Introduction to Internet Technology and Web Programming

Computer Science 103
Boston University
Vahid Azadeh-Ranjbar

These notes are based on the lecture notes provided by Professor Susan Worst.

Lecture 2-1

BU CS 103, Spring 2019

Today's Materials

- Review from last lecture
- Domain names
- Internet security (Part 1)
  - ✓ Malware
  - ✓ **File system security (lab 0 assignment)**
- Introduction to HTML

Deadlines

- Lab 0 Assignment: Wednesday, January 30 at 6PM

Exams

- The questions are mostly short answer, fill-in-the-blank and HTML coding.
- Coding includes short coding (just one to two lines of code) and long coding (debug a given code)

Domain names

Shared Web Hosting @ http://cs-people.bu.edu

Every webpage, folder or file in web has a specific address which is called **URL**.

What Is a URL?

- URL stands for "**Uniform Resource Locator**"
- A unique identifier for a page, image, or other files on the web
- What you enter in the address bar of your web browser
- Example: http://www.bpl.org/general/hours/index.php

Deconstructing a URL

- http://www.bpl.org/general/hours/index.php
  - **http://** is the protocol (Stands for?)
  - **www** is a hostname or subdomain (Stands for?)
  - **bpl.org** is the domain name
  - **/general/hours/** is the file path
  - **Index.php** is the page (file)

Deconstructing a URL

- http://www.bpl.org/general/hours/index.php
  - **http://** is the protocol (HyperText Transfer Protocol)
  - **www** is a hostname (Stands for?)
  - **bpl.org** is the domain name
  - **/general/hours/** is the file path
  - **Index.php** is the page (file)

Deconstructing a URL

- http://www.bpl.org/general/hours/index.php
  - **http://** is the protocol (HyperText Transfer Protocol)
  - **www** is a hostname (World Wide Web)
  - **bpl.org** is the domain name
  - **/general/hours/** is the file path
  - **Index.php** is the page (file)

How about **http://cs-people.bu.edu/vranjbar/cs103/lecture2-1.pdf**?

- What is the domain name?
  - **bu.edu**
- So what is **cs-people**?
  - It is called a subdomain.
  - Any domain name can have lots of subdomain.

How Do Domain Names Work?

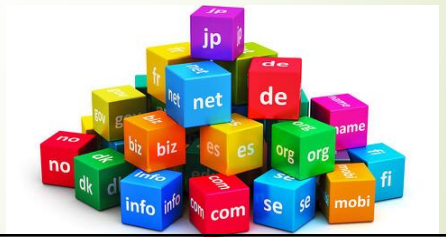
- Each computer on the Internet has an "IP (Internet Protocol) address" that distinguishes it from all other devices on the Internet
- Example: the IP address for the servers of BU is 128.197.236.4
- If you enter 128.197.236.4 into your browser's address bar, you will get to the BU home page http://www.bu.edu

From IP Address to Domain Name

- Because they consist only of numbers and periods, IP addresses are hard for humans to read and remember
- Domain names provide a friendly equivalent
- **Domain name server** translates domain names back into IP addresses

Top-Level Domain Names

- The right-most section of the domain name (the suffix or extension) is the top-level domain (TLD)
- For example in **cs-people.bu.edu**, "**.edu**" is the TLD.
- There are 2 kinds of TLDs:
  - ✓**Generic TLDs** (gTLDs) indicate the type of website
  - ✓**Country code TLDs** (ccTLDs) are particular to a geographic location



Generic Top-Level Domains

- There are 4 kinds of gTLDs:
  - ✓Unrestricted
  - ✓Restricted
  - ✓Sponsored
  - ✓New

Unrestricted gTLDs

- These gTLDs are all **unrestricted;** they can be used by anyone for any purpose.
- However, they are usually used for
  - ✓ **.com** = ?
  - ✓ **.org** = ?
  - ✓ **.net** = ?
  - ✓ **.info** = ?

Unrestricted gTLDs

- These gTLDs are all **unrestricted;** they can be used by anyone for any purpose.
- However, they are usually used for
  - ✓ **.com** = Company
  - ✓ **.org** = ?
  - ✓ **.net** = ?
  - ✓ **.info** = ?

Unrestricted gTLDs

- These gTLDs are all **unrestricted;** they can be used by anyone for any purpose.
- However, they are usually used for
  - ✓ **.com** = Company
  - ✓ **.org** = Organization, usually nonprofit
  - ✓ **.net** = ?
  - ✓ **.info** = ?

Unrestricted gTLDs

- These gTLDs are all **unrestricted;** they can be used by anyone for any purpose.
- However, they are usually used for
  - ✓ **.com** = Company
  - ✓ **.org** = Organization, usually nonprofit
  - ✓ **.net** = network
  - ✓ **.info** = ?

Unrestricted gTLDs

- These gTLDs are all **unrestricted;** they can be used by anyone for any purpose.
- However, they are usually used for
  - ✓ **.com** = Company
  - ✓ **.org** = Organization, usually nonprofit
  - ✓ **.net** = network
  - ✓ **.info** = Information Sites

Restricted gTLDs

- **Restricted** gTLDs are meant for a particular purpose. When registering one, you need to agree that you meet the criteria. You may be contacted upon to prove it later.
- Examples:
  - ✓ **.biz** = businesses
  - ✓ **.pro** – for licensed professionals (e.g. accountant, plumber)
  - ✓ **.name** – for individual people, or fictional characters to which the registrant has rights

Not Available to You and Me: Sponsored Generic TLDs

- **Sponsored** gTLDs are administered by organizations that tightly control their use. Only eligible entities are allowed to register them in the first place.
  - ✓ **.edu**: colleges, universities. (Where?)
  - ✓ **.gov**: U.S. federal government entities
  - ✓ **.mil**: (?)
  - ✓ **.museum**: Museums
  - ✓ **.travel**: travel and tourism sites
  - ✓ **.xxx**: "Adult" sites

Not Available to You and Me: Sponsored Generic TLDs

- **Sponsored** gTLDs are administered by organizations that tightly control their use. Only eligible entities are allowed to register them in the first place.
  - ✓ **.edu**: colleges, universities (within the US)
  - ✓ **.gov**: U.S. federal government entities
  - ✓ **.mil**: (?)
  - ✓ **.museum**: Museums
  - ✓ **.travel**: travel and tourism sites
  - ✓ **.xxx**: "Adult" sites

Not Available to You and Me: Sponsored Generic TLDs

- **Sponsored** gTLDs are administered by organizations that tightly control their use. Only eligible entities are allowed to register them in the first place.
  - ✓ **.edu**: colleges, universities (within the US)
  - ✓ **.gov**: U.S. federal government entities
  - ✓ **.mil**: (U.S. military)
  - ✓ **.museum**: Museums
  - ✓ **.travel**: travel and tourism sites
  - ✓ **.xxx**: "Adult" sites

"New" gTLDs

- In 2013, ICANN (**The Internet Corporation for Assigned Names and Numbers**) created a new process to dramatically expand the number of generic top-level domains
- More than 1300 new gTLDs available or in process since then
- Examples: .guru, .gallery, .bike, .rocks, .dell, .stockholm

Country Code TLDs (ccTLDs)

- ccTLDs are particular to a geographic location

.ca = Canada
.uk = United Kingdom
.mx = Mexico
.de = (?)

.us = United States
.cn = China
.in = India
.jp = Japan

Country Code TLDs (ccTLDs)

- ccTLDs are particular to a geographic location

.ca = Canada
.uk = United Kingdom
.mx = Mexico
.de = Germany

.us = United States
.cn = China
.in = India
.jp = Japan

Policies for Common ccTLDs

ccTLD	Who may register?
.us	United States citizens, residents, or organizations, or foreign entities with a bona fide presence in the United States. Source: usTLD Nexus Requirements Policy
.ca	Canadian citizens, residents, businesses, or organizations Source: Canadian Internet Registration Authority
.cn	Businesses and organizations only (not individuals). Business/organization does not need to be located in China. Source: Neustar
.de	Either the domain holder or the administrative contact must be domiciled in Germany. Source: DENIC Domain Guidelines
.in	Available to anyone worldwide. Source: Registry.in

Country Code TLDs (ccTLDs)

- Each country controls how its ccTLD is used. Some countries have chosen to allow people outside the country to use theirs such as:
  - ✓ .io (British Indian Ocean Territory)
  - ✓ .me (Montenegro)
  - ✓ .co (Colombia)

Internet Security

Internet Security

- Two reasons to talk about this:
  - ✓ When building a website, it's your responsibility to keep your visitors from being at risk
  - ✓ When using a website hosted by others, it's important to protect yourself

Taking Care of Your Website and Yourself

- Internet security has two parts:
  1. Protecting your computers and data from malware and intrusions
  2. Protecting your credentials from hackers and cybercriminals
- Internet security is also important to protect your files in your webhost (lab 0 assignment: **File system security**).

Malware (malicious software)

- Malware is "software designed to interfere with a computer's normal functioning".

  – Merriam Webster (merriam-webster.com)

How Does Malware Get Onto Our Computers?

- Different types of malware enter by different routes.
- Knowing these types of malware and their routes can help you act to protect yourself.
  - ✓ **Worms**: enter through vulnerabilities in software
  - ✓ **Viruses**: enter via infected files (such as email attachments)
  - ✓ **Trojan horses**: enter through files that we willingly download or open because we think they are useful

Malware via Software Vulnerabilities: **Worms**

- A worm travels from computer to computer on a network by exploiting vulnerabilities (bugs) in software.
- It does not require human action to spread.

What Makes You Vulnerable to Worms?

- Not automatically updating the operating system on your computer or phone
- Not applying updates for popular software like
  - ✓ **Microsoft Office**
  - ✓ **Firefox**
  - ✓ **Safari**
  - ✓ **Google Chrome**
  - ✓ **WordPress**
  - ✓ **Adobe Reader**
  - ✓ **Flash Player**

Protect Yourself from Worms

- On your own computer or phone
  - ✓ Keep ALL your software up to date
  - ✓ Whenever possible, turn on automatic updates
  - ✓ Get rid of software you do not need
  - ✓ Do not use apps or software that is no longer being updated

Malware from Infected Files: **Virus**

- A virus is:
  - ✓ a specific type of malware, not a term for all malware
  - ✓ "A parasitic application that can self-replicate."
  - ✓ "Requires a carrier that it can hide inside of."

—NakedSecurity.sophos.com

- Two great hiding places: **email attachments** and **Microsoft Office (Word or Excel) macros**

Protect Yourself Against Viruses

- Run anti-virus software – even if you have a Mac
- Keep anti-virus software up to date
- Do not open email attachments that you aren't expecting, or from people you don't know and trust
  - ✓ If you are not sure, call the person/company and ask before opening

It's Easier To Be Fooled Than You Think...

County Court		
County Court	Notice of appearance in Court #00000979470	3/24/2016
FedEx 2Day A.M.		
FedEx 2Day A.M.	Shipment delivery problem #00451395	7/16/2016
FedEx International Economy		
FedEx International Economy	We could not deliver your parcel, #0000917546	8/21/2016
FedEx International Ground		
FedEx International Ground	Problems with item delivery, n.0000376130	7/19/2016

Malware from Downloading: **Trojan Horses**

- A Trojan horse is malware "disguised as useful files or applications that entice you into executing them."

—NakedSecurity.sophos.com

Trojan Horse Example

- Fake anti-virus software has become a common Trojan horse technique. Affects both Windows and Mac systems
- A pop-up window tells you that your computer is infected – "Click here to scan or fix".
- Tricks you into purchasing something you do not need (stealing your money).
- May also install other types of malware.

—NakedSecurity.com

Another Trojan Horse Example

- Some web pages use JavaScript to download malware when you visit the page; no clicking required.
- This is called a "drive-by download."
- This is the reason to avoid pages with suspicious URLs.

Protect Yourself from Trojan Horses

- Read reviews from **trusted sources** (e.g. CNET, Mac World) before you download any software.
  - ✓ Trusted source means something that cannot be faked or manipulated
  - ✓ Are user reviews trusted sources?
- Never download software from a pop-up window or by clicking on a link. Always go to the vendor's website for software.
- Be suspicious of pop-ups, software offers, warning messages, and email attachments.
- Never visit a site marked as malicious in search engine results

Step 2: Protect Your Credentials

- We will talk about later during the semester.

File system security
(lab 0 assignment)

- Every file in World Wide Web has an **owner** also has a **group**.
- Any system user who isn't the owner and doesn't belong in the same group is determined to be **others**.
- There are two different format to show the access rights of owner, group members and others:
  1. Symbolic Notation (lab 0 assignment)
  2. Numeric Notation

Secure Shell (SSH) So far …

- ✓ We learned that to manage our webhost, we can use SSH commands such as:
  - ls Show directory contents (list the names of files).
  - cd Change Directory (move forward or backward between folders).
  - mkdir Create a new folder (short for make directory).
  - touch Create a new file (such as text files or html files).
  - rm Used to remove a chosen file/directory.
  - pwd Shows your current location in the file system.
  - cp Used to duplicate files and folders.
  - mv works the same way as **cp**, but instead, it deletes the original file)

Symbolic Notation
(lab 0 assignment)

How to find what is the access rights for each group?

Through terminal/Secure Shell App, instead of ***ls*** use ***ls -l***

For example:

ls -l test2.txt

drwxr-xr-x 2 vranjbar cs-faculty 4096 Sep 8 06:50 temp1

Here, you will see a 10 symbol string consisting of the symbols d, r, w, x and – i.e. ***drwxr-xr-x***

Symbolic Notation
(lab 0 assignment)

drwxr-xr-x 2 vranjbar cs-faculty 4096 Sep 8 06:50 temp1

1. The first character shows if the file is a directory or it is not.
2. The left group of 3 gives the file permissions for the user that owns the file (or directory) (vranjbar in the above example);
3. The middle group of 3 gives the permissions for the group of people to whom the file (or directory) belongs (cs-faculty in the above example);
4. The rightmost group of 3 gives the permissions for all other people in The World Wide Web.

❖ Read Lab 0 assignment for more information.

How change an access right?

To change an access right we need to use ***chmod*** command

Symbol	Meaning
u	user
g	group
o	other
a	all
r	read
w	write (and delete)
x	execute (and access directory)
+	add permission
-	take away permission

Syntax example:

chmod o-rwx test1.txt

chmod a+r test1.txt

- Read Lab 0 assignment for more information.

For Next Time …

- The notes for Lecture 2-1 can be found on Blackboard:

  Course Documents → 2- Lecture Notes → Lecture 2-1

- Read the assigned reading for Wednesday, found on Blackboard:

  Course Documents → 2- Lecture Notes → Lecture 2-2