CS 7260: Internet Architectures and Protocols January 22, 2007

Problem Set 1

Instructor: Prof. Nick Feamster *College of Computing, Georgia Tech*

This problem set has three questions, each with several parts. Answer them as clearly and concisely as possible. You may discuss ideas with others in the class, but your solutions and presentation must be your own. Do not look at anyone else's solutions or copy them from anywhere. (Please refer to the Georgia Tech honor code, posted on the course Web site).

Turn in your solutions in on **February 5, 2007** in class. Beware that, while several of these problems may bear resemblance to last year's problems, they have been modified/augmented/improved in various ways.

1 Naming

In the first part of this question, you will perform some hands-on DNS queries using `dig` and play with DNS lookups from various applications to understand more about the DNS. In the second part of this question, you will implement a variation on a stub DNS resolver.

RFC 1035 may be helpful for answering some of these questions.

Update: The first part of this problem only appears to work on Georgia Tech's Sun cluster machines.

1. In this question, we'll warm up by learning a few things about Georgia Tech's DNS setup.

   (a) Run "`traceroute ai`" from some machine on the Georgia Tech campus network. Now run "`traceroute ai.`" from the same machine. Include the output from each run in your problem set writeup. Why are the two traceroutes running traces to different machines?

   (b) What are the authoritative nameservers for `gatech.edu`? How long will your resolver cache the records pointing to these nameservers?

   What are the College of Computing's authoritative nameservers (*i.e.,* , for the domain `cc.gatech.edu`)? Give two benefits of topologically diverse authoritative nameservers. Why do NS records return names, rather than IP addresses?

   (c) What is another "canonical name" for the College of Computing's Web server?

   (d) What is the primary mail exchanger for `cc.gatech.edu`?

2. Now that you've had some experience playing with `dig`, in this part of the problem, we'll implement a stub resolver that performs iterative DNS queries. Most of the time, stub resolvers send queries with the "RD" (Recursion Desired) bit turned on. In this problem, you are *not* allowed to use the recursion bit.

   Of course, you are welcome to solve this problem any way you like. If you prefer, you may use the Ruby skeleton code that I have provided at `http://www.cc.gatech.edu/classes/AY2007/cs7260_spring/psets/ps1/aux/dns-resolv-rb.tgz`. This may save you the trouble of figuring out which modules to use, instrumenting your own performance measurements, etc.

(a) Why do stub resolvers typically set the RD bit?

(b) Implement a stub resolver that performs only iterative queries to resolve A records. To answer the next question, you'll want to make it possible to provide an option to your program to control the root nameserver.

Your resolver need not do anything special as far as caching, etc., but you should handle timeouts (*e.g.,* , quering the next preferred authoritative nameserver if the first does not respond).

Just make sure you can (1) point it at different root nameservers and (2) measure the time taken to resolve a query (the skeleton code is instrumented for this).

(c) Use your query to resolve (1) `www.cc.gatech.edu` and (2) `www.nytimes.com` at the following nameservers.

- a.root-servers.net (`198.41.0.4`)
- f.root-servers.net (`192.5.5.241`)
- m.root-servers.net (`202.12.27.33`)
- k.root-servers.net (`193.0.14.129`)
- a.gtld-servers.net (`192.5.6.30`)

(i) Through what sequence of nameservers was each query referred? How long did each referral step take? Based on this, what fraction of DNS query time is saved by caching at local resolvers?

(ii) What is the first referral when you send a query `www.cc.gatech.edu` to `a.gtld-servers.net`? Is the answer the same everytime? Why or why not?

(iii) How do stub resolvers typically choose root nameservers?

(d) Visit the site `http://www.traceroute.org/`, which maintains many "looking glass servers" from which you can run traceroutes.

(i) Find a looking glass server in London and traceroute to k.root-servers.net. Next, find a looking glass server in Sweden and perform a traceroute from to k.root-servers.net. What do you notice? Briefly explain the mechanism that explains this phenomenon.

(ii) Perform a traceroute from k.root-servers.net to Georgia Tech. Where does the traceroute go? What determines the ultimate destination for traffic destined to k.root-servers.net?

(iii) Describe two benefits that result from the phenomenon you observe.

Please hand in your code to this problem as well.

2 Understanding IS-IS Using Packet Traces

Obtain the IS-IS packet traces from the Abilene network for January 2, 2007. For example, the trace from the Atlanta router is located at `http://ndb2-blmt.abilene.ucaid.edu/isis/2007.01/ATLA/isisd.20070102.gz`. Seven of the 11 Abilene backbone routers capture such traces. You will need all seven IS-IS traces for this day to answer this question.

1. List all of the different types of IS-IS messages that appear in and eplain the purpose of each message.

2. What is the refresh interval for the link state announcements on the Abilene routers? What are the advantages of setting this value to a small value? What are the disadvantages?

3. Link state announcements are not only useful for routing, but also potentially for monitoring the status of the network. Using LSAs, one can, for example detect the existence of link or node failures on the network.

4. Compute (1) the expected propagation time of an LSA for each pair of routers in the network (since there are 7 routers capturing packet traces, your answer should be a 7 × 11 matrix; please order the rows and columns alphabetically) and (2) the distribution of propagation times for all LSAs (your answer should be a single CDF).

5. Based on the LSA propagation times that you computed in the previous question, compute the expected time that it would take for the network converge to a new state, given a failure of some link in the network. Based on your answer from the previous question, does the location of the link failure in the network have any bearing on the expected convergence time?

3 Understanding BGP using table dumps

For this question, you will need to download the Routeviews routing table from `http://www.cc.gatech.edu/classes/AY2007/cs7260_spring/psets/ps1/aux/oix-full-snapshot-2007-01-22-2200.dat.bz2` This file contains a Cisco BGP4 routing table snapshot, taken at Oregon Route Views (`http://www.routeviews.org/`) on January 22, 2007. (*Beware:* This is a text file that is 13MB, compressed. You should be able to analyze it without uncompressing it using, for example `bzcat`.)

If you are curious about what other snapshots look like, you can find daily snapshots at `http://archive.routeviews.org/`

1. Find the routing table entry for the Georgia Tech campus network.
   (a) What is the IP address of the best next hop from this router to Georgia Tech? How does this router know how to reach that next hop IP address?
   (b) From the routing table file, what is the AS number for Georgia Tech?
   (c) How many routes are there to get from this router to Georgia Tech?
   (d) What is the best route to Georgia Tech? Why was this route selected as the best route?[1]
   (e) How many ASes must a packet traverse between the time it leaves the router and the time that it arrives at Georgia Tech?
   (f) The next-hop IP addresses in this table are *not* routers in the local network. Explain (1) where the next-hop IP addresses are in this collection setup (2) how those IP addresses are reachable from the measurement box.
   (g) What are the AS numbers of all of Georgia Tech's upstream providers? What ISPs do each of these AS numbers correspond to? (*Hint:* You can discover this information using a whois query, similar to the one from L2.)

[1] If you're interested, see the L4 notes or for an overview of the BGP decision process. Note that the process is slightly vendor-specific.

(h) In paths where Georgia Tech uses Cogent (AS 174) as an upstream, the AS path ends with five instances of the same AS number. Why? What is the likely relationship between this AS number and Cogent?

(i) Consider AS 5650. This AS has chosen the "backup" path through Cogent to reach Georgia Tech. But, a closer inspection in the routing tables shows that AS 5650 actually connects directly to an provider through which it could get a much shorter AS path to Georgia Tech. Name one of these providers. List *two* reasons why AS 5650 may not be selecting a route through that provider.

(j) Why does the routing table have two paths from TeleGlobe (AS 6453) to Georgia Tech? Why are the two paths different?

(k) Look at all of the routes for which the AS path contains the sequence `11537 10490`. What do the ASes that appear first in those AS paths all have in common? Why wouldn't the ASes that select paths that don't have `11537 10490` in them not be selecting those paths?

(l) Use `traceroute` to measure route from some machine at Georgia Tech to the router that took the snapshot. Please include the output of your traceroute with your problem set.

Is the sequence of ASes from Georgia Tech to the router the same as the reverse route in the trace data? Why might the reverse path differ? (Please list reasons other than the fact that your traceroute was performed at a different time as the table snapshot!)

2. Look at the routing table entry for 12.106.30.0/24. This entry has several routes marked with a "d", for "damped". Give a short, one-to-two sentence explanation for (1) why routers damp routes and (2) why routers keep damped routes. To answer this question, you may want to look at RFC 2439.

3. Several of the IP prefixes in the table are formatted as w.x.y.z/m. The mask field, m, specifies the length of the network mask to use when matching input destination addresses to entries in the table.

(a) Write down the bit-wise operation to determine whether a destination address, A_i, matches a prefix A/m in the routing table. A_i and A are 32 bits each.

(b) Find the first "Class C" CIDR address in the table (address prefix $\geq$ 192.0.0.0). How many class C networks does this address correspond to? What is the maximum number of routing table entries that this single CIDR address saves? Why is it that we can only infer the maximum, and not the actual, number of addresses that this CIDR address saves?

(c) In the table, there are examples of groups of prefixes that have the same advertised AS path, but show up as separate entries in the routing table.[2]

(i) Provide an example of non-contiguous prefixes (and the corresponding AS path) for which this is true. Why might non-contiguous prefixes have the same AS path?

(ii) Provide an example of contiguous prefixes (and the corresponding AS path) for which this is true. This practice is often called *deaggregation*. Why might this be done?

[2] For both parts of this problem, it's sufficient to find the existence of one AS path that is advertised more than once. It is *not* necessary to find two prefixes for which *all* advertised paths are the same.

4. RouteViews makes available table snapshots from 1997 to present. Suppose you had access to all of these snapshots, as well as some routing table snapshots from pre-CIDR. For each of the following pieces of information available in the table snapshot, what information might you be able to infer about the evolution of the Internet?

   (a) Only the destination addresses.

   (b) Only the lines marked `*>`.

   (c) Only the paths, with best next-hops marked.

4 Network Operator for a Day with *rcc*

This problem deals with BGP and IGP routing configuration. In this example, you will examine a set of network wide routing configurations, run *rcc* on this set of router configurations, and write some configuration to fix the errors uncovered by *rcc*.

To work on this problem, you will need the following three resources:

- A machine that runs mysql server and client.
- The mysql tables that represent the Abilene "intermediate configuration representation": `http://www.cc.gatech.edu/classes/AY2007/cs7260_spring/psets/ps1/aux/abilene-rcc.tgz`. This set of tables is the output of running rcc over the configurations. (We have saved you the trouble.)
- Abilene router configuration files: `http://www.cc.gatech.edu/classes/AY2007/cs7260_spring/psets/ps1/aux/abilene-pset.tgz`
- The *first* routing table dump (or dumps, if you need them) on January 22 from the Abilene backbone network. BGP table dumps are available at: `http://ndb2-blmt.abilene.ucaid.edu/bgp/`.

1. To perform this part of the problem, you will need to install the mysql tables from the above URL and execute some SQL queries at the command line. Where appropriate, please include the SQL queries you used to find your answer, as well as the answer itself.

   (a) Other than the sessions to private AS numbers, what are the ASes with the most number of eBGP sessions. (*Hint:* the columns you will need for this query are `asn` and `ebgp`.)

   (b) At what routers does Microsoft have eBGP sessions to Abilene? (*Hint:* You will first have to figure out Microsoft's AS number(s)!)

   (c) Note that Microsoft is corporate, but Internet2 is supposedly a research and education network; why might Microsoft have eBGP sessions to Abilene?

   (d) What prefixes that are advertised by Microsoft are reachable from Abilene? (Your answer to this question help you validate your hypothesis to the previous question.) Which routing table did you look at to answer this question (and does it matter)?

2. Observe the output of running rcc's verifier at: `http://www.cc.gatech.edu/classes/AY2007/cs7260_spring/psets/ps1/aux/rcc-html/`

(a) Click on “IS-IS Errors” and then on “MTU Mismatch Checks”. What is an MTU mismatch, and why could it cause a problem? The pair of interfaces in question start with `ge-*`, which typically stands for “gigabit ethernet”. Which value is likely the correct value for the MTU?

(b) Under “BGP Errors”, click on “Information Flow”. These warnings indicate places where an import or export policy was configured in different ways on different routers for the same neighboring AS.

What is a reasonable explanation for why “anomalous import” (*i.e.,* , different import policies on different neighboring routers) might be a reasonable thing for an operator to do?

(c) Under “BGP Errors”, click on “iBGP Signaling”. What is meant by an “iBGP Signaling Partition”, and why is it bad?

This configuration error is caused because there is a line (or three, depending on how you count) of configuration missing from one of the router configurations. What lines need to be added to which configuration file?