9/4/2018

**CS 146**

Intro to the Web II: DNS & URL

What is Domain Name Resolution?

* In the early days of TCP/IP, users had to remember the IP address of every computer on the network
* In response, programmers developed a system of associating a user-friendly name to a computer that computers on the network could then associate with an address

Hostname System

* Simple name resolution technique developed early in the history of TCP/IP
* Each computer is assigned an alphanumeric name called a **hostname**
* If the operating system encounters an alphanumeric name where it is expecting an IP address, the operating system consults a **hosts file**

Moving On

* The hosts file system worked well (and still does) on small local networks
* This system becomes inefficient on larger networks
* The host-to-address associations have to reside in a single file, and the search efficiency of that file diminishes as the file expands
* In the ARPAnet days, a single master file (SRI international) called hosts.txt maintained a list of name-to-address associations, and local administrators had to continually update hosts to stay current
* Furthermore, the hosts name space was essentially flat

Introducing Domain Name System (DNS)

* DNS is the name resolution method used on the Internet and is the source of common Internet names such as [www.reddit.com](http://www.reddit.com)
* DNS divides the namespace into hierarchical entities called **domains**
* The **domain name** can be included with the hostname in what is called a **fully qualified domain name (FQDN)**
  + For instance, a computer with a hostname maybe in the domain whitehouse.gov would have the FQDN maybe.whitehouse.gov

Restriction on Hostnames

* Hierarchy goes right to left with a maximum of 127 levels
* Each level can have up to 63 characters
* Total name including delimiting dots cannot exceed 255 characters
* Even though any valid ASCII character could be used, it is usually limited to case-insensitive letters, numbers, and hyphens ‘-‘

Domain Name System – Types of Top Level Domains

* Generic top-level domain (gTLD)
  + Unrestricted (.com, .net, .org, and .info)
  + Sponsored. (.gov, .mil, .edu, etc.)
* Country code top-level domain
  + (.us, .ca, .uk, .au)
  + Internationalized Domain Names
* arpa

Authoritative Domain Name Server

* An authoritative domain name server is the one sending an authoritative response
  + It provides original and definitive answers to DNS queries
  + It does not provide just cached answers that were obtained from another name server
  + Therefore, it only returns answers to queries about domain names that are installed in its configuration system
* They are like the Yellow Pages
* Every domain registered is associated with a primary name server and at least one secondary NS

Recursive Domain Name Server

* When your browser sends out a DNS query – assuming the browser doesn’t already have the mapping stored in its cache – it is sent to a recursive DNS server
* Recursive servers are the part of the DNS that provides the required information to web clients
* They are usually managed by ISPs or the organizations that own the domain from which the connection is being made
* Authoritative responses have TTL (time-to-live), so the recursive servers must not send back outdated information to the client
* They are like the phone operators looking up a phone number from multiple phone books

**URL**

Uniform Resource Locators

* The domain identifies the server from which we are requesting resources
* Since the DNS system is case insensitive, this part of the URL is case insensitive
* Alternatively, an IP address can be used instead of the domain
* Port
  + The optional port attribute allows us the specify connections to ports other than the defaults
  + Add a colon after the domain, then specify an integer port number
* Path
  + Familiar concept to anyone who has ever used a computer file system
  + The root of a web server corresponds to a folder somewhere on that server
  + The path is optional. However, when requesting a folder or the top-level page, the web server will decide which file to send you
* Fragment
  + A way of requesting a portion of a page
  + Browsers will see the fragment in the URL, seek out the tag anchor in the HTML, and scroll the website to it