

# Introduction to the Internet

Tak Auyeung, Ph.D.

October 6, 2005

- 20050905 2254 TA: add Firefox discussion 3.7
- 20050822 2205 TA: add material, starting at section 2.3
- 20050818 0004 TA: add activities for today's lab
- 20050816 2344 TA: work on the bulk of chapter 2, not done yet!
- 20050816 0134 TA: document creation

# Contents

0.1	Copyright Notice . . . . .	6
<b>1</b>	<b>Week 1: Operating Systems and Infrastructure</b>	<b>7</b>
1.1	SLO (Student Learning Outcome) . . . . .	7
1.2	Windows . . . . .	7
1.3	Internet Structure . . . . .	7
1.3.1	Basic Idea . . . . .	7
1.3.2	Basic Terms . . . . .	7
1.3.3	Simplistic View of the Internet . . . . .	8
1.3.4	Basic Internet Infrastructure Terms . . . . .	8
1.4	DNS . . . . .	9
1.4.1	What's in a Domain Name? . . . . .	9
1.4.2	DNS Queries . . . . .	9
1.4.3	DNS Caching . . . . .	9
1.4.4	I Want My Own Domain Name . . . . .	9
1.4.5	I Want to Manage My Domain Names . . . . .	9
1.5	The Next Step . . . . .	10
1.6	Lab Activities in 152B 8/16 . . . . .	10
1.6.1	Log in . . . . .	10
1.6.2	Start Internet Explorer . . . . .	10
1.6.3	Go to My Website . . . . .	10
1.6.4	Navigate to the Classnotes . . . . .	10
1.6.5	Navigate to this Page . . . . .	11
1.6.6	Move around . . . . .	11
1.7	Lab Activities in 152B 8/18 . . . . .	11
1.7.1	Sign up for a ZIP account . . . . .	11
1.7.2	Moodle Registration . . . . .	11
1.7.3	Forum Participation . . . . .	11
<b>2</b>	<b>Week 2: Internet Services and Connections</b>	<b>13</b>
2.1	SLO . . . . .	13
2.2	Services . . . . .	13
2.2.1	Terms . . . . .	13
2.2.2	Webpage distribution . . . . .	13
2.2.3	Newsgroups . . . . .	14
2.2.4	File Transfer . . . . .	14
2.2.5	Remote Command-Line Access . . . . .	14
2.2.6	Email . . . . .	14
2.2.7	Peer to Peer (P2P) . . . . .	15
2.2.8	Online Chatting/Instant Messaging . . . . .	15
2.2.9	SSH . . . . .	15
2.2.10	SSL . . . . .	15
2.3	Connecting to the Internet . . . . .	16
2.3.1	Basic Concept . . . . .	16
2.3.2	Dial-up . . . . .	16

2.3.3	DSL . . . . .	16
2.3.4	Cable Modem . . . . .	17
2.3.5	Fiber Optics . . . . .	17
2.3.6	Selecting an ISP . . . . .	17
<b>3</b>	<b>Week 3: WWW and Browsing</b>	<b>19</b>
3.1	SLO . . . . .	19
3.2	Warning . . . . .	19
3.3	Terms . . . . .	20
3.4	Basic Browser Operation . . . . .	20
3.5	Advanced Browser Operations . . . . .	20
3.5.1	Bookmark/Favorite . . . . .	20
3.5.2	Bookmark Organization . . . . .	20
3.5.3	Searching in a Webpage . . . . .	21
3.6	Search Engines . . . . .	21
3.6.1	Simple Search Query . . . . .	21
3.6.2	Advanced Search Query . . . . .	21
3.7	Alternative Browser . . . . .	21
3.7.1	Firefox . . . . .	21
<b>4</b>	<b>Week 4 and 5: Email</b>	<b>23</b>
4.1	SLO . . . . .	23
4.2	Email Definition and Protocols . . . . .	23
4.2.1	No Direct Delivery . . . . .	23
4.2.2	One Middleman . . . . .	23
4.2.3	Two Middlemen . . . . .	24
4.2.4	Protocols . . . . .	24
4.3	Email Clients . . . . .	24
4.3.1	Account Name . . . . .	24
4.3.2	Email Address . . . . .	24
4.3.3	Username . . . . .	24
4.3.4	Server Type . . . . .	24
4.3.5	Server (Incoming) . . . . .	24
4.3.6	SMTP Server (Outgoing) . . . . .	24
4.3.7	Options . . . . .	25
4.4	WebMail . . . . .	25
4.4.1	Continuous Internet Connection . . . . .	25
4.4.2	Editing Options . . . . .	25
4.4.3	Sending Options . . . . .	25
4.5	Address Book . . . . .	26
4.6	Security . . . . .	26
4.6.1	Privacy . . . . .	26
4.6.2	The Importance of Privacy . . . . .	26
4.6.3	Authenticity . . . . .	27
4.6.4	PGP and GPG, Key-pair System . . . . .	27
4.7	Key-pair Encryption . . . . .	27
<b>5</b>	<b>Week 6: Webpage Authoring</b>	<b>29</b>
5.1	SLO . . . . .	29
5.2	What is a Web Site? . . . . .	29
5.2.1	Free Web Space . . . . .	29
5.2.2	Dedicated Paid Web Space . . . . .	29
5.2.3	Custom Domain Name . . . . .	29
5.2.4	Evaluate Web Hosting Companies . . . . .	30
5.3	Tools to Create Webpages . . . . .	30
5.3.1	Word Processor . . . . .	30

5.3.2	Text Editor . . . . .	30
5.3.3	Specialized HTML Editors . . . . .	30
5.4	Creating Web Pages . . . . .	31
5.5	Uploading . . . . .	31
5.5.1	Web-based Interface . . . . .	31
<b>6</b>	<b>Authoring Web Pages</b>	<b>33</b>
6.1	Getting Nvu . . . . .	33
6.2	Some Concepts . . . . .	33
6.3	Using Nvu . . . . .	33
6.3.1	Starting up . . . . .	33
6.3.2	Simple editing and formatting . . . . .	34
6.3.3	Hyperlinks . . . . .	35
6.3.4	Including images . . . . .	36
6.4	Site Management . . . . .	37
6.4.1	Getting started . . . . .	37
6.4.2	Image files . . . . .	38
6.4.3	Uploading files . . . . .	38
<b>7</b>	<b>Week 8: FTP, RSS and Forums</b>	<b>39</b>
7.1	FTP . . . . .	39
7.1.1	Prerequisites . . . . .	39
7.1.2	GUI FTP clients . . . . .	40
7.1.3	But, don't use FTP! . . . . .	40
7.2	Forums . . . . .	40
7.3	RSS . . . . .	41
<b>8</b>	<b>Week 8, Part II</b>	<b>43</b>
8.1	At Work . . . . .	43
8.1.1	Email . . . . .	43
8.1.2	Chatting . . . . .	43
8.1.3	Advertisement/Sales . . . . .	43
8.2	At Home . . . . .	43
8.2.1	Email . . . . .	43
8.2.2	Chatting . . . . .	44
8.2.3	Multimedia . . . . .	44
8.3	Malware . . . . .	44
8.4	DRM . . . . .	44
8.5	Trusted Computing . . . . .	44
8.6	Web Applications and Thin Clients . . . . .	45
	<b>Index</b>	<b>47</b>

## 0.1 Copyright Notice

All materials in this document are copyrighted. The author reserves all rights. Infringements will be prosecuted at the maximum extent allowed by law.

You are permitted to do the following:

1. add a link to the source of this document at [www.mobots.com](http://www.mobots.com)
2. view the materials online at [www.mobots.com](http://www.mobots.com).
3. make copies (electronic or paper) for *personal* use only, given that:
  - (a) copies are not distributed by *any* means, you can always refer someone else to the source
  - (b) copyright notice and author information be preserved, you cannot cut and paste portions of this document without also copying the copyright notice

# Chapter 1

## Week 1: Operating Systems and Infrastructure

### 1.1 SLO (Student Learning Outcome)

- Get (re)familiarized with the windows environment
  - Start and close applications
  - Understand and navigate the file system
  - Check whether a computer is connected
- Basic understanding of the structure of the Internet
- Understand the purpose of DNS

### 1.2 Windows

You are not required to use Windows in this class. You can use Mac OS (on a Machintosh), Linux (on any machine), FreeBSD (on any machine) or any operating system that allows you to connect to the internet. However, let's face it, Windows is what most people use.

This is not CISC 300 (computer familiarization) or CISC 320 (operating system). We only need to review enough operating system concepts to get you started with Internet related applications. If you have trouble keeping up with the class, please come see me during my office hours, or you may want to consider taking CISC 300. If you feel that you want to know more about Windows, please consider taking CISC 320.

Instead of focusing on Windows techniques here, I'll distribute discussions of Windows operations throughout the notes.

### 1.3 Internet Structure

#### 1.3.1 Basic Idea

The term "Internet" means it is a network of networks. In other words, the Internet connects smaller networks together. This is true! When you log in to an ISP, you are logging into a small network. Useful network resources, such as Google (the search engine), lives in its own small network. These two small networks are connected by the Internet. The Internet allows you computer connect to the Google machines.

#### 1.3.2 Basic Terms

In this section, we touch on some terms that are useful in the description of the Internet. We'll connect all the terms in the next section.

- **IP** : Internet Protocol. For now, look at it as a language that allows network connected machines (not necessarily computers) exchange information.
- **Network interface**: A network interface is a physical socket where message can be sent or received. A networked computer typically has one network interface, but it can also have multiple network interfaces.
- **IP address** : Each network interface acquires an IP address somehow. This allows a network interface be uniquely identified among other network interfaces.

### 1.3.3 Simplistic View of the Internet

In the most simplistic form, the Internet is a whole bunch of interconnections that allow a network interface communicate with another network interface, as long as one knows the IP address of the other. This simplistic view is sufficient for most people who want to use the Internet for purposes of communication.

In other words, as long as you know where point B is, you can travel from point A to point B using the Internet. The infrastructure of the Internet consists hubs, switches, routers and firewalls, with cables of all kind interconnecting these devices. For most of us, the infrastructure is not very important. Nonetheless, it is still a good idea to understand a part of it.

### 1.3.4 Basic Internet Infrastructure Terms

- **NIC**: Network Interface Card.
- **OSI**: Open System Interconnection. This is a standard/model for networking. The OSI model slices network protocols into 7 layers, from physical (electrical voltages and other physical properties) to application (logical/abstract).
- **MAC Address**: Media Access Control Address. Each interface should have a unique MAC address to uniquely identify itself. In reality, MAC addresses can be cloned. That is, one NIC can assume the MAC address of another NIC. MAC addresses are used mostly in OSI layer 2 protocols.
- **TCP/IP**: Transfer and Control Protocol/Internet Protocol. TCP implies IP, but not the other way round. In order to connect to the Internet, a computer must understand how to speak the language: TCP/IP. IP specifies layer 3 in the OSI model, while TCP builds on top of IP at layer 4 in the OSI model.
- **LAN**: Local Area Network. This is a bunch of computers linked by a network that is relatively local (close to each other).
- **WAN**: Wide Area Network. This is a bunch of computers that are not close to each other, linked together by a network. A WAN may link multiple LANs together.
- **hub**: A hub connects multiple network interfaces together. A hub only works for certain types of LANs, such as ethernet. Logically, a hub allows every network interface listen to every other network interface connected to the hub.
- **switch**: A switch connects multiple network interfaces together. Unlike a hub, however, a switch does not broadcast messages to all interfaces. Instead, a switch understands enough protocol (OSI layer 2, and a little bit of layer 3) to direct traffic as needed. Consequently, when a message is sent, only the intended network interface receives it. All other network interfaces connected to the same switch do not receive cross traffic not intended for them.
- **router**: A router connects multiple network interfaces like a switch or a hub. However, a router understands higher level protocols (OSI layers 3 and 4). As a result, a router differs from a switch because it can connect network interfaces with different IP address “families”. As the name implies, a router can also decide how to forward a packet of information based on traffic pattern and static routing information. Some routers are intelligent enough to masquerade an entire LAN as a single IP address (via NAT, network address translation).
- **firewall**: A firewall is a programmable filter that operates at layers 2 to 7 in the OSI model. In short, a firewall determines what type of traffic is permitted across a single connection point between networks. A “pure” firewall is completely transparent in a network. However, most firewalls are integrated into routers or gateways.

- **gateway:** A gateway is a special kind of router (and often doubles, or triples as a firewall). Sometimes, a gateway is called an “edge router”. A residential gateway is a combination of router, switch and firewall. Essentially, a gateway in a network is “the guy to ask if the destination is not in the local network.”

## 1.4 DNS

At this point, we mentioned MAC address and IP address. A MAC address is a 48-bit number, while an IP address (according to IPv4) is a 32-bit number. It is difficult for people to remember web site addresses as numbers.

Consequently, DNS was invented so that we can associate IP addresses with symbolic names that we can easily remember. The analogy of a world without DNS is that your cell phone has no address book. You *have* to remember exact phone numbers to reach others. A world with DNS, on the other hand, is a world with caller ID and phone-books. You just say “Johnny”, and your cell phone automatically looks up the actual phone number.

DNS also provides one level of abstraction. A web site can change the actual IP address. As long as the “phone book” is updated accordingly, everyone can still locate the web site.

### 1.4.1 What’s in a Domain Name?

Let’s use `power.arc.losrios.edu` as an example.

`edu` is called a top-level domain. This part identifies certain attributes of a domain name. For example, `edu` means educational institute, `com` means commercial entity, `org` means non-profit entity and etc.

`losrios` is a subdomain of `edu`. `arc` is a subdomain of `losrios.edu`. However, `power` is a hostname in the domain `arc.losrios.edu`.

### 1.4.2 DNS Queries

Everytime you use a domain name, the computer needs to translate it into an IP address. It is difficult for one single computer to perform all the translations of all the possible domain names in the whole world! As a result, there are many DNS servers forming a hierarchy to translate domain names to IP addresses.

Everything starts with root DNS servers. A DNS root server does not know everything, but it knows who else may know something. For example, a DNS root server may know who else knows about the top level domain `.edu`. The DNS server (no longer root!) who knows `.edu` knows who else knows `.losrios.edu`, and that server in return knows who else knows `.arc.losrios.edu`, which returns the IP address of `power.arc.losrios.edu`.

### 1.4.3 DNS Caching

It is a big hassle to have to query multiple servers just to resolve a domain name *everytime* the domain name is used. Most DNS servers cache frequently and/or recently used entries. This speeds up DNS query turnaround time, and it also saves a lot of network “bandwidth”.

### 1.4.4 I Want My Own Domain Name

You can have it, as long as no one else has claimed ownership. Domain name registration is the process to pay and inform a registry how to translate a domain name. Needless to say, most registries own DNS servers that are close to the root servers.

What if you are not satisfied with `www.gabethegreat.org`, and want to specify subdomains such as `cars.gabethegreat.org`, `boats.gabethegreat.org` and etc.? You have some options. If you have a single hosting computer, and it supports “virtual hostname”, then the differentiation of the two domain names can be done in software. You do have to set up the DNS records so that both domain names point to the same hosting machine.

On the other hand, if you want requests to the two domains be directed to two separate machines, then you’ll need multiple DNS records.

### 1.4.5 I Want to Manage My Domain Names

If you *really* want to do this, you can use a DNS service provider (often using a web interface). Or, if you are brave, you can also run your own DNS server. In either case, you need to inform your domain name registry the address(es) of the DNS servers that handles your domain names.

[www.dnsexit.com](http://www.dnsexit.com) is a DNS service provider that offers free web-based DNS services and configuration. There are many other companies offering similar services, you should shop around for the best option.

*Warning:* if someone's livelihood depends on a website, think very carefully before you decide to manage the DNS records, or to run a DNS server yourself! If you mess up the configuration, or if the DNS server goes offline, a website can become completely inaccessible. It is often best to let your domain name registry and your hosting company handle the DNS configuration.

Most hosting companies have affiliation with domain name registries. When you sign up for a domain name, you can also immediately contract server space and the necessary DNS configuration to get your web site going.

## 1.5 The Next Step

This chapter is rather dry and contains lots of technical terms. What are you going to do with this?

For one thing, I hope you'll be able to read the labels of many networking products. For example, you can read the box of a residential gateway/firewall, and see how much you understand. You can even try to compare two products, and see if you can identify the differences.

The terms you learn from this chapter also enable you better understand the instructions to configure a computer for network operation. This may be as simple as getting a computer to talk to other computers in a local area network. It can also be more involved, such as getting a computer to connect to the Internet via broadband or dial-up services.

The whole DNS discussion gives you some understanding of the basic operations of DNS servers. You can apply this knowledge when you want to register a domain name and try to get your own website up and running.

Obviously, I cannot cover all the topics in full detail. If you are interested to know more about the infrastructure of the Internet (or just networking, in general) or DNS, you can search for information online using search engines. Personally, I find <http://en.wikipedia.org> to be an *excellent* source of technical computer information without ads, pop-ups or spyware.

## 1.6 Lab Activities in 152B 8/16

### 1.6.1 Log in

I'll disclose the login password in the class. The login name is the number on the monitor of each work station. Before you can do anything else, you need to first log in to a workstation.

### 1.6.2 Start Internet Explorer

In the lab, Internet Explorer can be started from a quick start button. Locate a little icon to the right of the start button with a blue lower case "e". One click of this button start Internet Explorer.

### 1.6.3 Go to My Website

At the address bar, overtype the address with the following:

```
http://www.drtak.org
```

Then, press the ENTER key. This should bring you to my website.

### 1.6.4 Navigate to the Classnotes

Once you are at my website, click "teaches". Then, click "American River College". After that, click the link with the name of this class.

The syllabus and classnote are both available in two formats. The PDF (portable document format) file is suitable for printing, while the HTML version is suitable for online viewing. Click on the HTML version of the classnotes.

Note that the lab no longer has free printing! If you want to print something, you need to get a pay-per-print (GoPrint) card. Your student ID card can be used for this purpose. You also need to "charge" the card with money first. You can do this with the "paystation" in the lab.

### 1.6.5 Navigate to this Page

You can either go to this page directly from the table of contents, or you can keep clicking the “Next” button until you get to this page.

### 1.6.6 Move around

Go ahead and move around at my website. Click buttons, use the back button on the browser to backtrack. It is important that you understand how to navigate my instructional web pages.

## 1.7 Lab Activities in 152B 8/18

### 1.7.1 Sign up for a ZIP account

ARC offers every enrolled student a ZIP account. I know, you really don’t need another free email account. However, a ZIP account also links to your midterm/final grade. It is just a useful tool.

To sign up for a ZIP account, you *must* use a computer on campus. Once you register successfully, you can check email on any computer with Internet access.

Go to <http://zip.arc.losrios.edu> to get started. You’ll need your SS#.

### 1.7.2 Moodle Registration

I don’t do Blackboard. Instead, I use Moodle as an online teaching resource. You may want to refer to my Moodle tutorial for step-by-step instructions. This tutorial is located at [http://www.drta.org/teaches/moodle\\_tutorial](http://www.drta.org/teaches/moodle_tutorial).

When you enroll in this class, you’ll need to enter a password. For this semester, it is “webbie”.

### 1.7.3 Forum Participation

After you self enroll, participate in the forum titled “Why are you taking this class?” There is no right or wrong answer. It is worth 10 points (not much). It is mostly just an exercise to get you to use Moodle and get familiarized with the interface.



# Chapter 2

## Week 2: Internet Services and Connections

### 2.1 SLO

This chapter (week) introduces various services that you can access over the Internet, as well as setting up connections to the Internet. By the end of this week, you'll know some common uses of the Internet and how to access the Internet.

### 2.2 Services

The term “service” has a special meaning in networking. Let us first go over some terms, then we can explore some commonly available services that you can access over the Internet.

#### 2.2.1 Terms

- **Port:** a port is a logical component of a network interface. A port is numbered from 0 to 65535. When a TCP connection is made, it must specify the ports on the two ends.
- **Service:** a service is *how* a port is used. there are some conventions. For example, port 80 is the HTTP port, port 22 is the SSH port, port 21 is the FTP port, and so on.
- **Server:** a server is a machine that listens for requests at designated ports. When a request arrives, a server interprets the request, and generates a reply.
- **Client:** a client is a machine that originates requests to servers. Typically, a user's machine contain many different client programs.

#### 2.2.2 Webpage distribution

One important use of the Internet is the distribution of web pages. A web page is a text document that is written in HTML (HyperText Markup Language) .

In order to distribute webpages, servers and clients communicate in HTTP (HyperText Transport Protocol) . As with most protocols, a client generates an HTTP request, and sends that to a server. When the server receives the request, it interprets the request and usually replies with an HTML document. The HTML document (in plain text) is, then, interpreted by the client and display on a screen.

HTTP clients are often called web browsers . HTTP servers are often called web servers or web hosts. Well known HTTP clients include Microsoft Internet Explorer, Mozilla, Firefox and Opera. Well known HTTP servers include Microsoft Information Interchange Server and Apache.

### 2.2.3 Newsgroups

In the past, before web-based discussion forums became popular, many people participate in Internet newsgroups. A newsgroup is like a discussion board or bulletin board. Unlike web-based discussion forums, which require only a web browser, newsgroups require special software.

The protocol for news article distribution is called NNTP (Network News Transfer Protocol) . As you may expect already, a computer requires an NNTP client program to connect to an NNTP server to collect articles and to post new articles. An NNTP client is often called a newsreader.

Although newsgroups are giving way to web-based discussion groups, they are still very much alive and active. Most communication programs such as Outlook, Mozilla and Thunderbird allow users subscribe to newsgroups, view articles and post articles.

### 2.2.4 File Transfer

Although HTTP *can* be used to transfer files, another protocol is used more frequently for raw file transfer. FTP (File Transfer Protocol) has been used for many years for transferring files.

All versions of Windows include a command-line based FTP client called `ftp`. This program allows a user log in to a remote machine and transfer files to/from the remote machine. In general, a user needs to have a username and password on the remote machine to gain access to its files. However, anonymous login is very common with public FTP servers intended for the distribution of public data.

A command-line based FTP client can be difficult to use for beginner users. Fortunately, most web browsers (Internet Explorer, Mozilla, Firefox and Opera) also understand how to communicate in FTP with FTP servers. A user can drag and drop files to get them transferred.

### 2.2.5 Remote Command-Line Access

Next to FTP, telnet is an ancient protocol implemented by early networked computers. Telnet allows one compute gain remote command-line access to another computer. This was important in the early days when most personal computers were slow, and computations were performed on large mini and mainframe computers.

However, even today, telnet is still commonly used by system administrators. Telnet allows a system administrator remotely log in to another machine and perform maintenance and/or repair operations.

All versions of Windows include a telnet client. However, only server versions of Windows include telnet server programs.

### 2.2.6 Email

Another popular use of the Internet is the distribution of electronic mail (email). The distribution of email, unlike other services, involves more than one server and one client.

When a message is sent, an Email client (such as Outlook, Mozilla and Thunderbird) uses SMTP (Simple Mail Transfer Protocol) to transmit the message to an SMTP server. The SMTP server is usually not the destination! The SMTP server is like a local post office that collect mail from the neighborhood. It collects messages to be sent to their destinations.

Because we usually cannot predict when messages are created, an SMTP server must be on all the time. Upon the reception of a message to be sent to its actual destination, an SMTP server attempts to the talk to the intended destination SMTP server. Yes, we now have to SMTP servers! If all goes well, the destination SMTP server accepts the message, and stores it locally.

Note that the destination SMTP server is unlikely to be the machine of the recipient. In other words, Uncle Pat's PC probably does not run an SMTP server. When the destination SMTP server receives a message, it stores the message locally until "someone comes to pick it up". In our example, Uncle Pat needs to pick up the message from the destination SMTP server.

When Uncle Pat starts up Outlook, Mozilla or Thunderbird, the communication program attempts to log on to the destination SMTP server and retrieves messages. There are two protocols for picking up mail.

The older standard is POP3 (Post Office Protocol version 3) . The newer standard is IMAP (Internet Message Access Protocol) . Without going into details, let's just say that IMAP allows more flexibility and is generally speaking more efficient.

### 2.2.7 Peer to Peer (P2P)

The Internet is increasingly used to distribute files. Although NNTP, FTP and HTTP can be used to distribute files, all of these protocols require special server programs be set up on machines that contain files to be distributed.

Peer to peer programs serve as clients and servers at the same time. In other words, the same program can receive a file from another machine, and at the same time transmit a file to another machine. Such programs are designed to install on personal computers easily, without the difficulty associated with setting up HTTP, FTP and NNTP server.

Furthermore, most peer to peer protocols are designed to handle partial file transfer so that huge files can be transferred in sections.

A particularly interesting P2P protocol is BitTorrent . This protocol allows a file be transferred in pieces from multiple “nearby” peer computers. The result is reliable and efficient file transfer that requires minimal overall Internet bandwidth (because peers are close in network distance). BitTorrent also reduces the network load of serving peers because each only needs to supply portions of files to served peers. Special BitTorrent tracker machines keep track of active peers so that file transfer remain optimal as peers come and go.

### 2.2.8 Online Chatting/Instant Messaging

Online chatting become popular as more people have access to the Internet. Big players include Microsoft Network (MSN), America Online and Yahoo. There are many small players in this field, however. In fact, anyone who runs a server capable of IRC (Internet Relay Chat) can facilitate chatrooms.

The use of chatrooms is best known for online dating and sex predators. However, chatrooms are also useful for serious and legitimate activities. For example, technical chatrooms are available for installing programs, operating systems, as well as to configure networks and diagnose computer related problems.

Instant messaging often includes the capability of voice communication and file transfer.

There are many different protocols for online chatting. IRC is the open and public protocol. MSNP (Mobile Status Notification Protocol) is used by MSN Messenger. AOL Instant Messenger uses OSCAR (Open System for CommunicAtion in Realtime), while AIM Express uses TOC (Talk to OsCar, a limited protocol based on OSCAR). ICQ also uses OSCAR.

Although big players like AOL and MSN try to ensure loyalty using proprietary chatting clients, there are many programs that can handle all mentioned protocols. Furthermore, many of these programs are open source, which means they are free to download, install and redistribute.

### 2.2.9 SSH

Telnet and FTP are functional protocols from the past, before computer and network hacking became common. Because username, password and content are transmitted in plain text, telnet and FTP are both easy targets for packet sniffers. A packet sniffer is a machine that passively listens to packets on a network that are intended for other machines.

SSH (Secure SHell) is a misnomer. Although it may have been originally invented for “shell access” (same as remote command-line access), SSH has evolved to become a much more flexible protocol.

SSH uses encryption to keep sniffers at bay. Even though a sniffer can listen to SSH packets, the packets have little meaning to a sniffer because of the lack of the deciphering key. In other words, in the eyes of bystanders, SSH packets look like jibberish.

In order for SSH to work, a user must have an account on the remote machine (SSH server).

SSH can replace telnet and FTP for remote shell access and file transfer. However, SSH can do much more. The port forwarding feature of SSH allows just about any protocol to be tunneled through an SSH connection. Protocol tunneling is a topic that is out of the scope of this class, however.

### 2.2.10 SSL

SSL is a protocol that encapsulates another protocol. This means SSL is like an envelop that contains some content, possibly another envelop. SSL stands for Secure Socket Layer . The purpose of SSL is to allow authentication and encryption between two network interfaces.

In theory, SSL can encapsulate any protocol, including telnet, FTP and all the protocols mentioned in this chapter. In reality, however, SSL is mostly used to encapsulate HTTP (to become HTTPS). It is also used to encapsulate SMTP, NNTP, POP3 and IMAP.

Like SSH, SSL packets look like gibberish to bystanders due to encryption. *Unlike* SSH, SSL does not require a corresponding user account on the machine that listens to SSL requests. This makes SSL perfect for secure HTTP connections because HTTP connections do not require accounts on HTTP servers.

## 2.3 Connecting to the Internet

There are many ways to connect a computer to the Internet. This section discusses several common methods used by end users. Corporations and large organizations have other means to connect to the Internet for the necessary speed (bandwidth), security and reliability.

### 2.3.1 Basic Concept

The main infrastructure of the Internet consists of high-end routers. Such routers communicate with each other at high speeds, typically more than several Gb/s (giga bits per second). One giga bits is one billion bits, which translates to about 100 million bytes (MB). For comparison purposes, an audio compact disk (CD) has about 700MB of data on it. This means with these high end routers, one can transmit the entire content of an audio CD in a few seconds!

As much as we want to have this kind of connection speed to the Internet, the cost is prohibitive to individual users. Big corporations, on the other hand, can afford to have high speed connections. For the rest of us, we rely on ISPs (Internet Service Providers) to gain limited access to the Internet.

From the business perspectives, an ISP connects to the “main trunk” of the Internet by leasing high speed network connections from major telecommunication companies (such as SBC, MCI, Surewest and etc.). Then, an ISP resells Internet connections to end customers like us. The following sections describe how individual end users can access the Internet via ISPs.

### 2.3.2 Dial-up

This is the most conventional and slowest of all connections. An ISP runs a modem bank consisting of many modems connected to a telephone network. The computer of an end user also needs a modem to connect to the modem bank.

A modem is a “modulator-demodulator” device. The technical detail is rather boring and not important to an end user. It is important to realize, however, that a modem needs access to a regular telephone line. In addition, a modem needs to monopolize a telephone line when it connects to another modem. This means when a modem is using a phone line, it *cannot* be used for voice or fax.

The top connection speed of a modem is about 56kbps (56 thousand bits per second) . This is rather slow when compared to the other methods. However, the cost to subscribe to a modem (dial-up) based ISP is inexpensive at about US\$10 per month.

Dial-up ISPs try to make it as simple as possible to use their services. As a result, setting up a computer for dial-up access is often very simple. An ISP mails out a CD to a customer, and the customer needs to install some software on a computer (with a modem). Then, the user only needs to supply some form of ID (registered name, customer ID, etc.), authentication (a password), and the software connects to the ISP.

One beauty of dial-up (besides the modest cost) is there is no hardware to install besides the modem. In other words, no one from any telecommunication company needs to install anything at the point of service.

### 2.3.3 DSL

DSL (digital subscriber line) transmits information over regular phone lines. However, DSL differs from dial-up in many ways.

First, DSL does not monopolize a phone line. This means that you can use voice or fax (only one) while the computer is connected to the Internet at the same time. This is a very useful feature, as people are spending more and more time online.

Second, DSL is faster than dial-up. A “basic” DSL subscription provides 128kbps (128 thousand bits per second) of outgoing speed, and up to 1Mbps (1 million bits per second) incoming speed. Even basic DSL is fast enough for music and low-quality video streaming.

Subscribing to DSL involves the installation of some hardware. This means that a customer may need to pay an installation fee. The software configuration to connect via DSL is usually as simple as that of a dial-up connection (if not simpler).

### 2.3.4 Cable Modem

Cable modem transmits information over the cable used conventionally for cable television program delivery.

In many ways, cable modem and DSL are similar. The main difference is that one relies on a regular phone line (which most household already has), while the other one depends on cable TV connections, which may not be installed at some households.

The communication speed of cable modem is often comparable to a similarly priced DSL connection. The installation of hardware and configuration of software for cable modem connection are both comparable to that of DSL.

### 2.3.5 Fiber Optics

Some regions are now routed with fiber optic connections. A single fiber is capable of at least 1Gbps (one billion bits per second) to the telecommunication company. This is *much* faster than what is possible with DSL and cable modem.

While a fiber optics connection can be used for Internet access, it can also be used to deliver telephone and cable program all at the same time.

Although a fiber optics connection is more expensive at US\$50 per month for Internet access, it is much faster than DSL or cable modem at the same price range.

The downside of fiber optics connection is that it is only available to certain regions.

### 2.3.6 Selecting an ISP

This section does not guarantee that you'll make the right choice. Its purpose is simply to make you think about certain aspects of connecting to the Internet, and compare the relative advantages and disadvantages of different ISPs.

The first question for individual users is budget. With a budget of US\$10 per month, there is really no choice but to use dial-up. Dial-up connections are slow, and a modem ties up the phoneline. However, for occasional Internet connections that require little connection speed, dial-up access is a cost effective and realistic choice.

Among dial-up ISPs, a customer should select based on toll-free/local number availability, modem bank availability, price and service reliability. For beginners, user friendliness and customer support are also important factors.

DSL, cable modem and fiber optics are often lumped into the same category of "broadband". Selecting amongst these technologies depend much on availability and the cost of package deals. For example, some remote locations do not have fiber optics nor DSL, so cable modem is the only practical choice.

For a household already installed with phoneline, cable and fiber optics, the choice depends on usage and cost of packaged combinations. In other words, one needs to evaluate the total cost of entertainment, phone services *and* Internet access. This cost varies from one household to another. A household that wants paid entertainment may prefer cable modem or fiber optics, while a household that does not want paid entertainment may prefer DSL.



# Chapter 3

## Week 3: WWW and Browsing

This chapter introduces the World Wide Web and the world of web pages.

### 3.1 SLO

After reading this chapter, and attending the lectures, you'll be able to:

- understand terms related to the World Wide Web
- use a browser, and navigate to web sites
- understand what a search engine is
- query search engines for webpages related to specific topics
- use bookmarks to remember the addresses of web sites
- understand HTTPS and some security issues while accessing the WWW

### 3.2 Warning

The World Wide Web (WWW) is a fun place, and it can certainly be used for all kinds of research activities. However, keep in mind that the WWW is also a dangerous place. Your computer can be infected by viruses and spyware with little or no symptoms!

A virus is a computer program that satisfies the following criteria:

- gets installed without authorization or notification
- gets installed as the result of some user action (other than just turning on the computer)
- has means to propagate itself to other computers from an infected computer

The effect of a virus varies depending on the variant. Some relatively harmless viruses simply propagate by sending itself as an attachment to everyone on email contact lists. Others, however, can cause a system to crash, lose data or otherwise misbehave, all in addition to propagating itself to other systems.

A spyware program is a computer program that satisfies the following criteria:

- gets installed without explicit authorization or notification
- has means to collect information from a computer
- has means to send collected information to a “mother ship” computer

Depending on the program, a spyware program can acquire browsing habit, get email addresses from address books, read content on a hard disk and etc. Unlike viruses, however, spyware does not self propagate, nor do they damage the infected system.

### 3.3 Terms

- URL : Uniform Resource Locator. This is a form of address for webpages in the WWW. An URL always indicates where a document is located
- URI : Uniform Resource Identifier. This is a more general form of URL. Every URL is a URI, but not vice versa. An URI always identifies a resource, but it may not necessary indicate how to get it.
- HTTP : HyperText Transport Protocol. This is the grammar of the language understood by browsers (Internet Explorer, etc.) and web servers (Apache, etc.).
- hyperlink : a hyperlink is either an image, a part of an image, or some text that is “clickable”. When a hyperlink is clicked, it directs a browser to a designated URL.

### 3.4 Basic Browser Operation

The most basic way to start using a browser is to enter an URL in the address bar. If the URL corresponds to the address of a valid webpage, the browser downloads the webpage and all the necessary components (such as images), and renders the page on the screen.

Navigating a web page is much like navigating in any graphical program. Use the scroll bars to scroll horizontally and vertically. However, a web page often contains hyperlinks. A user can click on a hyperlink to access another webpage.

Every time a hyperlink is clicked and another webpage is loaded, the browser remembers how it got to the new webpage. The “back” button (often using the icon of a big left arrow) allows a user to “backtrack” to the originating webpage. Most browsers support a fairly large number of backtrack levels.

Sometimes, it is a hassle to have to press the “back” button many times. Most browsers include a small down arrow as a part of the “back” button. Pressing this down button pops up a list of steps that a user can backtrack. This feature allows a user quickly locate where he/she want to backtrack, and get there without having to press the back button many times.

As long as a user does not press any hyperlink, or enter new URLs at the address bar, he/she can also move forward to reverse backtracking.

The third button, a crooked arrow, is the refresh button. Pressing this button forces a browser to reload a webpage from its origin. This is helpful for reloading webpages that change periodically (such as that of a news site).

The fourth button is the stop button. Pressing this button aborts the loading of a webpage.

### 3.5 Advanced Browser Operations

#### 3.5.1 Bookmark/Favorite

Most people have relatively short lists of frequently visited URLs. Instead of having to type the URLs every time, it is convenient to let a browser maintain a list of frequently used URLs.

This feature is called “bookmark” in most browsers, but it is called “favorite” in Internet Explorer. When you want a browser remember the URL of the current webpage, you go to the “Bookmark” or “Favorite” menu, and select “add” or “new”. Most browsers pop up a dialog box to ask for more details, such as the name of the bookmark, and sometimes where you want to file the bookmark.

When you want to revisit a bookmarked URL, you go to the bookmark (or favorite) menu, and the title of the URL should be listed as one of the menu items.

#### 3.5.2 Bookmark Organization

If you only have 10 to 20 bookmarks, there is no need to organize them. However, if you have 100 bookmarks (which is not uncommon), you may want to consider organizing bookmarks into folders.

Most browsers allow you to organize bookmarks. You can create folders, (and folders within folders), and move bookmarks into/out of folders.

### 3.5.3 Searching in a Webpage

Some webpages are fairly long, and it may difficult to look for a particular part of the document. Most browsers let you search for a particular word/phrase in the webpage.

In Internet Explorer, you can select the “Edit” menu, then use the “Find” menu item. The shortcut is control-F. Press function key F3 for subsequent searches of the same word/phrase.

In most other browsers, you can use control-F, the slash (/) key, or just start typing when viewing a webpage to search for a word/phrase. You can also use F3 to look for subsequent occurrences of the same word/phrase.

## 3.6 Search Engines

A search engine is not a component of a browser. Instead, a search engine is a web site on the Internet. Search engines specialize in the quick location of webpages based on search criteria.

There are many search engines, but Yahoo and google are the two most commonly used ones. Their features are often comparable.

### 3.6.1 Simple Search Query

Once you are at Google.com, you can just type the keywords that should be contained in the retrieved web pages. For example, you can search for “best web browser”, and then google returns a list of webpages that contain the words “best”, “web” and “browser”. Google also displays portions of the text of each webpage so that you can see how the words are used.

This technique works in *many* cases, but there are certain cases in which it does not work well.

For example, let’s say I want to find webpages that talks about golf clubs (as in sports equipment). A search of “good golf club” returns links to clubs (equipment) as well as clubs (social gathering).

### 3.6.2 Advanced Search Query

Google and Yahoo support advanced search query. Although you *can* enter a complex search query by hand, it is much easier to do it by clicking the “Advanced Search” link.

Getting back to our example, we want to rule out webpages about social gathering clubs. After some analysis, I found that most pages for social gathering have the words clubhouse, area, acre, cart and course.

In the advanced options, we can specify that we do *not* want the returned webpages to contain the words ‘clubhouse’, ‘area’, ‘acre’, ‘acres’, ‘cart’, ‘course’, ‘carts’ and ‘courses’. The refined search is better.

To enhance our search, we can add ‘shaft’ to the search words. This really helps to narrow down the choices.

There are many options in the Advanced Search screen of Google. You can experiment with different options and see how you can fine tune your search.

## 3.7 Alternative Browser

Internet Explorer is a browser that is included in most versions of Windows operating system (up to Windows XP and 2003). While Internet Explorer is a capable web browser published by the largest software company (as of today), it is not available to other operating systems. Furthermore, Internet Explorer also lack certain features that can be found in other web browsers.

If you are both a beginning computer user and a beginning web surfer, you may want to skip this section. However, if you already have experience with operating a computer in general, this section can provide much information.

### 3.7.1 Firefox

Firefox is a “standalone” web browser from Mozilla.org. It is the product of an open source effort. This means Firefox is free to download. Furthermore, it also means the source code of Firefox is open for anyone to download, review, analyze and modify.

Firefox has many advantages. One advantage of a lesser known web browser is that fewer exploits are written for it. Even though there are now exploits for Firefox, they are still relatively minor compared to the numerous exploits that target Internet Explorer.

Unlike Internet Explorer, Firefox has a small footprint. This means it does not require much resource to install, and it does not require much to run.

Because Firefox is not published by a commercial corporation, it does not need to align itself to web-based advertisements. As a result, Firefox has more anti-ad features than Internet Explorer, and possibly *any* web browser. If you are unhappy about all the ads and pop-ups that constantly clutter up your screen, Firefox may be your best hope to regain control.

Last, but not least, for the not-so-casual web surfer, Firefox has many extension modules. Such modules extend the functionality of Firefox. You should visit <http://www.mozilla.org> and navigate to the extensions of Firefox. While some of these extensions are cosmetics, others can be very helpful.

### Downloading and Installing Firefox

Go to <http://www.mozilla.org> and navigate to download it. The web site recognizes your operating system and language settings so that you don't need to customize.

### Getting Extensions

Under the "Tools" menu, select "Extensions". You can now click on "Get More Extensions" to let Firefox navigate to a web page that catalogs extensions. You can also update and uninstall extensions that you already have installed.

I find the following extensions particularly helpful.

- Mozilla Calendar: this is a full featured calendar/todo tool. You can keep your calendar local to a storage device, or you can upload the calendar to web storage so you can share it with others.
- Adblock: admittedly, this is not a tool for beginning users. However, once mastered, this is an excellent tool for removing all the annoying ad banners, side columns and etc. Don't worry, if you make a mistake, you can always undo it.
- ReloadEvery: this is a cool little extension to keep reloading a webpage at a specific period. It can be used to keep a page current (for a news website, or eBay). Or, you can use it to keep yourself logged in at a log-in-required web site. Be warned, however, that your session can no longer be timed out. If you leave your computer and forget to logout, some one can walk in and continue your session!

# Chapter 4

## Week 4 and 5: Email

Email is one of the most original and common use of the Internet. This chapter discusses many issues regarding email.

### 4.1 SLO

- Email definition
- Email protocols
- Email clients (Outlook, Thunderbird)
- Web-based Email
- Email address maintenance
- Email Security

### 4.2 Email Definition and Protocols

Email, or electronic mail, refers to messages delivered over the Internet via certain protocols. As with regular mail, email messages are written and sent from one end, and it is received and read at the other end.

#### 4.2.1 No Direct Delivery

The first approach is to deliver email directly from the sender's computer to the receiver's computer. But this approach has several problems.

Both machines must be switched on when the message is delivered. This is not a practical assumption. It is like asking you to be home whenever mail is delivered.

So the big question is, how do we fix this problem?

#### 4.2.2 One Middleman

The second approach is to use a single middleman, let's just call that the email server. The sender writes a message, and send that to the email server. The receiver connects to the Internet, and queries the email server, and download any message that is stored in the server.

This approach seems to be the solution, but it has a problem.

What if the network has a problem, and the message cannot be delivered to the email server (from the sender's machine)?

The sender's machine needs to remain connected to the Internet to reattempt email delivery to the server. This, once again, is an unreasonable demand on a user's machine.

### 4.2.3 Two Middlemen

The third approach is to use two middlemen.

The sender's machine first connects and sends the message to the sender's email server. The sender's email server is often a server owned by the sender's ISP. This means connectivity to the sender's email server is very reliable.

The sender's email server stays on and connected to the Internet all the time. If somehow the Internet is broken, the sender's email server attempts to resend the message. The sender's own computer can disconnect from the email server.

Of course, the sender's server does time out after a few days. At that point, it replies to the sender to indicate a delivery problem.

Once the sender's server delivers the message to the receiver's server, the sender's server is done. The delivered message stays on the receiver's server until the receiver's machine connects to it. The receiver's machine then retrieves all message from the receiver's server.

### 4.2.4 Protocols

- sender's machine connects to the sender's email server via SMTP
- the sender's email server connects to the receiver's email server via SMTP
- the receiver's machine connects to the receiver's email server via POP3 or IMAP

## 4.3 Email Clients

An email client program serves to send and reserve messages. Consequently, it needs to understand SMTP, POP3 and IMAP. In the old days, email clients are command line oriented. Of course, email clients these days are fully GUI (graphical user interface) based.

Microsoft Outlook is email client that comes with Windows. However, you can download and install free alternative email clients. Mozilla Thunderbird is such an alternative.

Regardless of the email client, you need to configure it to send and receive email.

### 4.3.1 Account Name

Most email client asks for an account name. This is just a name you associate with the account. It can be any name you choose. It is best, however, to use a name that tells you the actual account.

### 4.3.2 Email Address

Next, you need to supply an email address. This email address may not be the same as your actual email address for this account. However, This will be the reply-to address of all messages that you send.

### 4.3.3 Username

This is the actual username for this email address.

### 4.3.4 Server Type

You have to select the server type. It should be either POP3 or IMAP. This is a piece of information that you should get from your ISP.

### 4.3.5 Server (Incoming)

This is the name and address of the POP3 or IMAP server. This is from which messages are received. This is often named `mail.your-isp.com`, `pop3.your-isp.com` or `imap.your-isp.com`.

### 4.3.6 SMTP Server (Outgoing)

This is the name and address of the SMTP server for sending messages. It may, or may not, be the same as the incoming server. This is usually named `mail.your-isp.com` or `smtp.your-isp.com`.

### 4.3.7 Options

Most email clients have many options to select.

- Delete messages from the server (Leave messages on server). This is an option that lets you specify if you want to delete messages from a server. Most of the time, you want to delete messages from the server once they are retrieved. However, if you want to be able to retrieve messages from the server from different clients, you need to make sure messages do not get deleted.
- Check for new messages at startup. Normally, you want your email client to check email messages as soon as it is started. The only exception is that when you want to control when you want to connect to the incoming email server.
- Check every  $x$  minutes. This sets up the email client program to check the incoming email server periodically.
- Automatically download new messages. This option is often enabled by default. However, you can disable this, and just have the email client notify you about new messages, but not to download them. This can be helpful if you have a slower connection, and don't want to download messages unless you have better access to the network.
- Fetch headers only. This feature can be combined by not automatically download new messages. You can download just the headers so that you know who sent you what messages. You can then delete all the junk mail without downloading the contents.
- Use secure connection (SSL). This feature should only be enabled if you know that your ISP supports SSL.
- Use secure authentication. Again, this feature should only be enabled if your ISP supports secure authentication.

## 4.4 WebMail

As computers become available at home and at work, it becomes increasingly cumbersome to configure email client programs on different machines to access email. To solve this problem, many ISPs offer web interface to email accounts.

Web-based email access is easy because there is no need to set up the account. The ISP does everything for a user. However, there are some issues related to web-based email.

### 4.4.1 Continuous Internet Connection

With an ordinary email client like Outlook or Thunderbird, one only needs to go online when he/she is fetching incoming messages, or sending outgoing messages. There is no need to connect to the Internet when a user is reading incoming messages or composing outgoing messages.

A web-based email system, however, really needs continuous Internet connection. This means that Outlook and Thunderbird (traditional email clients) are more suitable for dial-up accounts. With 24/7 broadband access, however, web-base email is easier to set up.

### 4.4.2 Editing Options

Most email clients like Outlook and Thunderbird provides a whole suite of editing options. For example, text can be formatted almost like with a word processor. Content can be spell-checked. You can usually include inline graphics, and attach additional files relatively easily with an email client program.

The editing features of a web-based email interface are often more limited.

### 4.4.3 Sending Options

When a message is sent, Outlook and Thunderbird allows the sender select certain options. Such options include "receipt requested" and etc. Some web-based email interfaces support that, too. However, most web-based interfaces do not support all the options offered by email client programs.

## 4.5 Address Book

Most email clients and web-based interfaces support address books to a certain degree. Some web-based interfaces support minimal address book features:

- email address
- display name

The address book of an email client is often a bit richer. Each entry has the following fields, in addition to the basic one described above:

- first name
- last name
- business address
- home address
- phone number
- fax number
- etc.

Email client programs also have the ability to import entries into an address book, or export entries from an address book. This is useful for the integration with mail merging or PDAs (personal digital assistants).

## 4.6 Security

As usual, security is a big topic.

### 4.6.1 Privacy

Unless you encrypt your messages, there is no privacy to speak of, period.

Your ISP (or the organization running the email servers) can save, log and review each and every email message that you send and receive. It is as simple as that. Furthermore, your ISP can also potentially scan your messages for keywords so that affiliated “advertisers” can spam you based on your messages.

Don’t forget, too, that in most cases there are to ISPs involved in the transmission of email messages. The sender’s ISP and the recipient’s ISP can *each* scan, log and review a message.

### 4.6.2 The Importance of Privacy

A common view is that if one does not do anything wrong, there is no need for privacy. Well, that explains why many government documents are guarded, does it?

Privacy (or confidentiality) is fundamental to the success of a business, especially when there are plenty of competitions. For instance, company A may have a strategic plan to share among its board of directors. If a spy of a competitor pays an ISP to log outgoing messages from these board members, then this competitor can learn all the strategic secrets, and beat company A in the market.

Similar situations exist even in schools and colleges. If a bad professor worries about complaint messages, he/she can pay the system administrators bribe and have such messages blocked or at least logged (so that the professor can retaliate).

### 4.6.3 Authenticity

Are you sure that message is *really* coming from a friend of yours?

Faking the origin (and sender) of an email message is *easy*. This is especially the case when the faker is using the same ISP as the fakee (the one being faked).

Of course, ordinarily, no one will fake an email message from your relatives. However, combined with the lack of privacy, messages with faked origins can be very dangerous. Let's consider the following scenario.

Mary plans to purchase a new car, but her CD (certificate of deposit) does not mature until 3 months later. She contacts her brother, John, to borrow some money and plans to return that when her CD matures. Her message to borrow money is intercepted by her ISP (or her brother's ISP). When John replies and asks of the routing number of the account to wire money, the ISP blocks the message so Mary never receives it. Furthermore, the ISP fakes a reply, and send the routing number of a temporary account set up just to receive John's money in this scam.

This *can* happen.

What is the lesson? If you need to send confidential information, or need to verify the authenticity of senders and recipients, do not use *plain text* email!

### 4.6.4 PGP and GPG, Key-pair System

PGP is Pretty Good Privacy. GPG is GNU Privacy Guard. Both consist of a suite of programs that ensures the privacy and security of content transmission over the Internet (or data in storage).

Fundamental to both systems is the key-pair concept. Let's worry about the mechanism of the key-pair system later. Right now, let's focus on the application of PGP or GPG.

GPG can encrypt messages. This means a message can only be viewed by its intended recipient. This is important, because it means an ISP can no longer understand a message in transit.

GPG can also sign messages. This means a message can be digitally signed so that the recipient can authenticate the origin of a message. Both plain and encrypted messages can be signed.

Thunderbird has extensions to integrate GPG nicely into the email client. It can encrypt, decipher, sign and authenticate messages conveniently. Furthermore, the GPG extension can also maintain key-rings, so that you don't have to manually track the public keys of recipients.

## 4.7 Key-pair Encryption

This paragraph is intended as a very concise and incomplete guide to encryption. You really need to take a few CISS courses to fully understand the concepts and applications.

Key-pairs are used in many encryption schemes because it elegance and flexibility. A key is, essentially, a large bit pattern. A key pair consists of two such keys, but these two bit patterns have a special property. A message encrypted with one can be deciphered *only* with the other one.

Let's say we have key A and key B in a key pair. This means messages encrypted by A can only be deciphered with key B, and vice versa. This is an interesting property, because we can designate one key as a private key, and the other key as a public key. The public key can, then, be broadcasted to the world. This means *everyone* can use the same key to encrypt message intended for a particular recipient. However, no one with just the public key can decipher such encrypted messages.

This also means that it is not possible to encrypt a message intended for a group of people. Because each recipient in the group has his/her own public key, the group cannot share the same encrypted message. The original message must be encrypted by the public key of each recipient, and sent only to the owner of the public key.

This takes care of the encryption application of a key-pair.

Occasionally, one does not want to encrypt a message, but rather to sign it and authenticate its origin.

Interestingly, signing a message works for a group of recipients. In other words, a message can be sent to a group of people, and everyone in the group can authenticate the origin of the message.

Because the ability to encrypt and decipher is symmetric between the private key and public key, here is how digital signatures work.

- The sender composes a message, with or without attachments
- A checksum algorithm computes a "digest" bit pattern from the message. The checksum algorithm must be known for all the recipients.

- The digest is encrypted using the private key of the sender. The private key encrypted digest is included as a special part of the message.
- The public key of the sender can be optionally sent along with the message. Most of the time, the public key is published at key-ring web sites or other means of broadcasting.
- A recipient receives the message, and acquires the public key of the sender by some means (see the previous point).
- The recipient computes the digest of the message body using the same checksum algorithm used earlier.
- The recipient uses the public key to decipher the private key encrypted digest from the message, and compares that against the computed digest of the message.
- If the deciphered private key encrypted digest matches the computed digest, then the message is authenticated.

The advantage of signing a message (as opposed to encrypting) is that the sender does not need to know the public keys of all the intended recipients, and all the recipients can still authenticate the origin of the message. Of course, signing a message is only applicable when the content of the message itself is not confidential.

# Chapter 5

## Week 6: Webpage Authoring

Let's talk about something that is more interesting. It is relatively inexpensive to get some "web space" to store your very own web pages. This chapter explores various options to create web pages.

### 5.1 SLO

- Identify key elements of webpage creation.
- Evaluate different methods to generate webpages.
- Compare relative advantages and disadvantages of different methods to generate webpages.

### 5.2 What is a Web Site?

A "web site" is loosely defined to be an URL (web address) that contains a set of related webpages that are interlinked by hyperlinks.

#### 5.2.1 Free Web Space

It is relatively inexpensive, or even free, to get web space to publish your webpages. There are one main disadvantages of free or low-cost web space. Most free web space have mandatory advertisements, which can be distracting to the reader.

Use a search engine to look for "free web space" or "free web hosting", and you will find many links. I cannot be responsible for the integrity (or there lack of) of these companies. The reason why companies can offer free web space is advertisement. Afterall, if you host something that is interesting from this kind of web hosting company, their advertisers will be willing to pay more to advertise. Most free web hosting companies determine the advertisements (on the banner and on the sides) depending on the content of the hosted pages.

For beginners who are just experimenting, free web hosting can be a real bargain. Some of these web hosting companies offer up to 100MB (100 mega bytes) of storage. That's a lot of space for beginners.

#### 5.2.2 Dedicated Paid Web Space

If you cannot stand the advertisement banners, you can pay a web hosting company so that you have complete control over the contents. Web hosting fees vary, but most companies offer starter plans at US\$12 per *year*. The availability of network bandwidth and inexpensive storage drive the price down.

#### 5.2.3 Custom Domain Name

To publish contents in the World Wide Web, you can optionally use your own domain name. You need to first register the domain name (assuming no one else owns it) with a registry, then inform a web hosting company that you want to host your web site using their space.

Some free web hosting companies welcome the use of custom domain names. This is not really surprising, as there is no extra effort required.

A custom domain name can be very helpful in terms of building a business. This is because you can pick a name that is easy to remember, or pick a name that matches that of your business.

## 5.2.4 Evaluate Web Hosting Companies

With so many web hosting companies out there, how do you select one?

**Reliability.** This is important if your business or livelihood depends on the web site. For experimenting beginners, this is really not much of a big deal. An uptime ratio of 99.9% versus 95% has very little impact on an experimental website just for friends and family. On the other hand, it makes a huge difference for a business-oriented website. Website down time means loss of business *and* loss of credibility.

**Speed.** There are many factors determining the apparent speed of a web site. A web hosting company that is close to the “trunks” of the Internet can deliver contents more quickly. However, script-based web sites also require processing power on the server.

**Ease of Access.** This is no big deal for professionals. It does make quite a bit of difference to beginners. Many web hosting companies use web-based interfaces for customers to maintain their websites. This means you don’t need to learn FTP, telnet and all those specialized tools. You can upload, move, rename and delete individual files using a browser.

**Shell Account.** This is a feature that is only useful for serious professional web developers. A shell account allows a web master perform complex operations to maintain a website. For security reasons, most web hosting companies do *not* offer shell access anymore.

**Scripting.** This is a feature that is only useful for serious professional web sites. Most beginners only use static web pages. For web sites that use scripts, the support of various scripting languages (and their modules) is crucial.

**eCommerce.** Some web hosting companies offer shopping cart and credit card processing services. Again, this is only useful for commercial or professional web sites.

## 5.3 Tools to Create Webpages

### 5.3.1 Word Processor

Word processors like Microsoft Word and OpenOffice Writer can export a document as a web page. This means if you know how to use a word processor, you know how to write webpages.

The main advantage of this approach is that there is no learning curve for most people. Of course, you need to learn how to specify hyperlinks in a document to fully utilize the capabilities of HTML.

Word processors are most suitable for creating single web pages. Word processors do not include tools for site management. As a result, if your website include many interlinked web pages, you may want to consider some other tools

### 5.3.2 Text Editor

Yes, you can create a web page using Windows Notepad! An HTML file is a plain text file that any text editor can edit.

To use a text editor to maintain HTML documents, you must have a good understanding of HTML itself. This is out of the scope of this course. To learn more about HTML, you should take CISW 300 and CISW 310.

### 5.3.3 Specialized HTML Editors

Most HTML editors do not require that you understand HTML. Some can interchange between HTML and WYSIWYG modes (making them great HTML learning tools).

For those who do not want to learn HTML *at all*, NVu (<http://www.nvu.com>) is a webpage (HTML) editor derived from Mozilla Composer. Another free HTML editor is Amaya from W3C (the same organization that sets HTML standards). The address is <http://www.w3.org/Amaya>. Both Nvu and Amaya are available for different platforms, including Windows, Linux and MacOS.

There are more powerful specialized editors that can manage web sites (not just web pages). Microsoft Frontpage and Macromedia (now Adobe) Dreamweaver are two such products. On one hand, these programs are very powerful

and flexible, making them suitable even for commercial web sites. On the other hand, they can be very complex and difficult for beginners. If you are interested in such products, take CISC 306 (Dreamweaver).

For this class, NVu and Amaya are reasonable choices because they are free, and it is also WYSIWYG (well, usually).

## 5.4 Creating Web Pages

To learn more about creating web pages, you can take CISC 306 (Dreamweaver), CISW 300 and CISW 310. This section is merely a starting point for very simple web pages.

A web page can be very simple, or it can be very fancy. For this class, we'll just start with simple ones.

Both Nvu and Amaya are WYSIWYG, they resemble word processors. The only difference is that you can insert hyperlinks in an HTML editor, the file produced is an HTML file. As mentioned, you can also use OpenOffice Writer or Microsoft Word to export ordinary documents as HTML files.

You can include pictures in an HTML document. However, unlike a word processor file, images in an HTML document are not stored in the document itself. Images are stored *external* to HTML documents.

As a good practice, it is often best to use a folder to store your HTML documents and related images. This way, it is easier to keep track of all the files that are related.

## 5.5 Uploading

Uploading is the process to send files from a personal computer to a server (or any remote computer). Once you have authored HTML documents and collected your image files, it is time to send those files to a web host.

For more details, you should take CISW 300, but here we have some simple instructions.

### 5.5.1 Web-based Interface

Although *real* webmasters do not use web-based interfaces unless there is no choice, web-based interfaces are more friendly to beginners. Do not expect to use a web-based upload interface in CISW classes!

Most web hosts these days offer web-based administration. This means you can authenticate yourself, then proceed to maintain your web site using nothing more than a browser.

There is usually an interface to maintain files stored on the server.



## Chapter 6

# Authoring Web Pages

This chapter is a step-by-step guide to web page authoring using Nvu.

### 6.1 Getting Nvu

You can get the latest version of Nvu from <http://www.nvu.com>. Click “Downloads” from the top bar, and select your operating system (probably Windows). I suggest that you download and run the `exe` file for ease of installation.

The installation of Nvu should be fairly straightforward, accepting defaults all the way through the installation should be okay.

### 6.2 Some Concepts

Before we start talking about how to use Nvu to author HTML pages, let us first discuss some of the related concepts.

HTML (hypertext markup language) is a computer language that is used to describe how a web page is displayed. The actual details of HTML is out of the scope of this class. You should take CISW300 (and then CISW310) for a better understanding of HTML.

HTML is a “plain text” language. This means that you can, if you so want to, author a web page using Notepad (or any other simple text editors). However, this requires that you understand HTML tags, which is out of the scope of this class.

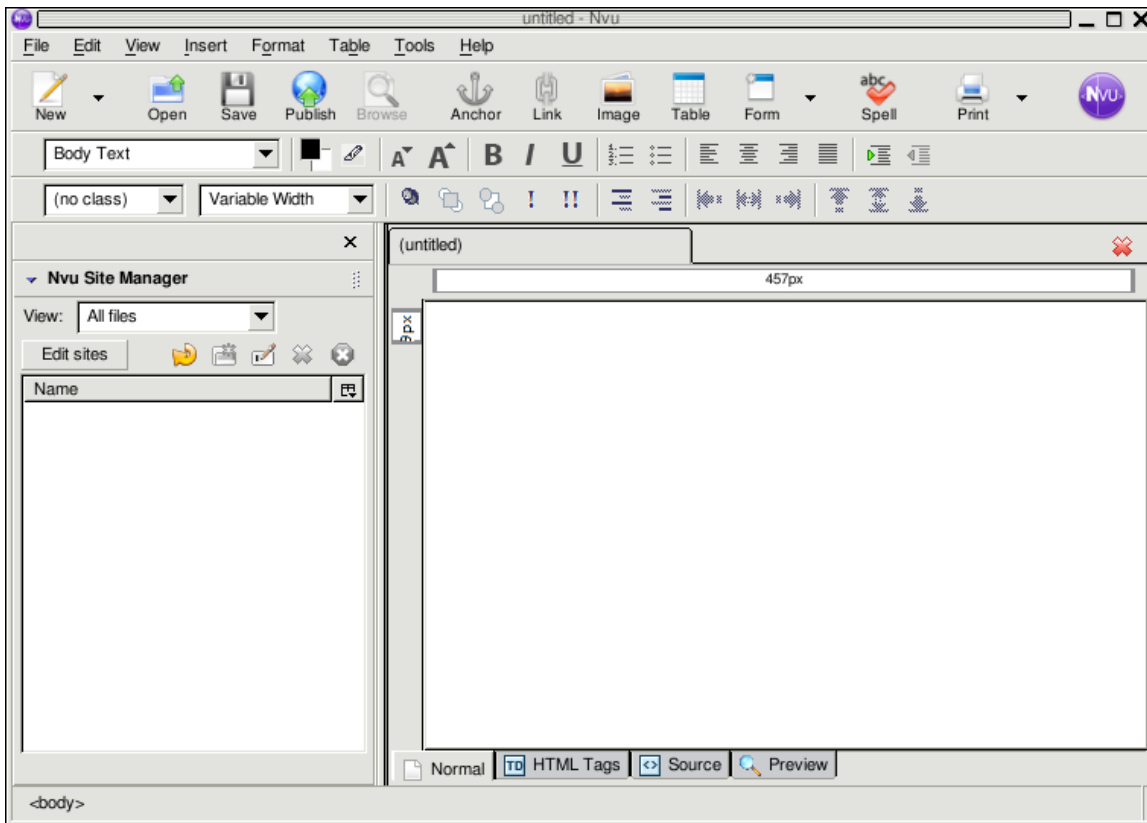
To avoid learning HTML in this class, we use editors that resemble word processors. Nvu and Amaya are two free editors that resemble word processors for HTML documents.

### 6.3 Using Nvu

#### 6.3.1 Starting up

When you start up Nvu, it pops up a “tips” dialog box. You can permanently disable it by unselecting the checkbox labeled “Show tips at startup”.

The initial screen of Nuv should look similar to the following:



At the top is the usual menu, followed by a row of tool icons under the menu. Under the tool icons are the usual word processor tool buttons.

On the left hand side is the “Nvu Site Manager” pane. This pane is important if you want to maintain a web site with multiple pages. To the right of the site manager pane is the main editor window for initial HTML files.

At the bottom of the editor are view tabs. “Normal” means WYSIWYG (what-you-see-is-what-you-get). “HTML tags” means the document is presented in a structured mode. “Source” means the document is displayed “raw” in plain text, with no special structuring for tags. “Preview” does not allow editing, but it renders an HTML document as if it is viewed by a web browser.

You probably want to use the “Normal” view most of the time. However, if you are curious about the actual HTML code, or need to insert raw HTML code (for free web space), you also need to use the “Source” view.

### 6.3.2 Simple editing and formatting

In the main text area, you can type and edit like a word processor. The buttons are fairly self explanatory.

I’ll explain the buttons/tools:

- Paragraph Style. This is a drop down box on the left side of the text formatting tool bar. If you click on the drop-down box, you can select one of the styles. The header styles are for section headings.
- Text/background Color. This appears as a square on top of another. You can change the color of text or background using this tool. Click on the square on top to change text color, and the square at the bottom to change background color. Note that the background color applies to the entire document.
- Highlight color. This changes the background color of just portions of the document.
- Increase/decrease font size. These two tools appear as the letter “A” with a down or up arrow. You can change the font size on the fly using this tool.
- Bold. This tool appears as the letter “B”. It changes the font to use bold face.
- Italic. This tool appears as a slanted letter “I”. It change the font to use italic style.

- Underline. This tool appears as an underlined letter “U”. It adds underline to the text.
- Numbered list. This tool appears as three rows numbered 1, 2 and 3. It switches numbered list mode. You can click on it to toggle between plain and numbered modes.
- Unnumbered list. This tool appears as three rows with bullets. You can click on it to toggle between plain and bulleted list modes.
- Paragraph alignment tools. These tools appear as left aligned, centered, right aligned and justified lines.
- Indentation tools. These tools appear as a right/left pointing triangle to the left of some lines. They can be used to control indentation of regular text, or nested levels of numbered or bulleted lists.

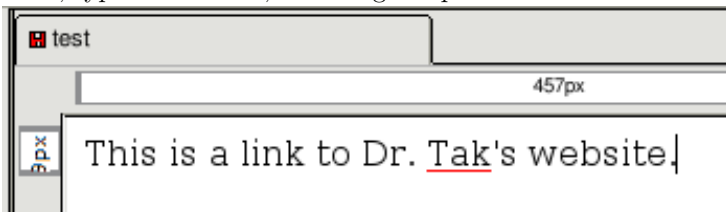
When you are done editing, you should save the file. Click on “file” on the menu, then select “save”. If this is the first time you save the file, you need to find a folder and give the file a name.

### 6.3.3 Hyperlinks

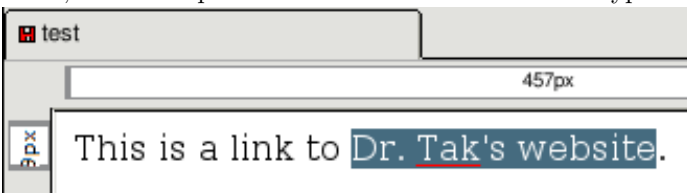
One feature of HTML is the use of hyperlinks. A hyperlink often appears as underlined blue text in a document. Click on a hyperlink lets a browser open another document.

You can insert hyperlinks in two ways. You can type normal text first, then convert portions of it to a hyperlink. This is how you do it:

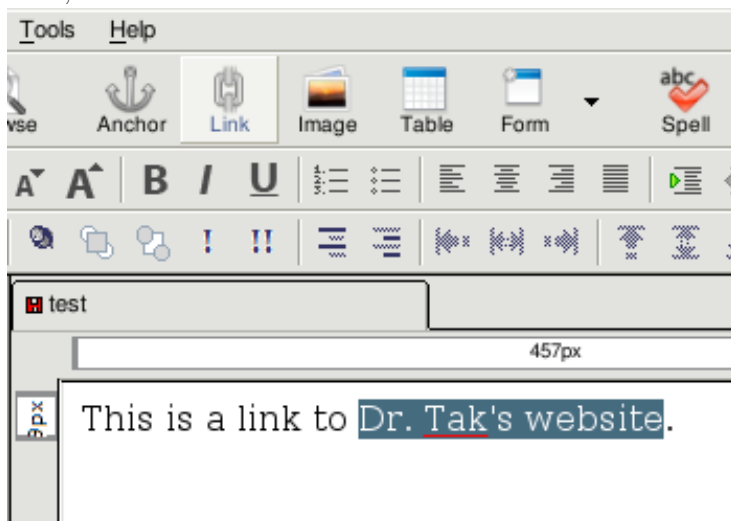
First, type all the text, including the portion that will be changed to a hyperlink.



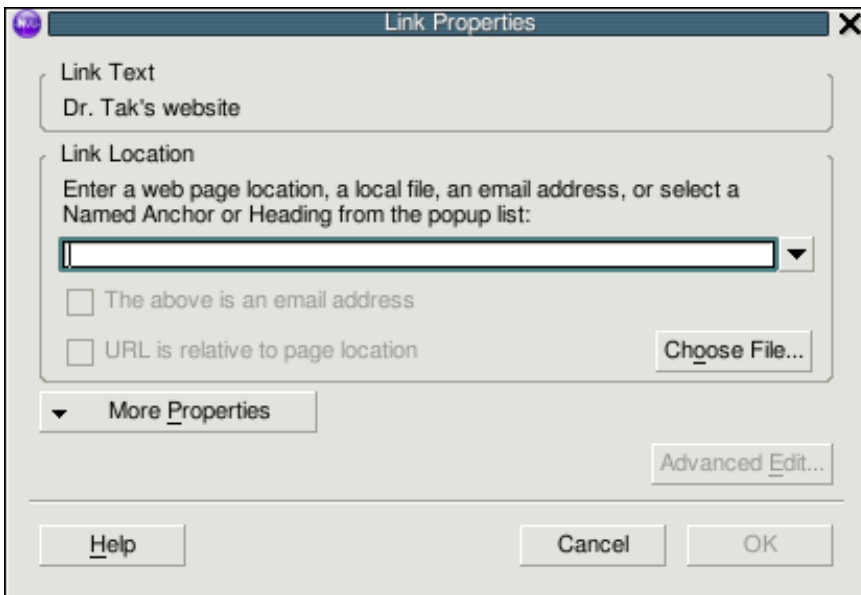
Then, select the portion that will be turned into a hyperlink.



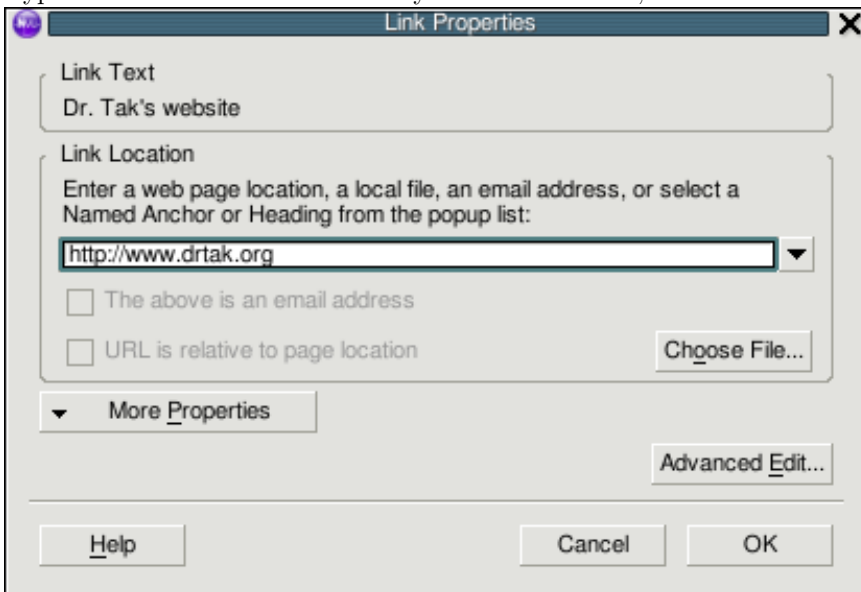
Next, click the link tool icon.



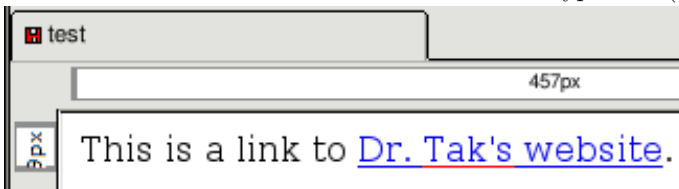
You'll be presented with dialog box as follows.



Type in the URL of the document you want to link to, then click “OK”.



You should see that the selected text is now a hyperlink (underlined blue text).



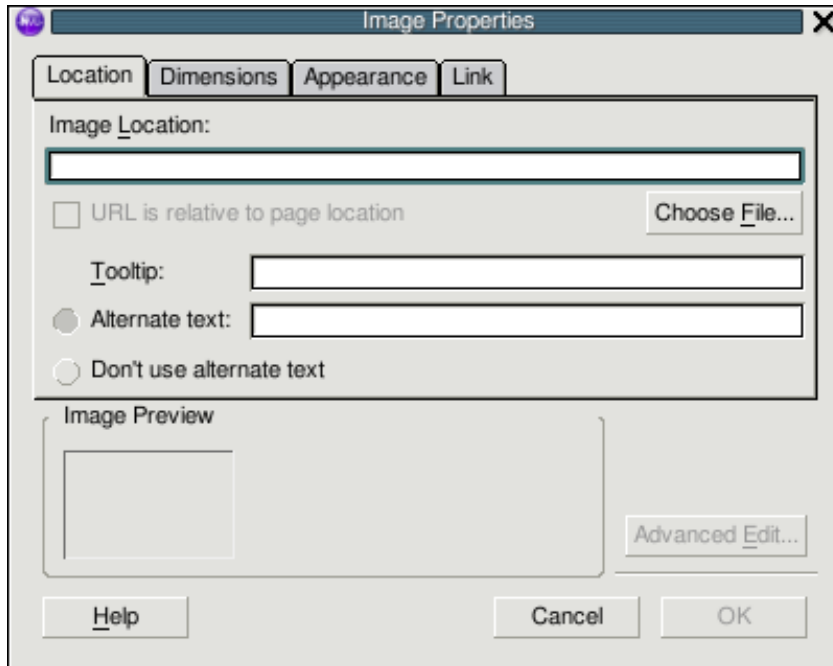
If you want to change the hyperlink to point elsewhere, you can right click on a link, then select “link properties”. This pops up the same dialog box, and you can change the link.

To turn a hyperlink back to regular text, select the entire hyperlink (or the portion to change), right click, then select “remove links”.

### 6.3.4 Including images

It is common for HTML documents to include images. Remember, however, an image in an HTML document is a file that is independent of the HTML document. In other words, you need to make sure an image is available online (either in your web space or elsewhere).

To add an image, click the Image tool, which opens a dialog box.



You need to indicate the location of the image file. This gets a little tricky. If the image file is somewhere else online, it is simple. You just need to enter the URL of the image file, such as “[http://images.photo.net/green\\_hills.jpg](http://images.photo.net/green_hills.jpg)”. However, if you will be supplying the image files, you need to do the following:

- use a folder to contain your HTML documents and images
- refer to the image file

We’ll talk about this more when we talk about site management.

Don’t forget to type the “alternate text” of your image.

Note that an image can serve as a hyperlink, too. Click the “Link” tab of the Image Property dialog box, and you can enter the URL of the linked document.

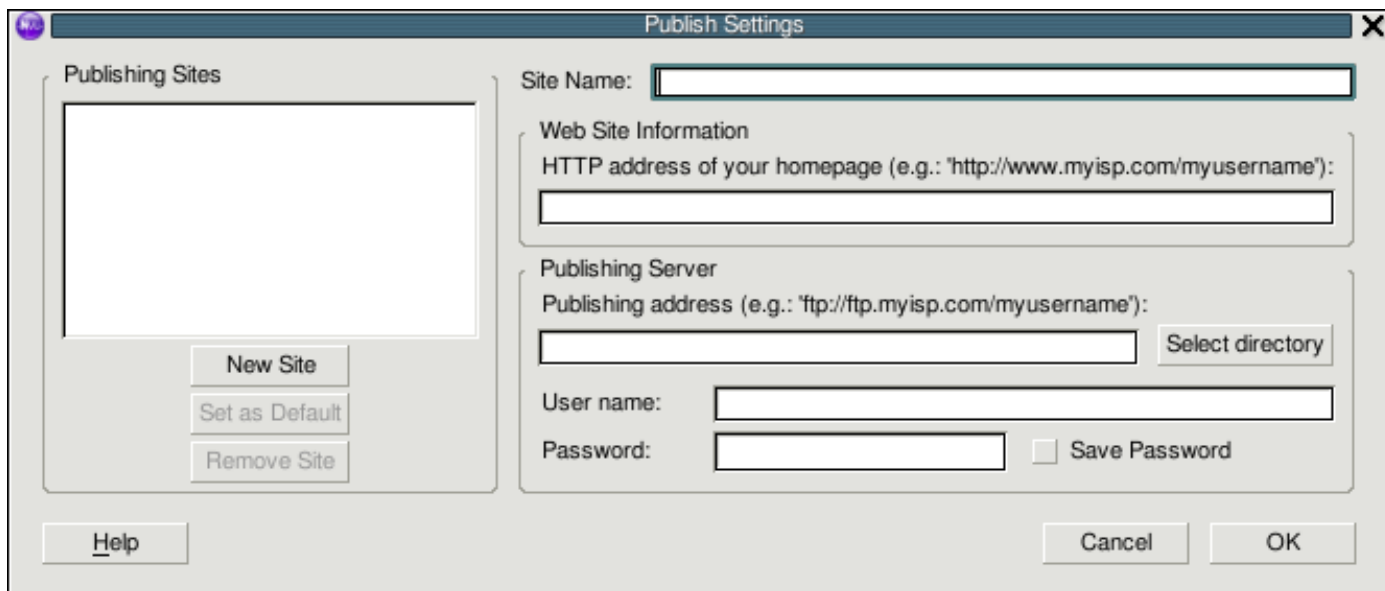
## 6.4 Site Management

Nvu has a site manager that makes managing a web site with multiple pages a little easier. This section talks about site management, as well as including image files in HTML documents.

### 6.4.1 Getting started

First, you need to use a folder to contain all the files you plan to upload to your web site. This is a task you need to do outside of Nvu with the operating system. Let’s call your folder “testsite” in our examples.

Next, click “Edit sites” in the Nvu Site Manager pane. A dialog box appears.



The “Site Name” can be anything you choose, it is merely a label to remind you what this site is about.

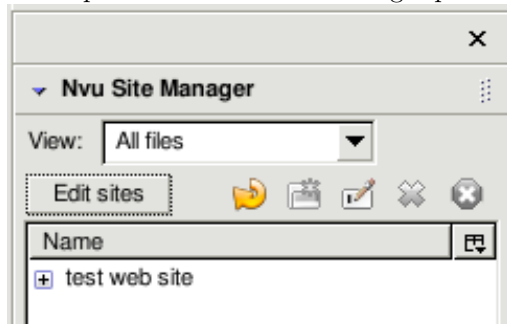
The “Web Site Information” portion is more important. You should use the URL of your web space.

The Publishing Server portion is also important. For our example, we just need to select the folder that is on the same computer. Click “Select directory” so you can navigate to the folder you want to use.

Unless you are using an FTP server to store your files, there is no need to enter “User name” or “Password”.

Click “OK” when you are all done.

This updates the Nvu Site Manager pane. If you see a “+” symbol, you can click on it to expand the item.



## 6.4.2 Image files

Unless you have many image files, it is okay to store image files along with your HTML documents. You can copy image files to the folder that represents your website using Windows Explorer or equivalent tools of your operating system.

Once you copy files to the web site folder, you can collapse the folder and re-expand it. It should now show the image files you copied earlier.

Once you have image files in the same folder as an HTML document, you can refer to the image file (when you insert the image) using the file name alone.

To change properties of an image in the HTML document, right click on the image, then select “Image Properties”. Note that you can change the displayed size of an image using the “Dimensions” tab. This makes it possible to make a huge image file appear smaller.

You can also resize the displayed dimensions using the resize tool.

Use the “Align Text to Image” drop down box to control how text should flow around the image.

## 6.4.3 Uploading files

Most free web spaces do not allow FTP. Instead, you need to use an online web interface for file transfer. Be sure to upload everything in your folder, including image files!

# Chapter 7

## Week 8: FTP, RSS and Forums

We are nearing the end of the 9 weeks, so just hang on!

### 7.1 FTP

FTP (file transfer protocol) is also the name of programs that can be used to transfer files. All versions of Windows since Windows 95 includes ftp as a program.

#### 7.1.1 Prerequisites

In order to use FTP to transfer files between computer A and computer B, certain requirements must be met first.

- A user must have an account on the server machine.
- The server must have FTP enabled.
- The client must have an FTP client program.

The command line version of FTP works as follows:

- type `ftp` on a command line
- supply the username and password when prompted
- once you are logged in,
  - use `put` to send a file to the server
  - use `get` to receive a file from the server
  - use `binary` to switch to binary file mode
  - use `text` to switch to text file mode
  - use `mget` to receive multiple files from the server
  - use `mput` to send multiple files to the server
- to log out, use the command `quit`

(Illustrate in class.)

In most cases, a PC is a client, making a connection to a remote computer that is the server.

### 7.1.2 GUI FTP clients

Although the command line `ftp` program is functional, it requires quite a bit of command line expertise. As a result, most users prefer GUI based FTP clients.

Interestingly, most web browsers (such as Internet Explorer, Firefox and Opera) are FTP clients. To connect to an FTP server, for example, `ftp.remotemachine.com`, enter the following in the address bar:

```
ftp://ftp.remotemachine.com
```

The browser will, then, prompt for the username and password. Once logged in, you can use drag-and-drop (from a Windows Explorer or other GUI file system tool) to transfer files and folders.

Some programs are specialized for FTP. `WSFTP` is the most recognized one, although it is not free. For a free program that provides easy file transfer operations, try `filezilla` from <http://filezilla.sf.net>. Note that Filezilla is available only for Windows platforms.

### 7.1.3 But, don't use FTP!

Yes, now I tell you.

The reason is that FTP is a very insecure protocol. Usernames and passwords are transmitted in plain text. This means any eavdropper can easily grab usernames and passwords.

For more secure file transfer between a PC and a remote server, use SSH instead. Although this can be done in a command line, it is better to use a GUI tool. `WinSCP` from <http://winscp.net> is an excellent free program for transferring files using SSH.

## 7.2 Forums

Since you are already using Moodle, forums require little introduction. It should be noted that most forums are web hosted. This means you don't need any special software (other than a web browser) to access forums.

The characteristics of a forum are as follows:

- Pull technology. Subscribers need to check back periodically for new articles.
- Public. Articles are visible to all subscribers.
- Discussion capable. Subscribers can follow up articles or feedback of each other.
- Nested display. Follow up articles are listed in a “nested” fashion so a reader can easily tell which article is in response to which other one.

Note that online forums have, for the most part, taken the place of newsgroups local to a community. If you access newsgroups via <http://news.google.com>, then the “global” newsgroups are access as online forums.

Despite the fact that forums are “pull” type technology, many forum engines (including that of Moodle) offer email notifications. Such notfications come in roughly three types.

- Per-posting notification. A reader gets a email notification for *each* new posting. This type of notification is suitable for relatively inactive forums.
- Per-day full notification. A reader gets a email that lists all the articles posted in a day. This type of notification is, again, only suitable for relatively inactive forums.
- Per-day digest. A reader gets a email that is a summary (topic title and author only) of all the articles posted in a day. This type of notification is suitable for busy forums so that a reader does not get flooded by the bulk of all messages.

Email notifications effectively convert forums from a “pull” technology to a “push” technology.

## 7.3 RSS

Although forums and newsgroups are great resources, it is cumbersome to have to check back periodically for new messages. RSS has many “real” names, but the most common understanding is “Really Simple Syndication”.

RSS (often called RSS feeds) enables an RSS subscriber periodically check and download links to web published articles. There are many RSS subscription programs. Mozilla Firefox and Mozilla Thunderbird are two programs that can subscribe to RSS feeds.

The content of RSS feeds varies. It is often used in news sites and other web sites that publish new articles periodically. RSS is a push type technology. It is also a one-way technology. Discussions are not permitted by RSS feeds.



# Chapter 8

## Week 8, Part II

After all the discussion in previous chapters, this chapter serves as a quick summary of the course. It also points to several directions that the Internet may take in the near future.

### 8.1 At Work

#### 8.1.1 Email

Email is, by far, the most common use of the Internet at work. Businesses, large and small, are now using email as an inexpensive alternative for written communication. Despite the availability of chat rooms and other more “real-time” technologies, email remain as a mainstream communication medium.

The main advantage of email is flexibility. A receiver does not have to be online when a message is sent. This means there is great flexibility as far as when a recipient reads a message. This property is important at the work place.

Although a lot of work places still use conventional email clients (such as Thunderbird and Outlook) for checking email, an increasing number of companies are switching to web-based interfaces. A web-based interface is more flexible, making it possible for employees to check company message from home without special software or configuration on home computers.

The most important threat posed by email include trojan horses, spam, phishing and spoofed (faked) messages. We will discuss possible solutions in another section.

#### 8.1.2 Chatting

Chat rooms/chatting are used increasingly to replace/enhance over-the-phone support services. The driving force is mainly cost. It is less expensive and far more flexible to distribute chatting to remote corners of the world (compared to phone calls).

Chat rooms are also used by open source and other non-profit organizations for self-supported groups. Gurus and experts can participate in chat room voluntarily to answer questions from beginners. Such chat rooms are already readily available for many Linux distributions and open source programs.

#### 8.1.3 Advertisement/Sales

Many companies already have web presense for the purposes of advertisement and sales. A web-based ordering system is inexpensive in the long run because it is automated. It is also valuable from the perspective of customers because an automated system is often more accurate.

### 8.2 At Home

With increasingly inexpensive broadband services, the Internet is accessible to many homes.

#### 8.2.1 Email

Email is an important application of the Internet. For the same reasons at work, email is a basic but important capability.

### 8.2.2 Chatting

Online chatting is very popular, especially for the younger (I-cannot-wait-for-a-reply) generation.

### 8.2.3 Multimedia

This is only becoming more popular now that we have inexpensive broadband (high bandwidth) access. Individuals can now purchase and download music. Many TV stations record and post segments of news cast. Radio stations also stream programs using the Internet.

Multimedia distribution will become more important in the near future, and it will help drive the demand for more bandwidth. See the section on DRM and the future of multimedia distribution over the Internet.

## 8.3 Malware

The popularity of computers and the Internet makes malware an increasing threat. Malware refers to any program that is written with a malicious intent. It includes viruses, worms, spyware and other categories of software.

In order for malware to cause problems, one or more of the following must be true:

- there are intrinsic weaknesses in the operating system or application programs
- an end user can be tricked to open files that contain malware

Currently, companies focus on passive defense against malware. Virus scanners, for example, are passive defense because it is only effective when a virus reaches a computer. A virus checker should be the last line of defense against viruses and worms. The same applies to anti-spyware programs.

The proper response to malware includes intrinsically secure software (operating system and applications) as well as better end user education. There has been some progress in both directions. See the trusted computing section for more information.

## 8.4 DRM

DRM (digital rights management) is a combination of policies, software and hardware that ensures copyrighted material cannot be copied without proper licenses.

DRM is pushed by the music and movie industry, as well as big software publishers. On the surface, it is a legitimate defense against illegal copying/replication of copyrighted material (music, movies, etc.). However, DRM can also go to the extreme and severely limit what a consumer can do.

The most important question is: what are you purchasing? When you purchase a music CD, you are not purchasing the hardware and also the right to do whatever you want with it. What you are really purchasing is a license to play the music *distributed* by the CD according to a set of rules set by the RIAA (recording industry association of America). What rules? Well, you cannot play a CD in public. You cannot make copies of a CD, and etc.

DRM is not in the near future, it is here and now! Some CDs are already DRM protected so that consumers cannot extract the tracks into MP3 files. While this prevents “pirates” from distributing tracks of a music CD over the Internet to others, it also prevents the rightful “owner” of a CD from copying a track to an MP3 player for jogging.

DRM and the Internet are closely related because the Internet is an important medium for distributing material that should be protected by DRM (according to the music and movie industries).

## 8.5 Trusted Computing

The Trusted Computing Group (TCG) is a consortium of companies from the hardware, software, music and movie industries. The Trusted Computer Initiative (TCI) aims to make sure a user’s computer can be trusted (that it cannot violate copyright laws and etc.).

TCI has always been rationalized as a way to make the Internet safer for end-users. In reality, however, there are other means to improve the security of the Internet. For example, software publishers can release source code for security auditing. Email messages can be encrypted and/or signed. Even online chatting sessions can be encrypted and/or signed to improve computer/Internet security.

The TCG has already proposed to change networking protocols to include hardware based signing. This makes every packet of information traceable down to a single computer. With a combined effort from hardware manufacturers, software publishers and ISPs, every packet can be tracked to a machine and probably its ownership. While this helps to deter criminals from using computers for criminal activities, it also have lots of implications regarding privacy for law abiding citizens. It even permits easy control of network traffic based on the originating and destination machines.

## 8.6 Web Applications and Thin Clients

Imagine the following in the near future:

- A diskless PC connects to a HDTV, and costs \$200 or less.
  - Something like this already exists, check out <http://www.projectblackdog.com>
  - In large scale production, it may cost about \$100 for a thin client PC.
- Broadband access is integral to telephone service.
  - Surewest already distributes services over fiber optics. The subscription fee is artificially priced to be comparable with DSL and cable modem.
  - Verizon already distributes Internet access wirelessly at 300kbps (comparable to DSL).
- Application programs and storage space reside on remote servers.
  - Google and Sun are already considering making an online office suite.
  - Inexpensive hardware and open source solutions makes this possible even for small organizations.

There are many advantages. For example, you do not need to purchase, install and upgrade software packages. You can, instead, subscribe to services that include applications and storage space. You don't have to worry about backing up your files, either, because the storage resides in a remote server that is backed up more often than you eat meals.

Because of the availability of open source software and inexpensive hardware, the pricing of web-based application and storage has to be very competitive. I think it will cost far less to subscribe to a web-based application and storage solution than to purchase and upgrade software on a local computer.

Regarding the privacy of your documents, there are several possible solutions.

- Download application code from the remote computer, but execute the code locally on a thin-client computer
- Before sending documents to remote storage servers, encrypt the document first. Keep the private local to the thin-client computer or a key card.
- Install a third party security hardware filter to ensure no unencrypted documents/content is transmitted.



# Index

- BitTorrent, 15
- cable modem, 17
- client, 13
- digital rights management, 44
- DNS
  - queries, 9
  - root servers, 9
- DNS:caching, 9
- domain
  - top-level, 9
- domain name
  - registration, 9
- DRM, 44
- DSL, 16
- Firefox, 21
- Firefox extensions
  - Adblock, 22
  - Mozilla Calendar, 22
- Firefox extensions:ReloadEvery, 22
- firewall, 8
- FTP, 14
- gateway, 9
  - residential, 9
- hostname
  - virtual, 9
- HTML
  - distribution, 13
- HTTP, 13, 20
  - client, 13
  - server, 13
- hyperlink, 20
- IMAP, 14
- Internet Service Provider, 7
- IP, 8
  - address, 8
- IRC, 15
- ISP, 7
- LAN, 8
- MAC Address, 8
- modem, 16
  - bank, 16
  - speed, 16
- network
  - interface, 8
- NIC, 8
- NNTP, 14
- OSI, 8
- P2P, 15
- Peer to peer, 15
- POP3, 14
- port, 13
  - forwarding, 15
- Really Simple Syndication, 41
- router, 8
- RSS, 41
- server, 13
- service, 13
- SMTP, 14
- spyware, 19
- SSH, 15
- SSL, 15
- subdomain, 9
- switch, 8
- TCP/IP, 8
- telnet, 14
- upload, 31
- URI, 20
- URL, 20
- virus, 19
- WAN, 8
- web browser, 13
- World Wide Web, 19
- WWW, 19