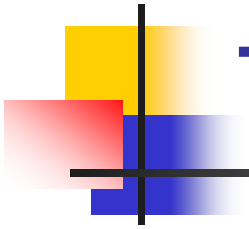
Computer Network Programming

The Transport Layer

Dr. Sam Hsu
Computer Science & Engineering
Florida Atlantic University



The Transport Layer

- The Big Picture
- Overview of TCP/IP protocols
- TCP Packet Exchanges
- TCP State Transition Diagram
- SCTP Packet Exchanges
- SCTP State Transition Diagram
- TCP and UDP Output
- Buffer Sizes and Limitations
- Port Numbers
- Internet Services

The Big Picture

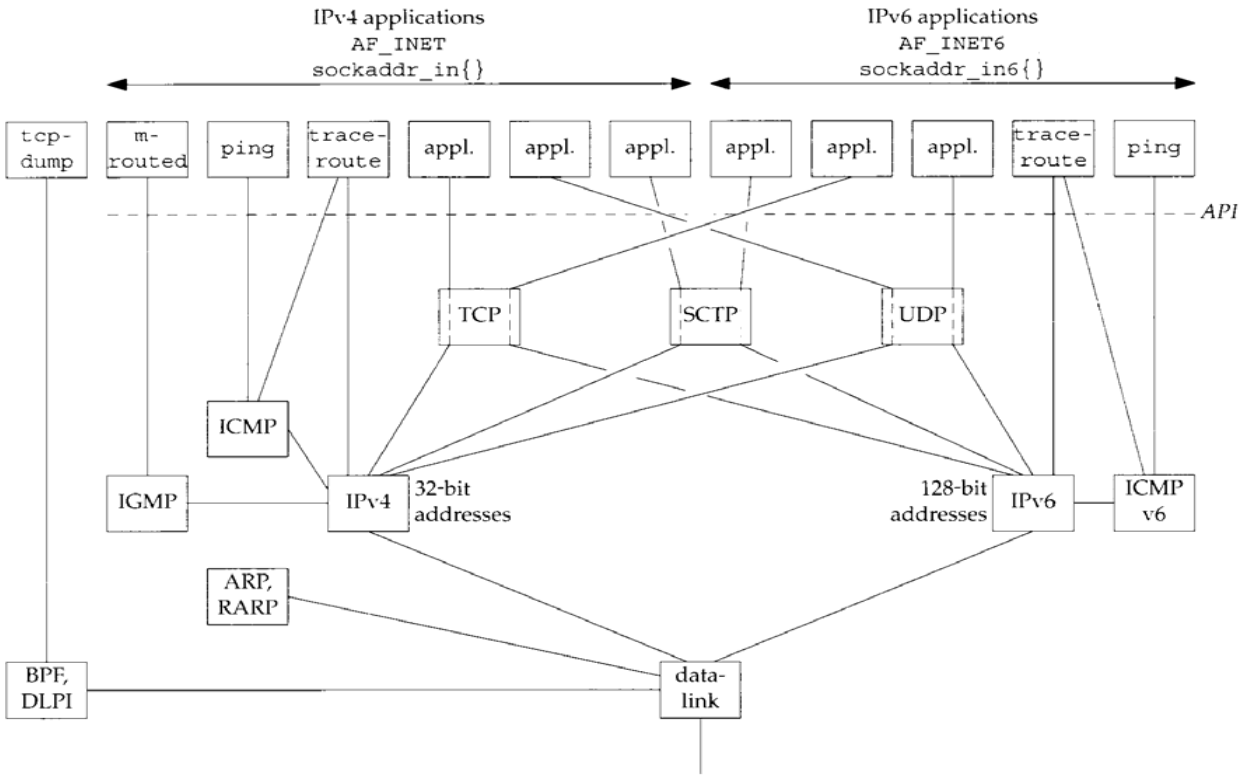
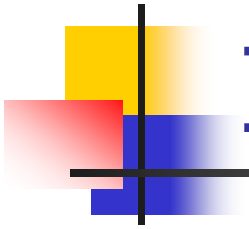
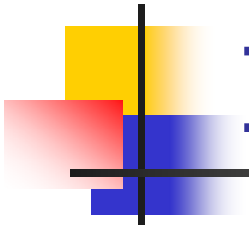


Figure 2.1 Overview of TCP/IP protocols.



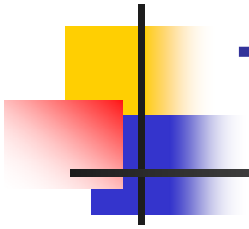
IPv4

- IPv4: *Internet Protocol version 4*
  - IPv4, often denoted as IP, has been the workhorse protocol of the TCP/IP protocol suite since the early 1980s.
  - IPv4 is a connectionless protocol.
  - IPv4 uses 32-bit addresses.
  - IPv4 packets are called *datagrams*.
  - IPv4 provides packet delivery service for TCP, UDP, SCTP, ICMP, and IGMP.



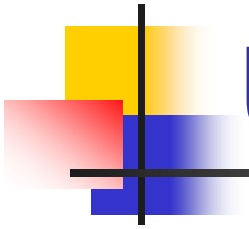
IPv6

- IPv6: *Internet Protocol version 6*
  - IPv6 was designed in the mid-1990s as a replacement for IPv4.
  - IPv6 is a connectionless protocol.
  - IPv6 uses 128-bit addresses.
  - IPv6 has built-in security features.
  - IPv6 provides QoS support.
  - IPv6 packets are called *datagrams*.
  - IPv6 provides packet delivery service for TCP, UDP, SCTP, ICMPv6.



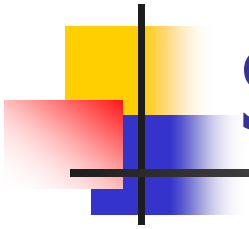
TCP

- TCP: *Transmission Control Protocol*
  - TCP is a connection-oriented protocol.
  - TCP provides a reliable, full-duplex byte stream service.
  - TCP features include acknowledgments, timeouts, retransmissions, end-to-end flow control, etc.
  - TCP packets are called *segments*.
  - TCP can use either IPv4 or IPv6.



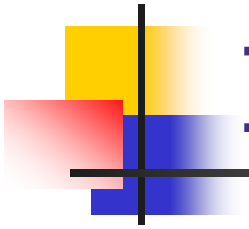
UDP

- UDP: *User Datagram Protocol*
  - UDP is a connectionless protocol.
  - UDP does not guarantee packet delivery.
  - UDP does not provide acknowledgments, timeouts, retransmissions, end-to-end flow control, etc.
  - UDP packets are called *datagrams*.
  - UDP can use either IPv4 or IPv6.



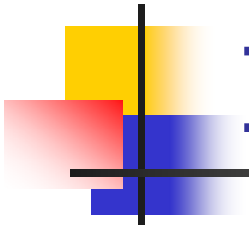
SCTP

- SCTP: *Stream Control Transmission Protocol*
  - SCTP is a connection-oriented protocol.
  - SCTP provides a reliable, full-duplex, message-oriented service.
  - SCTP provides *associations* between clients and servers.
  - SCTP is *multihomed*.
  - An SCTP packet is composed of a common header and *chunks*.
  - SCTP can use either IPv4 or IPv6, or both simultaneously on the same association.



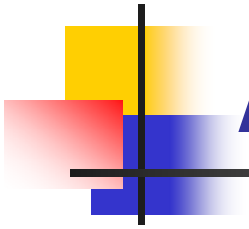
ICMP

- ICMP: *Internet Control Message Protocol*
  - ICMP provides error reports and control information between routers and hosts.
  - ICMP is message-oriented.
  - ICMP messages are transmitted via IP datagrams.
  - ICMP can be either ICMPv4 or ICMPv6.



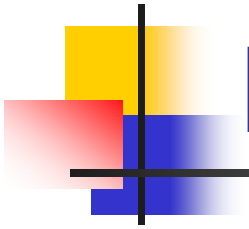
IGMP

- IGMP: *Internet Group Management Protocol*
  - IGMP is used by routers and hosts that support multicasting to exchange group membership information.
  - IGMP is message-oriented.
  - IGMP messages are transmitted via IP datagrams.



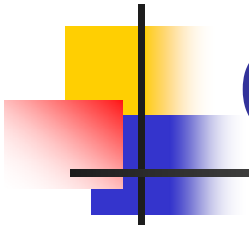
ARP/RARP

- ARP: *Address Resolution Protocol*
  - ARP maps an IPv4 address to an underlying hardware address.
  - ARP is normally used on broadcast networks such as Ethernet, token ring, and FDDI
- RARP: *Reverse Address Resolution Protocol*
  - RARP maps an underlying hardware address to an IPv4 address.
  - RARP is usually used by a diskless host to obtain its IP address from its server during booting.



BPF/DLPI

- BPF: *BSD Packet Filter*
  - BPF is an interface that provides access to the datalink layer.
  - BPF is normally found on Berkeley-derived UNIX.
- DLPI: *Datalink Provider Interface*
  - DLPI is an interface that provides access to the datalink layer.
  - DLPI is normally provided on SVR4 UNIX.

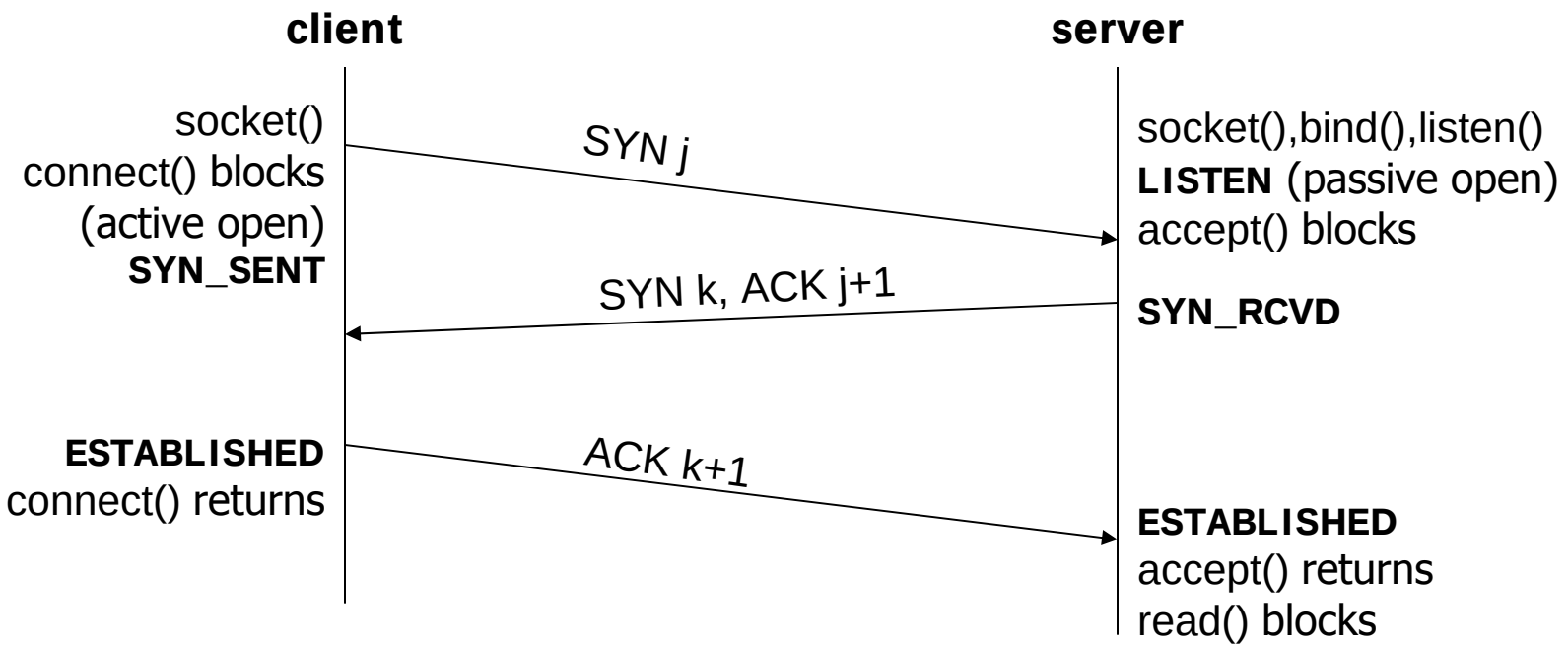


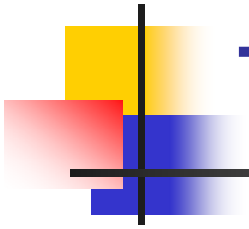
Comparing TCP/UDP/SCTP

Features	TCP	UDP	SCTP
Connection between client and server	Yes, one-to-one	No	Yes, one-to-one, one-to-many (multi-homed)
Data	Byte-stream, grouped into segments	Datagram	Byte-stream, grouped into chunks
Reliability	Yes, cumulative ack, time-out, retransmission	No	Yes, cumulative ack, time-out, retransmission
Sequencing	Yes	No	Yes
Flow/Congestion control	Yes, window-based	No	Yes, window-based
Full-duplex	Yes	Yes	Yes

TCP Connection Establishment

- Three-way Handshake



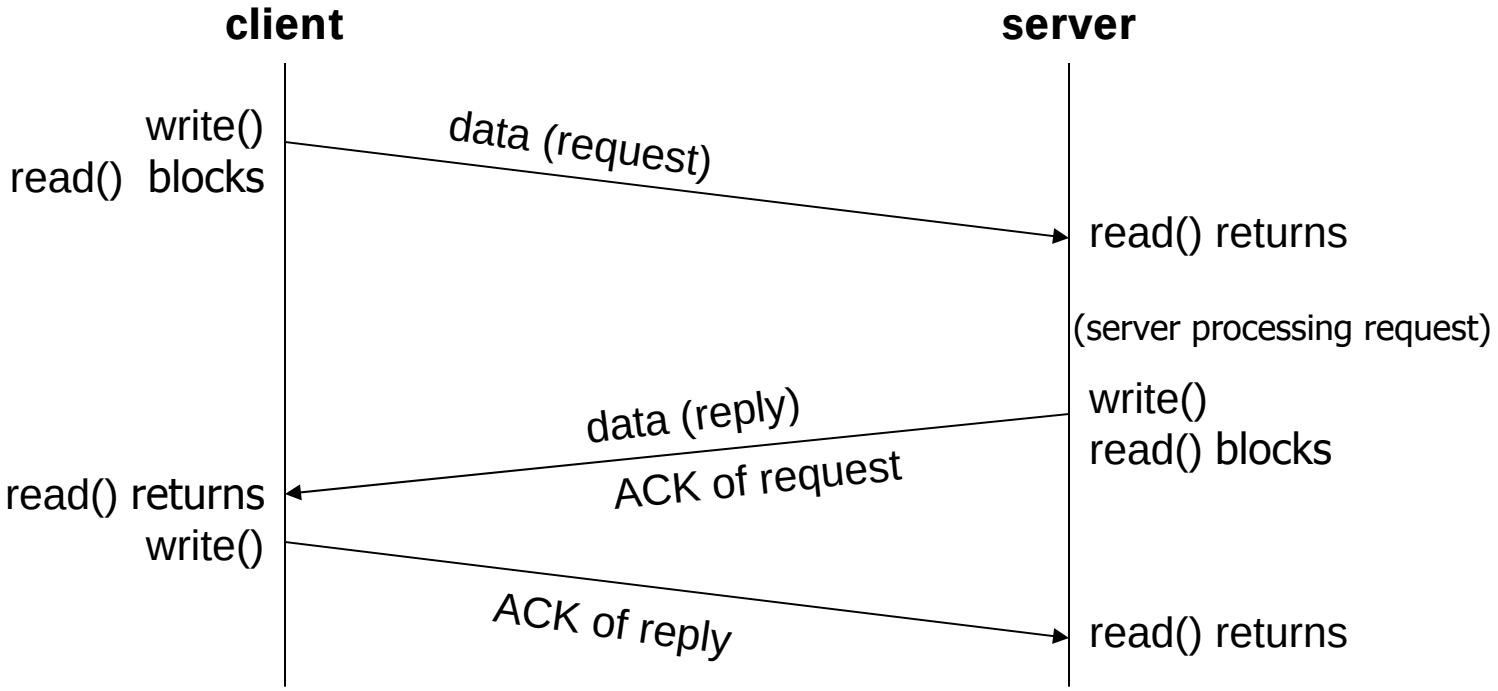


TCP Connection Options

- Three common TCP options (sent in SYN):
  - MSS option: The ***m****aximum* ***s****egment* ***s****ize* that can be received.
    - 65,535 (16 bits for the widow size field in the TCP header) is the upper limit.
  - Window scale option: To allow the advertised widow size to be scaled (left-shift) 0-14 bits.
    - MSS can thus reach up to 65,355 x 2^{14} bytes.
  - Timestamp option: May be used for high-speed connections to prevent possible data corruption caused by old, delayed, or duplicated segments.

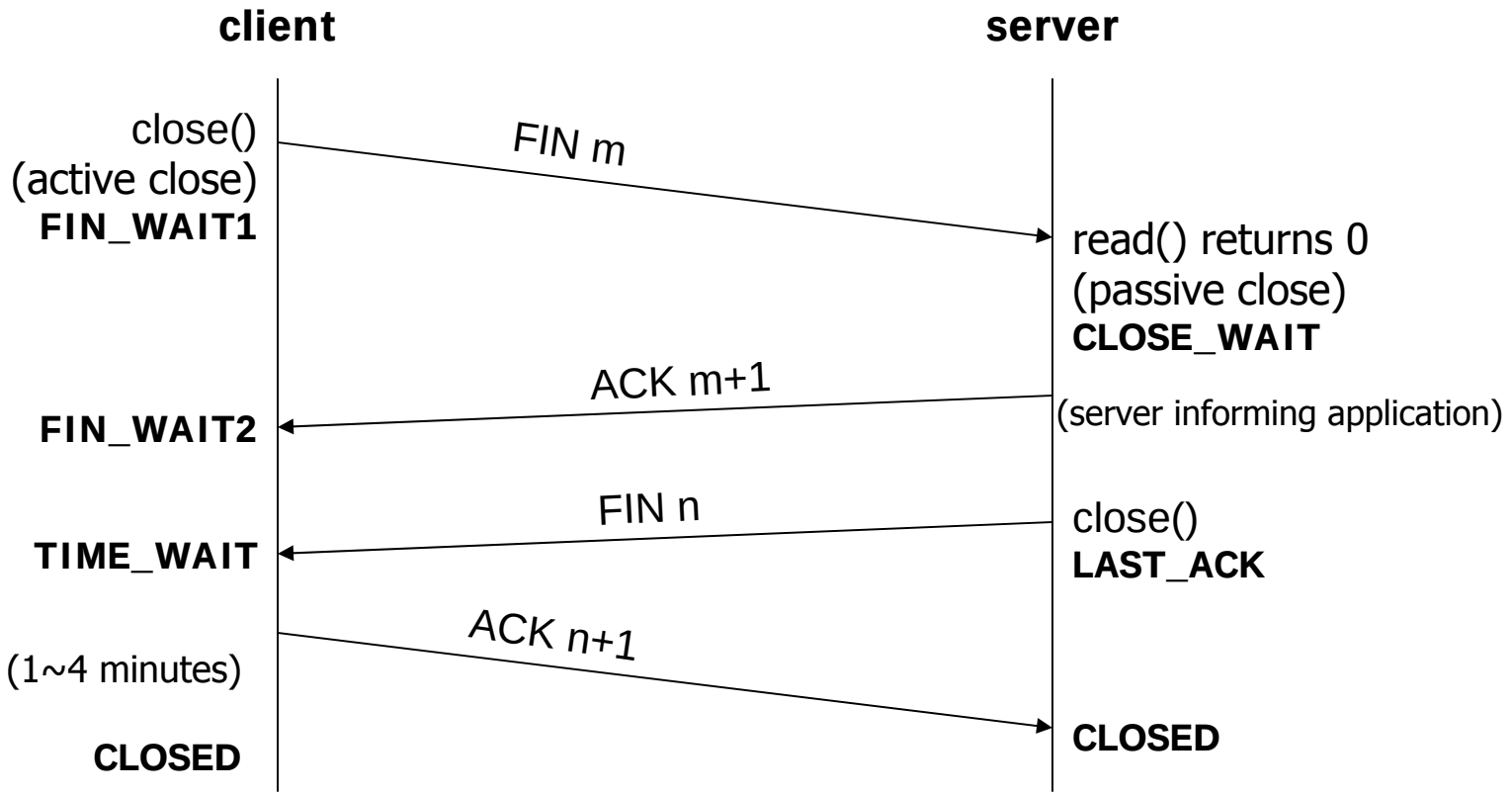
TCP Data Transfer

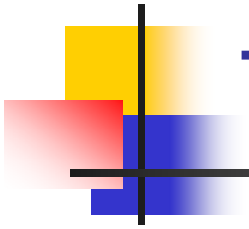
- Packets exchanged for data transfer



TCP Connection Termination

- Four-way handshake





TIME_WAIT State

- Its duration is twice the MSL (***m****aximum* ***s****egment* ***l****ifetime*), sometimes called 2MSL.
  - MSL may be 30 seconds, 1 minute, or 2 minutes.
- Two reasons for the TIME_WAIT state:
  - To implement TCP's FDX connection termination reliably.
    - To give TCP enough time to handle the loss of any of the last four segments.
  - To allow old duplicate segments to expire in the network.
    - To prevent old duplicates being misinterpreted as belonging to a new incarnation of the same connection.

TCP State Transition Diagram

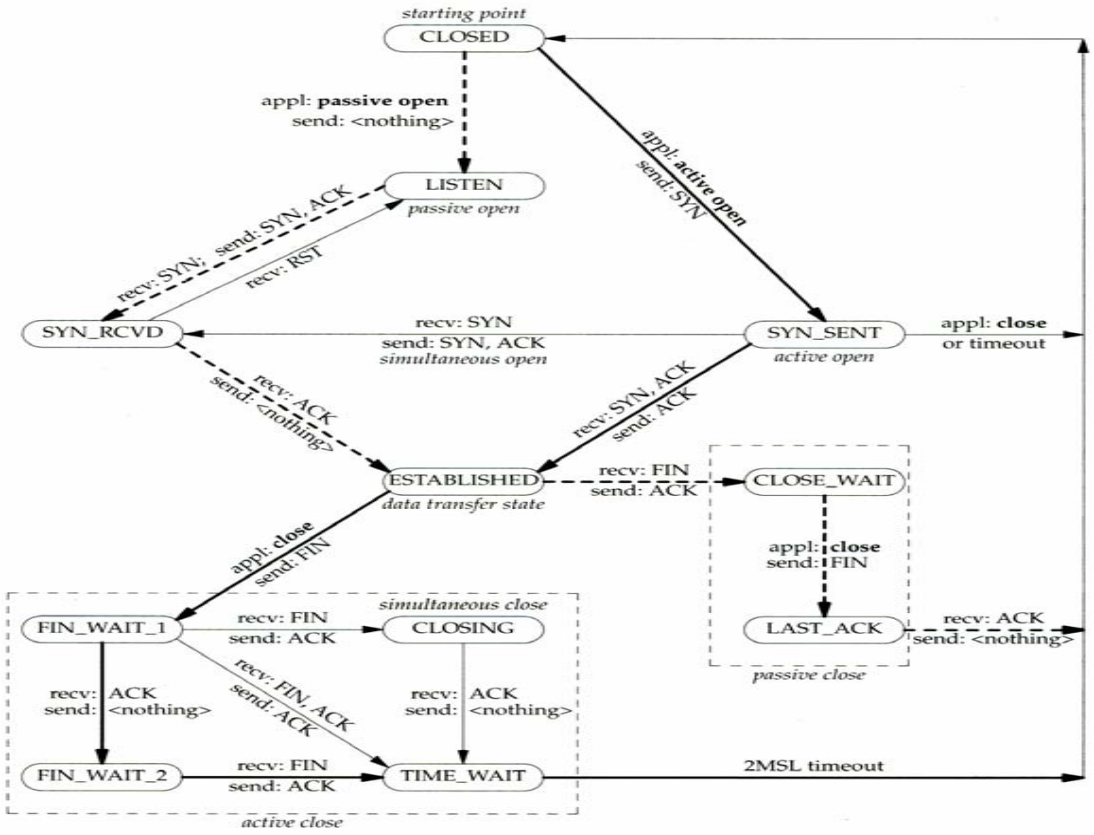


Figure 2.4 TCP state transition diagram.

TCP Packet Exchanges

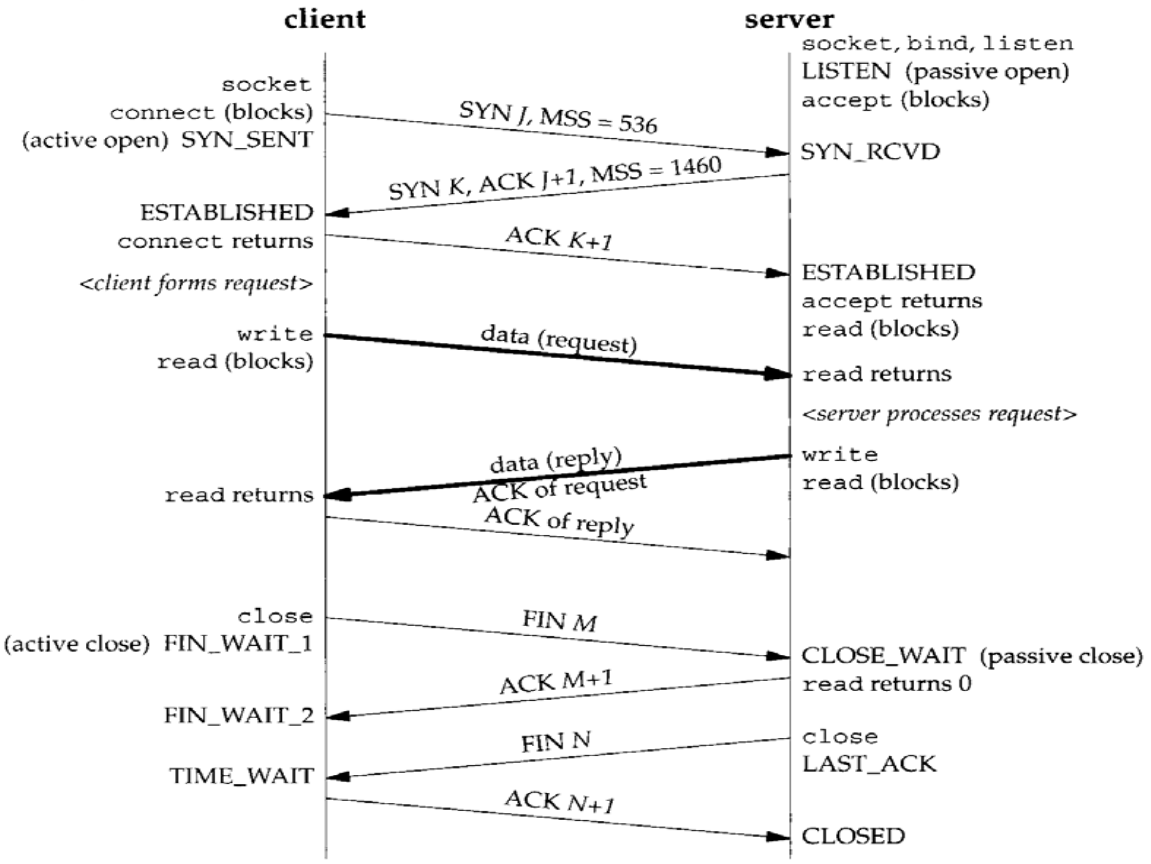
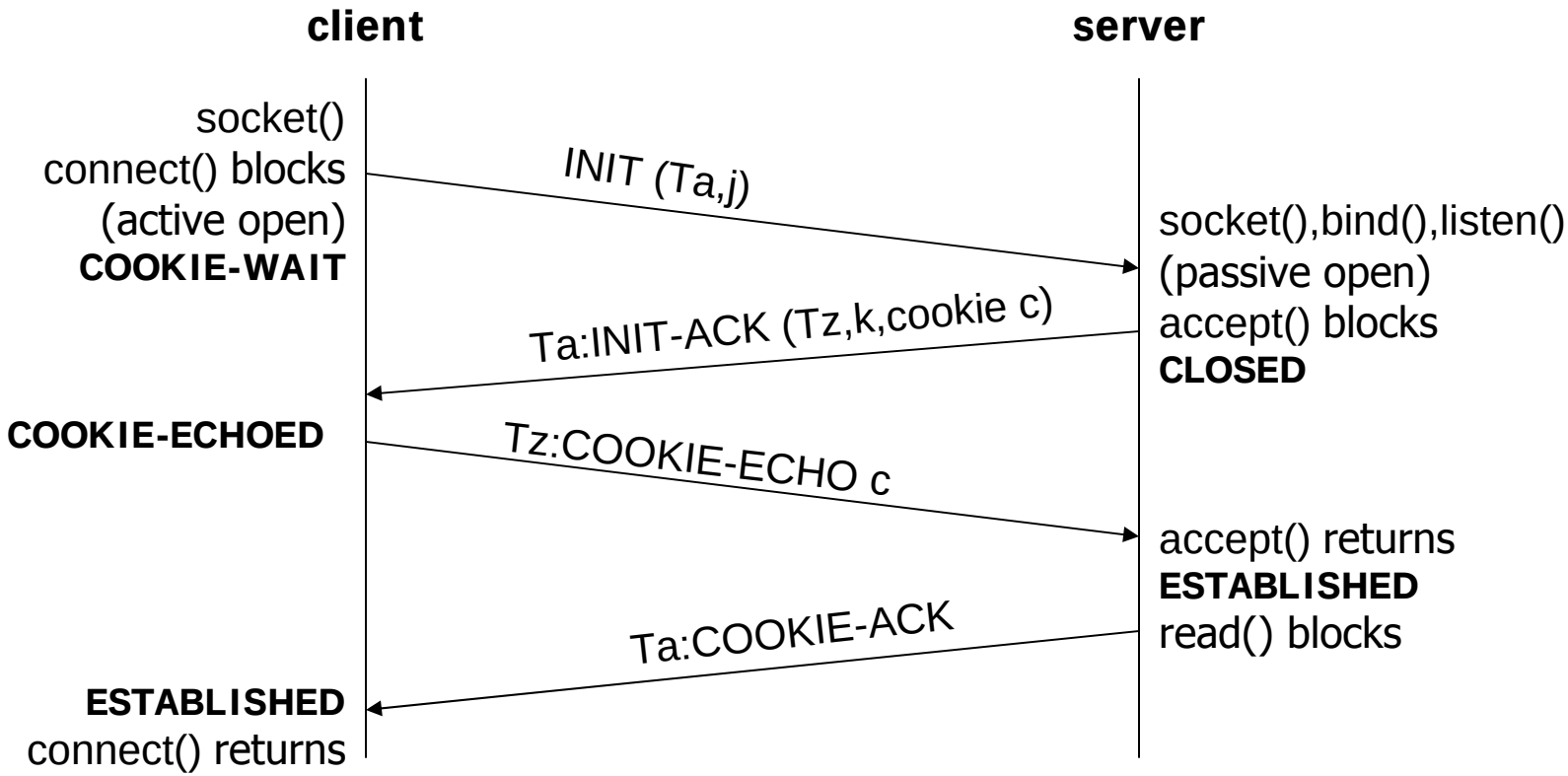
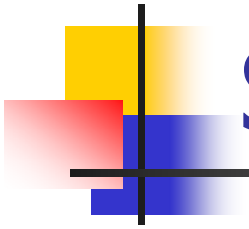


Figure 2.5 Packet exchange for TCP connection.

SCTP Association Establishment

- Four-way Handshake



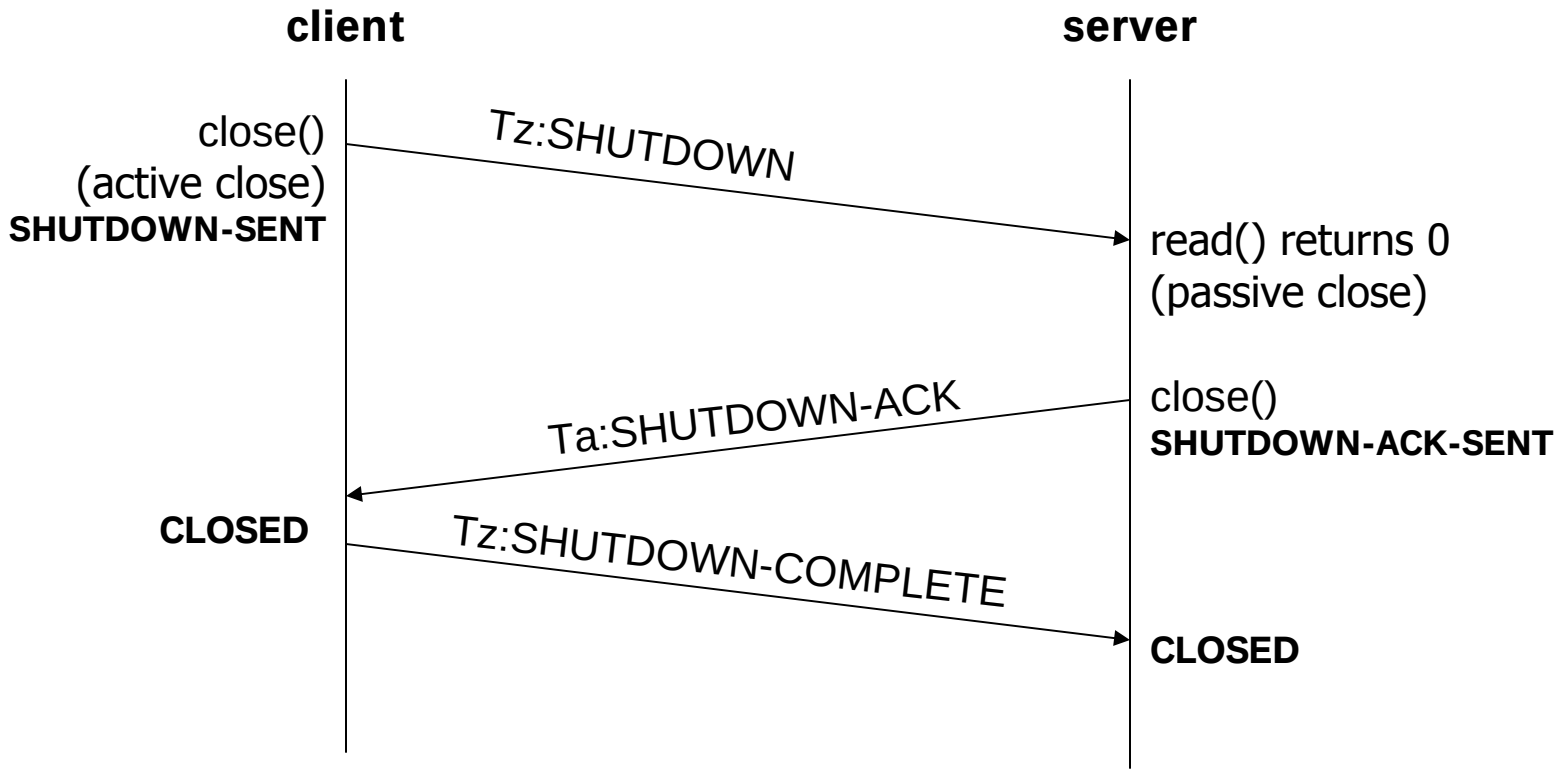


SCTP Options

- Two extensions are under development.
  - The dynamic address extension
    - It allows cooperating SCTP endpoints to dynamically add and remove IP addresses from an existing association.
  - The partial reliability extension
    - It allows cooperating SCTP endpoints, under application direction, to limit the retransmission of data.
    - When a message becomes too old to send, the message may be skipped and thus no longer sent to the peer.

SCTP Association Termination

- Three-way handshake



SCTP State Transition Diagram

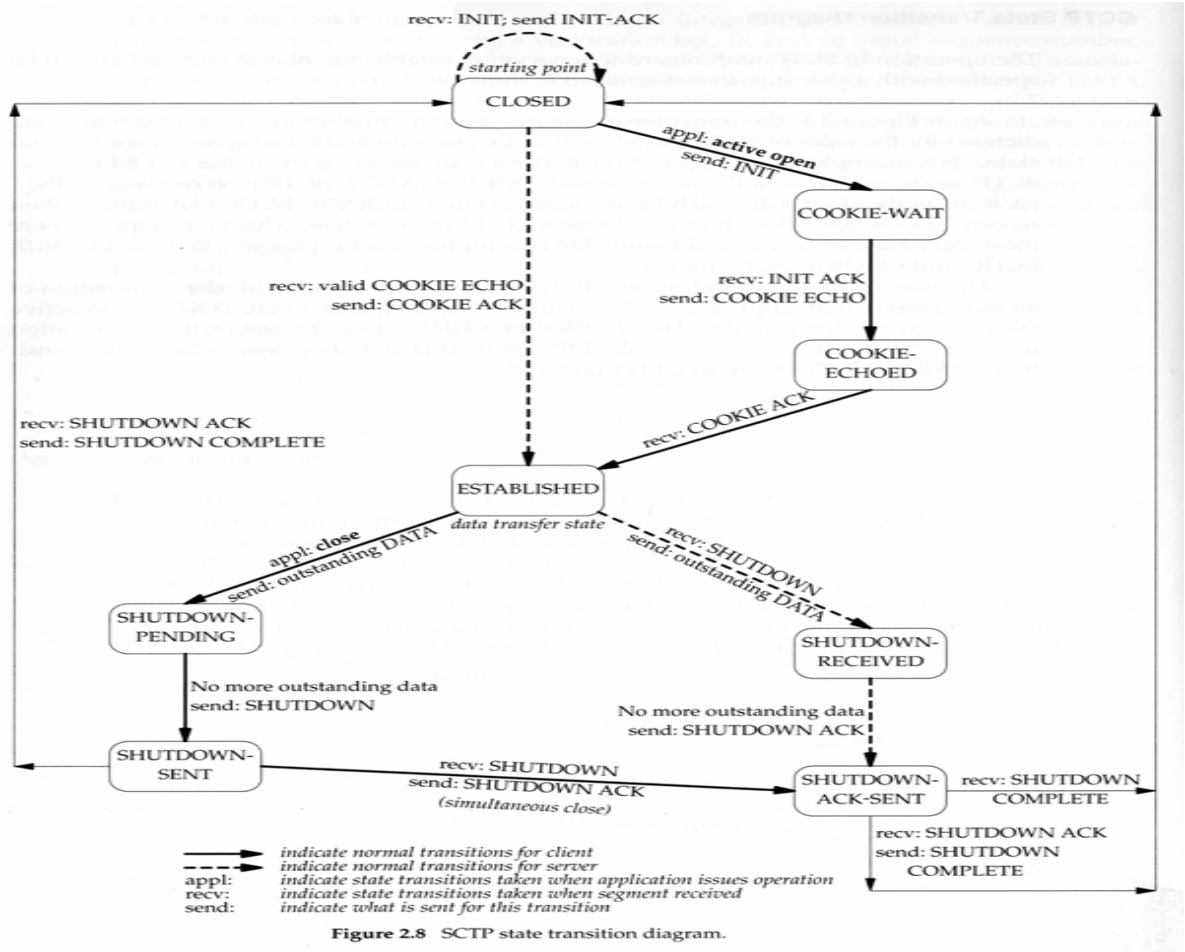


Figure 2.8 SCTP state transition diagram.

SCTP Packet Exchanges

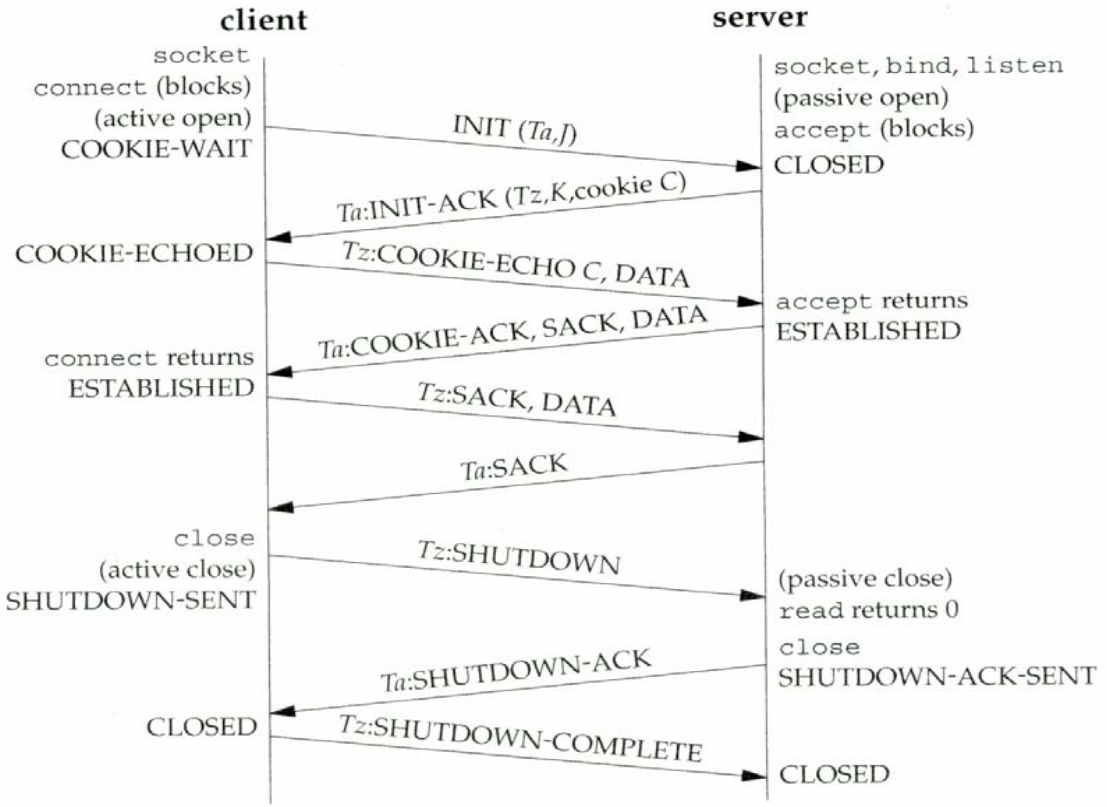
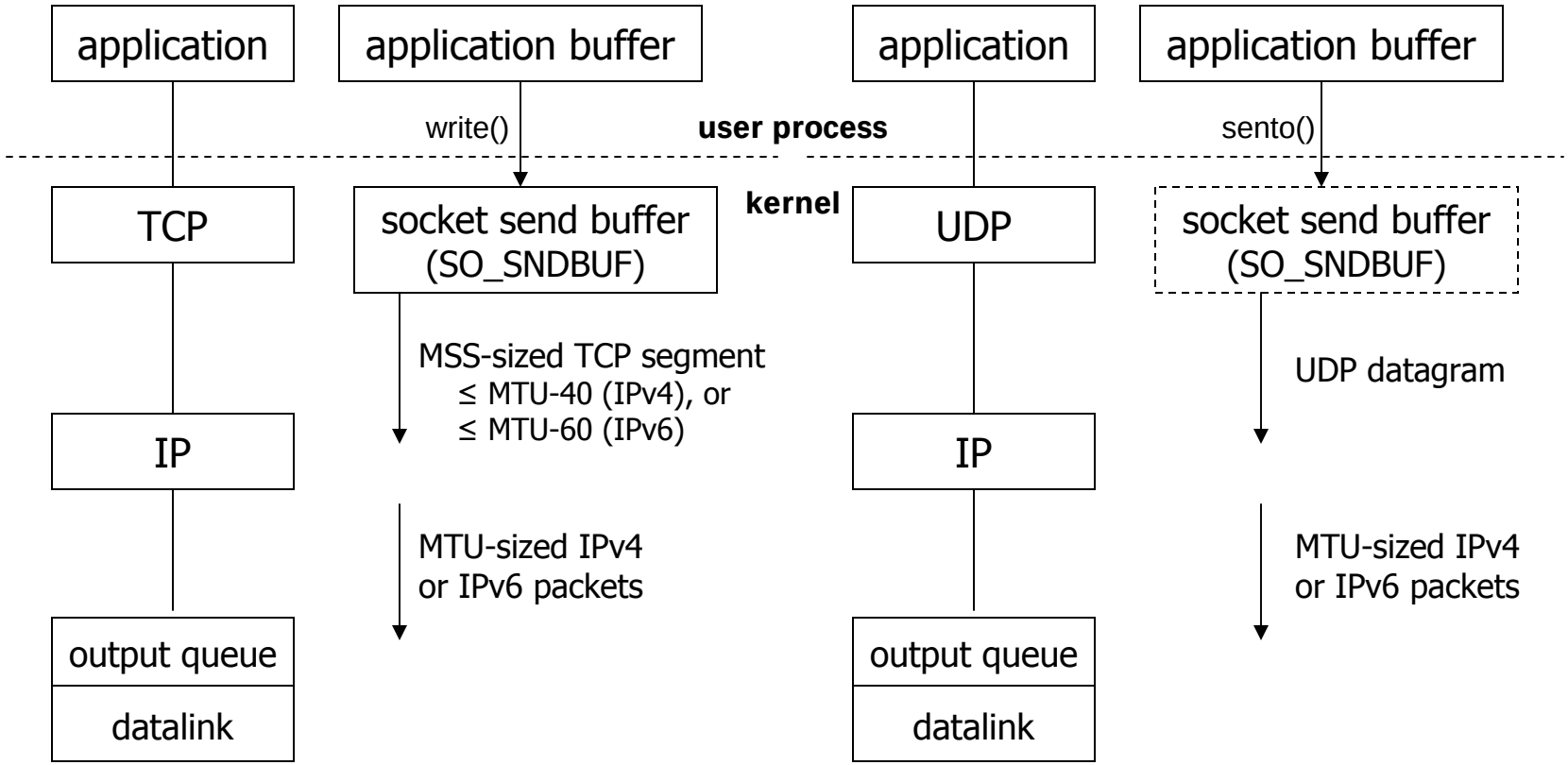
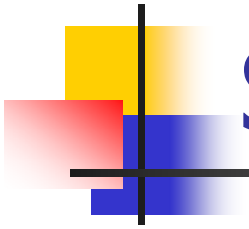


Figure 2.9 Packet exchange for SCTP association.

TCP Output and UDP Output



Ref: UNP, Stevens et al., vol 1, ed 3, 2004, AW, pp. 58-59.



Size Matters

- Maximum IP datagrams:
  - IPv4: 65,535; IPv6: 65,575 (without jumbo payload option)
- Link MTU (***m**aximum **t**ransmission **u**nit):*
  - Ethernet: 1,500; FDDI: 4,325; PPP: configurable
- Path MTU: The smallest MTU in the path between two hosts.
  - 1,500 is the popular one today.
  - Path MTU may be asymmetric between two hosts, due to possibly different route in each direction.
- TCP MSS (***m**aximum **s**egment **s**ize)*
  - It is subject to the reassembly buffer size at two hosts.
  - However, it is often the link MTU minus IP & TCP headers.

Allocation of Port Numbers

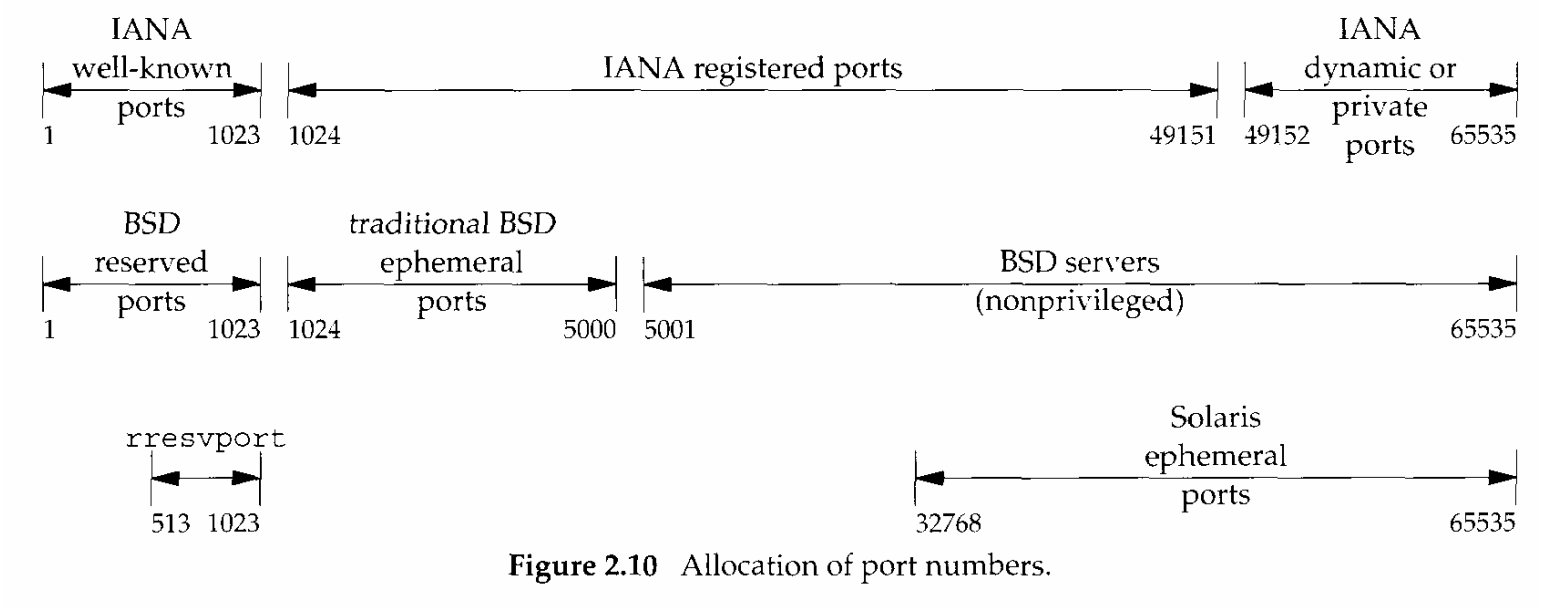
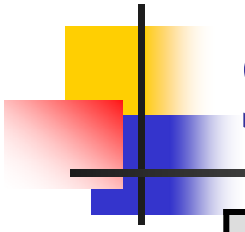


Figure 2.10 Allocation of port numbers.

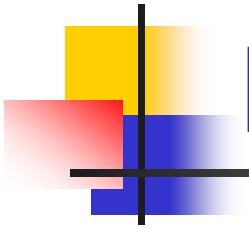


Standard Internet Services

Name	TCP Port	UDP Port	RFC	Description
echo	7	7	862	Sever returns whatever the client sends.
discard	9	9	863	Sever discards whatever the client sends.
daytime	13	13	867	Server returns the time and date in a human-readable format.
chargen	19	19	864	TCP server sends a continual stream of characters, until the connection is terminated by the client. UDP server sends a datagram containing a random number of characters (0-512) each time the client sends a datagram.
time	37	37	868	Server returns the time as a 32-bit binary number. This number represents the number of seconds since midnight January 1, 1900, UTC.

Protocol Usage by Common Internet Applications

Application	IP	ICMP	UDP	TCP	SCTP
ping		•			
traceroute		•	•		
OSPF (routing protocol)	•				
RIP (routing protocol)			•		
BGP (routing protocol)				•	
BOOTP (bootstrap protocol)			•		
DHCP (bootstrap protocol)			•		
NTP (time protocol)			•		
TFTP			•		
SNMP (network management protocol)			•		
SMTP (electronic mail)				•	
Telnet (remote login)				•	
SSH (secure remote login)				•	
FTP				•	
HTTP (HyperText Transfer Protocol)				•	
NNTP (network news)				•	
LPR (remote printing)				•	
DNS			•	•	
NFS (network filesystem)			•	•	
Sun RPC			•	•	
DEC RPC			•	•	
IUA (ISDN over IP)					•
M2UA, M3UA (SS& telephony signaling)					•
H.248 (media gateway control)			•	•	•
H.323 (IP telephony)			•	•	•
SIP (IP telephony)			•	•	•



Reading Assignment

- Read Chapter 2.