Learning Goals

- [CT Building Block] Define spam, phishing schemes, and cookies and give examples of each
- [CT Building Block] Tell whether a communication technology (Internet, radio, LAN, etc.) is synchronous or asynchronous
- [CT Building Block] Explain the roles of Internet addresses, domain names, and DNS servers in networking
- [CT Building Application] Explain how data is transferred from one location to another across networks, such as the Internet
- [CT Application] Understand some of the design features of TCP/IP networks, such as packets, routing, domain names, and hierarchical structure
- [CT Application] Explain the importance of headers
- [CT Impact] Describe some of the impacts of using cookies
- [CT Application] Describe why bias may exist in Internet items and news stories
- [CT Application] Evaluate the credibility of items found on the Internet
What has the Internet done for me lately?

- The Internet is pretty complex. We’ll discuss its impacts and look at some of the basics that can help us understand it.
  - Spam
  - Phishing
  - Cookies
Phishing

- Phishing: “the activity of defrauding an online account holder of financial information by posing as a legitimate company.” [Google Search]
- There are several ways in which Phishing occurs. Knowing more about the Internet can help avoid it.
A good place to start is to understand something about URLs and Domain Names.

Addresses are key to performing networking tasks:

- e-mail addresses: identify *people*
- URLs: identify *web pages*
- domain names: identify *computers*

Both e-mail addresses and URLs have domain names.
Breaking URLs down

Here’s a URL:

http://www.phdcomics.com/comics.php

- Protocol
- Domain
- File, directory, and additional information

Most of the time the protocol is http (Hyper Text Transfer Protocol), but it can be other things. https means that it’s the secure hyper text transfer protocol.
Domain names form a hierarchy

- Each “.” separates a different level
- The farther toward the end, the higher the level
- Example: www.ugrad.cs.ubc.ca

  - The name of the computer is **www**
  - Which is part of the undergrad domain (**ugrad**)
    Which is part of the Computer Science Department domain (**cs**)
  - Which is part of the University of British Columbia domain (**UBC**)
  - Which is part of Canada (**ca**)
Domain Names Pictorially

- These names form a hierarchy
- example names: cs.ubc.ca, google.com, ugrad.cs.ubc.ca
Do you remember the discussion in class about Google and fake news?

[Google search results]

Computational Thinking
http://www.ugrad.cs.ubc.ca/~cs100
cnn.com.de vs. www.cnn.com

cnn.com.de is not the same as www.cnn.com!
Fun Fact

• Did you know that you can look up who owns a website?
• “whois” is a command that lets you look up who owns a website.
• Here is one website that will run whois: http://whois.domaintools.com
Breaking down the files

After the domain name, comes the file organization. It’s in a hierarchy as well.

http://imgs.xkcd.com/comics/phishing_license.png

Protocol  Domain  Directory  file
Sometimes there can be additional information in a URL


So far we’ve covered everything through: http://www.phdcomics.com/comics/archive.php

Everything after a “?” is data needed to process the request. In this case, it’s the ID of the comic that the user has requested.
“Later, we found a few telemetry packets on our server and thought, wow, that could be another image.’

During operations, images were split into telemetry packets …

For the very last image the transmission was interrupted after three [of 6] full packets were received… just over half of a complete image. This was not recognised as an image by the automatic processing software, but the engineers … could make sense of these data fragments to reconstruct the image.”
From: Apple  <Do_not_reply@apple.com>
To: rap@cs.ubc.ca
Subject: unlock it now before you lose it

We notice that you didn't use your iTunes account for along time, therefore we are obligate to close your account according to our policy including your i-cloud email also.

but it is not too late, you can save your account I get the access back to our features and services just click On "Save it for me" and follow the instruction, we may ask you an extra information as security procedure to insure that are the account holder.

if you don't need this account, just ignore this message and the account will be terminated in just few days, and we thank again for using our service.

Regards.

Save it for me  Look at links before you click! Actual URL: niti2.net
E-mail has headers, too
(search “email headers [e-mail client]”)

Return-Path: <kovka24@host.dm80.ru>
Received: from mail-relay1.cs.ubc.ca (mail-relay1.cs.ubc.ca [142.103.6.79])
    by smtp.cs.ubc.ca (8.15.2/8.13.6) with ESMTPS id u7K257bU029713
    (version=TLSv1.2 cipher=DHE-RSA-AES256-GCM-SHA384 bits=256
    verify=NO)
    for <rap@smtp.cs.ubc.ca>; Fri, 19 Aug 2016 19:05:07 -0700
Received: from host.dm80.ru (free.msk.ispsystem.net [82.146.44.135] (may
be forged))
    by mail-relay1.cs.ubc.ca (8.15.2/8.15.2) with ESMTP id u7K2569w003434
    for <rap@cs.ubc.ca>; Fri, 19 Aug 2016 19:05:06 -0700
Received: from kovka24 by host.dm80.ru with local (Exim 4.80.1 (FreeBSD))
    (envelope-from <kovka24@host.dm80.ru>)
    id 1bavej-000Fao-F5
    for rap@cs.ubc.ca; Sat, 20 Aug 2016 09:05:05 +0700
Date: Sat, 20 Aug 2016 06:05:05 +0400
To: rap@cs.ubc.ca
From: =?UTF-8?Q?Apple_=? <Do_not_reply@apple.com>…
The short version

• **Never** submit your password to a site whose link you followed from e-mail
• In general, don’t click on unexpected links
• Not all Spam is a Phishing scheme. They may be just trying to sell you things… but if you have doubts, don’t click.

Fascinating discussion on spam, including quotes from computer science researcher Stefan Savage: http://www.npr.org/blogs/money/2013/01/15/169424047/episode-430-black-market-pharmacies-and-the-spam-empire-behind-them
Sometimes you have to click, but think first

From: UBC Broadcast Email <message@broadcast.ubc.ca>
To: "rap@cs.ubc.ca" <rap@cs.ubc.ca>
Subject: IT Alert: Recent fraudulent emails and ransomware targeting UBC

Follow the link below to view it online::
http://message.broadcast.ubc.ca/email/View.aspx?id=810026&q=1045734672&qz=9a7b7f

In this issue:

IT Alert: Recent fraudulent emails and ransomware targeting UBC
Where in the World is `rap@cs.ubc.ca`

- `rap` is a user ID; the computer addressed by `cs.ubc.ca` will figure it out for us.
- `cs.ubc.ca` is an address... but it’s designed for humans.
- The transport layer switches from human-readable **domain names** to machine-friendly “**IP addresses**”.
IP Addresses

- Each computer connected to the Internet is given a unique address called its **IP address** (short for Internet Protocol address)
- This address is either temporarily or permanently assigned by an Internet Service Provider (ISP)
- An IP address is a series of numbers and letters separated by dots (old style addresses have fewer #s than new ones)
- There are \(~4.3\) billion old style IPV4 addresses
- There are \(~3.4\times10^{38}\) IPV6 addresses
Domain Name Servers: From names to numbers

- Domain Name Servers (DNSs) keep a directory connecting domain names to IP addresses.
- Every computer connected to the Internet needs the IP address of its “nearest” DNS.
- This DNS is used to resolve, or translate, a domain name to an IP address.
- DNS names need to be constantly updated.

You can play with this:
http://ip-lookup.net/domain-lookup.php
Servers vs. client

- A **server** is a machine that “serves” content. It’s where the HTML (or other website info) lives.
  - For example, the machine [www.cs.ubc.ca](http://www.cs.ubc.ca) lives in the ICICS/CS building. This machine is networked with a **file server** (i.e., a machine that serves up the files), which allows different machines on the cs.ubc.ca network to access the same files.
- Note that a server does **not** have to have “www” in its name – e.g., canvas.ubc.ca
Client/Server Structure

Most interactions over the Internet use the client/server interaction protocol:

- When you click a Web link, your computer gets the page for you...beginning the client/server interaction
- Your computer is the client computer and the computer with the Web page is the server (Web server)
- The client, gets services from the server

When the page is returned, the operation is completed and the client/server relationship ends
The client/server structure is fundamental to Internet interactions.

A key aspect is that only a single service request and response are involved.

The *relationship* is very brief relationship, lasting from the moment the request is sent to the moment the service is received.
Many Brief Relationships

This approach means that the server can handle many clients at a time.

For example, between two consecutive client requests from your browser (getting a page and asking for another) that server could have serviced hundreds of other clients.

The server is busy only for as long as it takes to perform your request.
The Internet is asynchronous

**Synchronous communication:**
Both the sender and the receiver are active at the same time (think of talking on a telephone)

**Asynchronous communication:**
The sending and receiving occur at different times (think of email and answering machines)
TCP/IP Tablets of Stone

- Sooner or later, your partner received most of the tablets, but not necessarily in sequential order
- Nor do they necessarily take the same route
- The tablets are finally arranged in order
- The “tablets” are really IP packets
  - They hold: one unit of information, the destination IP, and their sequence number (which packet they are)
Packets Are Independent

Because each packet can take a different route, congestion and service interruptions do not delay transmissions

Each TCP/IP packet is independent

The TCP/IP protocol works under adverse conditions

If traffic is heavy and the packet progress is slow, the protocol allows the packet to be thrown away
That’s a bit about how the Internet works. What has the Internet changed?

Or maybe, what hasn’t the Internet changed?
Of course, there are downsides

We’ll discuss a bit about privacy, security and other related topics
You can protect yourself through *encryption*

- Encrypting data is taking data and changing it so that it’s hard to understand
- Encryption has been around for a long time!
- People used to do encryption by hand
- Then they developed machines (proto-computers) to do it so that it was harder to break
- The Germans did this with the Enigma machine in WWII. This was in turn broken using more proto-computers lead by Alan Turing as discussed in “The Imitation Game”:
  
  [https://www.youtube.com/watch?v=S5CjKEFb-sM](https://www.youtube.com/watch?v=S5CjKEFb-sM)
Encryption has many flavours

• You can encrypt things simply to obscure. Remember http://myaccount.google.com-securitysettingpage.tk/security/signinoptions/password?e=am9obi5wb2Rlc3RhQGdtYWlsLmNvbQ... ?

• That’s created by treating the 0’s and 1’s a bit differently (more on this later)

• Decode am9obi5wb2Rlc3RhQGdtYWlsLmNvbQ at http://base64decode.org → john.podesta@gmail.com
Encryption has many flavours

- That might be enough to fool naïve users, but it won’t protect your credit card info.
- For that, we need something more robust

https://www.youtube.com/watch?v=w0QbnxKRD0w
Okay, so that’s a problem. People can snoop to get your data.

But there’s another concern – sometimes websites you access (on purpose) can keep data on you, too.
Facebook

• How many of you have ever used Facebook?
• How many of you have ever paid money to use Facebook?
• Like many other webs services, Facebook makes money through ads
Facebook collects data that you give it

- Location
- Demographics
  - Age
  - Gender
  - Relationship status
  - Education
  - Type of work
- Interests (hobbies, pages they like)
- Behaviours
  - Devices they’re using, other info

https://www.facebook.com/business/a/online-sales/ad-targeting-details
Where does Facebook get this information from anyway?

- Some things you tell it
- Other things it finds out from: “… activities that people do on or off Facebook that inform on which device they're using, purchase behaviors or intents, travel preferences and more. Behaviors are constructed from both someone's activity on Facebook and offline activity provided by data from Facebook's trusted third-party partners.”
- We’ll talk a bit more about how it uses information from those sites later when we talk about data mining
- For now, let’s talk about how it gets that data from the other sites in the first place

C is for Cookie

- Cookies store information about your web browsing
- Some cookies save information about you for that website, e.g., that you’ve authenticated yourself, Fahrenheit vs. Celsius, language option.
- Third party cookies save information about you that is not used for that website
- Third party cookies are largely why we see targeted ads
Let's look at one example. Here's the CBC's privacy page:

http://www.cbc.ca/aboutcbc/discover/privacy.html

... "The ads appearing on this Web site are delivered to you by DoubleClick, our Web advertising partner. Information about your visits to this site, such as number of times you have viewed an ad (but not your name, address, or other personal information), is used to serve ads to you. For more information about DoubleClick, cookies, and how to 'opt-out', please visit http://www.doubleclick.net/us/corporate/privacy"
Learning Goals Revisited

• [CT Building Block] Define spam, phishing schemes, and cookies and give examples of each
• [CT Building Block] Tell whether a communication technology (Internet, radio, LAN, etc.) is synchronous or asynchronous
• [CT Building Block] Explain the roles of Internet addresses, domain names, and DNS servers in networking
• [CT Building Application] Explain how data is transferred from one location to another across networks, such as the Internet
• [CT Application] Understand some of the design features of TCP/IP networks, such as packets, routing, domain names, and hierarchical structure
• [CT Application] Explain the importance of headers
• [CT Impact] Describe some of the impacts of using cookies
• [CT Application] Describe why bias may exist in Internet items and news stories
• [CT Application] Evaluate the credibility of items found on the Internet