Data Communications and Networks

CSCI-GA.2262-001 – Spring 2016

Dr. Jean-Claude Franchitti

New York University
Computer Science Department
Courant Institute of Mathematical Sciences

Midterm Examination Solutions

This midterm exam covers the following data communications and networks topics covered in class: introduction and overview, the application layer, data encoding and transmission, data link control, and wireless and mobile networks.

As the exam is rather long, it is EXTREMELY IMPORTANT to use GOOD TEST TAKING STRATEGY. That is, first answer the problems that you can answer easily and quickly, and then go to the problems that are going to take you more work. Do not waste a lot of time at the beginning trying to figure out a problem that you find difficult. The answers do not have to be in order in your exam booklet.

General Directions:

This is an *open book* test; you may bring the course text book(s) or notes to the room. Computers are allowed for access to material stored on the hard drive(s). No internet access via computers or electronic devices is allowed. No communication with other students in the room is allowed.

Each question is preceded by a numeric score in parenthesis: that score indicates both the weight of the question, and a (high) estimate of how many minutes you should be spending on it. The total score is 100. The weight of each question is indicated in points and therefore each point counts for 1% of the total grade. The official exam period is 110 minutes.

Write your answers in your exam booklet(s). It is not expected that you will need additional sheets, but there will be a supply of extra booklets. Please number your booklets as needed. At the end of the exam period, insert all additional booklets and your copy of the exam specification in your first booklet and return the package to the instructor.

Please make sure that you also observe the following:

☐ Please provide answers to all ten (10) questions, unless stated otherwise you will not be penalized for providing a wrong answer.
☐ Place your name on your exam booklet(s) and on your copy of the exam specification.
☐ Please read the questions carefully and consider all hints and assumptions provided

Problems:

1. General Q&As – (6 points)

   Indicate whether each of the following statement is true or false (T/F) and explain your answers if/as needed:

   1.a. (1 point) If an Ethernet destination address is 08:07:06:05:44:33, then this is a broadcast address.

   Answer: F

   1.b. (1 point) In the Internet, the domain name space (tree) is divided into three different sections.

   Answer: T

   1.c. (1 point) During an FTP session the data connection is opened as many times as necessary.

   Answer: T

   1.d. (1 point) The RTS and CTS frames in CSMA/CA can solve the hidden station problem. The RTS and CTS frames in CSMA/CA cannot solve the exposed station problem.

   Answer: T

   1.e. (1 point) In slotted ALOHA, the vulnerable time is the same as the frame transmission time.

   Answer: T

   1.f. (1 point) In TDM, the transmission rate of the multiplexed path is usually equal to the sum of the transmission rates of the signal sources.

   Answer: F (greater than)

2. Computer Networks and the Internet – (9 points)

2.a. (6 points) Consider the scenario shown in Figure 1 in which a server is connected to a router by a 100Mbps link with a 50ms propagation delay. Initially this router is also connected to two routers, each over a 50Mbps link with a 200ms propagation delay. A 1Gbps link connects a host and a cache (if present) to each of these routers and we assume that this link has 0 propagation delay. All packets in the network are 20,000 bits long.

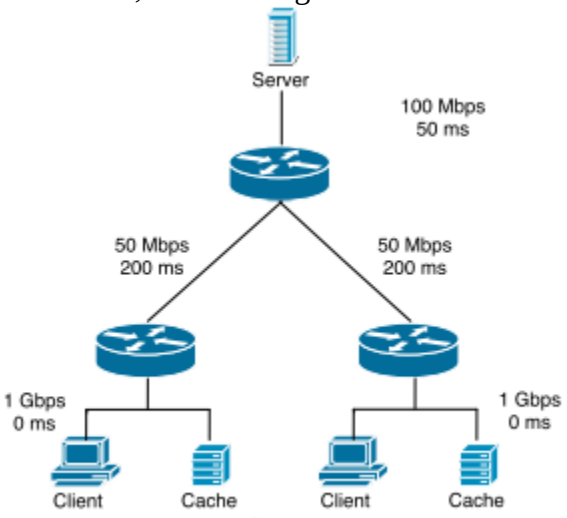


Figure 1

What is the end-to-end delay from when a packet is transmitted by the server to when it is received by the client? In this case, assume there are no caches, there is no queuing delay at the routers, and the packet processing delays at routers and nodes are all 0.

Assuming that client hosts send requests for files directly to the server (caches are not used or off in this case), what is the maximum rate at which the server can deliver data to a single client if we assume no other clients are making requests?

Answer: if all packets are 20,000 bits long it takes 200 usec to send the packet over the 100Mbps link, 400 usec to send over the 50 Mbps link, and 20 usec to send over the 1Gbps link. Sum of the three-link transmission is 620 usec. Thus, the total end-to-end delay is 250.62 msec.

2.b. (3 points) Explain why circuit switching DOES NOT exploit statistical multiplexing, and why do statistically multiplexed systems (like packet-switched systems) need to tackle stability issues unlike circuit-switched systems.

Answer: Circuit switching involves reserving some fixed bandwidth when the circuit is setup. But statistical multiplexing means sharing a link among various connections depending on the demands posed by the connections. Since circuit switching does not care if individual connections offer any demand and reserves bandwidth, it is clear that it does not use statistical multiplexing.

The statistically multiplexed systems provide gains by trading off delays and queuing of packets. If the system does not estimate the load of the constituent connections properly, there is a possibility that an excess traffic might be admitted. Such a situation can cause unbounded delay and queuing leading to system instability

3. Application Layer – FTP (10 points)

In the FTP architecture, the client and the server need to agree on the port numbers to use for the separate data-transfer connection. In one possible implementation approach referred to as "active mode," the client opens a socket with a dynamic port and sends the IP address and port number to the server (using the existing control connection) so the server knows what client address and port number to use for the data-transfer connection. For example, the client on IP address 192.168.0.1 that chose port 49150 for the data connection might send a command like "PORT 192.168.0.1 49150" over the control connection In an alternate implementation approach used to establish the data-transfer connection and referred to as "passive mode," the server selects a port number and instructs the client (using the existing control connection) to establish the data-transfer connection.

3.a. (5 points) When is it best to use passive vs. active mode and why?

Answer: Passive mode is easier than active mode in the presence of client-side NAT boxes. In passive mode, the server selects an address and port number, and sends them to the client over the control connection. The client-side NAT does not need to modify the address and port number used by the remote server. (That said, passive mode is challenging in the presence of a server-side NAT box, though this is a much less common configuration.)

3.b (5 points) How can you get around the issues with active mode identified in question 3a ?

Answer: The NAT box maps the IP address and port number of the client's data connection to new values. As such, the arguments the client sends in the PORT command (sent over the control connection) would not match the values the NAT box would use. In addition, the server would have trouble initiating a data connection to a client lying behind a NAT box, without the NAT box already having a table entry for the associated connection; since the client has not transmitted any packets on the data connection yet, the NAT box would not yet have an entry installed when the server sends the initial SYN packet. Worse yet, the IP address in the PORT command is a private, non-routable address, so the FTP server would not be able to direct a packet with this destination address to the appropriate place.

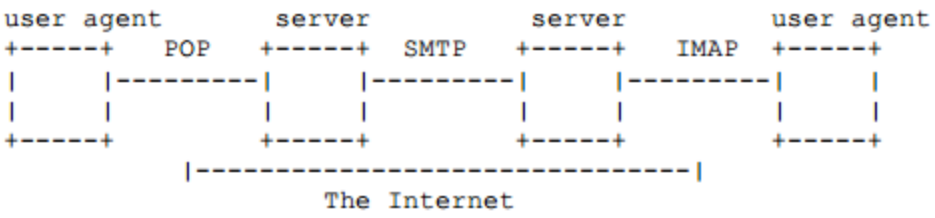
The NAT box would need to parse the messages sent on the FTP control connection (i.e., on port 21) to extract the arguments of the PORT command. The NAT box would need to create a table entry for the data connection and modify the arguments in the PORT message accordingly, so the FTP server can successfully create and use the data connection.

4. Application Layer – SMTP (8 points)

The following questions pertain to electronic mail applications in the Internet and the related SMTP protocol:

4.a. (4 points) Describe how email works. Describe the key components and flows. Identify key standards that apply. Use figures as needed. About 150 words should be sufficient.

Answer: Email is an asynchronous communications medium based on sent and received text messages (may include non-text attachments). The three major components of email are user agents, mail servers, and the SMTP (Simple Mail Transfer Protocol). SMTP is used to transfer messages between mail servers using TCP/IP (client/server). SMTP uses ASCII commands and headers. Commands are not authenticated. Between a mail server and a user agent, POP (Post Office Protocol), IMAP (Internet Mail Access Protocol), or HTTP (HyperText Transfer Protocol) is used to access received mail stored in inboxes on the mail server. Mail servers have a fixed IP address and are always powered-on. The user agents need not always be powered-on and also need not be fixed in location or IP address. The figure shows a user agent accessing a mail server with POP, two mail servers exchanging messages using SMTP, and another user agent accessing its mail server using IMAP.



4.b. (4 points) Electronic mail systems differ in the manner they handle multiple recipients. In some systems, the originating user agent or mail sender makes all the necessary copies and these are sent out independently. An alternative approach is to determine the route for each destination first. Then a single message is sent out on the common portion of the route and copies are only made when the routes diverge; this process is referred to as "mail-bagging". Discuss the relative advantages and disadvantages of the two methods.

5. Application Layer – HTTP and P2P – (11 points)

The following questions pertain to the Internet Directory Service and the World Wide Web and the related DNS and HTTP protocols.

5.a. (6 points) Assume two neighboring nodes in Circular DHT. Node A has ID 1100 and node B'sID is 1111. Another peer (neither A nor B) in the ring searches for key 1110. Is either A or B sending a reply message to the requesting node informing that it holds the <key, value> pair?

Answer: B

5.b. (5 points) Suppose that a browser on host A wants to retrieve an HTML document (D) and an embedded image (I), from host B. Assume that A does not initially know the IP address of B, but A's local name server S does know B's IP address. Also assume that the browser on A uses HTTP/1.0 (the non-persistent version).

Show the chronological sequence of transport layer segments (TCP or UDP) sent and the respective application layer data type (DNS or HTTP) by filing in the following table (add and fill in additional rows as needed). Also, for each TCP packet state when any of the SYN, FIN, and/or ACK bits in the TCP header are set to 1.

Answer:

Source	Destination	Transport Layer Protocol	Application Layer Protocol
A	S	UDP	DNS

S	A	UDP	DNS
A	B	TCP SYN=1	
B	A	TCP SYN=ACK=1	
A	B	TCP ACK=1	HTTP
B	A	TCP	HTTP
A	B	TCP FIN=1	
B	A	TCP FIN=ACK=1	
A	B	TCP ACK=1	
A	B	TCP SYN=1	
B	A	TCP SYN=ACK=1	
A	B	TCP ACK=1	HTTP
B	A	TCP	HTTP
A	B	TCP FIN=1	
B	A	TCP FIN=ACK=1	
A	B	TCP ACK=1	

6. Application Layer - DNS – (10 points)

The following questions pertain to the Internet Directory Service and the related DNS protocol.

6.a. (2 points) Before 9/11, the top-level domain (TLD) server for South Africa was located in New York City. Explain why the physical destruction on 9/11 disrupted Internet communication within South Africa (e.g., for a Web user in South Africa accessing a Web site in South Africa)?

Answer: Web transfers within South Africa rely on DNS to resolve domain names for South African Web site names (e.g., www.gov.za). The Web user's local DNS server needs to contact the TLD server for the .za domain as part of translating the Web server name into an IP address, making the Web download fail because the DNS information is not available.

6.b. (2 points) Following up on question 5a, explain why the effects in South Africa took place gradually, disrupting progressively more communication within the country in the hours (and even days) after connectivity to NYC was lost.

Answer: The local DNS servers in South Africa presumably had many name-to-address mappings cached already, particularly for popular sites. However, these cached entries had a time-to-live field that eventually expired, causing the local DNS servers to evict the expired entries. Once the expired entries were evicted, future requests from end hosts would fail due to the inability to contact the TLD server.

6.c. (2 points) How do the local DNS servers know the identity of the root servers? Why are most of the host-to-address queries seen by the root DNS servers for bogus or malformed names (like graceland.elvis or numeric top-level domains)?

Answer: Most queries do not require the local DNS server to contact the root servers, since typically the TLD servers for common TLDs like .com, .net, .edu, etc. are already cached (and have larger time-to-live values). Bogus or malformed names do not correspond to valid TLDs and, as such, are typically not cached, forcing requests to go to a root server.

6.d. (2 points) Who determines the value of the time-to-live field that determines how long DNS servers cache a name-to-address mapping? What are the pros and cons of using a small value?

Answer: The operator of the responding DNS server (e.g., the authoritative DNS server) assigns the TTL value. The advantages of a small TTL are: (i) rapid failover if the IP address associated with a name changes and (ii) enabling content distribution networks to exert fine-grain control (e.g., for load balancing) over which Web server replica handles the Web client requesst. The disadvantages of a small TTL are: (i) extra DNS requests (which place extra load on the network and the DNS servers) and (ii) extra latency for the user to wait for these DNS queries to complete successfully.

6.e. (2 point) A local DNS server typically discards cached name-to-address mappings when the time-to-live expires. Alternatively, the local DNS server could optimistically issue a new query for the cached domain name. Given one advantage and one disadvantage of that approach.

Answer: By “prefetching” the name-to-address mapping, the local DNS server can hide the DNS look-up delay, improving the performance experienced by the end user. However, prefetching introduces extra DNS queries and network load to look up name-to-address mappings that may never be needed.

7. Link Layer – Encodings and Error-Detection & Correction Techniques – (16 points)

   7.a. (5 points) For the bit stream 1010001, sketch the waveforms for NRZ-L, NRZI, Manchester, and Differential Manchester.

   7.b. (5 points) Compute the CRC using polynomial $P(x) = x^3 + x + 1$ and encode the bit stream given in question 7.a.

   Answer: CRC = 110

   7.c. (6 points) Explain how the receiver can estimate if only one or more bits are in error and what it does in each case.

   Answer: Re-compute to validate the message by adding the CRC value so that the remainder for the division is zero. To show that the bit in error can be detected, one or more bits will be in error if the division does not yield in a zero for the remainder.

8. Link Layer – Channel Partitioning and Random Access Protocols – (12 points)

   8.a. (6 points) In a CSMA/CD network with a data rate of 10 Mbps, the maximum distance between any station pair is found to be 2500 m for the correct operation of the collision detection process. What should be the maximum distance if we increase the date rate to 100 Mbps, 1 Gbps, and 10 Gbps?

   Answer: Let us find the relationship between the collision domain (maximum length of the network) and the data rate. We know that

   Tfr = (frame size) / (data rate) = 2 × Tp = 2 × distance / (propagation speed)

   or

   distance = [(frame size) (propagation speed)] / [2 × (data rate)]

   or

   distance = K / (data rate)

   This means that distance is inversely proportional to the data rate (K is a constant). When the data rate is increased, the distance or maximum length of network or collision domain is decreased proportionally. In Example 12.5, we mentioned that the maximum distance for a data rate of 10 Mbps is 2500 meters. We calculate the maximum distance based on the above proportionality relationship.

Data rate = 10 Mbps → maximum distance = 2500 m Data rate = 100 Mbps → maximum distance = 250 m Data rate = 1 Gbps → maximum distance = 25 m Data rate = 10 Gbps → maximum distance = 2.5 m

This means that when the data rate is very high, it is almost impossible to have a network using CSMA/CD.

8.b. (6 points) Consider five wireless stations, A, B, C, D, and E

- Station A can communicate with all other stations
- B can communicate with A, C, and E
- C can communicate with A, B, and D
- D can communicate with A, C, and E
- E can communicate with A, D, and B

When A is sending to B, what other communications are possible?
When B is sending to C, what other communications are possible?

Answer:

C

B A D

E

a) Nobody can talk. Everybody will hear A, so there will be interference.

b) E can talk to D. There will be no interference, because D can not hear B, and C can not hear E.

9. Link Layer – Bit Stuffing – (10 points)

Explain, using an example and the bit pattern 01111110 to represent the frame boundary in a bit stream being transmitted, how bit stuffing is used to preserve frame boundaries when transmitting binary data at the Data Link level of the protocol stack.

Answer: Suppose we choose the bit pattern 01111110 to represent the frame boundary in a bit stream that we are transmitting. To prevent the occurrence of this bit pattern in the payload and thus incorrect identication of a frame boundary

by the receiver we will modify the payload by inserting a 0 bit into the stream after every consecutive five 1 bits seen in the payload

Unstuffng at the receiving end will remove these zero bits and restore the format of the original frame payload.

As an example consider the following payload bit stream.

00111111110111111010001110011010

The transmitted stream would be

00111110111110101111010001110011010

preventing incorrect identification of a frame boundary in the highlighted section of the bit stream.

10. Wireless (8 points)

IEEE 802.11 WiFi uses three address fields in its frame format to forward packets from a host in a hotspot to the Internet and from the Internet to a host in a hotspot. Assume that a WiFi access point (AP) is connected to a router port via an Ethernet link. Describe the packet flow from WiFi host to Internet and from Internet to WiFi host. Carefully identify the contents of all WiFi and Ethernet addresses fields.

Answer: From host to AP to router port: Host to AP: WiFi frame has addr1 = AP MAC, addr2 = host MAC, addr3 = router MAC AP to router: Ethernet frame has DA = router MAC and SA = host MAC From router port to AP to host: Router to AP: Ethernet frame has DA = host MAC, SA = router MAC AP to host: WiFi frame has addr1 = host MAC, addr2 =AP MAC, addr3 = router MAC