CSEE 4119: Computer Networks	Thursday, March 13th, 2014

Midterm

P. Prakash, W. An, P. Nirantar, R. Yu, X. Zhu (TAs) A. Chaintreau (instructor)

Please indicate your uni: ____________________

- This exam should be answered in 75mn. It is graded on a total of 150pt.

General Review ($4 \times 12 = 48$ pt)

For each question, you have to circle all the correct answer(s) (there may be one, several or even zero) and cross all the incorrect answer(s). If you believe these answers are obvious you may not need to justify it, but for any question that may require justification, you can clarify your answer **on the answer sheet!** (Be brief, no more than a couple of sentences).

1. The 2013-2014 winter was the best in ten years to
   (a) Fly from La Guardia.
   (b) Snowboard in Morningside Park.
   (c) Invite your relatives in a tiny apartment.
   (d) Handpick apple to make Belgian cidre.

   Explanation: ____________________

2. The following are advantages of a packet switched architecture over a circuit switched architecture:
   (a) All bits are grouped into packets, so no information gets lost.
   (b) When users send information sporadically, a larger population can be served by the same link.
   (c) The transmission delay depends on the packet size but not on the number of routers that relay the information to the destination.
   (d) It permits to use proxy, which is important to scale content delivery on the web.

   Explanation: ____________________

3. The following are direct immediate advantages of organizing network into layers
   (a) A wireless links can divide its bandwidth equally among potential communication pairs.
   (b) It is guaranteed that the minimum overhead information is added to the actual useful data being sent.
   (c) The network edge does not need implements a physical and a link layer protocol.
   (d) When a new protocol becomes available, it guarantees all devices adopt it at the same time.

   Explanation: ____
   ____
   ____
   ____

4. In a peer-to-peer architecture,
   (a) Sockets are simplified, because in the absence of a server, no peer needs to bind to a specific port number.
   (b) Because it is important to handle peer churn, the application almost always uses a connectionless transport.
   (c) TCP is used because it guarantees that the receiving hosts reciprocate through its Tit-for-Tat mechanism.
   (d) When a DHT is implemented, we implement shortcuts to guarantee that a new host can join the system.

   Explanation: ____
   ____
   ____
   ____

Exercise 1: Cookies sync for advertising networks (30 pt)

On the web a content publisher (e.g., the New York Times) sells its advertising space not directly to advertisers but through a Service Side Platform (SSP) which aggregate many publishers. These ads slots are typically bought not by a company directly, or even an ad-network, but by a Demand Side Platform (DSP) which acts as an intermediary for multiple ad-networks. One of the dirty secret of the web is that each of these platforms maintains a cookie to track you whenever you interact with them (it allows them to "retarget" you after you have made an action), and these cookies are "synchronized", which means that SSP and DSP usually cooperate to figure out that two cookies they possess are actually capturing the same user.

Cookies synchronization roughly works as follows:

- During a "triggering action" (*i.e.*, you put a new digital camera in your cart in an online retailer), there is a small image (usually invisible to the eye) that is added by the content provider in the HTML of the page, and that image is hosted on the dsp1 domain. When making this call, the dsp1 domain is able to place a cookie 123 on this user. This cookie could useful later: imagine you do not actually buy the camera, the online retailer knows that you are interested and it would like to show you another ad to encourage you to finish your purchase.

- Later on, if you visit a website (*i.e.*, a blog about camera review), on which domain ssp2 is operating the ads. To retarget this user with an ad to this particular camera, dsp1 would like to recognize it using its cookie 123. Indeed, the blog website includes in its html a call to ssp2 to receive the ads: if this is the first time this user contact ssp2, the domain will set a new cookie 456, but it cannot request cookies from other domain, including dsp1.

One solution is to do a full **cookie synchronization**: Since ssp2 operates the ads, it can run a piece of javascript in the ads, that request the user to contact some of its customer DSPs (among which dsp1 will be present). The goal is that after cookie synchronization, ssp2 can advertise that it has an impression for user 456, and dsp1 will be able to know that this corresponds to user 123 in its domain, and hence it will bid and choose an ad for that user using previous actions it recorded.

1. $(\curvearrowright)$ Typically, the ssp2 will request in the javascript that the user requests the following url `www.dsp1.com/cookiesync?ssp2&456`.

   Explain how requesting the user to make this call allows the dsp1 to figure out that user with cookies 456 on ssp2 is the one with cookies 123 on dsp1?

Delayed synchronization Imagine that a new user is seen by ssp2 for the first time. There are 25 DSPs that could potentially be interested to serve that user (potentially using retargeting). Of course, their interests to bid higly on that user depends a lot on whether they can recognize and retarget it.

In practice, instead of doing a full cookie synchronization, ssp2 generally does a delayed synchronization: it runs a first auction before cookie synchronization, and then after the auction is run and at the same time as the ad is playing, it runs cookie synchronization.

2. $(\curvearrowright)$ Name some of the advantages and disadvantages of delayed synchronization in terms of performance and revenue? Describe what happens when a user visits ssp2 later on in time?

3. ($\curvearrowright$) You run ssp2 and a start-up approaches you claiming that it has a faster cookie synchronization technique. When inquiring about their product, they mention that they use UDP. Do you think that this claim is credible? What other performance should carefully check in this product?

Exercise 2: Pipelining and utilization (15 pt)

Consider a 100-Mbps link of 10-km in length with 5 nanoseconds per meter propagation delay, constant 4,000 bit data packets, and negligible processing and queueing delays at both the sender and receiver. Assume the sender always has data to send, that there are no losses or corruptions, and 1Mbps = 10^6bps.

1. ($\curvearrowright$) If pipelining is used with a window N packets, what is the smallest value of N that ensures at least a 80% utilization of the link.

Exercise 3: Difference of GO-BACK-N and SELECTIVE-REPEAT (20 pt)

In the diagram below, one row corresponds to a time slot during which one packet can be sent. Packet and ACKs propagates so that they are received in the following time slot in the other end. The RTT is hence 2 time slots. We assume that the timeout value is set to 5.5 (*i.e.*, after a packet is transmitted, unless an ACK is received the time out is triggered just before the 6th time slot). We assume Pipelining is used (window size 4, each packet size 1). Sequence numbers starts at 0. ACK indicates the sequence number of the packet they acknowledge. All transmissions are successful and without corruption except:

- The first packet with SSN=1 sent by the sender is lost.
- The first ACK sent by the receiver with SSN=3 is lost.

1. ($\rightarrowtail$) Assuming the Sender and Receiver implements GO-BACK-N, represent on the following diagram, for the fist 11 time slots:
   - The packet and ACK sent and received by each host during each time slot, with their SSN.
   - The set of packets currently in the window maintained by each host.

GO-BACK-N

Sender	Receiver
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9

2. ($\rightarrowtail$)Do the same diagram for SELECTIVE REPEAT

SELECTIVE-REPEAT

Sender	Receiver
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9	0 1 2 3 4 5 6 7 8 9

Exercise 4: Comparing protocols with loss on the reverse path (37pt)

Let us assume M packets are to be reliably transmitted using a reliable transfer protocol with a window size of M (i.e., all packets can be transmitted simultaneously). Assume that packets and ACKs can be lost, but not reordered. Also assume that the sender waits long enough so that it does not prematurely timeout and retransmit a packet (i.e., ACKs never arrive too late).

Last assume that each time the receiver gets a packet, it sends an ACK.

Consider the scenario where the M packets are sent, some of which are perhaps lost in transit. The receiver responds to arriving packets with ACKs, some of which are perhaps lost in transit. After waiting a sufficient time so that all ACKs that have not been lost arrive, the sender prepares a set of packets to retransmit. Let σ be this set of packets, and let $S = |\sigma|$.

1. ($\rightarrowtail$) For $M = 5$, give a sequence of data and ACK arrivals and losses, where S is larger when Go-back-N is used than when selective repeat is used. The same order of losses should occur in both scenarios.

2. ($\rightarrowtail$) For $M = 5$, give a sequence of data and ACK arrivals and losses where S is larger when selective repeat is used than when Go-back-N is used. The same order of losses should occur in both scenarios.

3. $(\curvearrowright)$ For general M, Suppose all data packets successfully reach the receiver, but each ACK is lost toward the sender with probability p. What is $E[S]$ for selective repeat?

   Full credits goes to closed form (no sums).

4. $(\curvearrowright)$ With the same assumptions as in question above, what is $E[S]$ for Go-Back-N.

   Full credit goes to closed form (no sums).

5. Which protocol is better for the type of environment when data is transmitted with little loss, but ACKs are sent over a very lossy channel? Explain your reasoning using either math from above or just general intuition (1 or 2 sentences).