Ch4.

4.1

4.1.1 Static channel allocation

- Static allocation
  - Poor fit to computer systems with extremely bursty data traffic.
  - Performance affected when peak traffic to mean traffic ratios are 1000:1
  - Most channels will be idle most of the time.
- Dynamic allocation tries to resolve static allocation problems.

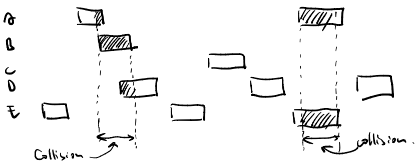
4.1.2 Dynamic Channel Allocation

- Independent traffic
- Single Channel
- Observable collisions
- Continuous or slotted time.
- Carrier sense or no carrier sense

4.2 Multiple access protocols.

4.2.1 ALOHA

Pure ALOHA: send the frame when ready, resend if the frame is destroyed.



在 pure ALOHA 中，能在任意时间传输 发生冲突就重传

pure ALOHA is free.

Slotted ALOHA add some limit.

like send frame at probability p.

Slotted ALOHA: required to wait to send a frame for the beginning of the next slot.

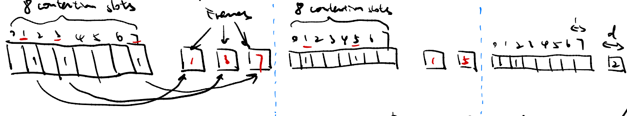
4.2.2 CSMA (Carrier Sense Multiple Access Protocols)

- Persistent and non-persistent CSMA.
  - 1-persistent CSMA: if the channel is idle, the station sends its data. Otherwise, the station just waits until it becomes idle.
  - Nonpersistent CSMA: if the channel is busy, the station does not continually sense it for the purpose of seizing it immediately.
  - P-persistent CSMA: if the channel is idle, it transmits with a probability p. With a probility, it defers until the next slot.
- CSMA with collision detection
  - Known as CSMA/CD (CSMA with Collision Detection)
  - Quickly detect the collision and abruptly stop transmitting.
  - Basis of the classic Ethernet LAN.

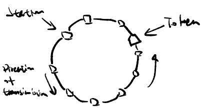
4.2.3 Collision-Free protocols.

- Bit-map protocol: - Basic bit-map method. - Reservation protocols.
- Token passing: Token bus.
- Binary countdown: As soon as a station sees that a high-order bit position that is 0 in its address has been overwritten with a 1, it give up.

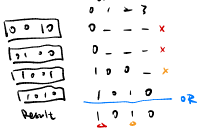
4.2.3.1. A Bit-Map Protocol



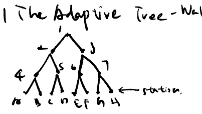
4.2.3.2. Token Passing



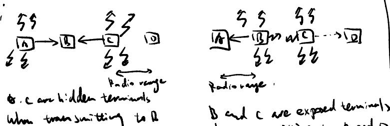
4.2.3.3 Binary Countdown



4.2.4.1 The Adaptive Tree-Walk Protocol



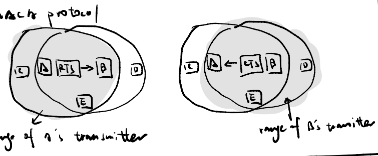
4.2.5 Wireless LAN Protocols



A, C are hidden terminals when transmitting to B

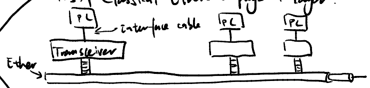
B and C are exposed terminals when transmitting to A and D

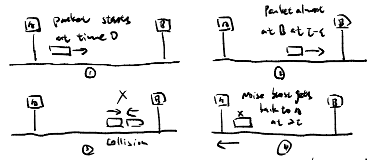
MACA protocol



4.3 Ethernet.

4.3.1 Classical Ethernet physical layer.





∴ A give up, wait later resend.

4.3.2 Classical Ethernet MAC Sublayer Protocol.

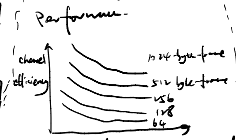
Ethernet (DIX) Frame formats

Byte	8	6	6	2	0–1500	0–46	4
	Preamble	Destination address	Source address	Type	Data	Pad	Checksum

IEEE 802.3

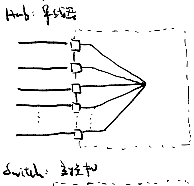
Preamble	SOF	Destination address	Source address	Length	Data	Pad	Checksum

4.3.3 Ethernet Performance.

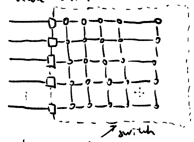


4.3.4 Switched Ethernet.

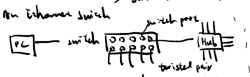
Hub: 集线器



Switch: 交换机



An Ethernet switch



Hub: all stations are in the same collision domain. They must use the CSMA/CD Algorithm to schedule their transmissions.

Switch: each port is its own independent collision domain. Collisions are now impossible, and CSMA/CD is not needed.

When a switch port receives an Ethernet frame from a station, the switch checks the Ethernet addresses to see which port the frame is destined for. The switch then forwards the frame over its high-speed backplane to the destination port.

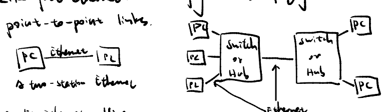
4.3.5 Fast Ethernet.

Name	Cable	Max. segment	Advantages
100 Base-T4	Twisted pair	100 m	Uses category 3 UTP
100 Base-Tx	Twisted pair	100 m	Full duplex at 100 Mbps (Cat 5 UTP)
100 Base-Fx	Fiber optics	2000 m	Full duplex at 100Mbps; long runs

The basic idea behind fast Ethernet was simple: keep all the old frame formats, interfaces, and procedural rules, but reduce the bit time from 100 nsec to 10 nsec.

4.3.6 Gigabit Ethernet.

Like fast ethernet, all configurations of gigabit Ethernet use point-to-point links.



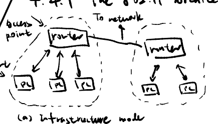
Gigabit Ethernet cabling

Name	Cable	Max. segment	Advantages
1000 Base-SX	Fiber optics	550 m	Multimode fiber (50, 62.5 microns)
1000 Base-LX	Fiber optics	5000 m	Single (10 μ) or multimode (50, 62.5 μ)
1000 Base-CX	2 Pairs of STP	25 m	Shielded twisted pair
1000 Base-T	4 Pairs of UTP	100 m	Standard category 5 UTP

4.3.7 10-Gigabit Ethernet.

Name	Cable	Max. segment	Advantages
10 GBase-SR	Fiber optics	Up to 300 m	Multimode fiber (0.85 μ)
10 GBase-LR	Fiber optics	10 km	Single-mode fiber (1.3 μ)
10 GBase-ER	Fiber optics	40 km	Single-mode fiber (1.5 μ)
10 GBase-CX4	4 Pairs of twinax	15 m	Twinaxial copper
10 GBase-T	4 Pairs of UTP	100 m	Category 6a UTP

4.4.1 The 802.11 Architecture and Protocol Stack.

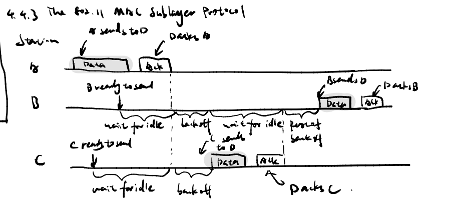


In infrastructure mode, each client is associated with an AP (Access Point) that is in turn connected to the other network. The client sends and receives its packets via the AP. Several access points may be connected together, typically by a wired network called a distribution system.

Part of the 802.11 protocol stack

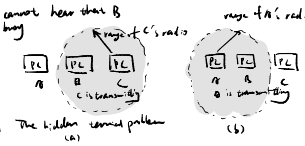
Upper layers							
Data link layer (MAC sublayer)	Logical link layer						
Physical layer	802.11 (legacy) Frequency hopping and infrared	802.11a OFDM	802.11b Spread spectrum	802.11g OFDM	802.11n MIMO OFDM	802.11ac MU-MIMO OFDM	802.11ax MU-MIMO OFDMA
Release date:	1997–1999	1999	1999	2003	2009	2013	2019

4.4.3 The 802.11 MAC Sublayer Protocol



A wants to send B but cannot hear that B is busy

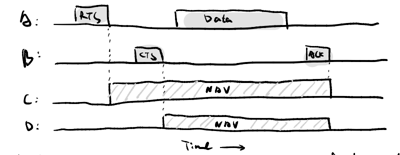
B wants to send to C but mistakenly thinks the transmission will fail



The hidden terminal problem (a)

(b) The exposed terminal problem.

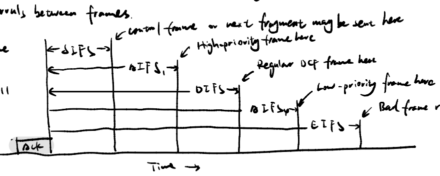
Solution①: With virtual sensing, each station keeps a logical record of when the channel is in use by tracking the NAV (Network Allocation Vector).



Virtual channel sensing using CSMA/CA.

However, while RTS/CTS sounds good in theory, it is one of those designs that has proved to be of little value in practice.

True solution: 802.11e. It works by extending CSMA/CA with carefully defined intervals between frames.



Interframe spacing in 802.11

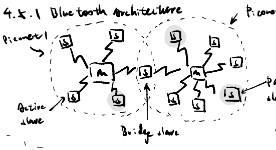
4.4.4 The 802.11 Frame Structure

Bytes	2	2	6	6	6	2	0–2312	4
	Frame control	Duration	Address 1 recipient	Address 2 transmitter	Address 3	Sequence	Data	Check sequence

Bits	2	2	4	1	1	1	1	1	1	1	1
	Version = 00	Type = 10	Subtype = 0000	To DS	From DS	More frag.	Retry	Pwr. mgt.	More data	Protected	Order

Format of 802.11 data frame.

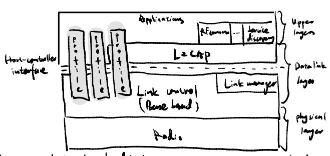
4.5.1 Bluetooth Architecture



Two piconets can be connected to form a scatternet.

The basic unit of a Bluetooth systems is a piconet, which consists of a controller node and up to seven active worker nodes within a distance of 10 meters. An interconnected collection of piconets is called a scatternet.

4.5.3 The Bluetooth Protocol Stack.



The Bluetooth protocol architecture

The link control (or baseband) layer includes elements of the physical layer. It deals with how the controller controls time slots and how these slots are grouped into frames.

The link manager handles the establishment of logical channels between devices, including power management, pairing and encryption, and provides reliability if needed.

L2CAP (Logical Link Control Adaptation Protocol) frames variable-length messages and provides reliability if needed.

4.5.6 The Bluetooth Frame structure

Bits	72	54	0–2744
	Access code	Header	Data (at 1x rate)

3	4	1	1	1	8	
Addr	Type	F	A	S	CRC	Repeated 3 times

72	54	16	0–8184	2
Access code	Header	Guard/Sync	Data (at 2x or 3x rate)	Trailer

(a) Basic rate data frame, top (b) Enhanced rate data frame, bottom.

It begins with an access code that usually identifies the controller so that workers within radio range of two controllers can tell which traffic is for them. Next comes a 54-bit header containing typical MAC sublayer fields. If the frame is sent at the basic rate, the data field comes next.

4.5.7 Bluetooth 5: Support for Internet of things device; Speed increased from 1Mbps to 2 Mbps; Message size has gone up from 31 bytes to 255 bytes; Range indoors has gone up from 10 m to 40 m; Power requirements have been reduced slightly; Slightly better security.

4.6 DOCSIS (Data over Cable Service Interface Specifications)

4.6.1 Overview

- MAC sublayer performs channel allocation, configuration of quality of service, unique forwarding model.
- Uses a standard MAC frame format.

4.6.2 Ranging

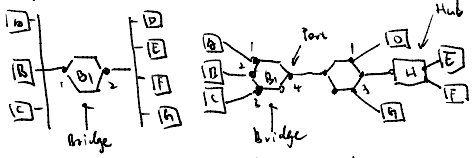
- Cable modem transmits a ranging request
- CMTS (headend) determines the network delay to the cable modem and performs necessary power adjustment

4.6.3 Channel bandwidth allocation

- Service flows
- Request-grant process and low-latency DOCSIS

4.7 Data link Layer Switching

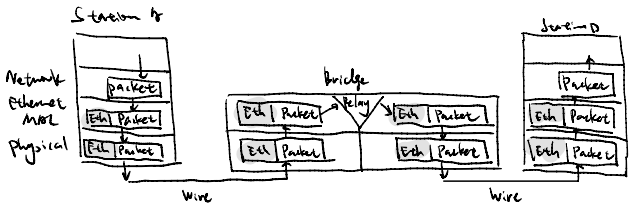
4.7.2 Learning Bridge



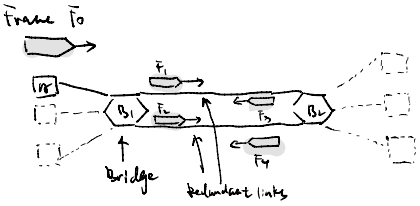
Bridge connecting two multidrop LANs

Bridges (and a hub) connecting seven point-to-point stations.

Protocol processing at a bridge:



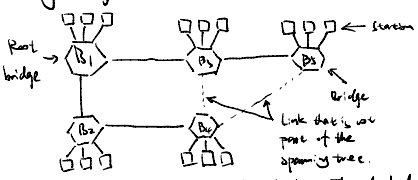
4.7.3 Spanning-Tree Bridges



Bridge with two parallel links.

Problem: B2 does not know that they are copies of the same frame, rather than two different frames sent one after the other.

Solution: For the bridges to communicate with each other and overlay the actual topology with a spanning tree that reaches every bridge:



A spanning tree connecting five bridges. The dashed lines are links that are not part of the spanning tree.

4.7.4 Repeaters, Hubs, Bridges, Switches, Routers, Gateways

Application layer	Application gateway
Transport layer	Transport gateway
Network layer	Router
Data link layer	Bridge, switch
Physical layer	Repeater, hub

Which device is in which layer.

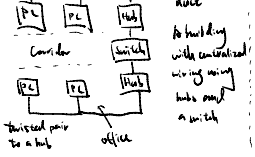
Packet (supplied by network layer)

Frame header	Packet header	TCP header	User data	CRC

Frame (built by data link layer)

Frame, packets and headers.

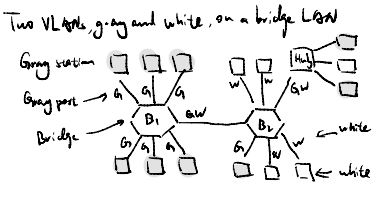
4.7.5 Virtual LANs



Cable drops

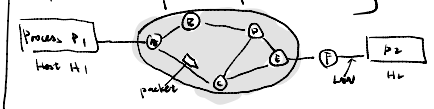
A building with centralized wiring using hubs and a switch

Two VLANs, gray and white, on a bridge LAN



Chp 5.1 Network Layer Design Issues

5.1.1 Store-and-forward packet switching

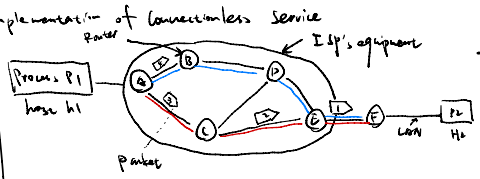


- A host with a packet to send transmits it to the nearest router; the packet is stored there until it has fully arrived, and the link has finished its processing by verifying the checksum; Then it is forwarded to the next router along the path until it reaches the destination host, where it is delivered. This mechanism is store-and-forward packet switching

5.1.2 Services Provided to the Transport Layer

- Services independent of router technology
- Transport layer shielded from number, type, topology of routers.
- Network addresses available to transport layer use uniform numbering plans — even across LANs and WANs

5.1.3 Implementation of Connectionless Service



Routing within a datagram network

A's table (initially)

A	-
B	B
C	C
D	B
E	C
F	C

Dest. Line

A's table (later)

A	-
B	B
C	C
D	B
E	C
F	B

C's table

A	B
B	B
C	-
D	E
E	E
F	E

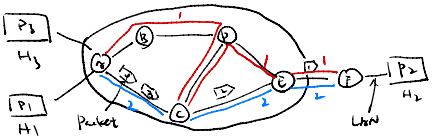
E's table

A	C
B	D
C	C
D	D
E	-
F	F

A decides to send packet 4 via a different route than of the first three packets. Perhaps it has learned of a traffic jam somewhere along the A-C-E path and updated its routing table.

The algorithm that manages the tables and makes the routing decisions is called the routing algorithm.

5.1.4 Implementation of Connection-Oriented Service



Virtual circuit Network

A's table

In		Out	
H1	1	C	1
H3	1	C	2

C's table

In		Out	
A	1	E	1
A	2	E	2

E's table

In		Out	
C	1	F	1
C	2	F	2

5.1.5 Comparison of Virtual-Circuit and Datagram Networks

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 failure	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

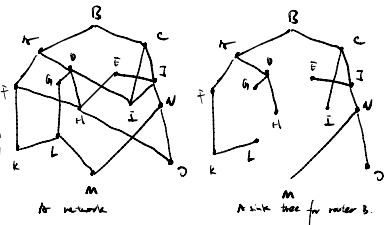
5.2 Routing Algorithms in a Single Network

Optimality principle; shortest path algorithm; Distance vector routing; Link state routing; Optimality principle; shortest path algorithm; Distance vector routing; Link state routing.

- The routing algorithm is that part of the network layer software responsible for deciding which output line an incoming packet should be transmitted on.
- Properties are desirable in a routing algorithm: correctness, simplicity, robustness, stability, fairness, and efficiency
- Nonadaptive algorithms do not base their routing decisions on any measurements or estimates of the current topology and traffic. Known as static Routing.
- Adaptive algorithms, in contrast, change their routing decisions to reflect changes in the topology, and sometimes changes in the traffic as well. Known as Dynamic Routing.

Network with a conflict between fairness and efficiency

5.2.1 The Optimality Principle



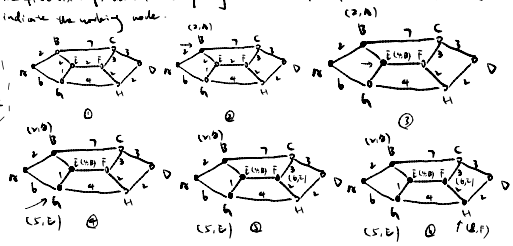
A network

A sink tree for router B

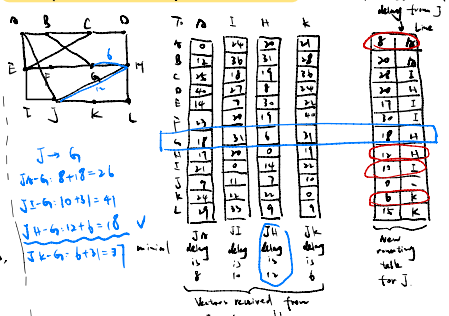
- One can make a general statement about optimal routes without regard to network topology or traffic. This statement is known as the optimality principle.
- As a direct consequence of the optimality, the optimality principle, the set of optimal routes from all sources to a given destination form a tree rooted at the destination. Such a tree is called a sink tree.
- Note that a sink tree is not necessarily unique; other trees with the same path lengths may exist.
- If we allow all of the possible paths to be chosen, the tree becomes a more general structure called a DAG (Directed Acyclic Graph)

5.2.2 Shortest Path Algorithm

The first six steps used in computing the shortest path from A to D. The arrows indicate the working node.



5.2.4 Distance Vector Routing



- A distance vector routing algorithm (AKA Bellman-Ford routing algorithm) operates by having each router maintain a table (i.e. a vector) giving the best-known distance to each destination and which link to use to get there.
- Drawback: although it converges to the correct answer, it may do so slowly.

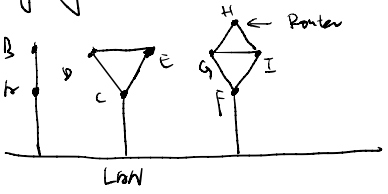
5.2.5 Link State Routing

The idea behind link state routing is fairly simple and can be stated as five parts.

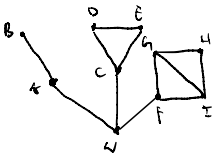
① Discover neighbors, learn network addresses; ② Set distance/cost metric to each neighbor ③ Construct packet telling all it has learned ④ Send packet telling all it has learned ⑤ Send packet to, receive packets from other routers ⑥ Compute shortest path to every other router

5.2.5.1 Learning about the Neighbors

- When a router is booted, it's first task is to learn who its neighbors are. It accomplishes this goal by sending a special HELLO packet on each point-to-point line.
- The router on the other end is expected to send back a reply giving its name. These names must be globally unique.

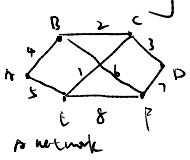


(a) Nine routers and a broadcast LAN

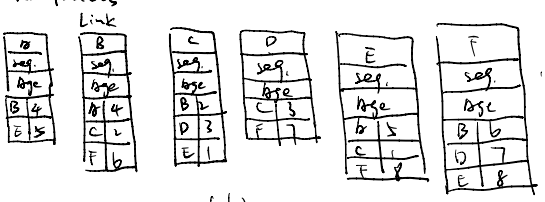


(b) Graph model of (a). Here we have introduced a new, artificial node N, to which A, C, and F are.

5.2.5.2 Building Link State Packets



(a) network



(b)

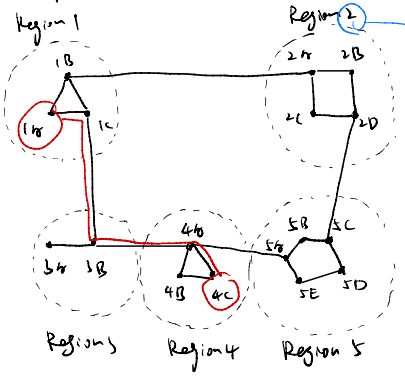
The link state packets for this network.

- The packet starts with the identity of the sender, followed by a sequence number and age and a list of neighbors. The cost to each neighbor is also given.
- The hard part is determining when to build packets. One possibility is to build them periodically, at regular intervals. Another possibility when some specific event occurs, such as a line or neighbor going down or coming back up again or changing its properties.
- Drawback: Not only is router memory consumed by ever-increasing tables, but more CPU time is needed to scan them, and more bandwidth is needed to send status reports about them.

5.2.6 Hierarchical Routing within a Network

- When hierarchical routing is used, the routers are divided into what we will call regions or areas.
- Drawback: There is a penalty to be paid: increased path length. For example, the best route from 1A to 5C is via region 2, but with hierarchical routing all traffic to region 5 goes via region 3, because that is better for most destinations in region 5.

e.g. if 1A → 5C then it can read from Hierarchical table 1A → 1C → ... → 5C



Full table for 1A

Dest.	Line	Hops
1A	—	—
1B	1B	1
1C	1C	1
2A	1B	2
2B	1B	3
2C	1B	3
2D	1C	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

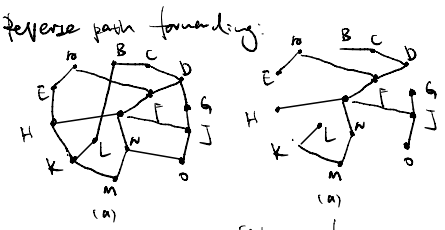
Hierarchical table for 1A

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

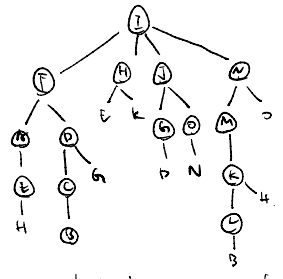
5.2.7 Broadcast Routing

- Sending a packet to all destinations simultaneously is called broadcasting.
- The principal advantage of reverse path forwarding is that it is efficient while easy to implement.
- It sends the broadcast packet over each link only once in each direction, just as in flooding, yet it requires only that routers know how to reach all destinations, without needing to remember sequence numbers or list all destinations.
- Drawback: Each router must have knowledge of some spanning tree.

Reverse path forwarding.



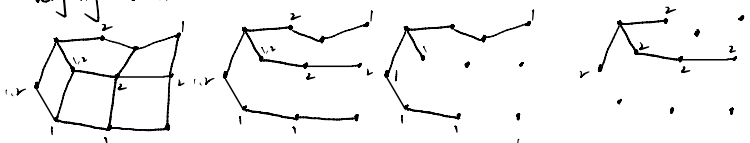
(a) A network. Sink tree for router I.



The tree built by reverse path forwarding from I.

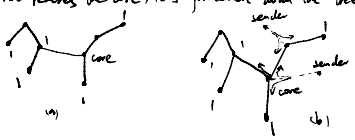
5.2.8 Multicast Routing

- Sending a message to such a group is called multicasting, and the routing algorithm used is called multicast routing.
- The solution is to prune the broadcast spanning tree by removing links that do not lead to members. The result is an efficient multicast spanning tree.
- Drawback: One potential disadvantages is that it is lots of work for routers, especially for very big networks.



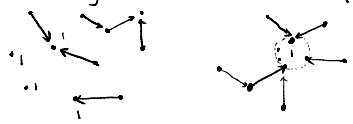
(a) A network. (b) A spanning tree for the leftmost router. (c) A multicast tree for group 1. (d) A multicast tree for group 2.

- An alternative design uses core-based trees to compute a single spanning tree for the group. All of the routers agree on a root (called the core) and build the tree by sending a packet from each member to the root. The tree is the union of the paths traced by these packets. Here, to send to the group 1, a sender sends a packet to the core. When the packet reaches the core, it is forwarded down the tree.



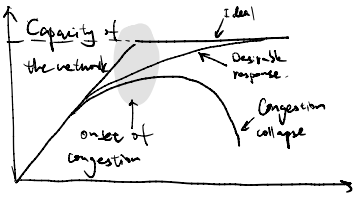
5.2.9 Anycast Routing

- In anycast, a packet is delivered to the nearest member of a group. Schemes that find these paths are called anycast routing.
- Suppose we want to anycast to the members of group 1. They will all be given the address "1", instead of different addresses. Distance Vector routing will distribute vectors as usual, and nodes will choose the shortest path to destination 1. That will result in nodes sending to the nearest instance of destination 1.



5.3.1 Congestion

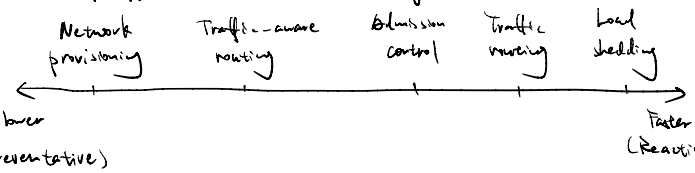
- Traffic management or engineering has to do with making sure the network is able to carry the offered traffic. It can be performed by devices in the network, or by the senders of traffic often through mechanisms in the transport protocol, which are often referred to as congestion control.
- Congestion management and control concerns the behavior of all the hosts and routers.
- Flow control relates to the traffic between a particular sender and a particular receiver and is generally concerned with making sure that the sender is not transmitting data faster than the receiver can process it.



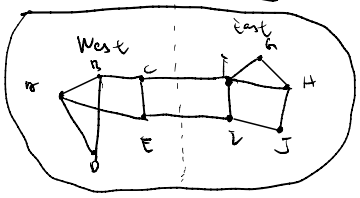
Performance drops significantly in the presence of congestion: packet loss rates increases, and latency also increases as router queues fill with packets.

5.3.2 Approaches to Traffic Management

Timescales of approaches to traffic and congestion management.



5.3.2.1 Traffic-Aware Routing

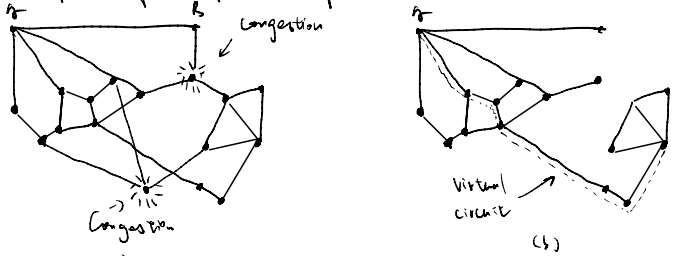


A network in which the East and West parts are connected by two links.

- The goal in taking load into account when computing routes is to shift traffic away from hotspots that will be the first places in the network to experience congestion.
- Set the link weight to be a function of the (fixed) link bandwidth and propagation delay plus the (variable) measured load or average queueing delay. Least-weight paths will then favor paths that are more lightly loaded, all else being equal.
- Disadvantage: routing tables may oscillate widely, leading to erratic routing and many potential problems.

5.3.2.2 Admission Control

- One technique that is widely used in virtual-circuit networks to keep congestion at bay is admission control. The idea is simple: do not set up a new virtual circuit unless the network can carry the added traffic without becoming congested.
- Admission control can be combined with traffic-aware routing by considering routes around traffic hotspots as part of the setup procedure.

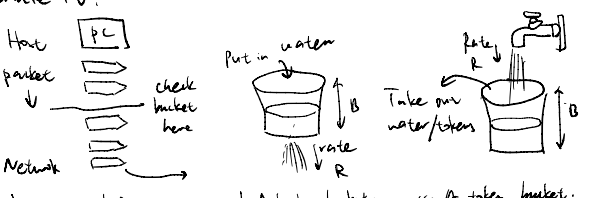


A virtual circuit from A to B is also shown.

(a) A congested network (b) The portion of the network that is not congested.

5.3.2.4 Traffic Shaping

Traffic shaping is a technique for regulating the average rate and burstiness of a flow of data that enters the network. The goal is to allow applications to transmit a traffic that suits their needs. When a flow is set up, the user and the network agree on a certain traffic pattern for that flow. In effect, the customer says to the provider "My transmission pattern will look like this: can you handle it?"

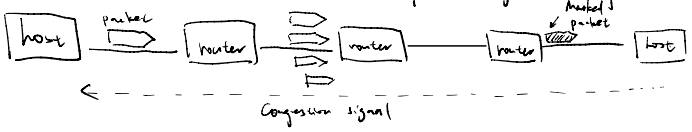


(a) Shaping packet (b) A leaky bucket (c) A token bucket.

- Network issues to address for quality of service
  - Addressing applications needs
  - Regulating traffic entering the network
  - Reserving resources at routers to guarantee performance
  - Safely accepting more traffic

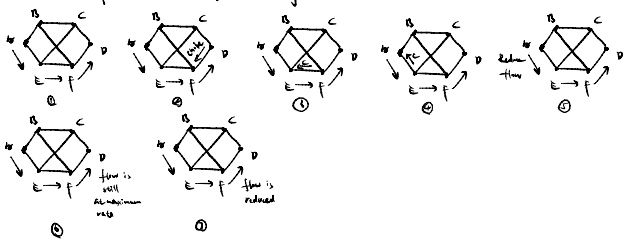
5.3.2.5 Explicit Congestion Notification

Instead of generating packets to warn of congestion, a router can tag any packet it forwards (by setting a bit in the packet's header) to signal that it is experiencing congestion. When the network delivers the packet, the destination can note that there is congestion and inform the sender when it sends a reply packet. The sender can then throttle its transmissions. The design is called ECN (Explicit Congestion Notification).

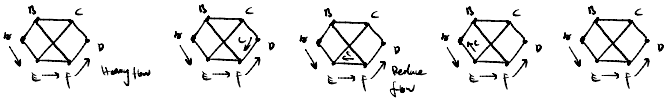


5.3.2.6 Hop-by-Hop Backpressure

A choke packet that affects only the source.



A choke packet that affects each hop it passes through.



An alternative approach is to have the choke packet take effect at every hop it passes through, as shown in the sequence of (b). Here, as soon as the choke packet reaches F, F is required to reduce the flow to D. Doing so will require F to devote more buffers to the connection, since the source is still sending away at full blast, but it gives D immediate relief.

The next effect of this hop-by-hop scheme is to provide quick relief at the point of congestion, at the price of using up more buffers upstream.

5.4.1 Application QoS Requirements

- A stream of packets from a source to a destination is called a flow.
- The needs of each flow can be characterized by four primary parameters: bandwidth, delay, jitter, and loss.
- Together, these four determine the QoS (Quality of service) the flow requires.

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

Stringency of applications' quality-of-service requirements.

5.4.2 Overprovisioning

- Helps to provide good quality of service
  - Ensure the network has the capacity for all traffic
  - Expensive solution.

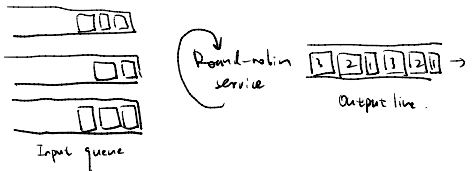
5.4.3 Packet Scheduling

- Algorithms that allocate router resources among the packets of a flow and between competing flows are called packet scheduling algorithms.
- Router resources reserved for different flows: Bandwidth, Buffer spaces, CPU cycles.
- Algorithms: FIFO, Fair queueing, weighted fair queueing.

5.4.3.1 FIFO

Each router buffers packets in a queue for each output line until they can be sent; They are sent in the same order that they arrived; This algorithm is known a FIFO.
Disadvantages: Processing packets in the order of their arrival means that the aggressive sender can hog most of the capacity of the routers its packets traverse, starving the other flows and reducing their quality of service.

5.4.3.2 Fair Queuing

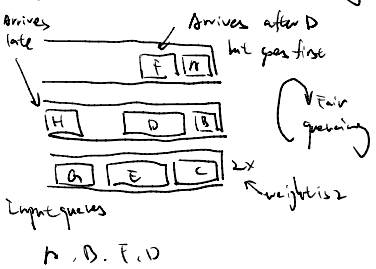


The essence of this algorithm is that routers have separate queues, one for each flow for a given output line; When the line becomes idle, the router scans the queues round robin. In this way, with n hosts competing for the output line, each host gets to send one out of every n packets; It's fair in the sense that all flows get to send packets at the same rate. Sending more packets will not improve this rate.

Dis: It gives more bandwidth to hosts that use large packets than to hosts that use small packets.

5.4.3.3 Improved Fair Queuing

- Round robin is done in such a way as to simulate a byte-by-byte round robin, instead of a packet-by-packet round robin.
- The trick is to compute a virtual time that is the number of the round at which each packet would finish being sent. Each round drains a byte from all of the queues that have data to send. The packets are then sorted in order of their finishing times and sent in that order.



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	16	4
F	10	6	16	5
G	11	10	19	6
H	20	8	28	8

A, D, F, D

Dis: It gives all hosts the same priority.

5.4.3.4 Weighted Fair Queuing

- In many situations, it is desirable to give, for example, video servers more bandwidth than, say, file servers.
- This is easily by giving the video server two or more bytes per round. This modified algorithm is called WFQ (Weighted Fair Queuing).

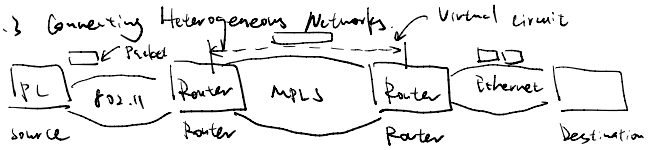
5.5.1 Internetworks

- When two or more networks are connected, they form an internetwork, or more simply an Internet
- How networks differ; Connecting heterogeneous networks;
- Connecting endpoints across heterogeneous networks; Supporting different packet size: packet fragmentation.

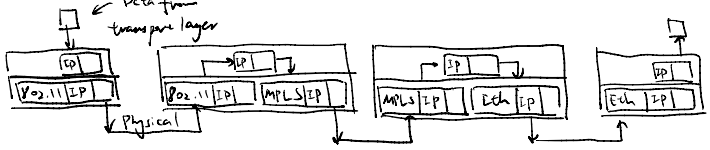
5.5.2 How Networks Differ

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.

5.5.3 Connecting Heterogeneous Networks



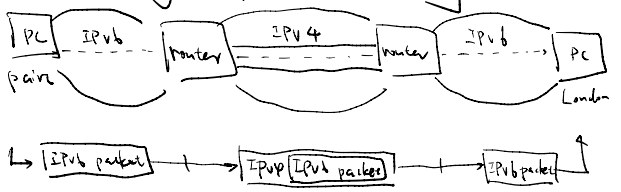
A packet crossing different networks



Network and link layer protocol processing

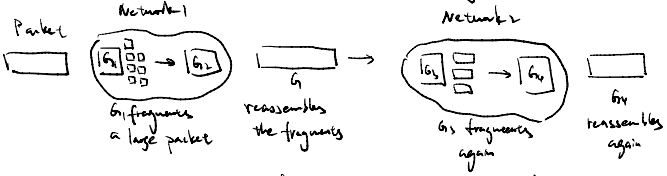
- Because different networks may, in general, have different forms of addressing the packet carries a network layer address that can identify any host across the three networks. ; • Once the packet has traveled along the virtual circuit, it will reach the Ethernet network; • The packet may be too large to be carried, since 802.11 can work with larger frames than Ethernet; • To handle this problem, the packet is divided into fragments, and each fragment is sent separately. When the fragments reach the destination, they are reassembled.

5.5.4 Connecting Endpoints Across Heterogeneous Networks

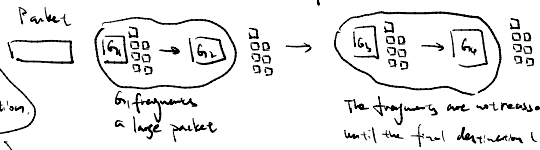


- A common special case is where the source and destination hosts are on the same type of network, but there is a different network in between.
- The solution to this problem is a technique called tunneling
- To send an IP packet to a host in the London office, a host in the Pairs constructs the packet containing an IPv6 address in London and sends it to the multiprotocol router that connects the Pairs IPv6 network to the IPv4 Internet.
- When this is between router gets the IPv6 packet, the router puts the IPv6 packet inside a IPv4 packet. When this wrapped packet arrives, the London router removes the original IPv6 packet and sends it onward to the destination host.

5.5.6 Supporting Different Packet Size: Packet Fragmentation

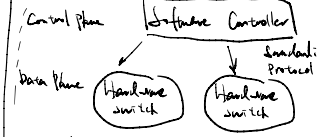


(a) Transparent fragmentation; is straightforward but has some problems.



(b) Non transparent fragmentation. The advantage of this is it requires routers to do less work.

5.6 Software-Defined Networking

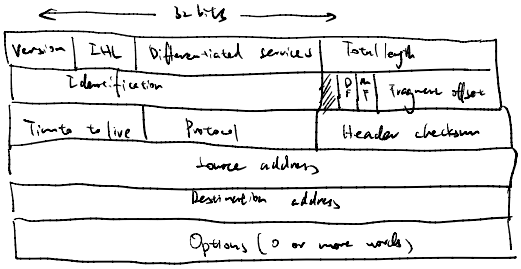


One of the main concepts driving SDN was to recognize that the control plane, the software and logic that selects routes and decides what to do with forwarding traffic, runs in software and can operate completely separately from the data plane, the hardware-based technology that is responsible for actually performing look ups on packets and deciding what to do with them.

5.7 The Network Layer in the Internet

Top 10 principles: Make sure it works; Keep it simple; Make clear choices; Exploit modularity; Expect heterogeneity; Avoid static options and parameters; Look for a good design; it need not be perfect; Be strict when sending and tolerant when receiving; Think about scalability; Consider performance and cost.

5.7.1 The IP Version 4 Protocol



5.8 Policy at the Network Layer

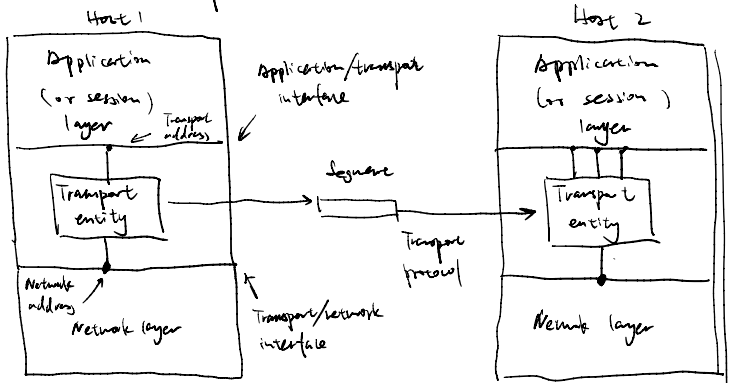
- Peering disputes
  - A breakdown in negotiations over paying for transit.
- Traffic prioritization
- Generally agreed upon bright-line rules
  - No blocking; - No throttling; - No paid prioritization;
  - Disclosure of any prioritization practices.

6.1 The Transport Service

- Services provided to the upper layers
- Transport service primitives
- Berkeley sockets
- Example of socket programming: An Internet file server.

6.1.1 Services Provided to the Upper layers

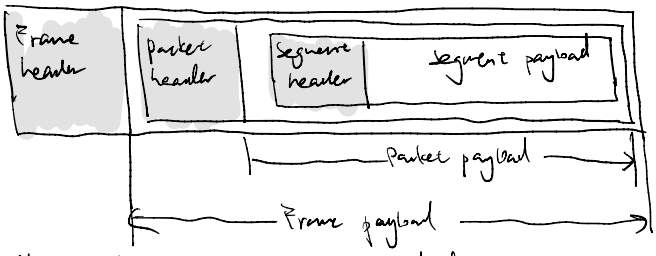
- The goal of the transport layer is to provide efficient, reliable, and cost-effective data transmission service to its users, normally processes in the application layer.
- The transport layer makes use of the services provided by the network layer. The software and/or hardware within the transport layer that does work is called the transport entity.
- The transport entity can be located in the operating system kernel, in a library package bound into network applications, in a separate user process, or even on the network interface card.



The network, transport, and application layers

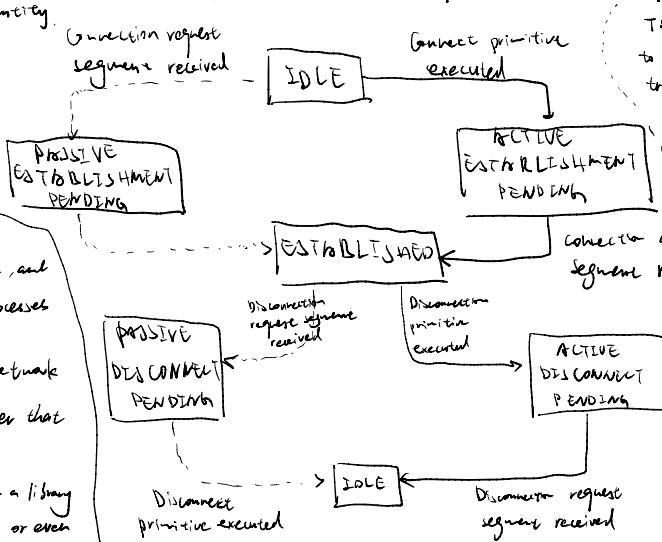
6.1.2 Transport Service Primitives

Primitive	Packet sent	Meaning
LISTEN	(none)	Block until some process tries to connect
CONNECT	CONNECTION REQ.	Actively attempt to establish a connect
SEND	DATA	Send information
RECEIVE	(none)	Block until a DATA packet arrives
DISCONNECT	DISCONNECTION REQ.	Request a release of the connection



Nesting of segments, packets and frames.

- We will use the term segment for messages sent from transport entity to transport entity.
- Segments (exchanged by the transport layer) are contained in packets (which are exchanged by the network layer). In turn, these packets are contained in frames (exchanged by the data link layer).
- When a frame arrives, the data link layer processes the frame header and, if the destination address matches for local delivery, passes the contents of the frame payload field up to the network entity. The network entity similarly processes the packet header and then passes the contents of the packet payload up to the transport entity.



A simple connection management scheme. The solid lines show the client's state sequence. The dashed lines show the server's state sequence.

6.1.3 Berkeley Sockets

Socket primitives were first released as part of the Berkeley UNIX 4.2BSD software distribution in 1983. They quickly became popular. The primitives are now widely used for Internet programming on many operating systems.

The socket primitives for TCP

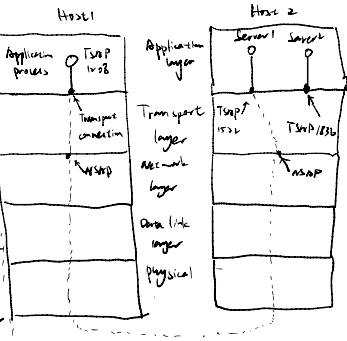
Primitive	Meaning
SOCKET	Create a new communication endpoint
BIND	Associate a local address with a socket
LISTEN	Announce willingness to accept connections; give queue size
ACCEPT	Passively establish an incoming connection
CONNECT	Actively attempt to establish a connection
SEND	Send some data over the connection
RECEIVE	Receive some data from the connection
CLOSE	Release the connection

6.1.4 An Example of Socket Programming: An Internet File Server.

6.2 Elements of Transport Protocols

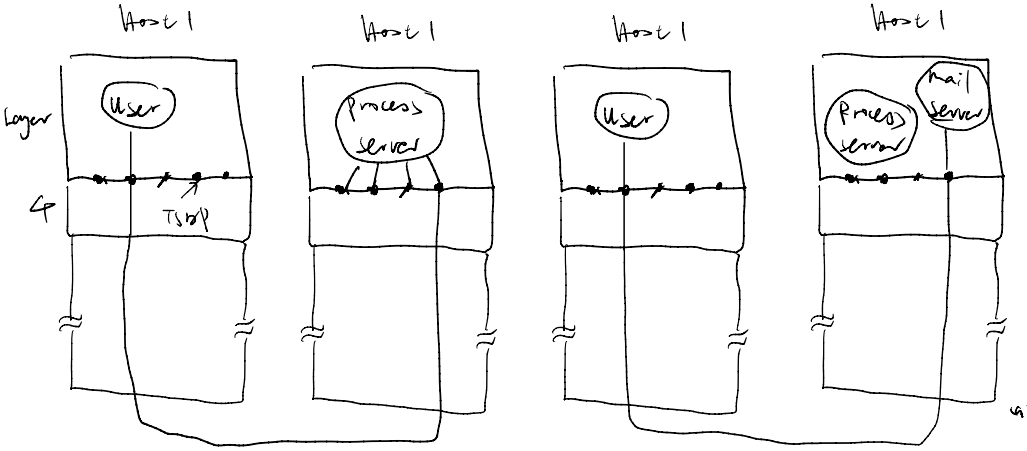
- Addressing
- Error control and flow control
- Multiplexing
- Crash recovery.

6.2.1 Addressing



TSAPs, NSAPs, and transport connections.

- The transport addresses to which processes can listen for connection requests are called ports.
- We will use the generic term TSAP (Transport Service Access Point) to mean a specific endpoint in the transport layer.
- The analogous endpoints in the network layer are called NSAPs. IP addresses are examples of NSAPs.

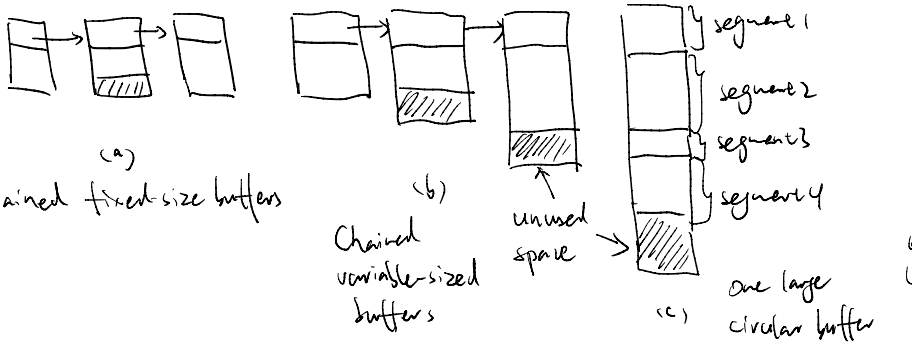


How a user process in host 1 establishes a connection with a mail server in host 2 via a process server.

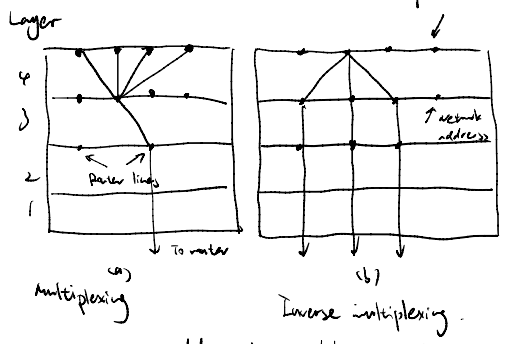
- Initial connection protocol: Instead of every conceivable server listening at well-known TSAP, each machine that wishes to offer services to remote users has a special process server that acts as a proxy for less heavily used servers.
- It listens to a set of ports at the same time, waiting for a connection request. If no server is waiting for them, they get a connection to the process server.
- The new server does the requested work, while the process server goes back to listening for new requests.

6.2.4 Error Control and Flow Control

- Error Control is ensuring that the data is delivered with the desired level of reliability, usually that all of the data is delivered without any errors.
- Flow Control is keeping a fast transmitter from overrunning a slow receiver.
- A frame carries an error-detecting code (e.g. a CRC or checksum) that is used to check if the information was correctly received.
- A frame carries a sequence number to identify itself and is retransmitted by the sender until it receives an acknowledgement of successful receipt from the receiver. This is called ARQ (Automatic Repeat reQuest)
- There is a maximum number of frames that the sender will allow to be outstanding at any time, pausing if the receiver is not acknowledging frames quickly enough. If this maximum is one packet the protocol is called stop-and-wait.
- The sliding window protocol combines these features and is also used to support bidirectional data transfer.



6.2.5 Multiplexing



(a) If only one network address is available on a host, all transport connections on that machine have to use it. When a segment comes in, some way is needed to tell which process to give it to. This situation called multiplexing.

(b) A host has multiple network paths that it can use. If a user needs more bandwidth or more reliability than one of the network paths can provide, a way out is to have a connection that distributes the traffic among multiple network paths on a round-robin basis. This is called inverse multiplexing.

6.2.6 Crash Recovery

Strategy used by receiving host

Strategy used by sending host	First ACK, then write: OK	OK	OK	First write, then ACK: OK	DUP	DUP
Always retransmit	OK	DUP	OK	LOST	OK	OK
Never retransmit	LOST	OK	LOST	LOST	DUP	OK
Retransmit in S0	OK	DUP	LOST	OK	OK	DUP
Retransmit in S1	LOST	OK	OK			

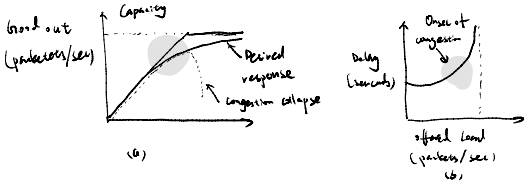
OK = Protocol functions correctly
DUP = Protocol generates a duplicate message
LOST = Protocol loses a message.

Each client can be one of two states: one segment outstanding S1, or no segments outstanding S0.

6.3 Congestion Control

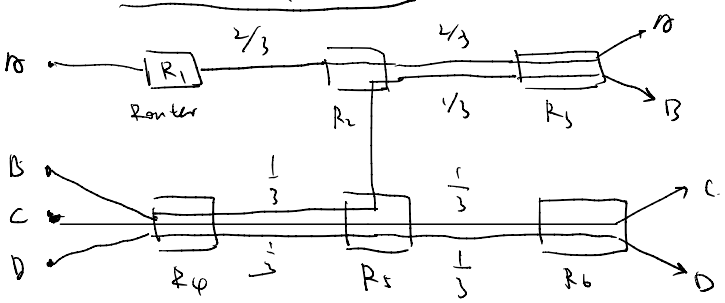
- If the transport entities on many machines send too many packets into the network too quickly, the network will become congested, with performance degraded as packets are delayed and lost.
- Controlling congestion to avoid this problem is the combined responsibility of the network and transport layers.
- Congestion occurs at routers, so it is detected at the network layer. However, congestion is ultimately caused by traffic sent into the network by the transport layer.
- The only effective way to control congestion is for the transport protocols to send packets into the network more slowly.

6.3.1 Desirable Bandwidth Allocation



6.3.1.2 Desirable Bandwidth Allocations

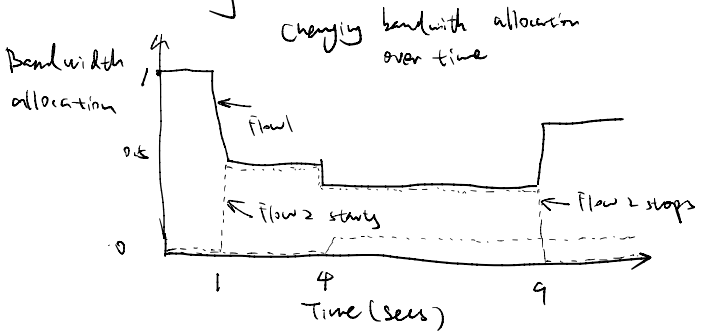
Min-Max Fairness



Max-min bandwidth allocation for four flows.

- The form of fairness that is often desired for network usage is max-min fairness. An allocation is max-min fair if the bandwidth given to one flow cannot be increased without decreasing the bandwidth given to another flow with an allocation that is no larger.
- Each of the links between routers has the same capacity, taken to be 1 unit, though in the general case the links will have different capacities.
- If more of the bandwidth on the link between R2 and R3 is given to flow B, there will be less for flow A. This is reasonable as flow A already has more bandwidth. However, the capacity of flow C or D (or both) must be decreased to give more bandwidth to B, and these flows will have less bandwidth than B. Thus, the allocation is max-min fair.

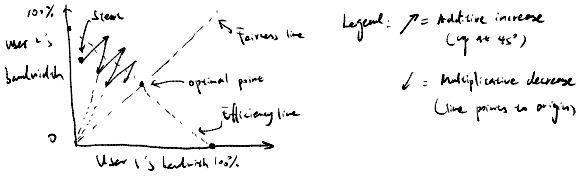
6.3.1.3 Convergence



Because of the variation in demand, the ideal operating point for the network varies over time. A good congestion control algorithm should rapidly converge to the ideal operating point, and it should track that point as it changes over time. Initially, flow 1 has all of the bandwidth. One second later, flow 2 starts. It needs bandwidth. At 4 seconds, a third flow joins. However, this flow uses only 20% of the bandwidth, which is less than its fair share (which is a third). Flow 1 and 2 quickly adjust, dividing the available bandwidth to each have 40% of the bandwidth. At 9 seconds, the second flow leaves, and the third flow remains unchanged. The first flow quickly captures 80% of the bandwidth.

6.3.2.1 Regulating the Sending Rate (AIMD)

Additive Increase Multiplicative Decrease (AIMD) control law.

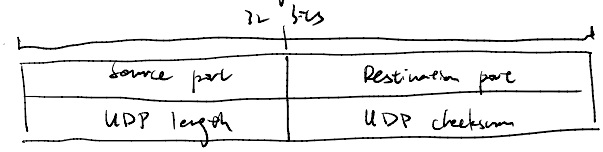


The users additively increase their bandwidth allocations and then multiplicatively decrease them when congestion is signaled. This behavior is the AIMD control law. It can be seen that the path traced by this behavior does converge to the optimal point that is both fair and efficient.

6.4.1 Introduction to UDP

- The connectionless protocol is UDP (User Datagram Protocol)
- It does almost nothing beyond sending packets between applications, letting applications build their own protocols on top as needed.
- UDP provides a way for applications to send encapsulated IP datagrams without having to establish a connection.

The UDP header



- The source port is primarily needed when a reply must be sent back to the source.
- By copying the source port field from the incoming segment into the Destination port field of the outgoing segment, the process sending the reply can specify which process on the sending machine is to get it.
- The UDP length field includes the 8-byte header and the data.
- An optional checksum is also provided for extra reliability.