SPRING 2006
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN-MADISON
PH. D. QUALIFYING EXAMINATION
Networking
Monday, January 30, 2006
3:00-7:00 PM

GENERAL INSTRUCTIONS:

1. Answer each question in a separate book.

2. Indicate on the cover of *each* book the area of the exam, your code number, and the question answered in that book. On one of your books list the numbers of all the questions answered. *Do not write your name on any answer book.*

3. Return all answer books in the folder provided. Additional answer books are available if needed.

SPECIFIC INSTRUCTIONS:

All six questions must be answered.

POLICY ON MISPRINTS AND AMBIGUITIES:

The Exam Committee tries to proofread the exam as carefully as possible. Nevertheless, the exam sometimes contains misprints and ambiguities. If you are convinced a problem has been stated incorrectly, mention this to the proctor. If necessary, the proctor can contact a representative of the area to resolve problems during the *first hour* of the exam. In any case, you should indicate your interpretation of the problem in your written answer. Your interpretation should be such that the problem is nontrivial.

Computer Sciences Department
Networking Qualifying Exam
Spring 2006
All six questions must be answered.

1. **Internet security**

   Botnets are collections of compromised systems under the control of a single entity. It is estimated that there are currently tens of thousands of botnets in the Internet, the largest of which are reported to have over 100,000 systems. As such, botnets represent one of the most significant threats in the Internet.

   i) Describe two possible ways in which botnets might be recognized and tracked by security analysts. Be sure to state all assumptions clearly.

   ii) Describe two possible approaches for defending networks from botnets.

2. **Internet address space**

   The exhaustion of the 32-bit IPv4 address space is the main driving force behind the adoption of IPv6 which uses 128-bit addresses. Network address translation (NAT) is used to alleviate the current shortage of IPv4 addresses. But address size and NAT also have other effects.

   i) For a worm that spreads using scanning, by picking a destination address uniformly at random from the address space, compare the following three cases: a network using IPv4, a network using IPv4 and some NATs, and an IPv6 network. Assume that the number of vulnerable computers and the network speed are the same in all three cases and compare the rate at which the worm spreads and the final number of computers infected. Give numeric results where feasible.

   ii) To avoid routing table growth, many organizations moving from one ISP to the other have to renumber their computers so that they use a prefix from the address space of the new ISP. Given the same three cases as above (i.e., IPv4, IPv4 with NATs, and IPv6), compare the amount of work required to renumber computers when such an organization moves from one ISP to the other.

3. **Caching and virtual machines**

   New proposals for mobile computing advocate using virtual machines: instead of the user carrying a laptop around, the user would use different physical computers in various places, and always download from servers in the network a virtual machine with all his personal files and personalized applications, and run it. Updates of hard disk blocks or evicted virtual memory pages would be sent back to the server, so that when the user logs in again from a different machine, he finds the virtual machine the way he left it. Since the size of these virtual machines can be gigabytes, one needs a fast network for this download to work.

   i) Assume that the typical network connection to the server is as fast as the hard disk. Explain why it still makes sense for the host computer to store (parts of) the virtual machine on the local disk.

   ii) Assume that the local hard disk is treated as a cache for disk blocks of the virtual machine, and the in-network server as the keeper of the authoritative copy. Discuss the differences between caching approaches for disk blocks of this personal virtual machine, and caching approaches for DNS records.

4. **TCP over wireless**

   TCP is one of the most widely used protocols on the Internet. While its performance is known to be fairly robust on wired networks, its original design had ignored potential interactions of TCP when operating on the wireless medium.

   i) Consider the following Internet path between a wired host A and a wireless host B, AP is a wireless access point. The A-AP part is a wired connection, while the AP-B part is wireless.

   A ---------- AP - - - - B
   If TCP (consider the Reno version) is used to transfer data between A and B, what implicit assumptions of the TCP will be violated, and hence lead to performance inefficiencies.

ii) Describe a potential mechanism to mitigate this inefficiency.

iii) Consider the following wireless topology with four nodes.

C ---- D ---- E ---- F
In this topology assume that each node is only in communication range of its immediate neighbor(s) in the linear arrangement and no other node. For example, C is in range of D alone; while D is in range of both C and E.

Assume that two TCP flows, one from C to D, and the other from E to F, were started simultaneously. Both flows were transferring files of the same size. Assume the underlying MAC protocol is based on the 802.11 standards and all the nodes were using the same wireless channel. In multiple experiments it was consistently observed that at the time the E->F flow had finished the transfer, the C->D flow had barely started, i.e., the C->D flow was completely starved. Explain how this could happen.

5. **Internet routing**

i) Consider an Autonomous System on the Internet which uses a shortest path routing algorithm. Assume that hop count is used as the routing metric. Any three nodes in this network will obey the triangle inequality. Explain why.

ii) Now consider two Autonomous Systems, each of which uses the shortest path routing algorithm to construct internal routes. Assume that the two Autonomous System has two transit links and they their inter-AS routing policy uses the 'hot potato' routing. How does such a routing policy lead to violation of the triangle inequality property between three nodes in this network (explain with a precise example).

Note: Three nodes, A, B, and C obey the triangle inequality if the distance between any pair of these nodes is less than or equal to the sum of the distances between the other two node pairs, e.g., d(A,B) + d(B,C) >= d(A,C), d(A,B) + d(A,C) >= d(B,C), and d(A,C) + d(B,C) >= d(A,B).

6. **Content distribution**

One of the major uses of the Internet is in distribution of content. Over the years, the nature of distributed content and mechanisms to achieve this goal has evolved. For example, web caching systems, peer-to-peer networks (e.g., Kazaa), and Content Distribution Networks or CDNs (e.g., Akamai), are all different mechanisms with fairly similar goals of distributing content.

i) Compare and contrast the main differences between these three approaches --- web caching systems, peer-to-peer networks, and CDNs.

ii) CDNs often use small TTL values for their DNS records. Why? What are the performance implications of such a choice on the DNS system?