Computer Science E-1
Spring 2010
Scribe Notes

Lecture 3: February 22, 2010
Andrew Sellergren

Contents

Computer Science E-1
Spring 2010
Scribe Notes

Lecture 3: February 22, 2010
Andrew Sellergren

1 The Internet (0:00–100:00)

1.1 DNS (0:00–20:00)

- We'll begin our discussion of the internet with a technical demonstration. Although you may be more familiar with Mac OS or Windows, there exist other operating systems which have interfaces that aren't quite as fancy. Some Linux operating systems, for example, have only a command line—a black screen with a single blinking prompt—with which you can interact. At this command line, if you type `traceroute cnn.com`, you'll be shown several lines of text, each of which represents a "hop" along the path from your computer to CNN's servers. In the first column on each of these lines is an IP address. On the first line of text, this IP address corresponds to the computer you ran the command from. Each computer that is hooked up to the internet is uniquely identified by an IP address of the form x.y.z.w, where x, y, z, and w are numbers between 0 and 255.[1] Each of the hops represented by these lines of text is a router that knows where to send data next based on its final destination.

- Examining these hops a little more closely, we can see that the first few routers belong to Harvard. Next we hop to routers that belong to Qwest, a very large Internet Service Provider (ISP), one in Boston (the `bst-` prefix) and one in Newark (the `ewr-` prefix). From there we seem to head to Washington and eventually Atlanta, where we might surmise that CNN's servers (or at least some of them) are located.

- The other columns in these lines of text are time measurements in milliseconds. As you can see, it only takes a matter of milliseconds for a request to go from a computer here in Cambridge, MA all the way to Atlanta, GA. Pretty phenomenal.

- If we change it up a bit and run `traceroute cnn.co.jp` to find the path between us and the Japanese version of CNN, we see that the first few hops are the same but the rest are much different. At some point we hit a server at `nox.org`, which belongs to the Northern Crossroads (NoX) which consolidates a lot of internet traffic in New England. Notice that between lines 9 and 10 or thereabouts, we see a large jump in the time values in the righthand columns. This makes sense since the servers we're hitting are now located on the other side of the Pacific Ocean. And, in fact, it's very impressive that data can travel across the Pacific Ocean in a matter of 100 milliseconds or so!

- A computer's IP address is much like a house's mailing address in that it encapsulates how it can be located so that information may be sent to it. However, we don't access websites by typing in numeric IP addresses.

[1] This is a small white lie. If multiple computers are hooked up to the same router, they will all appear to have the same external IP address. The router is reponsible for funneling traffic to each individual computer using that computer's internal IP address.

Instead, we type URLs into our browser. Behind the scenes, these URLs are converted to IP addresses via the Domain Name System (DNS). To find out what IP addresses a given URL will be converted to, we can use the `host` command on Linux or any number of web-based lookups. If we type `host cnn.com`, we'll get multiple IP addresses in response. Being very popular, CNN has multiple IP addresses so it can handle a large amount of web traffic. You can think of DNS as a phonebook. If you want to send something to a company, you won't necessarily know its mailing address off the top of your head. But if you know its name, you can look up its mailing address in the phonebook. The company's name is like a URL whereas the company's mailing address is like the IP address.

- If you open up the network settings on your own computer, you can view your current IP address and the DNS servers it will use to look up the IP addresses of websites you visit. In years past, you might have had to enter these values manually. Your ISP most likely has its own DNS servers which respond to most of your requests although there are larger servers higher up the chain which can be contacted if needed.

- Question: DNS is separate from the actual message which is sent between servers. DNS is a method for finding where to send the message. The content of the message is an entirely different story.

1.2 Modems and Routers (20:00–80:00)

- To make this discussion of the internet a little more concrete, we'll take a look at modems and routers. Whether you have DSL or cable internet, you have a modem that's connected to your computer (with an ethernet jack or perhaps wirelessly). This modem is plugged into a wall jack, either coaxial for cable or phone for DSL.

- Although both your computer and CNN's servers are connected to this vague entity called the internet, your computer doesn't know a priori how to send and receive messages from CNN's servers because it can't locate them by itself. With some configuration, however, your computer knows who to ask to find CNN's servers: the DNS servers. As we mentioned previously, this configuration used to take the form of users manually entering the IP addresses of the ISP's DNS servers. These days, however, dynamic host configuration protocol (DHCP) automates this configuration step. DHCP is also responsible for assigning a unique IP address to your computer so that the DNS servers and CNN's servers know where to send data.

- In this picture we're composing, the internet is more properly speaking the series of routers that bounce our message along. Generally, a message can be conveyed between two servers in 30 or fewer hops. The message contains not only the request but also a "return address," that is, the IP

address of the computer which sent the message so that the response can be properly directed.

- DHCP also provides an IP address for the primary router with which your computer will communicate. Every message that it sends and receives will pass through this hop first before jumping to other routers.

- Even today, you can configure your computer manually without using DHCP. You must be careful, however, that the IP address you choose hasn't already been assigned to another computer on the same network lest there be a conflict.

- Recall that the significance of the 0 to 255 range for each number in an IP address is that a single byte can represent these numbers. Therefore the whole IP address can be represented in 4 bytes. As a result, the number of unique IP addresses is actually around 4 billion. Given that there are 6 billion people in the world and there may be more than one computer per person (servers, for example, aren't associated with any person in particular), this maximum number of IP addresses may pose problems in the near future. IPv6, which will gradually replace IPv4, will solve this problem by expanding the number of unique IP addresses from 4 billion to 2^{128}.

- Another way of leveraging the limited number of IP addresses is to use one for more than one computer. Home routers make this possible by communicating with the internet via a single external IP address, but identifying each computer connected to it via different internal IP addresses. A useful analogy would be apartment numbers in an apartment building. The whole building has a single street address, but multiple apartments within it, each of which has its own internal number or address. Internal IP addresses are often of the form 10.1.x.y or 192.168.x.y.

- If you've ever made a typo while entering in a URL and been whisked away to an actual website, but not the one you were interested in, you've experienced firsthand the phenomenon of *squatting*. Squatters buy domain names that are very close in spelling to other popular domain names in the hopes that someone will accidentally navigate there or even buy the domain name from them. These sites are usually littered with advertisements to make money off visitors. ISPs are actually doing this on a much larger scale and have outraged the tech community for breaking the behavior of DNS.

- Routers are responsible for taking in data and sending it along to one of multiple destinations. They might also be called home routers or NAT (Network Address Translation) routers. If they offer wireless connectivity, then they are also *access points* (AP). These access points implement some variant of the 802.11 standard. You may have seen 802.11b or 802.11g or 802.11n on the box of a router you bought. These are the

newest amendments to the standard which allow for increased transmission speed. 802.11b specifies 11 megabits per second, 802.11g specifies 54 megabits per second, and 802.11n specifies 600 megabits per second. Cable modems generally have a maximum download speed of 10 to 12 megabits per second. Upload speeds tend to be more on the order of 2 megabits per second.

- Routers and modems may implement a *firewall* for security purposes. Generally, firewalls operate by allowing requests made by the user as well as the corresponding responses to go through, but blocking unsolicited connection attempts from outsiders. Firewalls may pose problems to technologies like Voice Over Internet Protocol (VOIP) which rely on one user being able to initiate contact with another via a phonecall. Fortunately, there are workarounds for this kind of issue. The Great Firewall of China is perhaps the largest example of a firewall since it affects an entire country with a population of over 1 billion. The Great Firewall is the Chinese government's initiative to filter ingoing and outgoing traffic to all end users.

- Firewalls operate on a relatively low technological level. They may block certain IP addresses or ranges of IP addresses, for example. Ranges of IP addresses are often associated with particular companies or geographic locations. MIT, for example, owns all the IP addresses of the form 18.x.y.z. Because all users in China have IP addresses within a well-defined range, the Chinese government can very effectively control the content which reaches them. Facebook and many other popular websites are blocked entirely. To get around this, you might think you could choose a wildly different IP address and enter it in manually. This generally won't work, however, as the particular router your connected to only accepts values within its predefined range. On the other hand, virtual private networks (VPNs) do successfully circumvent firewalls by providing the user with an IP address from a different network. If you connect to Harvard's network via VPN, your computer will adopt an IP address of the form 140.247.x.y and the rest of the internet will identify you as belonging to Harvard's network no matter where you are physically located. Similar in spirit, a *proxy* is a computer that acts as a middle man for all the data sent between you and the internet. A number of web-based proxies exist which are very easy to use.

- In the world of wireless, security takes the form of encrypting access points so that a specific key is required in order to connect to it. An early version of wireless security, WEP, has been replaced by WPA and WPA2 which are much harder to crack.

- Routers are connected to your computer via a Network Interface Controller (NIC) otherwise known as an ethernet card. This is a logic board which is responsible for relaying the data from the internet to your CPU where

it can be translated. NICs have an address of their own called a MAC address which identifies them to the device they're connected to. A quick troubleshooting note: many cable modems remember this MAC address, so if you connect a different computer directly to the cable modem, the internet connection might not immediately work. Usually power cycling (i.e. unplugging its power source and then reconnecting it) the cable modem will fix this problem.

- These days, routers offer very fast internal network speeds. That is, if two computers are connected to the same router, data can be transferred at upwards of 1 gigabit per second.
- Comcast and other ISPs might offer premium service with download speeds up to 50 megabits per second, but be wary that very few if any websites will actually support speeds like this. However, the increased upload speeds that these premium services offer might be worth the extra money if you spend a lot of time sending large amounts of data.
- Cellular devices offer much slower download speeds than wireless or wired connections. They leverage the existing cellular networks to send and receive data just as you would a phonecall.

1.3 Domain Registration (80:00–100:00)

- To go about getting your own website, you'll need to register a domain name of your own. Domain registrars like GoDaddy offer domain names for as low as $1.99 per year. Of course, the domain name you request must not be registered to anyone else and it must be one of the top-level domains (TLDs) available to the general public. TLDs are the suffix at the end of a URL, e.g. `.com`, `.org`, `.net`, of which a few, including `.gov` and `.edu`, are restricted. Some TLDs are actually country codes. For example, `.me` is the country code for Montenegro although it is used more colloquially for personal websites that have nothing to do with Montenegro. Likewise, `.tv` is technically affiliated with Tuvalo although it is used to reference the more familiar acronym TV for television.
- Once you've registered your domain name, you must tell the registrar which servers will actually host the content of your website. Only then will DNS servers know where to direct users who request your domain name.
- Question: do all DNS servers have the same information? No, in fact there is a hierarchy to these servers. If the low-level DNS servers do not have information on a domain name that's been requested, they will ask the higher-ups. Overall, this improves performance and minimizes response time. Another way these servers improve performance is by *caching* or remembering the information associated with a particular domain name. In this way, the low-level DNS servers won't have to continually ask the

higher-ups for information. Your computer itself also caches DNS responses.

- Question: what do you tell the domain registrar when it asks where you'll be hosting your site? Either you have your own server(s) that you maintain, in which case you provide *its* IP address, or you rent space on a shared server through HostMonster, Dreamhost, or even GoDaddy, in which case you provide the IP address of the shared server.
- Check out this Warriors of the Net video which will hopefully tie together a lot of what we've been talking about today.