

CSIS 625 Week 11

Transport Layer and TCP/IP

Copyright 2001, 2002, 2003 – Daniel R. Oelke

For use by students of CSIS 625 for purposes of this class only.

I. Overview

A. Transport Layer

1. Up to layer 4
2. Ports, Connections, etc

B. TCP/IP

1. Network Layer - IP
2. IP Addresses, Subnets,
3. Transport Layer - UDP or TCP
4. ICMP, Arp, etc

II. Transport Layer

A. Role of Transport Layer

1. End to end delivery of messages (not just packets)
 - a. Provides segmentation and reassembly of messages into packets
2. Addressing - addition of port number
3. Flow Control
 - a. Error control and flow control typically done using a sliding window mechanism.
 - (1) Sequence numbers with ACKs and NAKs
4. Ordered Delivery
5. Reliable Delivery
 - a. At least the TCP part of TCP/IP does
 - b. There is UDP/IP which is not reliable
6. Duplicate Detection
7. May be connection oriented (TCP) or connectionless (UDP)
 - a. Connection Oriented Transport protocol provides establishment, maintenance, and termination of a logical connection

B. Port numbers

1. Transport Layer adds to network address the SAP – Service Access Point
 - a. In TCP/IP and many protocols this is called the port number
2. Provides an additional level of addressing beyond the host.
 - a. Allows for an additional level of multiplexing
3. Typically identifies the service –
 - a. HTTP server
 - b. SMTP server
 - c. POP3 server
 - d. Telnet server
 - e. Etc.
4. How does a user application know what port number to use?
 - a. User “just knows” the number - it is a configuration option
 - b. Well known port numbers are used
 - (1) /etc/services on many systems

- (2) This is commonly used for servers
- c. A name server is used
- d. Another application on a well-known port spawns a child application on some other port (remote job management or FTP)
- 5. The combination of the network addresses and the port numbers uniquely identifies a connection.
 - a. Local Network Address
 - b. Remote Network Address
 - c. Local Port Number
 - d. Remote Port Number

C. Connection Establishment

1. Typically a three-way handshake
2. Initiator sends a SYN (Synchronize sequence number) packet
3. Receiver sends back a SYN packet that also acknowledges the initiators initial sequence number
4. Initiator sends an ACK packet to acknowledge the receiver's initial sequence number
5. Now either side may start sending data
6. If the SYN packets overlap - no problem both just send ACKs.
7. Connection Establishment Security concerns
 - a. The initial sequence number must be random to prevent session hijacking.
 - (1) If not, a malicious sender can create packets that look like they come from a trusted source and inject any data that they choose.
 - b. A malicious initiator can send a lot of initial SYN packets, but never finish the 3-way transaction
 - (1) This can cause resources on the receiver to be tied up until the three-way handshake times out.

D. Connection Termination

1. One side decides it is done and sends a FIN (Finish) packet to the other.
2. The other side responds with a FIN packet.
3. After receiving the corresponding FIN packet back the session is considered closed.
4. If you receive a FIN packet, it is considered closed after sending a FIN packet back.

E. Sequence numbers

1. Some systems use a sequence number per packet.
2. Some systems use an implicit sequence number for each byte.
 - a. This means that sequence numbers can increase a lot for every packet of data.
 - b. TCP uses this scheme
3. By ensuring sequence numbers occur in order we get:
 - a. Ordered delivery
 - b. Error control for lost or damaged packets
 - c. Flow Control
 - d. Duplicate detection

F. Retransmission strategy

1. A positive acknowledgement of each received segment is required
2. If an acknowledgement is not received after some time period, a retransmission of the segment occurs
 - a. May be lost data segment -or- lost ACK

3. Timeout for retransmission
 - a. May be a fixed value - but it is difficult (impossible?) to get a good value for all situations
 - (1) Too long means sluggish response to lost packets
 - (2) Too short means many retransmissions for packets that were delayed (not lost)
 - (3) Ideal timer is just a little longer than round-trip time
 - b. May be adaptive
 - (1) Difficult because transmission and processing delays can change widely and rapidly.

G. Duplicate detection & Out of Order Data Management

1. A receiver doesn't know if a duplicate is the first copy or second
 - a. The first copy may have been delayed causing the second copy to arrive before the first.
2. The receiver acknowledges the first copy received
3. The sequence number window must be large enough so that a packet will die before sequence numbers wrap around
4. If data is received out of order
 - a. Receiver may discard segment
 - (1) Simpler
 - (2) Results in retransmission of data
 - b. Receiver may hold segment and wait for missing segment
 - (1) more complex
 - (2) reduces amount of retransmission necessary
 - (3) May not help much as out of order reception is relatively rare.

III. TCP/IP

A. TCP/IP Introduction

1. TCP/IP is the protocol used for the Internet
2. Developed in the 70's for the US Department of Defense
 - a. Arpanet - Advanced Research Project Agency NETwork
3. TCP/IP Defines the network and transport layers
 - a. Assumes a connectionless, unreliable packet oriented data link and physical layer.
 - b. May use connection oriented or non-packet data link layers, but does not take advantage of their capabilities.

B. TCP/IP by the layers

1. ARP - Address Resolution Protocol - a layer-2 to layer-3 address mapping protocol
2. IP - Internetwork Protocol is the network layer
 - a. Best effort unreliable delivery
3. TCP - Transmission Control Protocol - a connection oriented transport layer
 - a. Stream of data that is guaranteed delivery in sequence
4. UDP - User Datagram Protocol - a connectionless transport layer
5. Applications do the rest
 - a. lately there are some presentation layer type protocols for encryption (SSH is the prime example)
6. DNS - Domain Name System
 - a. A way to map names to IP addresses
 - b. Example: www.stthomas.edu => 140.209.3.54

- a. Instead of looking at the first couple of bits and determining what the class is, and therefore what the Network portion is, now all systems use a subnet mask.
- b. Subnets were started before class notation was abandoned as a way to break down bigger networks.
- c. Subnet is a 32 bit number that when bitwise-and'ed with an address breaks it into a network portion and a host portion
- d. Subnets are generally set with only the most significant bits set to 1's.
- e. This allows for a simplification where the address is written with a slash indicating number of bits in subnetmask
 - (1) Example: 192.176.32.3/24 indicates that the subnet mask is 24 bits or 255.255.255.0. This indicates a network of 192.176.32.0
- f. Does not have to end up on even byte boundaries.

E. IP – Routing

1. An interface on an IP node is generally provisioned with
 - a. IP address
 - b. Subnet mask
2. A node may have more than one interface if it connected to more than one network
3. An interface may have more than one address and subnet mask provisioned for it.
4. A node keeps a table of IP addresses and subnet masks and the IP address of the machine to send it to for those packets.
5. When an IP packet goes out – the IP protocol stack looks at each of the routes, and then the addresses and subnet masks and sends it out the one that matches.
6. If none of the routes match – a special default route is used.
7. Default gateway – the IP address of the machine that gets packets by default.
 - a. Typically the default gateway is a router that forwards packets to the correct network

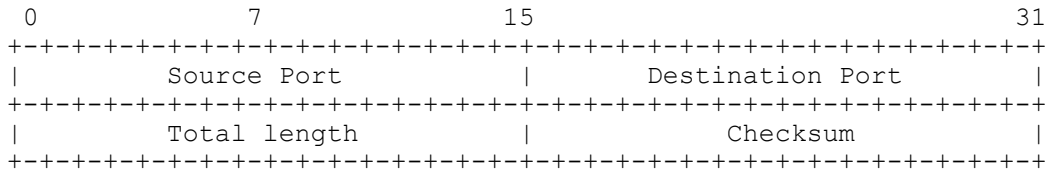
F. ICMP - Internet Control Message Protocol

1. Documented in RFC 792
2. Uses IP to transport messages, but is not a fully separate transport layer protocol because it is so integrated with IP
3. Reports some errors - but not everything so it isn't there to make IP reliable.
4. Does not send error messages when the source of the destination address isn't an individual address (multi-cast, loopback, etc)
5. Does not send error messages for ICMP messages (avoid the infinite loop)
6. ICMP - Types of messages:
 - a. Echo & Echo Reply -
 - (1) Used for “ping” command to see if a node is there
 - b. Destination unreachable
 - (1) A router in-between can't forward the packet because a link is down
 - (2) The end node doesn't have a service running on that port.
 - c. Source Quench
 - (1) Meant to be a way for the destination to tell the source to slow down
 - (2) Often not used
 - d. Redirect
 - (1) A router tells the previous node a better way to send the packet.
 - e. Time Exceeded

- (1) The TTL value of a packet counted down to zero before the packet could be delivered.
- (2) Used by the traceroute command.

G. Transport layer – UDP/IP

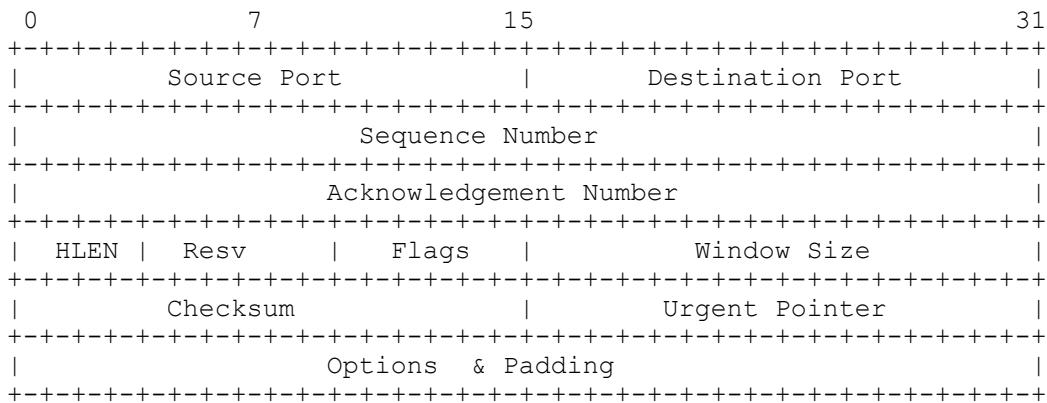
- 1. UDP is simple in that all it really has to support in addition to the IP header is port addresses.
- 2. Header format



- 3. Source and destination ports determine the service that is running them
- 4. Checksum protects the UDP header (not the packet data)

H. Transport layer – TCP/IP

- 1. TCP is connection oriented so it must provide connection setup and teardown as well as provide mechanisms for reliable packet delivery.
- 2. Header format



3. TCP Header

- a. Port Addresses – used to identify services
- b. Sequence Number & Acknowledgement number – used for sliding window flow control and error control
- c. HLEN – Header length in multiples of 4
- d. Resv – Reserved for future use
- e. Flags –
 - (1) URG – Urgent – there is urgent data in the data portion
 - (2) ACK – The acknowledgement field is valid
 - (3) PSH – Push – higher throughput is desired
 - (4) RST –
 - (5) SYN – Sequence number synchronization in the connection setup
 - (6) FIN – connection termination
- f. Window Size – The number of packets that can be sent.
- g. Checksum – error detection for the header (not the data)
- h. Urgent pointer – an offset into the data portion for where the urgent data is (if the URG flag is set).

4. TCP flow control

- a. TCP uses a modified sliding window technique - called a credit scheme
- b. Each ACK has both a number in the window for the bytes that are being acknowledged, and a number in the window that may be sent up to before acknowledgement.
- c. Slow Start - developed by Van Jacobson - 1988
 - (1) Exponentially increases the window size as data is successfully sent.
 - (2) Allows the amount of data being sent to grow up to the network capacity.
 - (3) Causes “slowness” for very short data transfers
- d. Dynamic Window Sizing on Congestion
 - (1) When a packet is lost, and retransmitted – the window size is cut dramatically and slow-start redone up to ½ the previous level.
 - (2) From that point on – a slower linear rather than exponential growth is taken.
- e. These methods when widely implemented allow the Internet to work even in the face of extreme loads.
- f. Fortunately few people have the ability to re-write their TCP/IP stack and defeat these mechanisms.

I. ARP - Address Resolution Protocol

1. Used as a way for IP to map an Ethernet Address to an IP address
2. When a node wants to send an IP datagram over an Ethernet network, it needs to know the MAC address of the destination.
3. An Ethernet broadcast is sent out asking who owns this IP address
4. The node with the address replies.
5. From the reply the original node gets the MAC address.
6. Now the IP packet can be sent over the Ethernet to the destination.
7. ARP Cache
 - a. The sender keeps a cache of recently resolved addresses so it doesn't have to ARP before sending every packet
 - (1) This cache can often be displayed using “arp” command
 - (2) This cache must time out if one node stops using an IP address and another starts.
 - b. When one node sends out an arp reply message, all nodes on a broadcast network may add it to their cache.
8. Proxy-ARP
 - a. Sometimes an administrator will want to merge two separate Ethernet networks to look like one for IP packets
 - b. A router can be configured so that it will send an ARP response on an interface for a whole range of IP addresses.
 - c. The router will then receive the packets, and forward them to the correct Ethernet network
 - (1) Will need to do an ARP request on that interface to find the actual node's MAC address.
 - (2) Router will typically be configured to proxy-ARP in both directions.

J. DNS - Domain Name System

1. A protocol and the whole system for mapping names of machines to IP addresses
2. The protocol is usually over UDP packets.
 - a. Unreliable, but since message is only one packet to the server and one packet in response it has lower overhead than TCP.

3. A node is typically configured with the IP address of one or more DNS servers.
 - a. If the first one fails to respond, the second one is used, etc.
4. Top Level Domain - the last set of letters after a period (.) in a domain name.
5. Root name server - the master domain name server for a given top level domain.
6. DNS Control
 - a. ICANN - Internet Corporation for Assigned Names and Numbers
 - (1) Created by US government as a way to sort out the management of DNS
 - (2) Very controversial in how it has been handling things
 - b. Each of the top level domains has a single database maintainer
 - (1) .com, .net, .org are all through Network Solutions
 - (2) .gov is controlled by the United States government
 - (3) Each country has a two letter top level domain
 - (a) (.us, .cc, .tv, .ru, .uk, .de, .au, etc.)
 - c. There may be multiple companies that register names into that database, but a single database exists.
 - d. Some people have started creating alternative name servers.