Doctoral Written Exam in Networking, Fall 2009

November 9, 2009

Answer all parts of all questions. There are four multi-part questions, each of equal weight. Turn in your answers by Thursday, November 12, at the same time you picked up your exam, to Jennifer Streubel at MCS-140G.

Brevity is the soul of wit. Be brief and to the point. This is a depth exam. So, your answers should show research maturity and depth. Do not pad your answers and do not write for the sake of writing.

You are free to consult any notes, books and papers during the examination, but make sure to include your references. You must develop your own solutions, which might use existing ideas and techniques, as long as you cite them and clearly explain how these existing ideas/techniques fit within your own solution.

You are not allowed to consult with any person during the examination.

If you have any doubt as to the interpretation of a question, make a reasonable assumption and explain your interpretation in your answer. No explanations will be given during the exam.
1. For each of the following four scenarios, formulate the problem and solve it using an appropriate analysis technique. Provide an explanation of the validity of any assumptions you make in your model and of the appropriateness of your chosen analysis technique. Also discuss how can you implement your solution over existing Internet or wireless technologies, commenting on the complexity of an implementation and how might this complexity be reduced if you could start with a clean-slate architecture.

(a) Consider a company that is connected to the Internet via two providers, i.e., the company is multi-homed. Suppose that the providers charge per-byte and provide different transit delays. For example, the lower-priced provider may guarantee delays that are less than 100ms, while the higher-priced provider may provide a delay bound of 20ms. Suppose the company has two commonly used applications that have different delay requirements — one application is more tolerant to delay than the other. On average, the applications generate a certain amount of traffic that has to be carried by either or both of the two connections (links). How can the company allocate the two applications’ traffic on the two links such that applications’ benefit is maximized while payments to the providers are minimized?

(b) Consider two computers contending to transmit over a shared time-slotted wireless channel using an 802.11 (WiFi) like protocol. If both computers send data at the same time slot, they both fail, and both must try again. If only one computer sends data, it succeeds. If neither send data, airtime is wasted. Assume each time slot has a certain utility, where collision yields the lowest utility value, followed by wasted time slot, then success yielding the highest utility value. What should be the transmission strategy of each computer? Now consider that in reality, a wireless transmission from a computer succeeds even in the presence of other competing transmissions, if it is a stronger signal, due to the so-called “capture effect”. What should be the transmission strategy of each computer?

(c) Consider a company that is connected to the Internet via two providers, i.e., the company is multi-homed. Suppose that the network of each provider routes traffic over paths with different propagation delays — one provider’s paths are much longer than the other due to the use of satellite links. Moreover, the provider with the longer paths (call it “provider A”) charges less, so the company should be using provider A as long as the performance requirements of its applications are met. Suppose the company’s most common applications all use a TCP-like transmission control protocol to adapt their data sending rate. Knowing provider A’s delays, how can the company determine if using provider A would support well these applications?

(d) Consider that you want to make a decision whether to use Delta-t as a transport protocol for your loss-intolerant transactional applications. You know that your provider’s network is usually very lossy so even small acknowledgment packets can get lost with a non-negligible probability. Thinking of single-message transactions, how can you estimate the rate at which you can successfully establish such Delta-t single-message connections, or equivalently the rate of successfully delivering these single messages, reusing the same connection ID?

You should start by reading the paper carefully.

(a) Provide a 1-2 page public review of this paper, making sure to comment on the value (or lack of value) in the problem considered, the system developed, and in the experimental results. Examples of public reviews written by TPC members are available for the papers in SIGCOMM 2008. Since your review is not actually going to be made public, you should feel free to discuss limitations of the work and provide a more critical assessment than those reviews, if you see fit. A good public review draws in an outside perspective without directly rehashing ideas already provided in the paper.

(b) The paper very briefly mentions some reasons why the authors decided not to employ chunk-level redundancy elimination, and instead chose packet-level redundancy elimination. Elaborate the benefits and drawbacks of these two approaches in detail, especially as relates to neighboring layers in the network stack, applications, and the network architecture. Provide settings and situations where each approach is superior.

(c) Critique the design decisions made for the network-wide optimization problem established in Section 4.2 and the assignment of caching responsibilities chosen in Section 4.3. In particular, argue whether the authors’ design has weaknesses (and ideally, provide suggestions for improvement if so), or whether the design meets all of the desiderata articulated in the paper.
3. You have been hired as a summer intern in a local startup that decided to deploy its services in the cloud. The subsystem that you are asked to design is a content distribution system that will serve some large static content to a potentially large number of clients. Noting that the cloud provider will charge your company based on the volume of outbound traffic from your subsystem, you proposed to your boss to consider a peer-assisted design, in which a single content server leverages the lateral connections between multiple clients to minimize the volume of outbound traffic.

The particular design you proposed is for the server and clients to use a P2P protocol such as bittorrent, with the server acting as a seed and the clients acting as leechers. In order to preserve the bandwidth of the server, the bittorrent software used by the server will be hacked so as to seed the swarm at a constant (low) rate, relying on the leechers to feed one another.

A fair assumption to make in such a setting is that clients will join the bittorrent swarm to get the content, and as soon as they receive the last piece of content, they leave the swarm (i.e., clients act as seeders only while they are actively downloading content).

Your boss loved the idea and realized that it could potentially be the basis for a patent. But, before rushing in to file one, she asks you to do some due diligence to identify prior academic research work that you think should be disclosed to the patent’s office because it is related to your invention, and to briefly explain how your invention is different.

(3.1) Write a maximum of two pages (+ appropriate references) in which you contrast your proposed design to others. You need to be brief and not just rephrase abstracts of papers. Also, notice that related work should not be limited to approaches leveraging P2P swarming, as techniques used in other settings may be quite applicable.

While the idea is attractive, your boss felt that before rushing to an implementation, it may be prudent to analytically evaluate a simple model of the proposed system. Being at a loss, you consulted your PhD advisor, who proposed that it may be a good idea to study the tradeoff between the rate with which the server (seed) injects traffic into the swarm and the time it takes for a client (leecher) to download the content. In particular, he proposed that you build a birth-death Markov chain that represents the evolution of the number of replicas (available in the swarm) for a single piece of the torrent. He also suggested that you make the following approximations and/or assumptions (not all of which may be necessary) and also encouraged you to make others if necessary:

- There is only one swarm to worry about.
- The content (torrent) consists of \( M \) pieces, where \( M >> 1 \)
- The number of leechers \( N \) in the swarm is constant, i.e., every departure is immediately replaced by an arrival.
- The churn (departure followed by an immediate arrival) rate is \( \mu \) leecher/sec.
- The server (seed) injects pieces into the torrent at a fixed rate \( \alpha \) pieces/sec.
- Leechers have identical upload bandwidth of \( \lambda \) pieces/sec and unlimited download bandwidth.
- The seed and leechers use a uniform piece selection strategy.

(3.2) What would be the states and transition probabilities and rates of your Markov chain? Please provide enough explanation.

(3.3) Either analytically or numerically evaluate the probability that a piece would be missing from the torrent.

(3.4) Using the Markov model (or not) can you provide an upper bound on the time it takes a client to download the content? Does your answer check for "obvious" parameterization - for example, if \( \lambda = 0, \alpha = 0 \), or \( N \) approaches infinity?
To help you with the evaluation of the system, your boss assigned an undergraduate intern to the project, tasking him with designing a simple simulator of the system. A few weeks later, your boss shows you the results that the undergraduate intern produced (Figure 1) and asks you if it makes sense.

(3.5) Write a couple of paragraphs that summarize your opinion. If the results make sense, explain what they mean/imply. If not, explain why you are doubtful.

As modeled, the server (seed) is operating in an “open loop” fashion, in the sense that it injects pieces into the swarm at a low fixed rate independent of whether or not the torrent is “healthy” (i.e., the number of replicas from each piece in the torrent is above some threshold). Clearly, if the torrent is healthy, then the server (seed) would be wasting its outbound bandwidth. To be more efficient, you suggested that the bittorrent software used by the server be further hacked so as to seed the swarm at a much higher rate, but only when it is necessary to do so - for example when a piece is missing. In other words, the server’s bandwidth would be taxed only as a last resort, when the torrent is not healthy.

When you described your idea in a company-wide meeting, members of the network security group raised concerns about this new design. In particular, they implied that this new design is susceptible to two types of attacks by competitors: The first would deplete the company’s bank account by virtue of what the company would owe the cloud provider, and the second would make the clients very unhappy with the service they get, including the possibility of a DoS attack.

(3.6) What type of attacks are they alluding to? How hard would it be to mount each one of these attacks? How hard would it be to make your system immune to such attacks? Explain and provide enough details.
4. This question is based on the paper “Spatio-Temporal Compressive Sensing and Internet Traffic Matrices,” in SIGCOMM 2009. You should start by reading the paper carefully.

The paper is concerned with replacing (inferring) missing values in a time-evolving traffic matrix (TM). In this question, we’ll ignore the time dimension and concentrate on the TM for just one time period. We’ll consider a TM $X$ where $X_{ij}$ is the number of bytes flowing from $i$ to $j$ over some specified time period.

(a) First, comment on how hard it is to infer a missing value in a traffic matrix. Consider that a model that is often used for a traffic matrix is the ‘gravity’ model in which $X_{ij}$ is given by $X_{i\ast}X_{\ast j}/X_{\ast\ast}$ where $X_{i\ast}$ is the total traffic entering the network at $i$, $X_{\ast j}$ is the total traffic leaving the network at $j$, and $X_{\ast\ast}$ is the total traffic flowing through the network. (For more background, you can consult “An Independent Connection Model for Traffic Matrices,” IMC 2006 — but note that we are concerned here with the gravity model, not the more complicated models described in that paper).

i. Assume that you are given a TM that is well modeled using the gravity model, and that one data value is missing. Can you estimate the missing value? Explain exactly how you would do that.

ii. In the best case, what is the maximum number of missing data values you could infer if the TM is generated according to a gravity model?

iii. If the TM is generated according to a gravity model, what is the rank of the resulting matrix?

iv. Assume you are given a $n \times n$ TM and told that the full TM has rank $\ell$ with $\ell < n$. In the best case, what is the maximum number of missing data values you could infer? Explain how you would do that.

(b) Next, consider the following variant of the problem, which is different from the paper. Instead of considering a TM that represents traffic flowing inside a single network, let’s assume the TM represents traffic flowing between autonomous systems in the Internet. So entry $X_{ij}$ now represents the traffic originating in AS $i$ and terminating in AS $j$.

i. Assume your viewpoint is at a single AS, say AS$_0$. An observable element of $X$ is one whose flow passes through AS$_0$. Are there any rows or columns of $X$ that are fully observable? Explain why or why not.

ii. Different ASes have different locations in the Internet (sometimes called “tiers” or “backbone/regional/access”) corresponding to how “central” the AS is in the Internet’s structure. ASes also take on roles such as provider, customer, and peer. Comment on how the kinds of elements that are observable in $X$ vary for ASes having different locations and roles. Under what conditions might an AS have multiple fully observable columns or rows?

iii. Now let us assume AS$_0$ wants to infer missing values in the TM $X$. What practical challenges might it encounter?

iv. Could an outsider (not in AS$_0$) determine which elements of $X$ are visible to AS$_0$ at a given time? What sources of data would be useful? How could the outsider go about making this determination?

v. Could the methods described in the paper be helpful for this problem? Explain why or why not.

vi. Suppose AS$_0$ were able to infer large parts or all of $X$. Could AS$_0$ use that information for any interesting purpose? Describe in detail as many as you can of the distinct uses that AS$_0$ might have for that information.