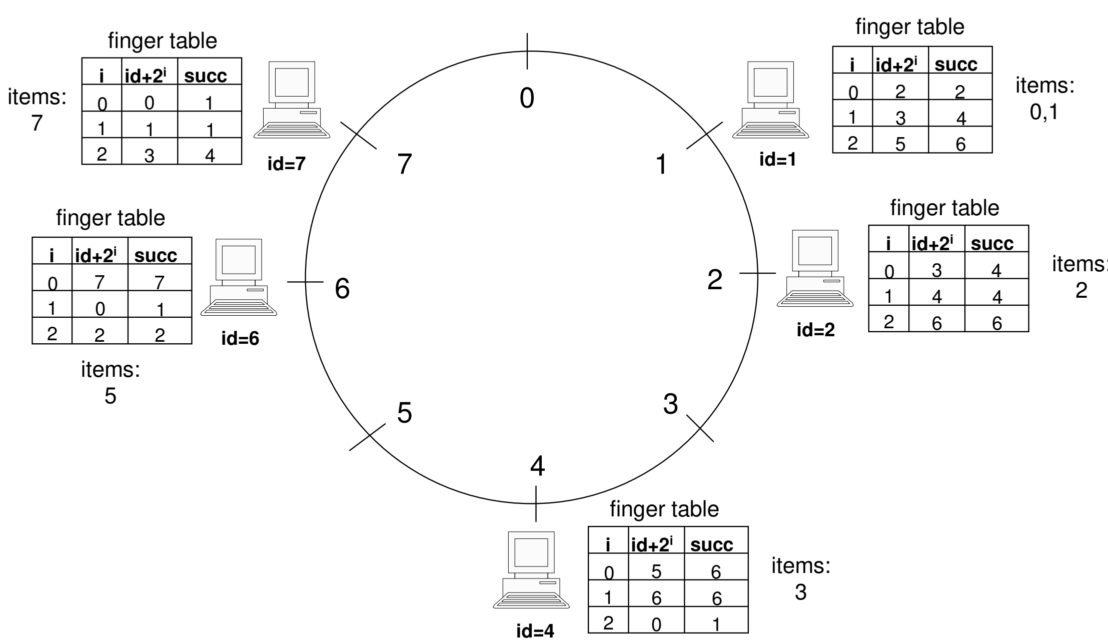
15-744 Computer Networks — Spring 2017
Homework 3
Due by 4/7/2017, 10:30am

(please submit through e-mail to zhuoc@cs.cmu.edu and srini@cs.cmu.edu)

Name:

A P2P and DHT

1. Srini, in fear that the RIAA will shut down his centralized P2P server (like Napster), sets up a Chord DHT for lookups and routing in his peer to peer network. Unfortunately (or fortunately, for you), Srini's P2P network is not very popular and only consists of five peers at the moment with finger tables and items illustrated below. For example, *node 4* has *item 3*.



Node 7 finger table (items: 7):

i	id+2^i	succ
0	0	1
1	1	1
2	3	4

Node 1 finger table (items: 0,1):

i	id+2^i	succ
0	2	2
1	3	4
2	5	6

Node 6 finger table (items: 5):

i	id+2^i	succ
0	7	7
1	0	1
2	2	2

Node 2 finger table (items: 2):

i	id+2^i	succ
0	3	4
1	4	4
2	6	6

Node 4 finger table (items: 3):

i	id+2^i	succ
0	5	6
1	6	6
2	0	1

(a) List the nodes that will receive a query from *node 1* for *item 7* (explain the path the query takes).

(b) List the nodes that will receive a query from *node 2* for *item 0* ((explain the path the query takes).

(c) Suppose node 4 crashes. *node 7* queries for *item 5*. List the nodes that will receive this query, assuming the the tables have had time to converge after noticing that node 4 has left.

2. Provide two reasons a company might prefer to pay Akamai to host their webpage instead of putting it onto a peer-to-peer network (such as Napster) for free.

3. Peer-to-peer file sharing systems need a way for users to locate other peers that have the content they want. List two advantages and disadvantages of using a centralized database to keep track of such information (e.g. Napster), as opposed to using a distributed hash table.

   Advantages of using centralized database:

   Disadvantages of using centralized database:

B CDN and CCN

4. Bovik is trying to figure out a scheme his clients should use, so that given a URL, they can find the appropriate CDN node to fetch the content from.

   Bovik has come up with a hash function h that takes a string and maps it to a real number in the range $[0, 1)$. Assume there are 3 CDN nodes with names such that $h(node_1) = 0.1, h(node_2) = 0.85, h(node_3) = 0.5$. When a client needs to fetch a URL and has to decide which replica to query, it picks $node_i$, such that the absolute value of the difference between $h(node_i)$ and $h(URL)$ is minimum. This scheme does *not* use circular mapping - it's just numeric closeness. This technique is "scheme 1"

   (a) Assuming all URLs are equi-popular, which node is likely to see the highest load?

   (b) Since load can be unevenly distributed in the above scheme, Bovik is not satisfied with the scheme. Instead, he thinks of a new arrangement. Let there be m CDN nodes in all; sort them using the $h(node_i)$ values. If the rank of a node is r, $(0 \leq r \leq m - 1)$, it is responsible for storing all URLs that map to the interval $[r/m, (r + 1)/m)$. This new scheme is called "scheme 2". Why might a CDN with a large number of nodes (that occasionally crash and are later repaired) choose scheme 1 over scheme 2?

5. Answer the following questions about CCN.

   (a) In CCN, how are the *interest* packets and *data* packets routed, respectively?

   (b) Compared to IP, why does CCN make it more difficult to target a *particular host* with a denial of service attack?

   (c) In class, we have learned how CCN could support VoIP applications. Will the same method work for an instant messaging application like GChat?

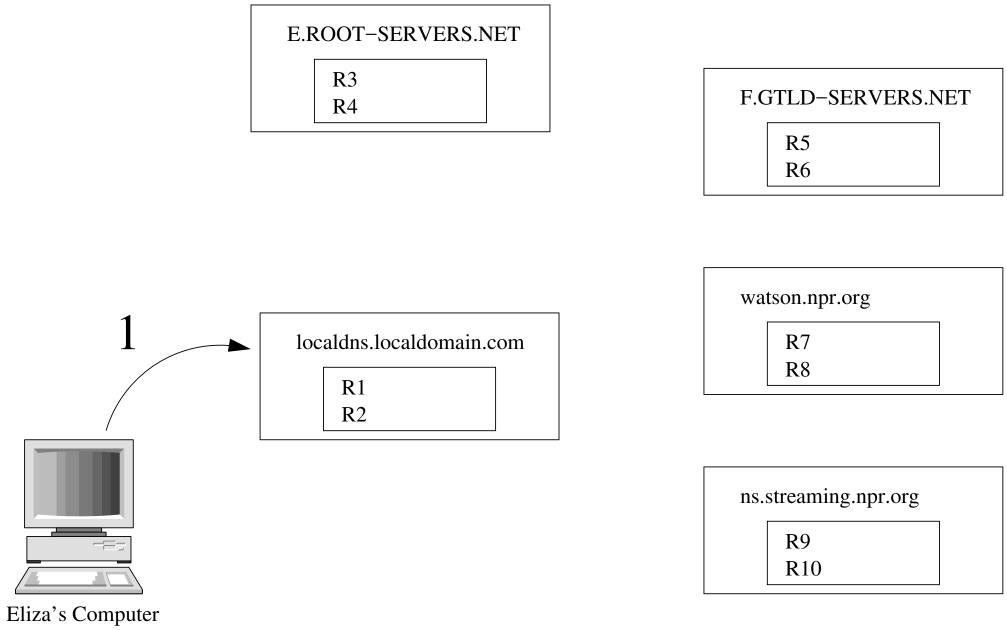
C DNS

6. Elisa wants to listen to the National Public Radio news over the Internet. She starts her favorite audio player and points it to ra1.streaming.npr.org. The audio player calls `gethostbyname()` with the given name to obtain the IP address of the server. As a result of the `gethostbyname()` call, the local resolver in Elisa's machine contacts the local DNS server to translate the host name into an IP address. The local DNS server performs an iterative lookup.

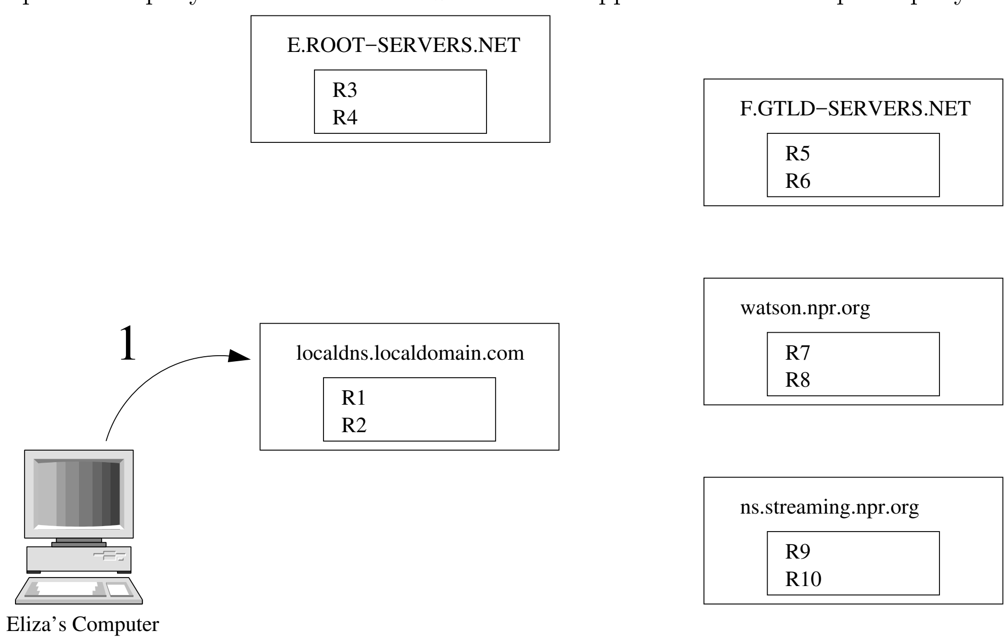
   The table below contains the DNS distributed database. A row corresponds to a DNS record. The records are grouped by DNS server.

Record #	Name	TTL (sec)	Type	Value
		`localdns.localdomain.com`		
R1	.	262542	NS	`E.ROOT-SERVERS.NET.`
R2	`E.ROOT-SERVERS.NET.`	348942	A	192.203.230.10
		`E.ROOT-SERVERS.NET`		
R3	org.	172800	NS	`F.GTLD-SERVERS.NET`
R4	`F.GTLD-SERVERS.NET`	172800	A	192.35.51.30
		`F.GTLD-SERVERS.NET`		
R5	`npr.org`	172800	NS	`watson.npr.org.`
R6	`watson.npr.org.`	172800	A	205.153.37.175
		`watson.npr.org`		
R7	`streaming.npr.org.`	172800	NS	`ns.streaming.npr.org.`
R8	`ns.streaming.npr.org`	172800	A	205.153.36.175
		`ns.streaming.npr.org`		
R9	`audio.streaming.npr.org.`	172800	CNAME	`ra1.streaming.npr.org.`
R10	`ra1.streaming.npr.org.`	10	A	205.153.36.175

(a) In the figure below, draw arrows to indicate the sequence of queries and responses exchanged among the different machines. Label each arrow with a sequence number and, if the message is a response, indicate which record(s) it contains. To make your sequence as simple as possible, assume the server includes both the A and NS records when applicable.



(b) Eliza repeats her query two minutes later. Show what happens for this subsequent query.



D Basic Tools

7. In this question you will use the unix utility `dig` to explore the contents of DNS messages.

   The format of a `dig` request is simple. Just type: *dig cs.cmu.edu* to perform a look-up for that DNS name. As you now know, DNS requests can do more than just ask for the IP address corresponding to a single DNS name. Type *dig cs.cmu.edu ANY* to see DNS records of all types that are associated with the domain 'cs.cmu.edu'.

   (a) What IP address did the computer you are logged into contact to make the DNS request? Where do you think this server is located?

   (b) List all of the different types of records received as a result of your query. For each record, explain its purpose, using one of the entries provided in the reply as a concrete example.

   (c) Use `dig` to find the TTL for the DNS mappings of 'www.cnn.com' and 'www.cs.stanford.edu'. What are they? If your boss asks you to provide two positive and two negative effects of having a short DNS TTL for the company's e-commerce site, what would you say?

   (d) Use `dig @ns1.google.com www.google.com` to find out the list of server IPs used for `www.google.com` if we delegate the DNS lookup to Google DNS `ns1.google.com`. Is the list the same as when you do `dig www.google.com`? If not, can you explain why?

   (e) DNS-based blackhole list (DNSBL) is an effort to keep a blacklist of locations in the Internet reputed to send email spam (see Wiki). `Dig` could also be used to check if an IP belongs to such a list. What is the command to check if your IP is in the list at `zen.spamhaus.org`?