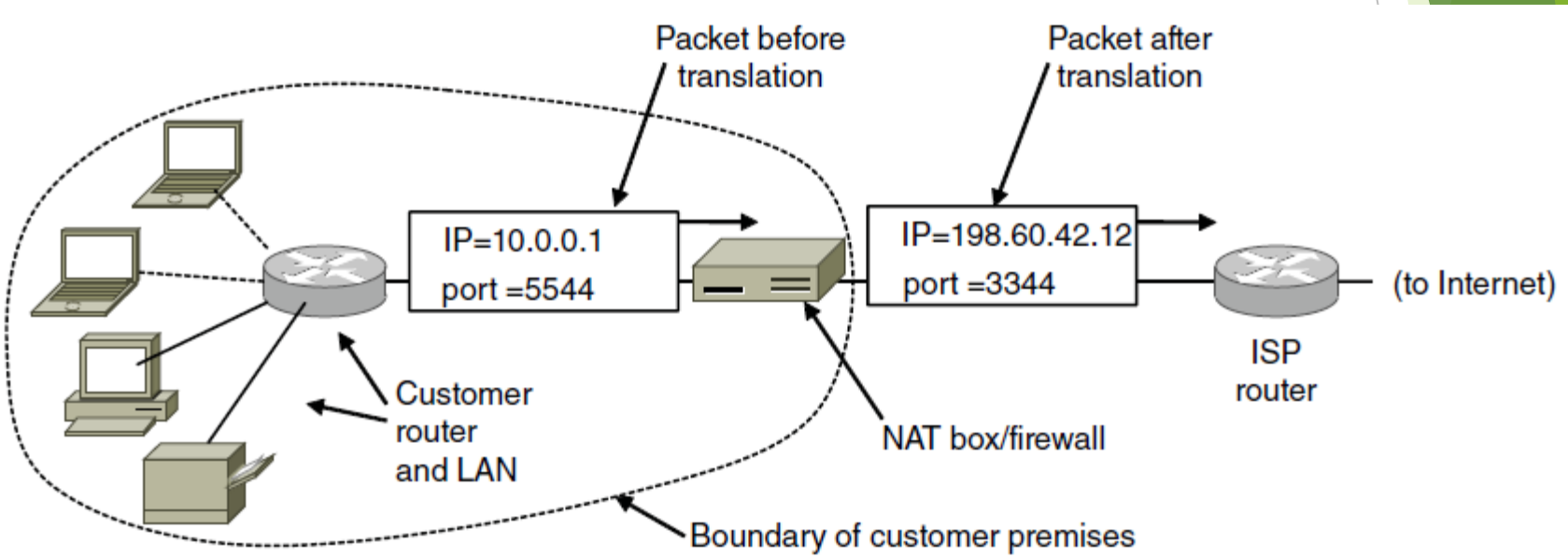


CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

IP Addresses (6) – NAT

NAT (Network Address Translation) box maps one external IP address to many internal IP addresses

- Uses TCP/UDP port to tell connections apart
- Violates layering; very common in homes, etc.



CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

IP Version 6 (1)

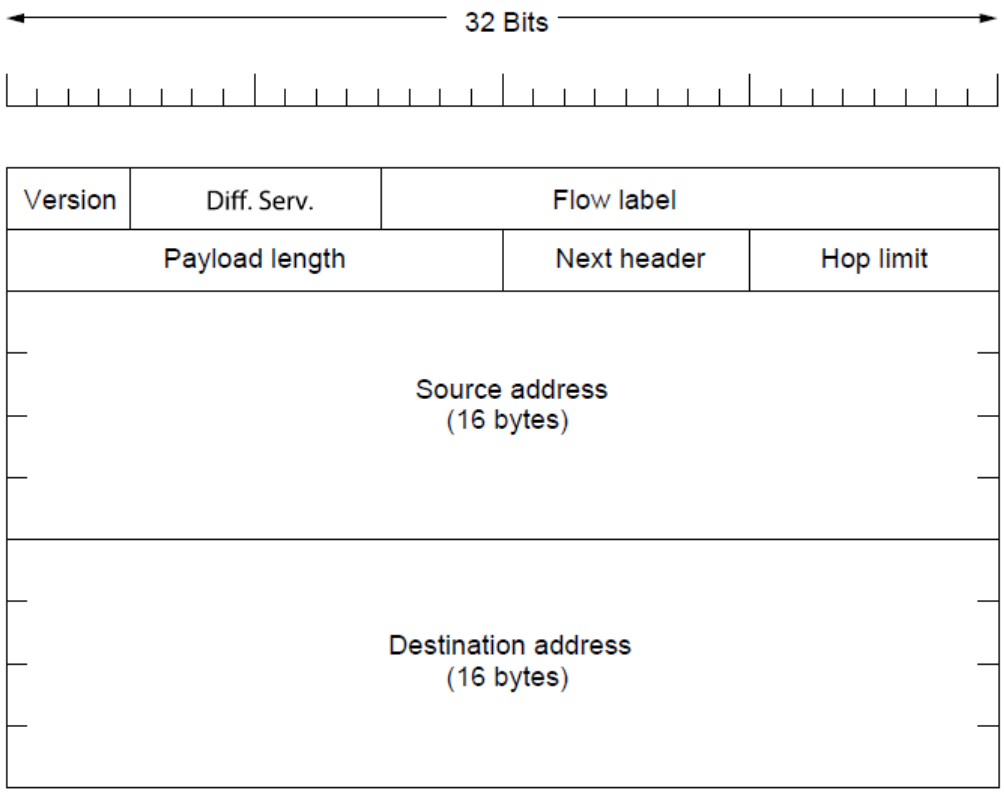
Major upgrade in the 1990s due to impending address exhaustion, with various other goals:

- Support billions of hosts
- Reduce routing table size
- Simplify protocol
- Better security
- Attention to type of service
- Aid multicasting
- Roaming host without changing address
- Allow future protocol evolution
- Permit coexistence of old, new protocols, …

Deployment has been slow & painful, but may pick up pace now that addresses are all but exhausted

IP Version 6 (2)

IPv6 protocol header has much longer addresses (128 vs. 32 bits) and is simpler (by using extension headers)



CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

IP Version 6 (3)

IPv6 extension headers handles other functionality

Extension header	Description
Hop-by-hop options	Miscellaneous information for routers
Destination options	Additional information for the destination
Routing	Loose list of routers to visit
Fragmentation	Management of datagram fragments
Authentication	Verification of the sender's identity
Encrypted security payload	Information about the encrypted contents

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Internet Control Protocols (1)

IP works with the help of several control protocols:

- ICMP is a companion to IP that returns error info
  - Required, and used in many ways, e.g., for traceroute
- ARP finds Ethernet address of a local IP address
  - Glue that is needed to send any IP packets
  - Host queries an address and the owner replies
- DHCP assigns a local IP address to a host
  - Gets host started by automatically configuring it
  - Host sends request to server, which grants a lease

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Internet Control Protocols (2)

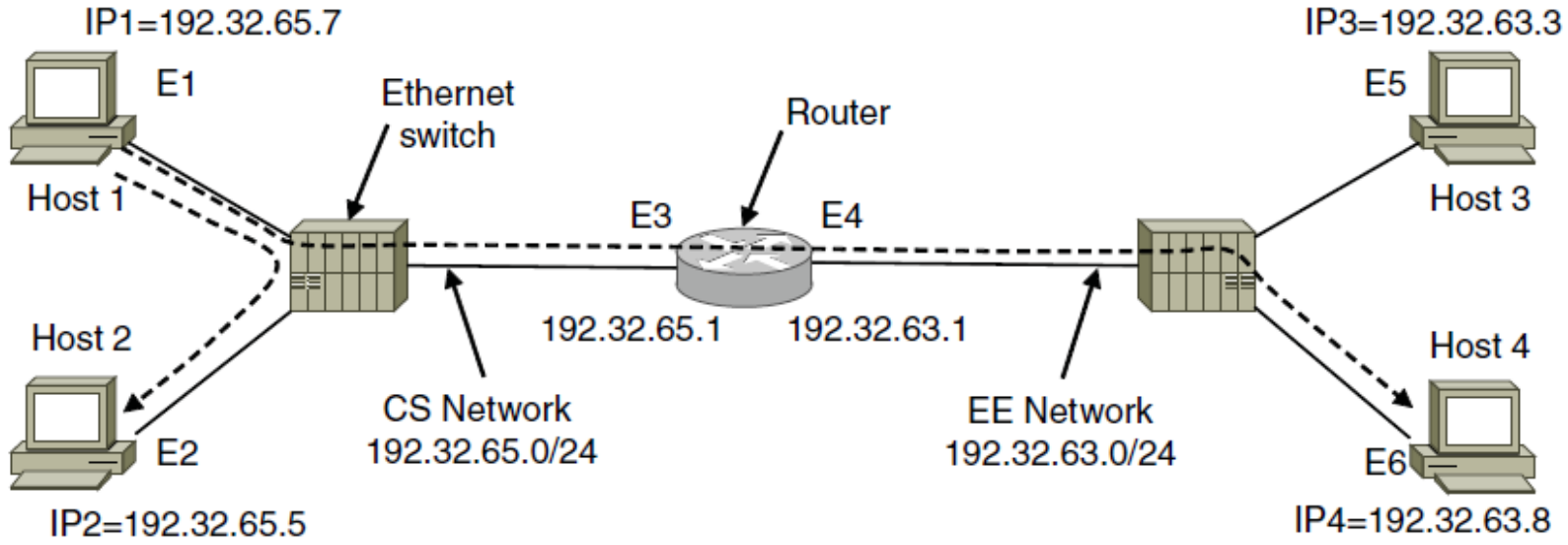
Main ICMP (Internet Control Message Protocol) types:

Message type	Description
Destination unreachable	Packet could not be delivered
Time exceeded	Time to live field hit 0
Parameter problem	Invalid header field
Source quench	Choke packet
Redirect	Teach a router about geography
Echo and Echo reply	Check if a machine is alive
Timestamp request/reply	Same as Echo, but with timestamp
Router advertisement/solicitation	Find a nearby router

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Internet Control Protocols (3)

- ARP (Address Resolution Protocol) lets nodes find target Ethernet addresses [pink] from their IP addresses



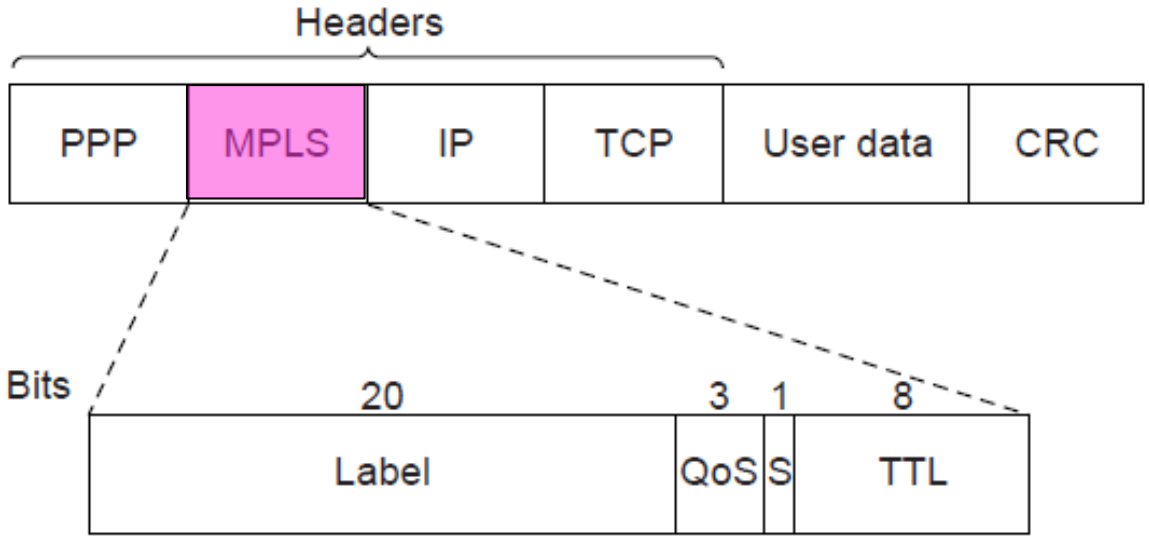
Frame	Source IP	Source Eth.	Destination IP	Destination Eth.
Host 1 to 2, on CS net	IP1	E1	IP2	E2
Host 1 to 4, on CS net	IP1	E1	IP4	E3
Host 1 to 4, on EE net	IP1	E4	IP4	E6

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Label Switching and MPLS (1)

MPLS (Multi-Protocol Label Switching) sends packets along established paths; ISPs can use for QoS

- Path indicated with label below the IP layer

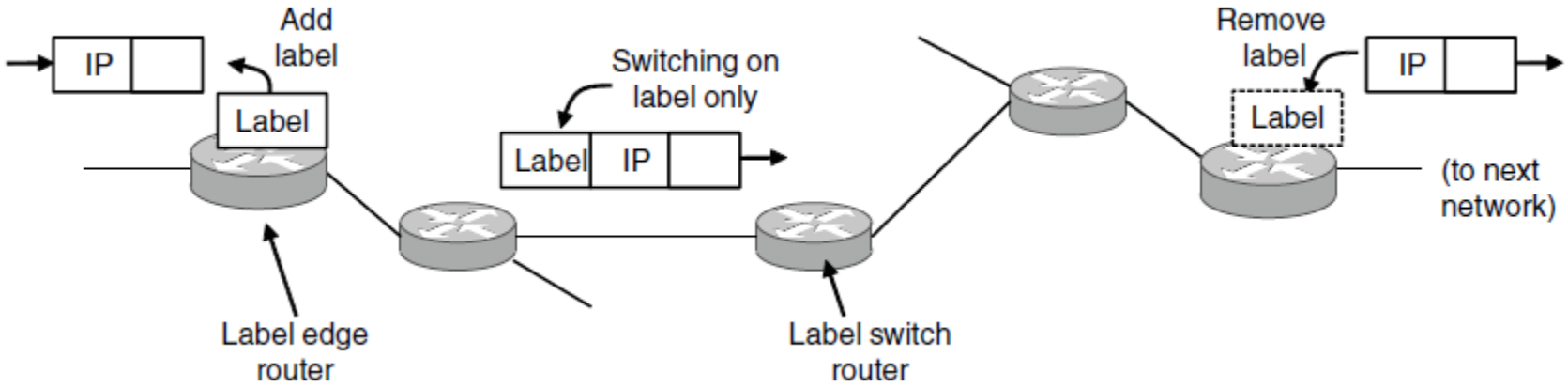


CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Label Switching and MPLS (2)

Label added based on IP address on entering an MPLS network (e.g., ISP) and removed when leaving it

- Forwarding only uses label inside MPLS network

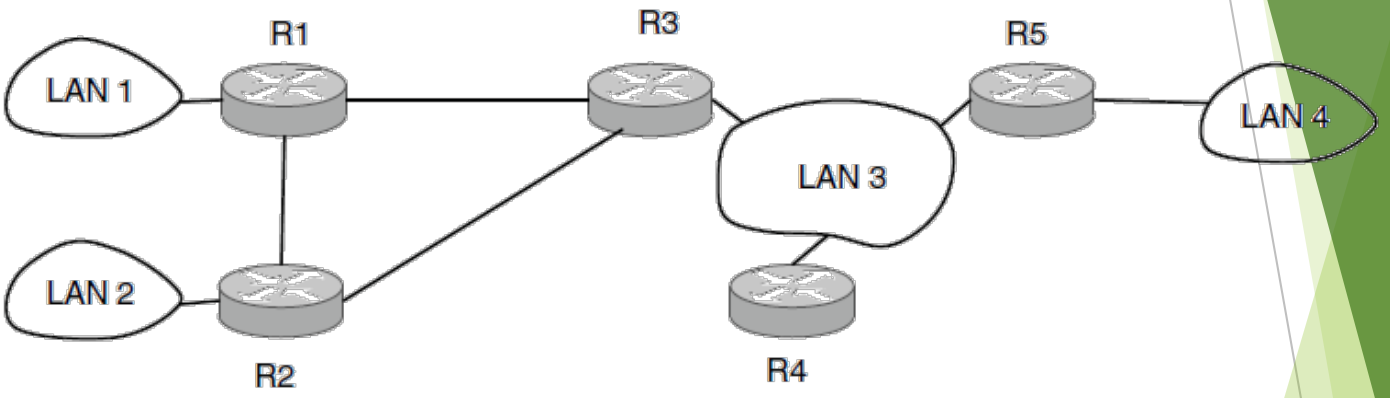


CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

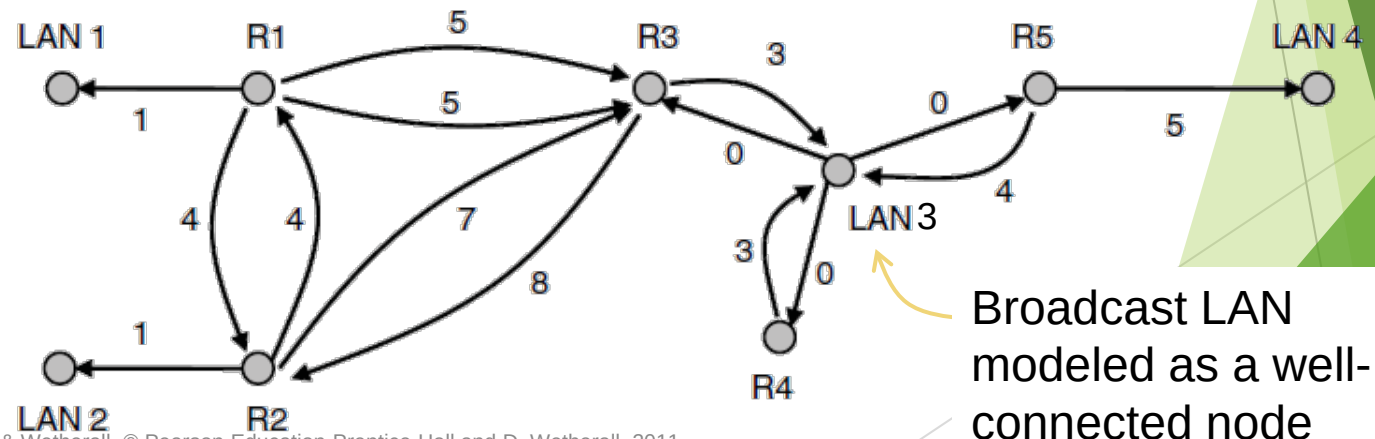
OSPF— Interior Routing Protocol (1)

- OSPF computes routes for a single network (e.g., ISP)
  - Models network as a graph of weighted edges

Network:



Graph:



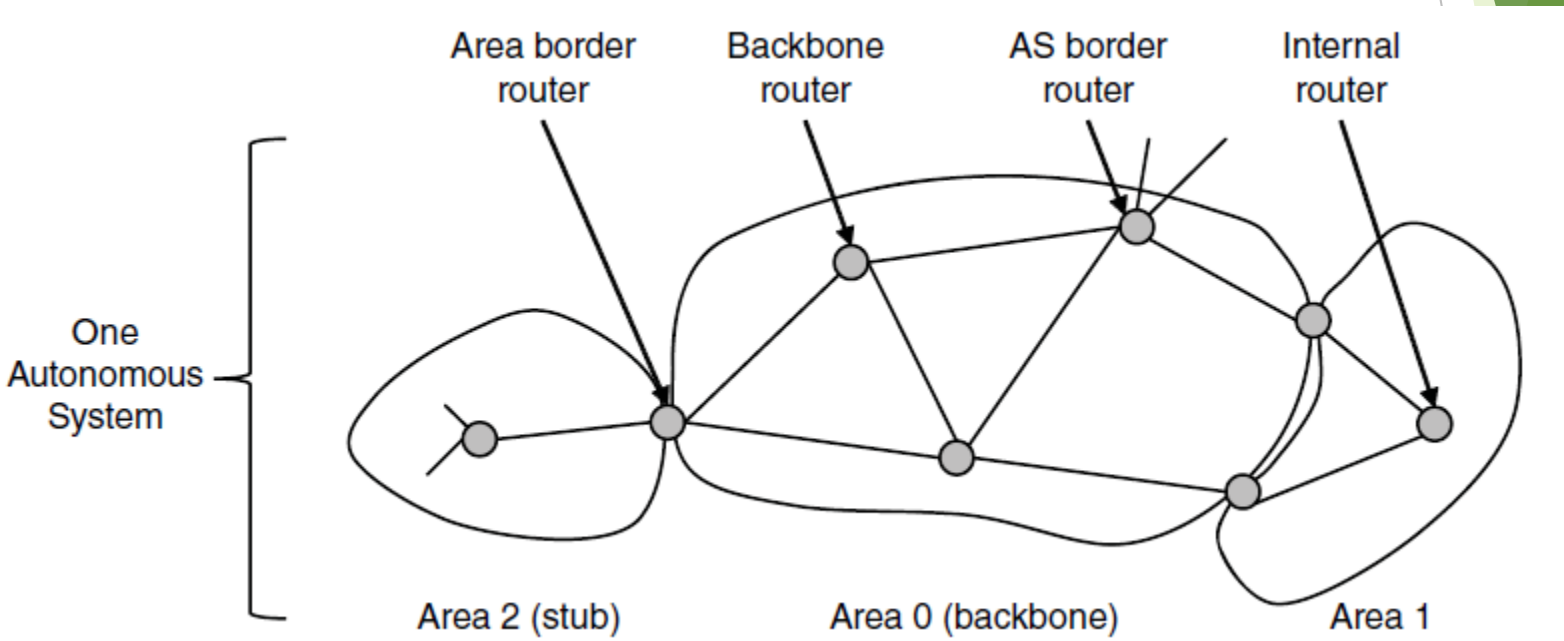
Broadcast LAN modeled as a well-connected node

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

OSPF— Interior Routing Protocol (2)

OSPF divides one large network (Autonomous System) into areas connected to a backbone area

- Helps to scale; summaries go over area borders



CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

OSPF— Interior Routing Protocol (3)

OSPF (Open Shortest Path First) is link-state routing:

- Uses messages below to reliably flood topology
- Then runs Dijkstra to compute routes

Message type	Description
Hello	Used to discover who the neighbors are
Link state update	Provides the sender's costs to its neighbors
Link state ack	Acknowledges link state update
Database description	Announces which updates the sender has
Link state request	Requests information from the partner

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

BGP— Exterior Routing Protocol (1)

BGP (Border Gateway Protocol) computes routes across interconnected, autonomous networks

- Key role is to respect networks' policy constraints

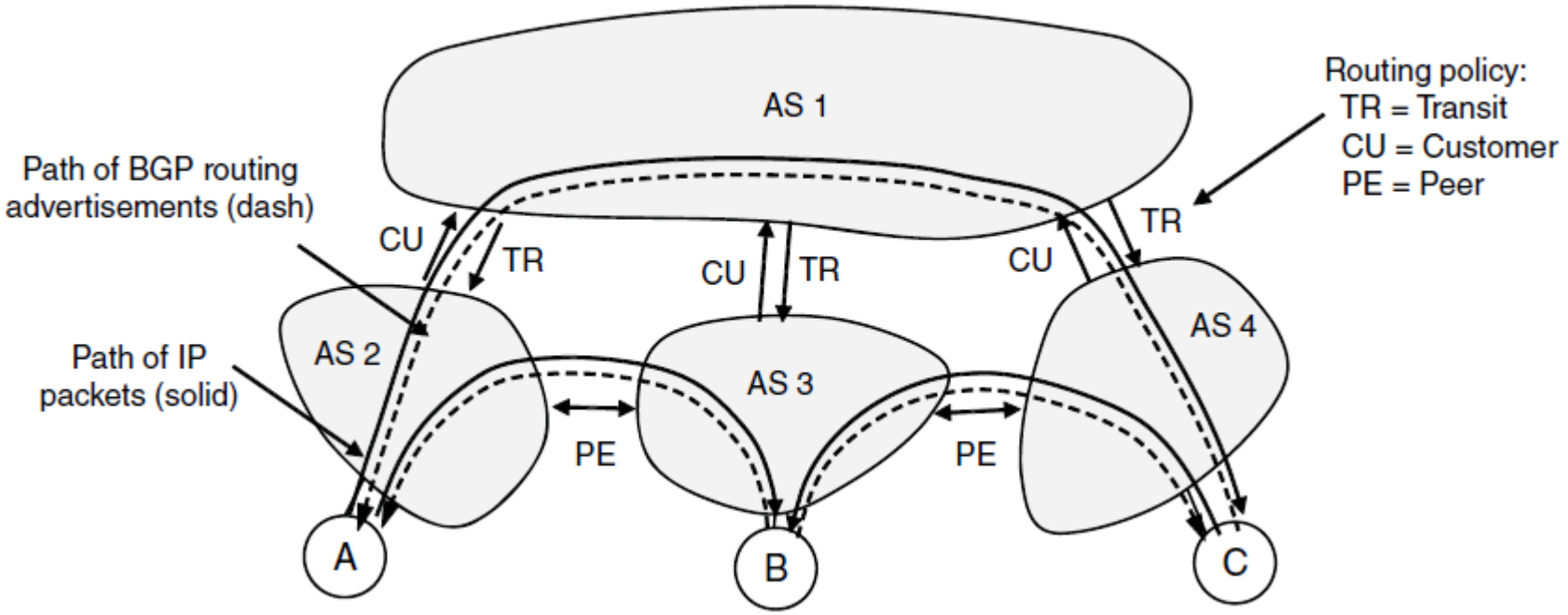
Example policy constraints:

- No commercial traffic for educational network
- Never put Iraq on route starting at Pentagon
- Choose cheaper network
- Choose better performing network
- Don't go from Apple to Google to Apple

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

BGP— Exterior Routing Protocol (2)

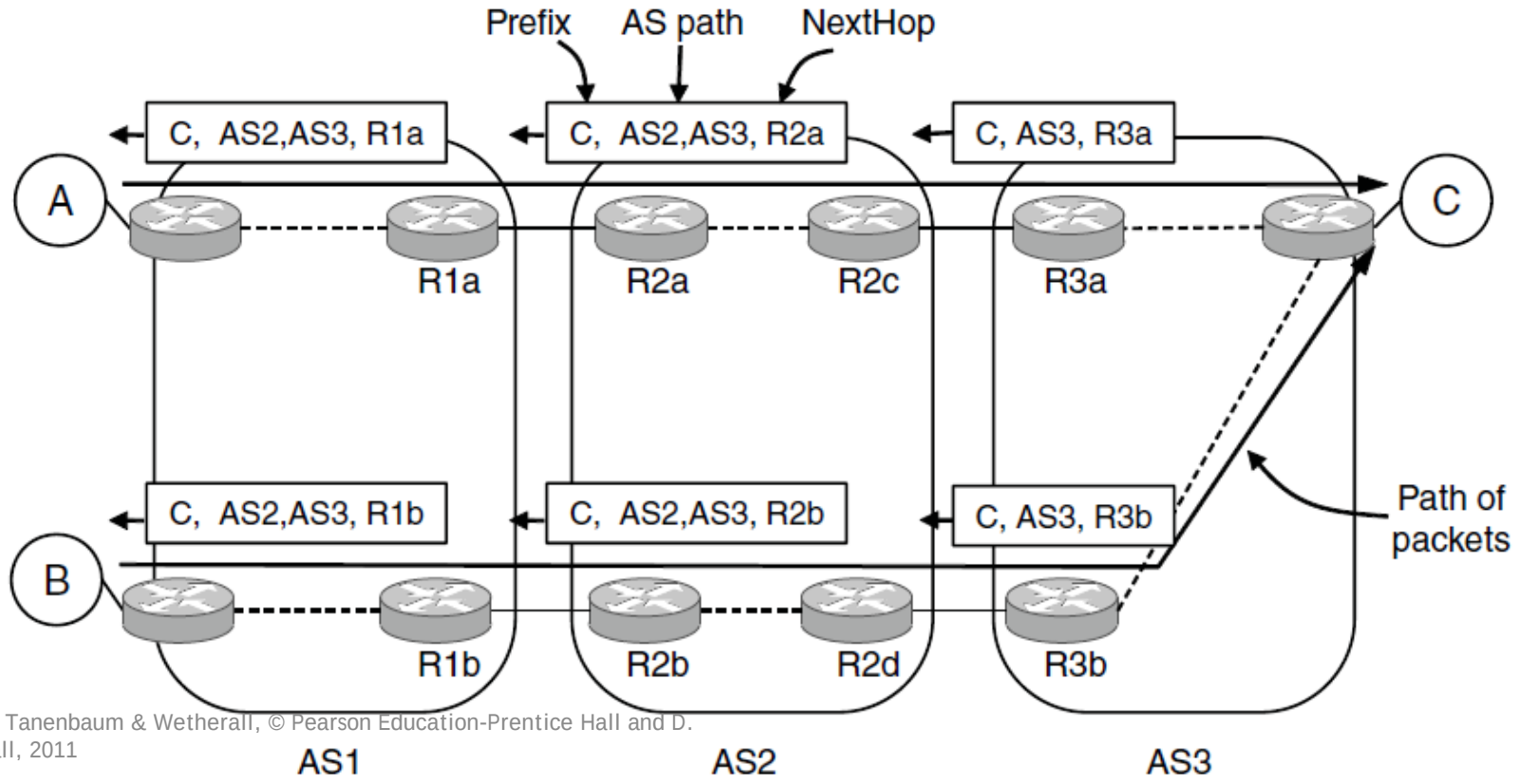
- Common policy distinction is transit vs. peering:
  - Transit carries traffic for pay; peers for mutual benefit
  - AS1 carries AS2↔AS4 (Transit) but not AS3 (Peer)



CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

BGP— Exterior Routing Protocol (3)

- BGP propagates messages along policy-compliant routes
  - Message has prefix, AS path (to detect loops) and next-hop IP (to send over the local network)



CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Internet Multicasting

Groups have a reserved IP address range (class D)

- Membership in a group handled by IGMP (Internet Group Management Protocol) that runs at routers

Routes computed by protocols such as PIM:

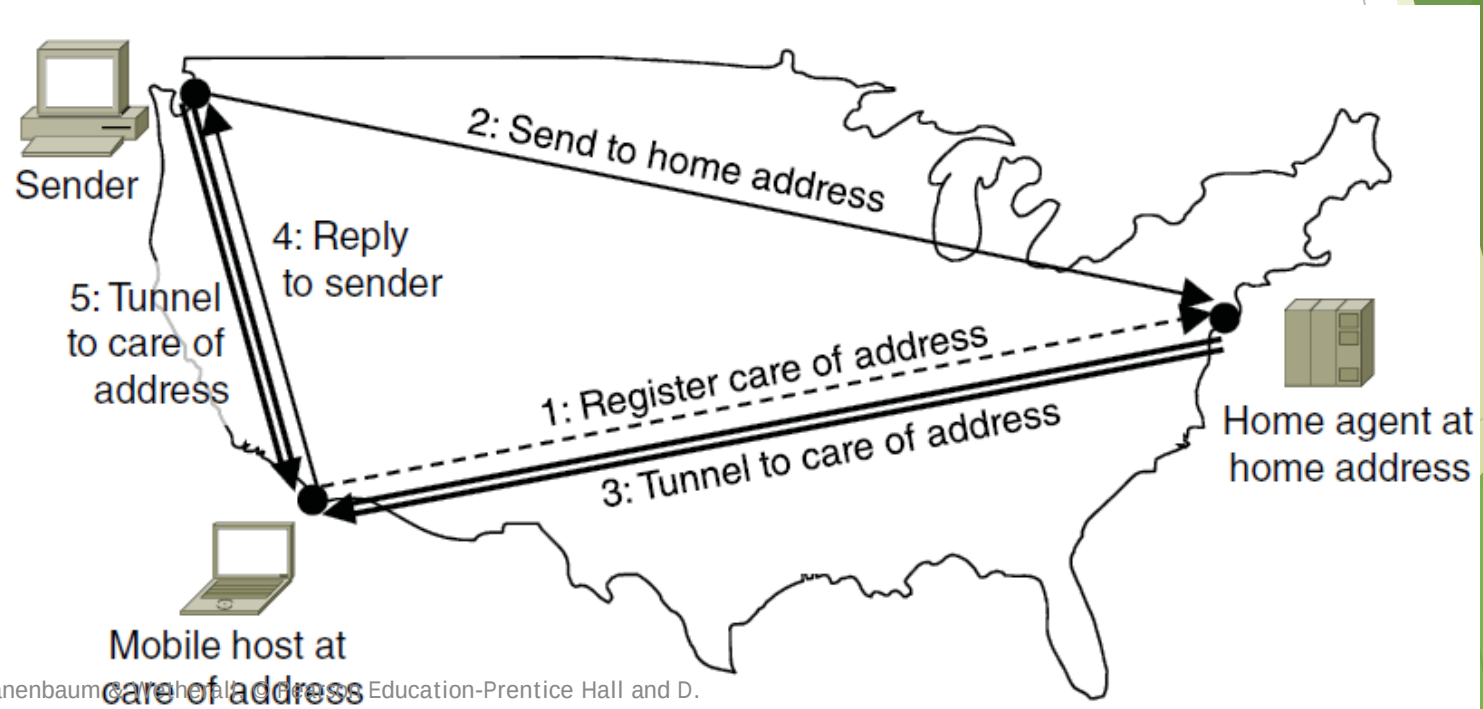
- Dense mode uses RPF with pruning
- Sparse mode uses core-based trees

IP multicasting is not widely used except within a single network, e.g., datacenter, cable TV network.

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011

Mobile IP

- Mobile hosts can be reached at fixed IP via a home agent
  - Home agent tunnels packets to reach the mobile host; reply can optimize path for subsequent packets
  - No changes to routers or fixed hosts



End

Chapter 5

CN5E by Tanenbaum & Wetherall, © Pearson Education-Prentice Hall and D. Wetherall, 2011