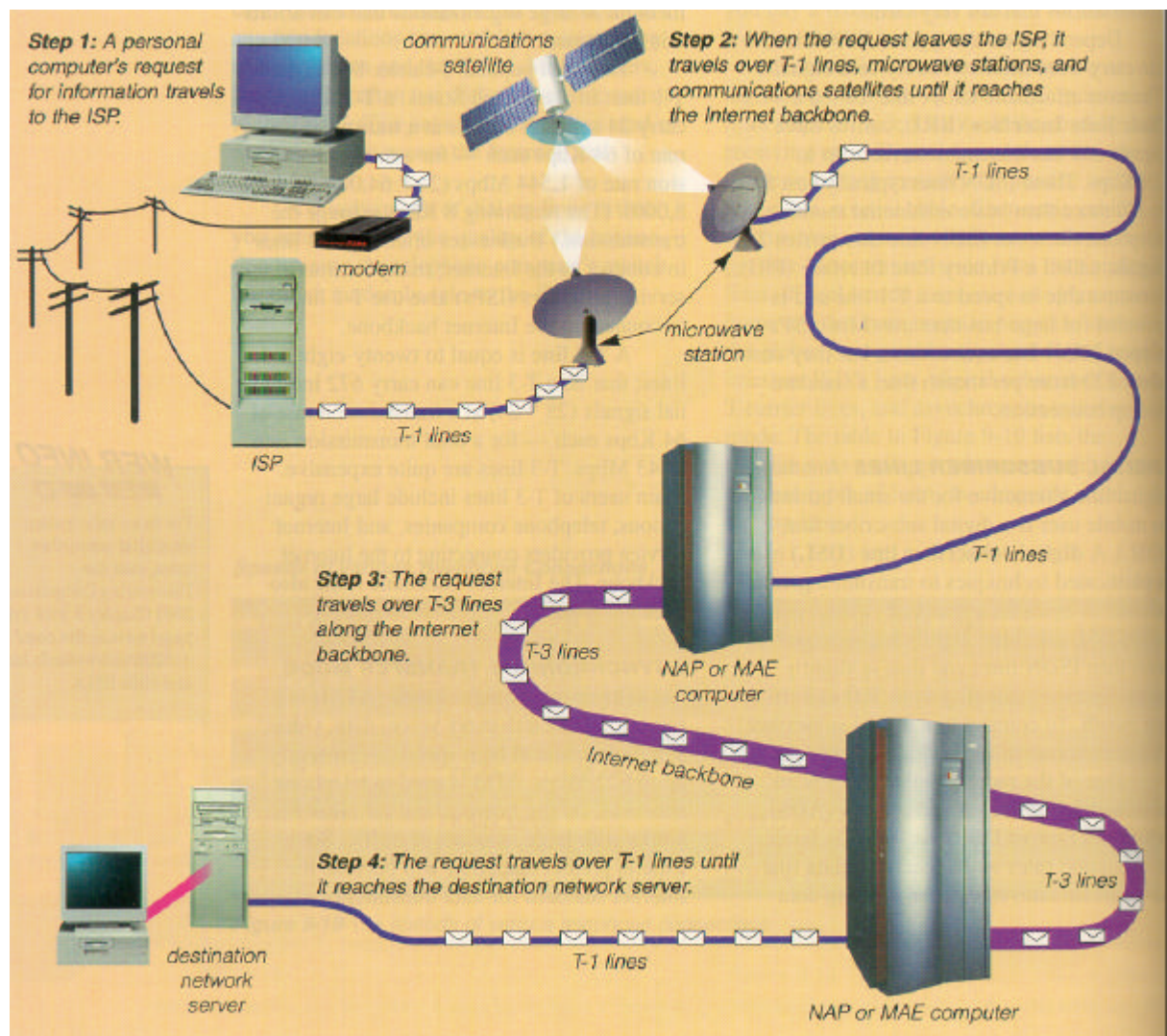
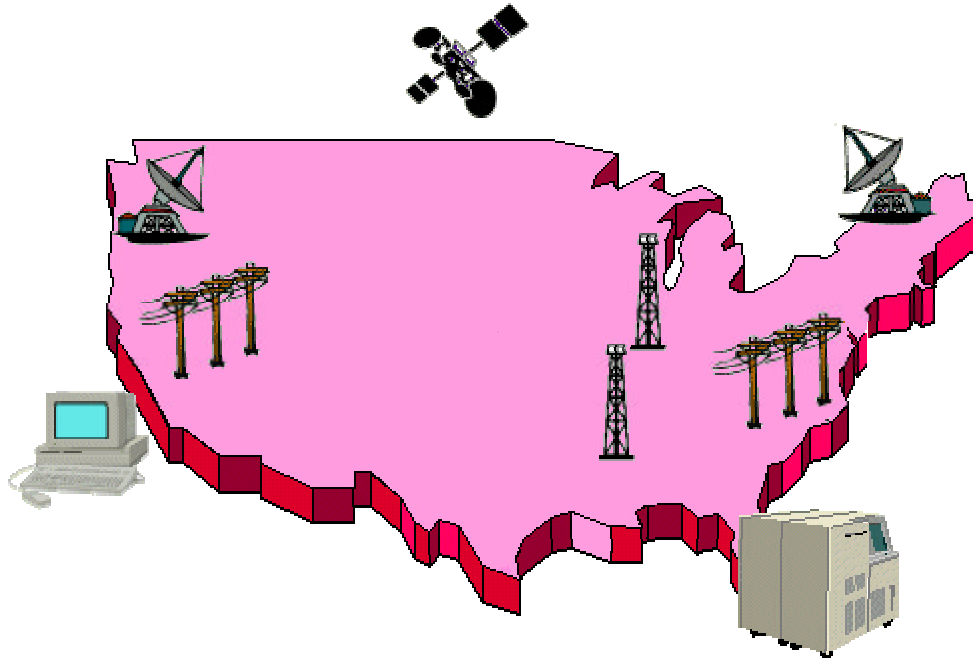


1. Introduction to the Internet

1.1 What is the Internet?

- ICQ, Internet Explorer, Email, Netscape?
- The Internet is a computer network allowing people to share information?
- The Internet is a worldwide network allowing people in different locations of the world to communicate?
- The Internet is a worldwide resource centre?





1.2 Where does Internet come from?

At 1960s, the Advanced Research Projects Agency of Defense Department in USA sponsored a project named the ARPANET. The goal is to build a network that could maintain itself under adverse condition so that it can carry military and government information during wars.

After several years working, they found that there was no single network that can serve for this purpose. They determined that it is far better to *develop technology* for connecting different types of networks to form a single large system. This is the fundamental of the “inter-networking” concept.

Under this project, a global backbone of computer networks has been formed which we now call it the Internet. The name “Internet” is derived from the work “**inter-networking**” for it is a collection of tens of thousands of networks.

<http://www.pbs.org/internet/timeline>

1.3 Why Computers need to be Networked?

A **network** is a collection of computers and devices connected together via communications media and devices such as cables, telephone lines, modems, or other means. When your computer is connected to a network, you are said to be online.

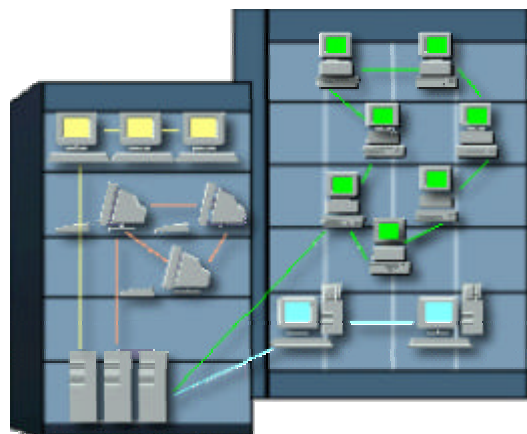
Computers are networked together so users can share resources, such as hardware devices, software programs, data and information. Sharing resources saves time and money. For example, instead of purchasing one printer for every computer in a company, the firm can connect a single printer and all computers via a network; the network enables all of the computers to access the same printer.

2. Types of Computer Networks

Network exists in a range of sizes, from a small network connecting two computers to a global network such as the Internet, which connects millions of computers around the world. Networks also can connect computers of all sizes, from handheld computers to supercomputers. Two types of networks more widely used in business are local area networks and wide area networks.

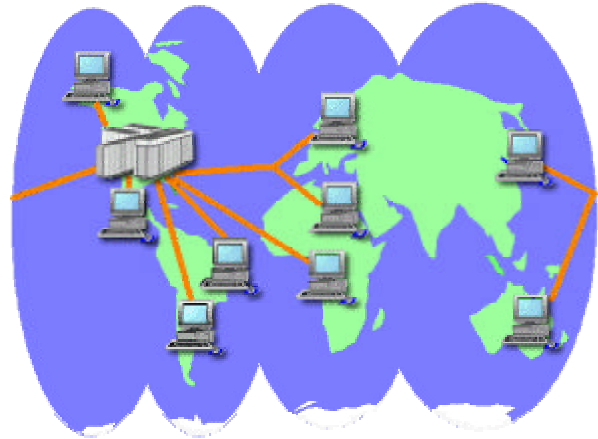
2.1 Local Area Network (LAN)

A **local area network** is a network that connects computers in a limited geographical area, such as a school computer laboratory, an office, a department or group of buildings. The maximum length of an individual LAN is usually limited to about 1 to 2 km. A LAN consists of a communications channel, networked computers and devices, a network interface card, and a network operating system.



2.2 Wide Area Network (WAN)

A **wide area network** is a network that covers a large geographical area (such as a city or country) using a communications channel that combines telephone lines, microwave, satellites or other transmission media. Today, a WAN typically consists of two or more LANs connected by routers that ensure that data, instructions and information are delivered to the correct destination. Computers often are connected to a wide area network via public networks such as the telephone system or by dedicated lines or satellites. The Internet is the world's largest wide area network.

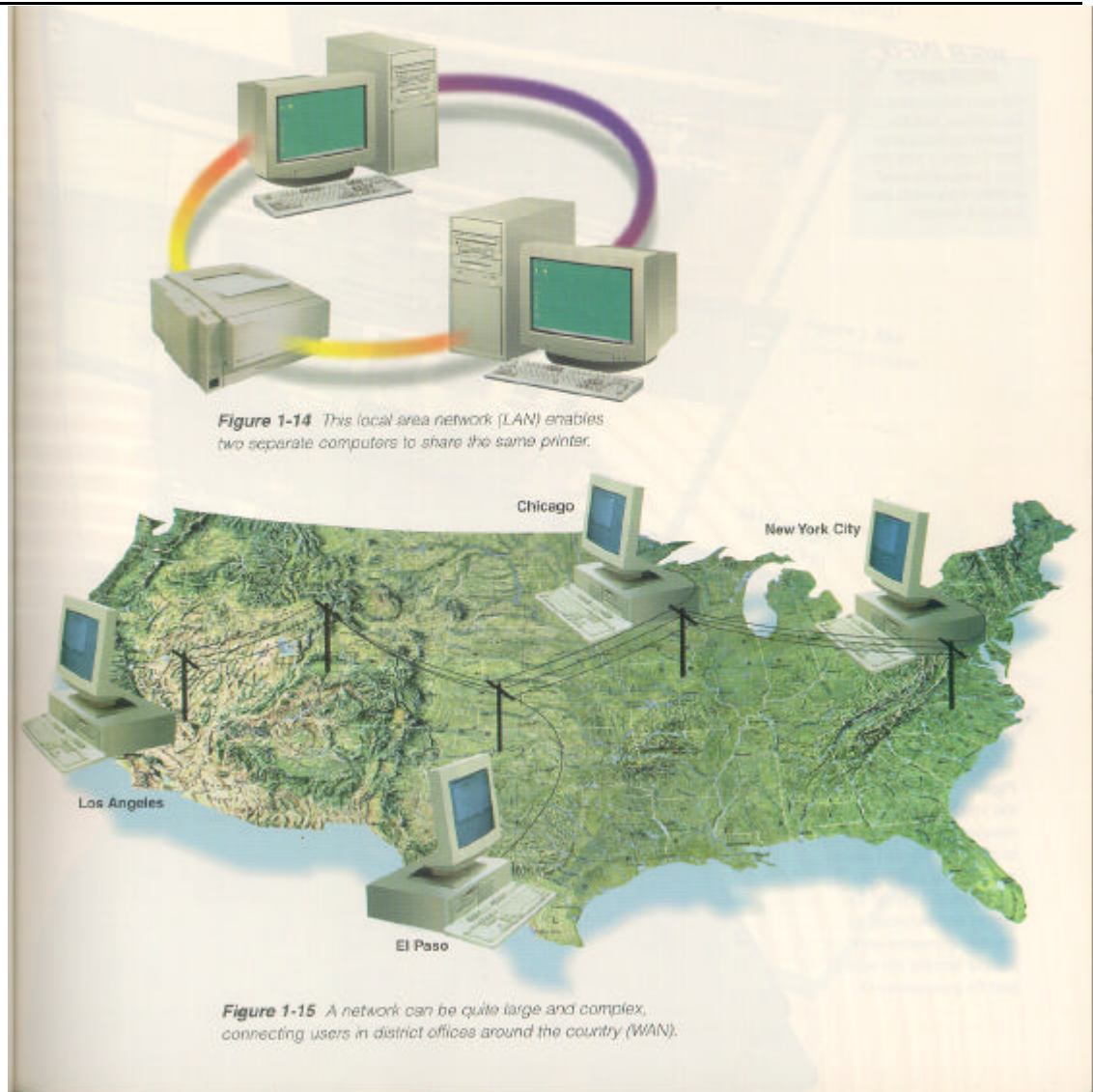


2.3 Metropolitan Area Network (MAN)

A **metropolitan area network** is a backbone network that connects LAN in a metropolitan area such as a city or town and handles the bulk of communications activity, or traffic, across that region. MANs also frequently provide a shared connection to other networks using a link to a WAN.

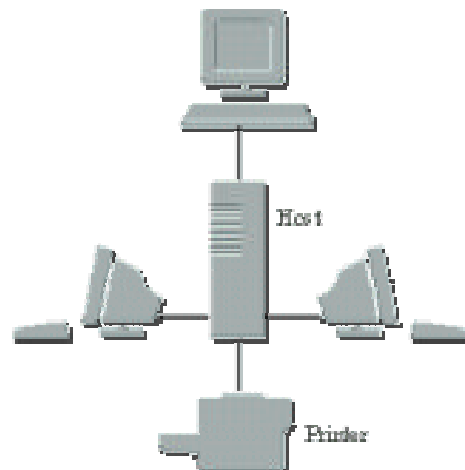
3. Network Topologies

The configuration or physical arrangement of the devices in a communications network is called the **network topology** or **network architecture**. Network architecture is similar to the architecture of a building. Like a blueprint which shows the physical layout of the building, a drawing of a network architecture provides a pictorial representation of the physical layout of the network. Three commonly used network topologies are star, bus and ring network.



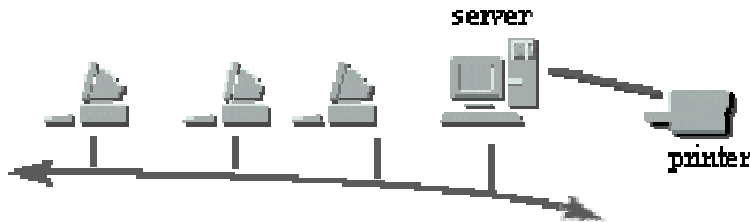
3.1 Star Topology

In a star network, computers are connected to a central computer which provides the major resources. The central computer that provides a common connection point for devices in the network is called the **hub**. All data that transfers from one computer to another passes through the hub.



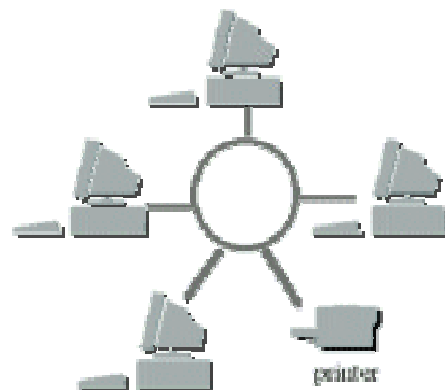
3.2 Bus Topology

A bus network consists of a **single central cable**, to which all computers and other devices connect. The **bus** is the physical cable that connects the computers and other devices. The bus in a bus network can transmit data, instructions and information in both directions. Only one device, such as a computer, can transfer items at one time, however.



3.3 Ring Topology

A ring network is designed so a cable forms closed ring or loop, with all computers and devices arranged along the ring. Data transmitted on a ring network travels from device to device around the entire ring, in one direction. If a device on a ring network fails, all devices before the failed device are unaffected, but those after the failed device cannot function.



4. Some Basic Concepts

4.1 Clients and Servers

As you have seen, the Internet is an extremely large WAN. Then, how can the network provide facilities to the users and how can the users get access to the facilities of the network? The answer is, these are done through using two types of programs, **clients** and **servers**, which are populated in the Internet. The Internet is

constructed in such a way that users can use client programs to talk to server programs that provide facilities.



As an end-user, you only need to learn to use the corresponding client program to access a particular kind of facilities.

Question: What is the difference between file-server model and client-server model?

4.2 Communication Protocols- TCP/IP

Internet is made up of many different types of computer networks. They all follow the same standard **protocols**. A protocol is a set of rules governing how things are to be done.

In the Internet, a collection of more than one hundred protocols are used and are collectively call TCP/IP. The name TCP/IP comes from the two most important protocols in inter-networking: *Transmission Control Protocol (TCP)* and *Internet Protocol (IP)*.

IP is regarding the rules of transmitting raw data (packets) from host to host. TCP sets rules on managing the correctness and the flow of the data. TCP and IP together set up rules for providing error-free channels from host to host.

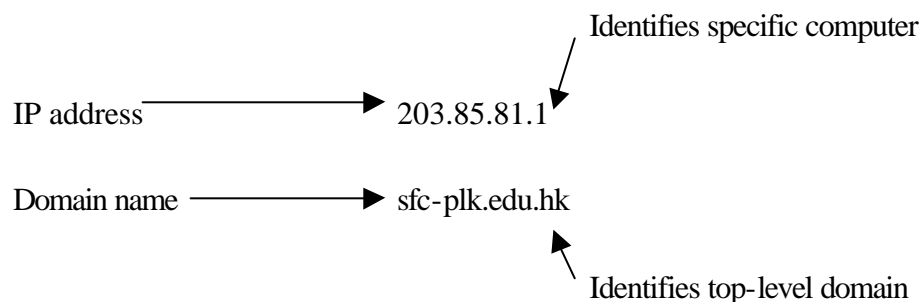
TCP/IP was developed in 1973 for use in the ARPANET. Since then, TCP/IP has developed into a protocol used by both WANs and LANs. In 1983, TCP/IP was adopted as the Internet standard; all hosts on the Internet are required to use TCP/IP.

4.3 Internet Addressing

The Internet relies on an addressing system much like the postal service to send data to a computer at specific destination. Each computer location on the Internet has a numeric address called an **IP (Internet Protocol) address**. The IP address consists of four groups of numbers, each separated by a period (such as 203.85.81.1).

Because these all-numeric IP addresses are difficult to remember and use, the Internet supports the use of a text name that represents one or more IP addresses. The text version of an IP address is called a **domain name** (such as www.sfc-plk.edu.hk).

The top-level domain is the rightmost sub-domain that gives us the most general information about the computer. Basically, there are two types of top-level domains. The older style is **Organizational Top-Level Domains**. The latest one is **Geographical Top-Level Domains**.



The part of the TCP/IP services that keeps track of addresses in the Internet the **Domain Name System (DNS)**. It performs, on behalf of users, the translation of domain names to and from IP addresses. Certainly, this is done automatically and even without letting you know.

<http://www.cisco.com/warp/public/787/17.html>

4.4 UNIX

UNIX is a family of operating system that are the most widely used ones in the computer world. Nearly all types of computers can run UNIX in which different variations of UNIX are designed for different computers according to the capacities.

Since most of the computers in the Internet use UNIX, many people consider the Internet and the UNIX not separable.

Learning UNIX is not a must for using the Internet but certainly makes you a more competent user.

5. Internet Resources

5.1 World Wide Web

Today, one of the more popular uses for communications is to access the World Wide Web. Recall that the **World Wide Web**, **WWW** or **Web** consists of a worldwide collection of electronic documents (Web pages) that have built-in links to other related documents. The major reasons why the WWW is so popular are that it is very easy to use and it is easy to share your own web information with all others over the Internet.

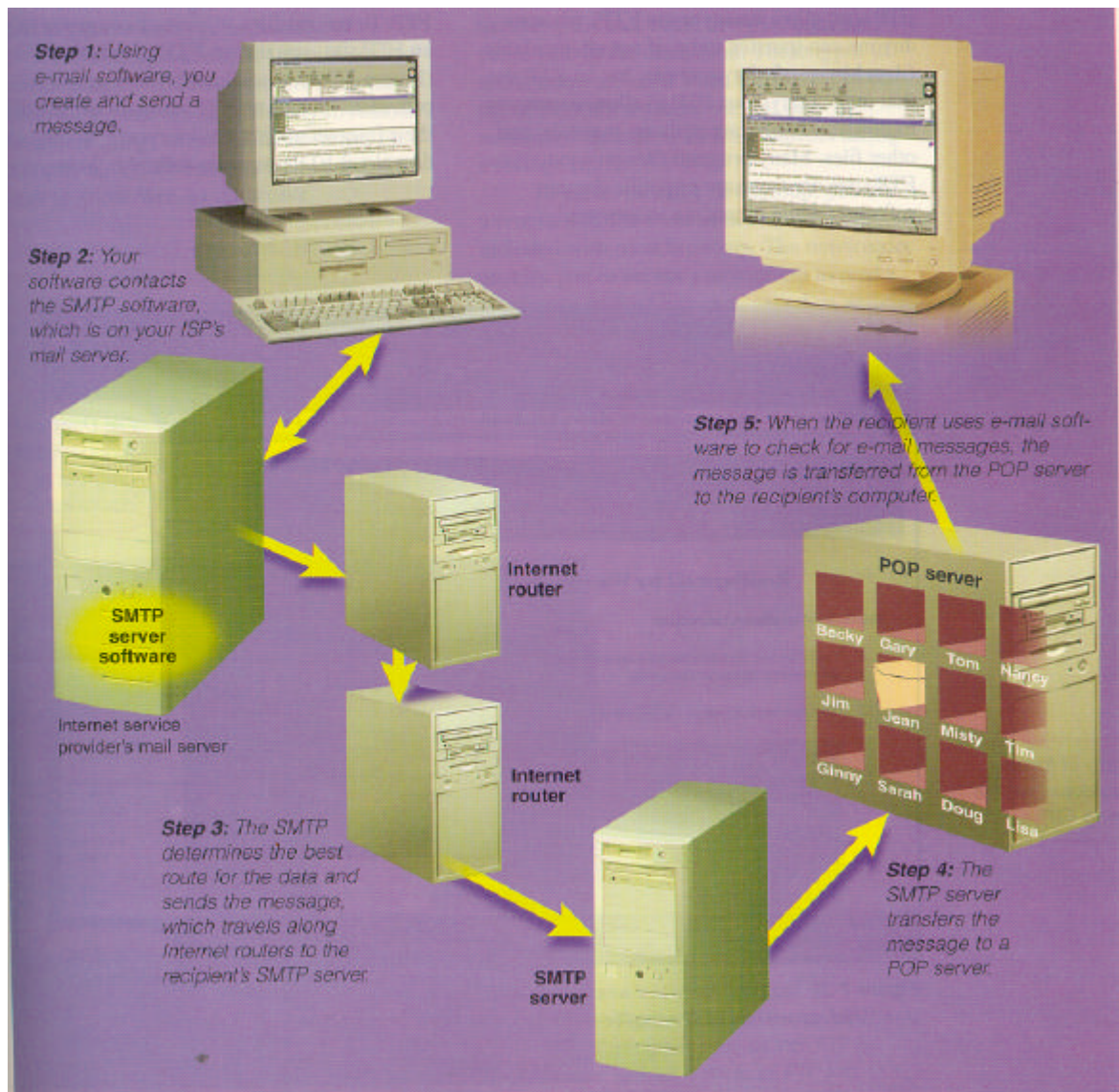
Each Web page on a Web site has a unique address, called a **Uniform Resource Locator (URL)**. A URL consists of a protocol, a domain name and sometimes the path to a specific Web page or location in a Web page. Most Web page URLs begin with **http://**, which stands for **hypertext transfer protocol**.

5.2 Newsgroups

A newsgroup is an online area in which users conduct written discussions about a particular subject. To participate in a discussion, a user sends a message to the newsgroup, and other users in the newsgroup read and reply to the message. The entire collection of Internet newsgroups is called **Usenet**, which contains thousands of newsgroups on a multitude of topics. Some major topic areas include news, recreation, business, science and computers.

5.3 Electronic Mail

Electronic mail, or simply called as **email**, is one of the three most popular Internet services. It is the exchange of text messages and computer files transmitted via a communications network such as a LAN or the Internet. Communications devices, such as modems, transfer the email messages to and from computers or terminals on the same network or a separate network.



5.4 Bulletin Board System

An electronic bulletin board system (BBS) is a computer that maintains a centralized collection of electronic messages. You access a BBS by using your computer to connect to the main BBS computer. Once you connect to the BBS, you can add or delete messages, read existing messages, or upload and download software. Others BBSs function as electronic meeting rooms for special-interest groups to share information. Because you can use the Internet to access many of the same services, however, BBS usage is declining.

5.5 Chat Rooms

A **chat** is a real-time typed conversation that takes place on a computer. **Real-time** means that you and the people with whom you are conversing must be online at the same time.

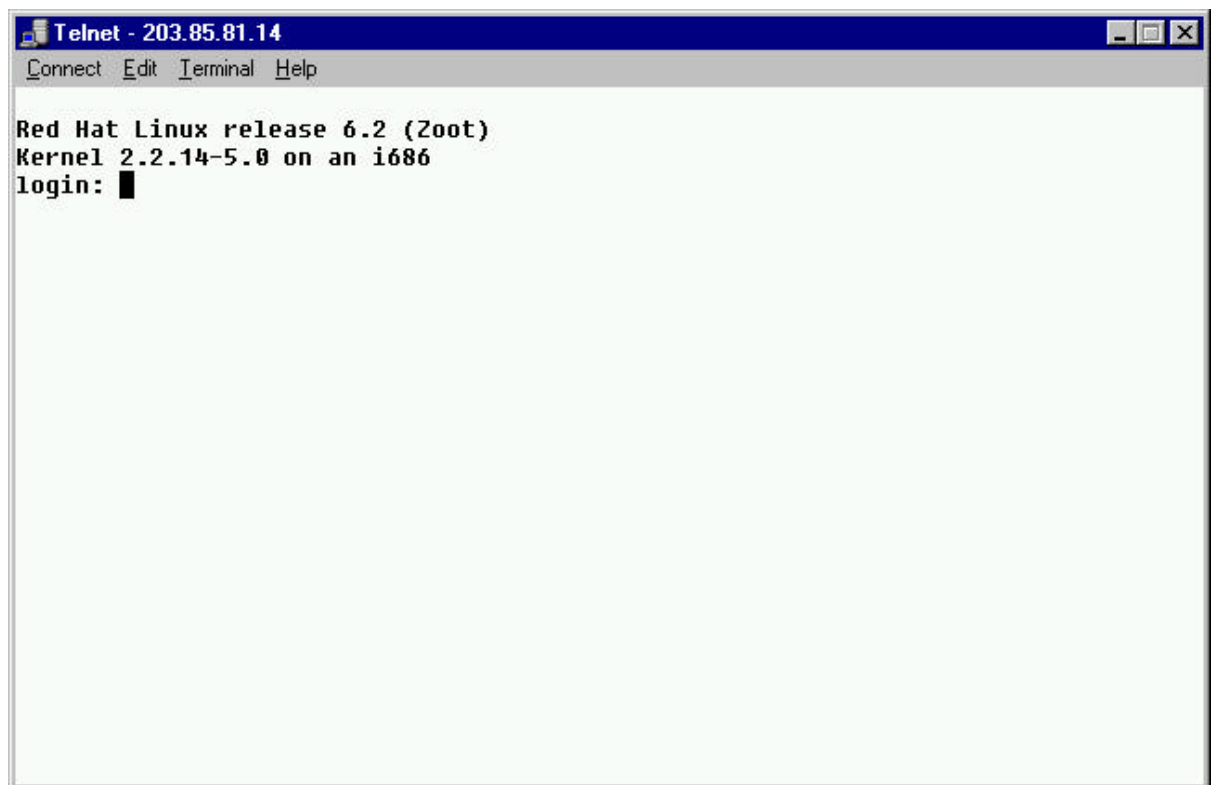
A chat room refers to the communications media, or channel, that permits users to chat with each other. Anyone on the channel can participate in the conversation, which usually is specific to a particular topic. Each discussion is assigned a different channel.

UNIX talk, IRC and nowadays ICQ are examples that provide online chat rooms. The “talk” command in UNIX environment was the orientation of yesterday’s IRC and nowadays ICQ. Back to several years before, people use UNIX talk to make online chat with others.

IRC (**I**nternet **R**elay **C**hat) is a service that enables an Internet user to participate in a conversation online in real time with other users. IRC was invented in 1988.

5.6 Telnet

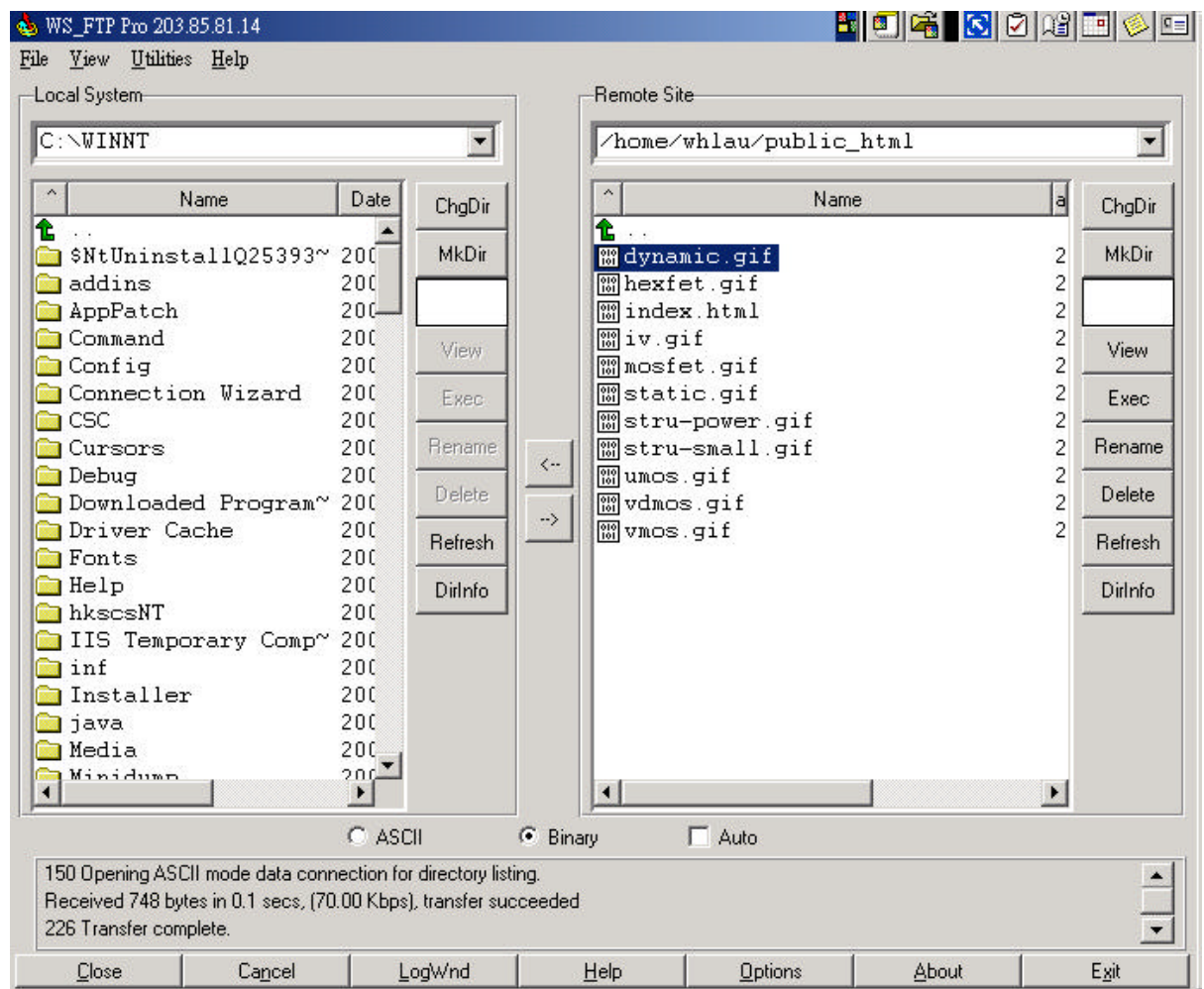
Telnet is a program or command that enables you to connect to a remote computer on the Internet. To make a Telnet connection to the remote computer, you enter a user name and password. Once connected, your computer acts like a terminal directly linked to the remote computer. Telnet access to many remote computers is free, while others are accessed for a fee. For convenience, some remote computers provide the Telnet program.



5.7 FTP

FTP (file transfer protocol) is an Internet standard that allows you to exchange files with other computers on the Internet. For example, if you click a link on a Web page that begins to download a file to your hard disk, you probably are using FTP.

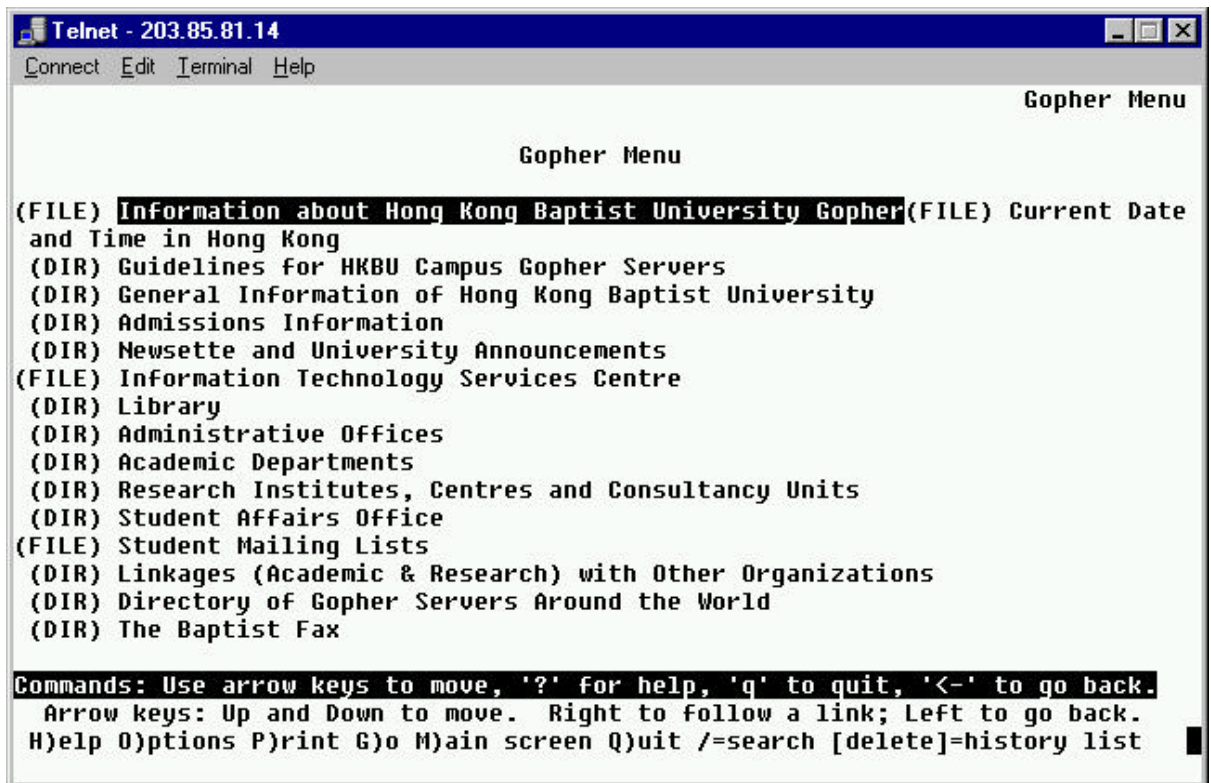
An **FTP server** is a computer that allows users to upload and download files using FTP. An **FTP site** is a collection of files including text, graphics, audio, video and program files that reside on an FTP server. Some FTP sites limit file transfers to individuals who have authorized accounts on the FTP server. Many FTP sites allow **anonymous login**.



5.8 Gopher

The ***gopher*** system, similar to the WWW, provides information to Internet users. All the information that is available via the gopher system is called a gopherspace. The way of getting information in the gopherspace is also the same as using the WWW: a user uses a client to connect to gopher servers all over the world to get the desired information.

The difference between the gopherspace and the WWW lies in the organization in them. In the gopherspace, the information is presented as a series of simple menus. Each gopher server has a main menu and a series of submenus. Your client will fetch the information to you when you select a menu item from a gopher server.

A screenshot of a Telnet window titled "Telnet - 203.85.81.14". The window has a menu bar with "Connect", "Edit", "Terminal", and "Help". The main content area displays a "Gopher Menu" with a list of items. The first item, "(FILE) Information about Hong Kong Baptist University Gopher", is highlighted. Below the list, there are instructions on how to use the menu: "Commands: Use arrow keys to move, '?' for help, 'q' to quit, '<-' to go back. Arrow keys: Up and Down to move. Right to follow a link; Left to go back. H)elp O)ptions P)rint G)o M)ain screen Q)uit /=search [delete]=history list".

```
Telnet - 203.85.81.14
Connect Edit Terminal Help

Gopher Menu

Gopher Menu

(FILE) Information about Hong Kong Baptist University Gopher
(DIR) Guidelines for HKBU Campus Gopher Servers
(DIR) General Information of Hong Kong Baptist University
(DIR) Admissions Information
(DIR) Newsette and University Announcements
(FILE) Information Technology Services Centre
(DIR) Library
(DIR) Administrative Offices
(DIR) Academic Departments
(DIR) Research Institutes, Centres and Consultancy Units
(DIR) Student Affairs Office
(FILE) Student Mailing Lists
(DIR) Linkages (Academic & Research) with Other Organizations
(DIR) Directory of Gopher Servers Around the World
(DIR) The Baptist Fax

Commands: Use arrow keys to move, '?' for help, 'q' to quit, '<-' to go back.
Arrow keys: Up and Down to move. Right to follow a link; Left to go back.
H)elp O)ptions P)rint G)o M)ain screen Q)uit /=search [delete]=history list
```


6. Electronic Mail (Email)

Email is the transmission of messages and files via a computer network. Email was one of the original services on the Internet, enabling scientists and researchers working on government-sponsored projects to communicate with colleagues at other locations. Today, email quickly is becoming a primary communication method for both personal and business use.

6.1 Simple Mail Transfer Protocol (SMTP)

The delivery of email in the Internet is standardized by the *Simple Mail Transfer Protocol (SMTP)* which describes how emails are to be delivered from one computer on the Internet to another one. Again, SMTP is part of the TCP/IP family of protocol.

6.2 Email Program (Email Client)

Using an email client, you can create, send, receive, forward, store, print and delete message. To receive message, you need an email address, which is a combination of a user name and a domain name that identifies a user.

Here are some popular email clients: Microsoft Outlook, Microsoft Outlook Express, Netscape Composer, Endora, Pine.

Preparing an email to your friends, just similar to that of sending letters. In a traditional letter, you have to provide the sender address, receiver address, content, stamp, etc. In an email system, you have to provide *your email address*, the *receiver email address* and the *content*. But there are also some differences from traditional postage system.

Email is a service that is absolutely free of charge no matter how many things you want to deliver. Second, a *subject* is allowed for you to add as an optional. It depends on whether you want to have a brief summary of what your email is. Third, you are allowed to make some *attachment(s)* on the email. Attachment means there is a

location, a space for you to put in the file, no matter in what format, to the receiver. By doing this, it shortens the distance between people as the communication way has been highly improved.

6.3 Email Server

When you receive an email message, the message is placed in your mailbox. A mailbox is a storage location usually residing on the computer that connects you to the Internet, such as the server operated by your Internet service provider (ISP). The server that contains the mailboxes often is called an email server.

Several years before, **POP3 (Post Office Protocol)** servers were nearly the only email servers in the computer world. And the users have to install an email client program or telnet directly to the email server in order to manage their mailbox (send/receive email).

However, as WWW becomes more and more popular, some companies provide their email service to their customers and do not request them to have a email client! They achieve this by providing a **Web-base** email service. And there is no need for the clients to have individual email client program in order to manage their mailboxes.

7. World Wide Web (WWW)

As we have already had an overview of the World Wide Web (WWW), we will here revise the basis of the WWW very briefly. The World Wide Web is the information resource centre of the Internet. The information stored in the WWW is in the form of pages called Web pages. You use a program called a browser to vie Web pages.

Because of the provision of graphical linkages, the WWW is usually treated as the graphical user interface of the Internet. Browsers display the links differently from surrounding text. For example, text links are displayed in blue and underlined, graphical links are displayed as 3-D buttons.

When the mouse cursor is above a link, it will change to a pointing finger. When you click on a link, a browser loads the corresponding Web page and displays it. This is called *following a link*, and the concept of using links to retrieve related pages of information is called *hypertext*.

7.1 WWW Clients (WWW Browser)

You know that a browser is a client program for accessing the WWW. The browser does contribute to the huge success of the WWW for it is really easy to use. Using a browser to view Web pages, you just need to use your mouse to click on the desired items and the browser will help you do the rest. You may not be able to find other program as simple as it is. Moreover, once you have learned to use one implementation of browsers, you will find no difficulties in using the others.

Through this section, the Internet Explorer 5.0 (IE5), one of the most popular browser nowadays, will be used to illustrate the necessary techniques and concepts of using a browser.

Back:	Go back to the previous page
Forward:	Go forward to the next page
Stop:	Stop loading a new page
Refresh:	Reload the current page
Home:	The first page of a set of Web pages to be loaded.
Search:	To use search engine
Favorites:	Display your favorite sites

7.2 Using the WWW

The WWW is a huge information centre. What can you get from it then? You can get the answer of this question by experiencing it. Here, we just quota a few example of services provided.

In Hong Kong, many popular newspapers and magazines are now “online”. For example, you may view the daily news on the WWW by visiting URL: <http://www.appledaily.com>. you may get the latest information on various topics on URL: <http://www.teleproperty.com>.

If you do not have much experiences of surfing the WWW, do it now. You will never regret about the time you spend on it.

7.3 Search Engine

There is a very popular saying about the WWW: “*You can get any information you need on the Net, if you can find it.*” This is to say that the WWW is so large that any kind of information can be found on it, but it is also so large that anyone may not be able to find the information one needs.

Why? Anyone can create and remove a Web site at any time and there is no ***central directory*** of Web sites. This is a direct consequence of the lack of a central authority.

To solve this problem, a good number of people have attempted to compile a Web directory and design a program to allow users to search the directory for a particular resource. The resulting tool is called a ***search engine***. One of the most popular search engines is – **Yahoo!**

8. Accessing the Internet

In this section, we will go through the ways of accessing the Internet. In particular, the most common way of accessing the Internet will be covered in detailed.

If you are a university student in Hong Kong and you have an account to access the university computers, accessing the Internet is very easy. Basically, the computers at any one of the universities in Hong Kong are connected to local networks. These networks are connected to form a large campus network that is connected to the Internet. You can then access the Internet through any one of these campus computers.

If you get a computer and a modem at home, you can even access the campus computers and hence the Internet at home. Remember the telnet programs (a server and a client) introduced in previous section. You can use a telnet client to emulate a campus computer at home.

However, accessing the Internet through a campus computer is not our purpose of discussion in this chapter. What we want to talk in details is how to access the Internet through an *Internet Service Provider (ISP)*.

8.1 Communication Media

8.1.1 Phone Line (Dial-up Line)

The most common way of accessing the Internet at home is to connect your computer with a remote Internet host via a modem and a phone line. Once connection is made, you will be ready to use the Internet. Then, which remote Internet host can you connect to? In most cases, the remote Internet host is provided by an ISP.

ISPs are organizations or companies offering public access to the Internet. They maintain computers that are connected to the Internet and ‘sell’ Internet access to the public. In order to be able to connect to a remote Internet host provided by a particular ISP, you must first arrange for an account from that ISP. Upon receiving your request, the ISP will give you:

?? a user name;

?? a password;

?? a phone number your communication program should dial in order to connect to the Internet host provided by that ISP.

In return, you have to pay the necessary fees to the ISP.

8.1.2 Dedicated Line- Leased Line

A **dedicated line** is a connection that always is established between two communications devices (unlike a dial-up line where the connection is reestablished each time it is used). Because dedicated line provided a constant connection, the quality and consistency of the connection is better than a dial-up line.

Businesses often use dedicated lines to connect geographically distant offices. Businesses can buy and maintain their own dedicated lines, or they can lease a dedicated line from a telephone or communications service company, in which case it is called a **leased line**. The cost of leased lines varies according to the distance between the two connected points and the speed at which the line transmits data. The charges for leased line, however, usually are flat fees, meaning you pay a fixed monthly amount for 24-hour a day access.

Leased lines can be either analog or digital. As with dial-up lines, analog leased lines require modems at both the sending and receiving ends. Digital leased lines are using any of the following transmission media: twisted-pair cable, coaxial cable, fiber-optic cable, microwaves or infrared. Because they provide faster transmission rates than analog lines, digital leased lines increasingly connect home and business users to networks around the globe. Four popular types of digital leased lines are ISDN lines, digital subscriber lines (DSL), T-carrier lines (T1, T3), and asynchronous transfer mode (ATM). The table below lists the transfer rates and approximate monthly costs of these types of lines, as compared with dial-up lines.

Type of Line	Transmission Rates	Approximate Monthly Costs
Dial-up	56Kbps	\$200
ISDN	128Kbps	\$300
ADSL	16Kbps to 8.45Mbps	\$1000
T1	1.544Mbps	\$8000 or above
T3	45Mbps	\$70000 or above

8.2 Hardware Requirement- Modem

Form the above discussion, we focus on how home user can access Internet. The basic hardware components include a computer, a modem and a phone line.

What is a modem? Why a modem needed in accessing the Internet via the phone line? How to choose a modem for yourself?

Inside a computer, data are stored in a digital format. Digital signals are employed to transmit data around the various units. However, only analog signals can be transmitted over the phone line. Hence something must be done if we want to transmit signals form one computer to another computer via the phone line.

Before putting the digital signal on the phone line, we must first convert it into an analog signal. This process is called **modulation** and the device used is called a **modulator**. After transmitting through the phone line, we have to convert the signal back to digital one so that the receiving computer can use it. This process is called **demodulation** and the device is called a **demodulator**.

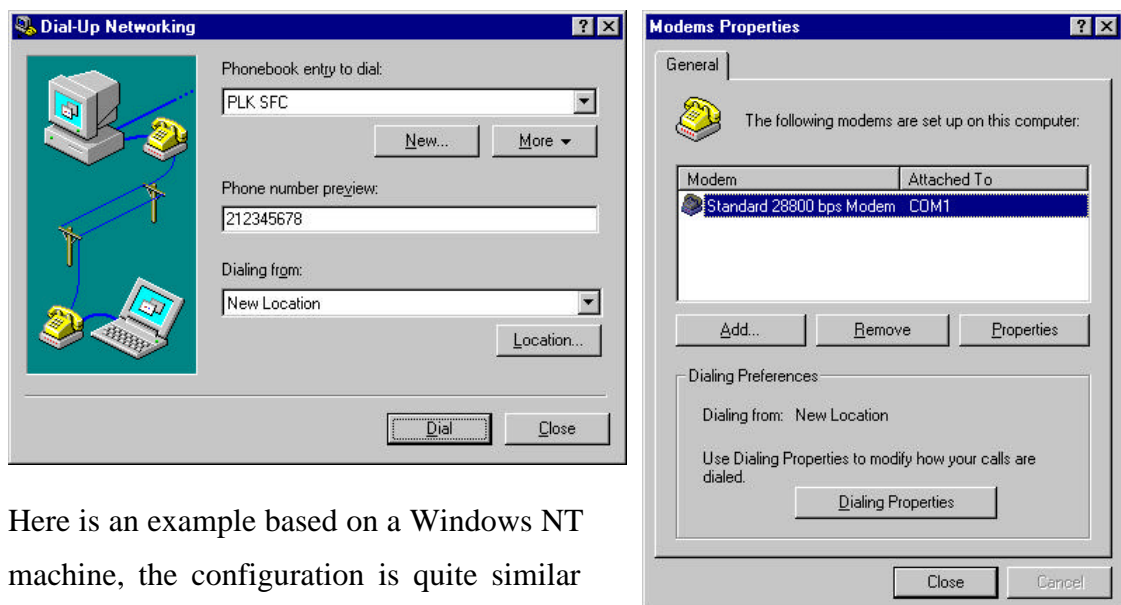
Since each computer using the phone lines to communicate needs to send and receive signals, both modulator and demodulator are required. One single device capable of carrying out both processes is designed and this is called a **modem** (**mo**dulator-**de**modulator). Now, you see that a modem acts as an interface between a computer and the phone line. It converts digital signal into analog signal and vice versa.

Modem speed is measured in term of '**bits per second**' or '**bps**', which means the maximum number of bits transferred in a second. Modems are manufactured in some standard speeds. Currently, common speeds of modems in market are 14.4Kbps, 28.8Kbps, 33.6Kbps and 57.6Kbps.

Some people have that mis-concept that the computer always occupies the phone line so that one has to spare a phone line solely for the purpose of Internet access. The fact is that the phone line is occupied only when you are connected to the Internet. You can use the phone line for other purposes when you are not accessing the Internet. The situation is just like using a phone line for a fax machine.

8.3 System Configuration

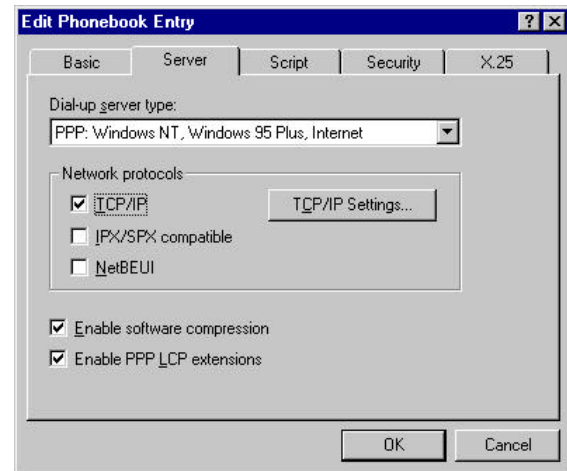
It is assumed that the modem is already installed successfully in your system. Here, we would discuss now to make a new connection.



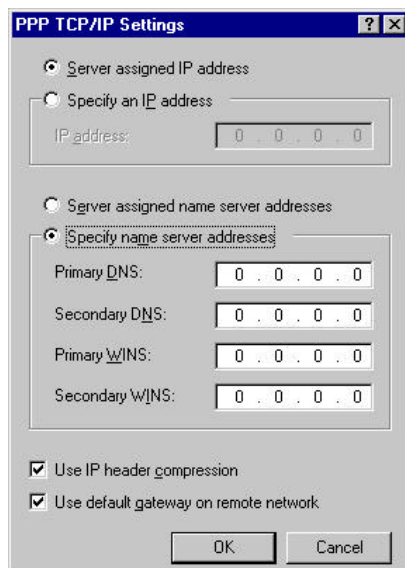
Here is an example based on a Windows NT machine, the configuration is quite similar no matter what kind of operating system you are using.

8.3.1 Protocols

One of the settings required you to configure you computer what kind of communication protocols that you want to use. You are allowed to use maximum three communication protocols for the connection, they are respectively TCP/IP, IPX/SPX and NetBEUI. What is the detailed of the last two protocols are out of our scope of discussion. You should still remember that the computers would use TCP/IP to communicate to each other over the Internet, that's why you have to enable this protocol.



8.3.2 DNS (Domain Name System)



Internet address or IP address (e.g. 203.85.81.1) is the only identification for the computers to identify each other. However, it is quite difficult and unreasonable for human being to remember these numbers. And we would only remember the domain name only. The function of DNS is to make a transformation between the IP address and the domain name. Nearly most of the ISPs get their own DNS and you are requested to enter this information into the corresponding boxes.

9. Creating Web Pages

9.1 Introduction

You have learned how to use a browser to navigate around the World Wide Web (WWW) through traversing hypertext links in Web pages. You are also aware that it is almost true that anyone can always find Web pages to get resources on any topics he may think of. There comes a question. Who produce the millions of Web pages on the Internet all over the world. You, as an Internet user, can also create your own Web pages.

Web pages are written in a special kind of language known as Hypertext Markup Language (HTML).

There are no secrets to build good Web pages. Good Web pages should be simple, interesting and useful. However, it is much more difficult to do than to say. Making good Web pages is never an easy task. We have three guidelines to follow in designing Web pages. They are *content*, *clarity* and *colour*.

9.2 Hypertext Markup Language (HTML)

A Markup language is a pseudo-computer language that combines plain text and formatting features together. Hypertext Markup Language (HTML) is a special kind of markup language for creating hypertext document. A hypertext document enables you to format information in a friendly, interactive and visually appealing way.

The next page shows a sample hypertext document:

```
<html>
  <head>
    <title>This is a Sample</title>
  </head>

  <body>
    
    <font size=6>This is my first homepage</font>
    <font size=6 color=blue><b>This is bold text.</b></font>
    <br>
    <font size=6 color=red><i>This is italic text.</i></font>
    <font size=6 ><u>This is underlined text.</u></font>
    <hr size=3>
    <a href="http://sfc-plk.edu.hk"></a>
  </body>
</html>
```

By a first glance, you may think that it is just a plain text document. Look at it more carefully, you can find that there are some special codes contained in it. Each special code is known as an **HTML tag** that contains **element name** enclosed by a pair of ‘<’ and ‘>’. The HTML tags tell Web browsers how to display and print this document. If you use a browser to view this document, it will displayed in a more beautiful way.

Compare the HTML document and the one displayed by the browser, you can note that the lines in the paragraph break in different places. This is because the browser will choose where to break lines according to the width of the browser window. In this case the Web page author cannot control how a browser displays his Web page. You will see some examples of this kind in the coming sections.

Examining the HTML document again, you may find that most of the HTML tags occurred in pairs: known as a **begin tag** and an **end tag**. The begin tag and the end tag of the same pair bear the same element name but the end tag contains a forward slash before the element name. On the other hand, some HTML tags are not used in pairs ,
 and <hr> are some tags of this kind.

You are reminded that the element name is case insensitive. This means that the tags <html>, <HTML>, <hTML> and <HtmL> are the same. Nonetheless, it is better to use them consistently.

An HTML document is a plain text file. Every character in it can be entered directly from you computer’s keyboard. These include all the alphabetic and numeric

characters, punctuation and other symbols that you can type directly into a file. Consequently, you can just use a simple editor (like **Notepad** in Windows and **Edit** program in DOS) to create your HTML documents. That they must have the extension `.htm` or `.html` is the only special requirement on HTML files.

Although no specialized programs are needed for creating HTML documents, there do exist a number of programs doing this job. These programs, often known as Web page development tools, greatly simplify the task of creating HTML documents. With them, you can even create your own Web pages without knowing anything on HTML. Nonetheless, it is better to start writing HTML with a plain text editor.

The basic structure of an HTML document contains two sections- the *head section* and the *body section*. The header section is usually very short for there is only a couple of things to be done here. The body section is much longer for it is a place for you to define the contents of your document.

9.3 Text Formatting

Here, we will look at some building blocks of HTML for formatting text in which the outcome is determined by the browsers.

9.3.1 Headings

A heading is a word or phrase that is highlighted to indicate the title of a section of text.

Six pairs of tags for headings of six levels are provided in HTML. Each pair of heading tags has general form `<hn>` and `</hn>`, where *n* is from 1 to 22. There is no precise definition on the style of each heading level, except that heading level 1 is the most prominent, and each successive level is slightly less prominent in some way.

Below is an HTML document for demonstrating how different browsers will display the headings.

```
<html>
```



```
<head>
  <title>Heading Practice</title>
</head>
<body>
  This is normal text.
  <h1>Heading 1</h1>
  <h2>Heading 2</h2>
  <h3>Heading 3</h3>
  <h4>Heading 4</h4>
  <h5>Heading 5</h5>
  <h6>Heading 6</h6>
</body>
</html>
```

Viewing this HTML document with two most popular browsers, the Internet Explorer and the Netscape, we can see that there are no major differences.

9.3.2 Paragraphing

As with any ordinary document, a unit of text that should logically be grouped together is called a *paragraph*. Most browsers align a paragraph of text to the left margin and wrap the text to the next line at a white-space character whenever necessary so as to fit the width of the browser windows.

Originally, each paragraph in an HTML document should be enclosed by the paragraph beginning tag `<p>` and paragraph terminating tag `</p>`. For example, here is an HTML document containing a sequence of five paragraphs in the body:

```
<html>
  <head>
    <title>Paragraph Practice 1</title>
  </head>
  <body>
    <p>This is a paragraph of normal text which

    is enclosed by a pair of paragraph tags.<p>
    <p>And this is the second paragraph of text.</p>
    <p>The third paragraph.</p><p>The fourth paragraph.</p>

    <p>The fifth paragraph.</p>
  </body>
</html>
```

You may note that the browser ignores any blank lines in the HTML document. Instead, the browser will automatically insert a blank line when it starts a new paragraph. Ignore the blank lines in the HTML document. You can see that

paragraphs are separated by both the tags `</p>` and `<p>`. it is easier to visualize and more productive if a single tag can be used to separate the paragraphs. Currently, nearly all browsers treat the `<p>` tag alone as a signal to start a new paragraph and ignore the `</p>` altogether.

Question: Think about how to present the above HTML document in a better way.

You can also set the alignment of the text in a paragraph by setting the align attribute of paragraph tag `<p>`. You have seen that the default alignment is left. You may also set it to be right or centered. In newer versions of HTML, you may even set it to be justified. Here is an example:

```
<html>
  <head>
    <title>Paragraph Practice 2</title>
  </head>
  <body>
    <p align="left">There are four paragraphs of text
    in this document. This is the first paragraph which
    is aligned to the left.
    <p align="right">This is the second paragraph which
    is aligned to the right.
    <p align="center">This is the third paragraph which
    is aligned at the centre.
    <p align="justify">This is the last paragraph. This
    paragraph is justified. That means it is aligned to
    both sides of the Windows.
  </body>
</html>
```

Try and see what is the look of the about HTML document.

9.3.3 Line Break

You see that browsers will ignore blank lines in HTML documents and automatically insert a blank between paragraphs in displaying the documents. However, that is not good enough and sometimes you need to have more control over your text.

Assume that someone wants to produce the following on the browser windows:

```
Common public transport in HK:
Bus
```

Taxi
Tram
Minibus

Example 1:

```
<html>
  <head>
    <title>Line Break Practice 1</title>
  </head>
  <body>
    <p>Common public transport in HK:
      Bus
      Taxi
      Tram
      Minibus
    </body>
</html>
```

Example 2:

```
<html>
  <head>
    <title>Line Break Practice 2</title>
  </head>
  <body>
    <p>Common public transport in HK:
    <p>Bus
    <p>Taxi
    <p>Tram
    <p>Minibus
  </body>
</html>
```

Example 3:

```
<html>
  <head>
    <title>Line Break Practice 3</title>
  </head>
  <body>
    <p>Common public transport in HK:
    <br>Bus
    <br>Taxi
    <br>Tram
    <br>Minibus
  </body>
</html>
```

You can see that the result of example 2 is quite similar to the target. What he needs is to have a line break between two points without any blank lines inserted. The line break tag `
` should be used for this purpose.

9.3.4 Rule Lines

A rule line is a horizontal line for separating logical sections of a document. The tag for inserting a rule line is `<hr>` and no corresponding terminator is needed. In encountering this tag, browsers will render a rule from the left margin to the right margin.

The default image for a rule line is just a single horizontal line. In newer versions of browsers, you can specify the thickness and the length of the line by using the `size` attribute and the `width` respectively. Consider the following HTML document:

```
<html>
  <head>
    <title>Rule Line Practice</title>
  </head>
  <body>
    A standard line
    <hr>
    A line of size=1
    <hr size=1>
    A line of size=3
    <hr size=3>
    A line of size=5
    <hr size=5 color=red>
    A line of size=8
    <hr size=8 color=blue width=75%>
  </body>
</html>
```

A size of 1 gives a solid black line, and larger values produce 3-D effects. The width of a rule is specified by the percentage of the full length that runs from margin to margin.

9.3.5 Preformatted Text

Sometimes we want the browsers to render our text exactly the same as the way it appears in our HTML documents. This is necessary when we want to place table of numbers, ASCII pictures, poems for which we want each stanza to be intact, or other documents for which we want to preserve the same formatting. To achieve this, we have to prevent any line breaks or use of special fonts by the browsers.

The kind of text is known as preformatted text in HTML and it is to be enclosed by the tags `<pre>` and `</pre>`. This pair of tags asks browsers to render the text enclosed exactly as it appears in the document. For example, an ASCII picture of a car is included in the following HTML document:

```
<html>
  <head>
    <title>Preformatted Text Practice</title>
  </head>
  <body>
    Hello world
    <pre>
      .
      ( )
      ( )
      ( ^ ^ )
      (___U___)
    </pre>
  </body>
</html>
```

Try to see what it is.

9.4 Character Styles

In the previous section, we have focused in HTML documents that set the general formatting of text. In this section, we will look at HTML facilities that format HTML document in the character-level.

9.4.1 Character-Level Attributes

By character-level attributes, we mean any special formatting that applies to a sentence, phrase, word or even a single character. Generally, there are two categories of character-level formatting elements: *font-style elements* and *information-type elements*.

A font-style element is an attribute to assign to any type of text regardless of its basic element type. For example, it is common to make a particular word in a paragraph

shown up in Bold or Italic. The following table shows some common font-style elements rendered in most browsers.

<code>...</code>	Bold	<code><big>...</big></code>	Big
<code><i>...</i></code>	Italic	<code><small>...</small></code>	Small
<code><u>...</u></code>	Underline	<code><sub>...</sub></code>	Subscript
<code><s>...</s></code>	Strike-through	<code><sup>...</sup></code>	Superscript
<code><tt>...</tt></code>	Teletype		

An information-type element also applies some kinds of formatting to any type of text. It is different from a font-style element in which the formatting is determined by HTML specifications but not by users.

For example, if you mark a word to be emphasized, most browsers today will just make it bold. Some years later, the trend may be that emphasized words are in bold and italics, the HTML specifications and browsers on that day will be changed accordingly. Browsers on that day will render your marked word to be in bold and italics. Hence, your HTML document will be rendered according to the future situation without making any changes on it. You see the advantage of information-type element. The following table shows some common information-type elements. Note that some browsers may not render the code as illustrated.

<code><cite>..</cite></code>	Citation	<code><kbd>..</kbd></code>	Keyboard
<code><code>..</code></code>	Code	<code><samp>..</samp></code>	Sample
<code><dfn>..</dfn></code>	Definition	<code>..</code>	Strong
<code>..</code>	Emphasized	<code><var>..</var></code>	Variable

Font-style elements apply physical formatting while information-type elements apply logical formatting to texts.

9.4.2 Font Attributes

With the font tag ``, we can set the font size and/or colour of the associated text.

Font Size

The default size of the text rendered by browsers is called the base font size. The default base font size is 3 and you can change the base font size by the tag `<basefont size=n>` where `n` is any one from 1 to 7.

We can change the font size of some of the text in an HTML document in two ways. First, change it to a value relative to the base font size. Second, set it to a specific value directly. In either way, we use the `size` attribute of the font tag ``.

Font Colour

We can use the `color` attribute of the font tag `` to set the colour of the associated text in an HTML document. Again, there are two ways of doing so.

First, you may set `color` with predefined color name, such as `Red`. Predefined colour names include: aqua, black, blue, gray, green, lime, maroon, navy, olive, purple, red, silver, teal, white, yellow, etc.

Second, you may also set `color` with a hexadecimal value which represents the red, green and blue content of the colour. The following table shows the hexadecimal colour values of some colours.

Colour	Hexadecimal Value	Colour	Hexadecimal Value
Black	000000	White	ffffff
	ff0000		00ffff
	00ff00		ffff00
	0000ff		ff00ff

Special Characters

You should be aware that some characters have special meaning in HTML and to browsers. The special characters which are the most frequently encountered by now are the less-than sign (<) and the greater-than sign (>). Some other examples are the quotation mark (") and the ampersand mark (&).

To display special characters in the browser window, some special constructs are needed. Each special construct is called an *entity* which has the form "&xxx;", where xxx is a string of characters that identify each particular entity. The following table shows a list of the most common entities you may use.

Entity	Character	Description
<	<	Less-than character
>	>	Greater-than character
&	&	Ampersand character
"	"	Double quote character

9.5 Lists

Three categories of list elements, each with several variations and options, are defined in HTML. They are unordered list, ordered list and definition list.

Unordered list and ordered list are very similar and they have a general format of HTML codes:

```
<!--start tab with optional attributes-->
    <lh>list head</lh>
        <li>item1
        <li>item2
        <li>item3
<!--end tab-->
```

9.5.1 Unordered List

Most browsers treat unordered list as bulleted list and use a heavy dot as the default bullet mark. The start tag and terminating tag for an unordered list are `` and `` respectively. Practise the following example.

```
<h3>Unordered List Practice</h3>
  <ul>
    <lh>list head 1</lh>
    <li>item1
    <li>item2
    <li>item3
  </ul>
  <ul type=square>
    <lh>list head 2</lh>
    <li>item1
    <li>item2
    <li>item3
  </ul>
```

9.5.2 Ordered List

An ordered list has the same characteristics as an unordered list except that the items in an ordered list are numbered but not bulleted as in an ordered list. The tags for ordered list are `` and ``. Here is an example.

```
<h3>Ordered List Practice</h3>
  <ol>
    <lh>list head 1</lh>
    <li>item1
    <li>item2
    <li>item3
  </ol>
```

9.5.3 Definition List

A definition list is for defining terms like the book glossary at the end of a book. A definition-list has two parts – the word or term to be defined and the defined and the defined of the word or term.

The tags for enclosing a definition list is `<dl>` and `</dl>`. As each item has two parts, there is a tag for each part. The first part (i.e. the word or term to be defined) is identified by the tag `<dt>` and the second part (i.e. the definition) by `<dd>`.

Most browsers render the terms to be defined on a line by themselves. The definitions are indented an on the next line.

```
<dl>
  <dt>Term 1
    <dd>Definition of term 1.
  <dt>Term 2
    <dd>Definition of term 2.
  <dt>Term 3
    <dd>Definition of term 3.
</dl>
```

9.6 Tables

Anything including text, data, pictures, graphics and anchors that HTML recognizes can be placed in a table. Creating table is easy with the pair of table tags <table> and </table>.

9.6.1 Defining a Table

```
<table border>
  <caption>This is the caption</caption>
  <tr>
    <th>Column 1 Head
    <th>Column 2 Head
  <tr>
    <td>Cell 21
    <td>Cell 22
  <tr>
    <td>Cell 31
    <td>Cell 32
</table>
```

The text above the table is called a **caption**. Most browsers render table captions at a position centered above the table.

The intersection of a row and a column is called a **cell**. Each cell store anything mentioned in the first paragraph of this section. The first cell of each column is called a **column head**. There isn't much difference between an ordinary cell and a column head except that the text in a column head is bold and centered in the cell by default. As there is usually more than one cell in a table, we use the row number and column number to specify a particular cell.

The grid separating the columns and rows is called the **border**. Without border attribute, the table will not have border.

Each row definition begins with the table row tag `<tr>` and doesn't require an ending tag. It is obvious that a new `<tr>` tag ends the previous row and starts a new row. After the `<tr>` tag, there are two tags for you to choose: **<th> for a header cell** and **<td> for a data cell**. Both tags mark the beginning of the cell content and do not require a closing tag for a similar reason.

9.6.2 Spanning Cells

To span cell across rows and/or columns, we use the `rowspan` and/or `colspan` attributes of the cell data tag `<td>` or the column head tag `<th>`. The parameters to these attributes are the number of rows and/or columns that the cell should span. A cell can span across multiple rows, multiple columns or both.

```
<table border>
  <tr>
    <th>Column 1 Head
    <th colspan=2>Cell spanned across 2 columns
  <tr>
    <td rowspan=3>Cell spanned across 3 rows
    <td>Cell 22
    <td>Cell 23
  <tr>
    <td>Cell 32
    <td>Cell 33
  <tr>
    <td>Cell 42
    <td>Cell 43
  <tr>
    <td>Cell 51
    <td>Cell 52
    <td>Cell 53
</table>
```

As the head of the second column spanned across 2 columns, we add the attribute `colspan=2` to the corresponding column head tag `<th>`. This also effectively **removes** the head of the third column from the table.

Similarly, you can use the same algorithm to consider the spanning in second row.

9.6.3 Column Width

We can specify an overall fixed width for a table by using the `width` attribute in conjunction to the `units` attribute to the table tag `<table>`. For example, `<table border units=pixels width=350>`.

9.7 Anchors

Many people agree that ‘links’ make hypertext called “*hyperlink*”. **Anchors** are then the most important components of HTML documents. Anchors are the cross-references that define the hypertext in HTML documents. With anchors, you can link a “*hot-spot*” in a Web page to other pages, to other part of the same page, to files and to even other Internet services such as the Internet mail service.

```
<a href="http://www.sfc-plk.edu.hk">PLK SFC</a>
```

An anchor consists of a pair of tags `<a>` and ``. The `href` attribute is the most important attribute and defines the object to which the anchor is linked. Any text enclosed by the anchor tags is marked as the hot-spot for that anchor. A hot-spot is usually rendered by browsers in a special format, like:

[PLK SFC](http://www.sfc-plk.edu.hk)

When the mouse cursor is over a hot-spot, the mouse cursor will be changed to a pointed-finger. When a user clicks on a hot-spot, the corresponding link defined with the `href` attribute will be activated.

9.7.1 Link to Other Pages

In linking to other pages, we specify the address in the `href` attribute of the anchor. For example:

```
<a href="http://www.sfc-plk.edu.hk/ca/index.html">CA</a>
```

9.7.2 Link to Other Part of the Same Page

We can also use an anchor to jump to another spot in the same document. This is useful when the document is really large. We have to define both the hot-spot and the destination. To define the destination, we use the `name` attribute of the anchor tag `<a>`. Try the following HTML document.

```
<a name="first">First Part</a>
<a href="#second">Go to Second Part</a>
<a href="#third">Go to Third Part</a>
<p>
<a name="second">Second Part</a>
<a href="#first">Back to First Part</a>
<a href="#third">Go to Third Part</a>
<p>
<a name="third">Third Part</a>
<a href="#first">Back to First Part</a>
<a href="#second">Back to Second Part</a>
```

The hot-spot and destination of other anchors are defined in similar ways.

9.7.3 Linking to Other System

You can use the `mailto`, `ftp`, `gopher`, `telnet`, etc links to link to the many other Internet services. Here are some examples:

```
<a href="mailto:whlau@sfc-plk.edu.hk">Mail to W. H. LAU</a>
<a href="ftp://ftp.microsoft.com">Connect to MS FTP Server</a>
<a href="gopher://sunsite.ust.hk">HKUST Gopher Server</a>
<a href="telnet://bbs.org.hk">BBS or HK</a>
```

9.8 Including Image & Sound

9.8.1 Displaying Images

You can use the image tag `` to display an image in your Web pages. The source of the image is specified with the `src` attribute. For example:

```

```

9.8.2 Images for Horizontal Rules

Newer versions of HTML enable you to specify an alternative image file for browsers to use for rule line. The `src` attribute is used again for this purpose. For example:

```
<hr src="http://www.sfc-plk.edu.hk/Badge.gif">
```

9.8.3 Backgrounds

You can use the background attributes of the body tag `<body>` to specify an image file. Browsers will replicate the image to create a column background for your Web page. For example:

```
<body background="http://www.sfc-plk.edu.hk/Badge.gif">
```

9.8.4 Sound

Web browsers do not yet have a facility for inline sound. Hence, browsers still rely on plug-in to handle sound. As different operating systems have their own sound applications. For example:

```
Click to listen school hymn:  
<a href="hymn.wav">WAV Format</a>  
<a href="hymn.aiff">AIFF Format</a>  
<a href="hymn.au">AU Format</a>
```

It will be even better if we provide a descriptive icon with sound links and an approximation of the size of the audio file.

9.9 Web Page Development Tools

A Web page development tool is a program releasing the burden of the users to write HTML codes in creating Web pages. It usually let the users create the Web pages using tools as they are using a word processor to make ordinary documents. Throughout the process, suitable HTML tags are inserted into the documents by the program.

There are plenty of Web page development tools available in the market such as Netscape Composer, FrontPage and Word. Some of them are even free to download from the Internet.

You may ask a question. As there are a lot of software that can generate the HTML coding for us, why we need to study HTML?

First, different Web development tools would have different coding method for the same action. So do not expect the software would give you a unique source code.

Second, since some effects cannot be generated or generated wrongly by these tools. Hence, we have to use the technique that we learnt to fine tune the HTML document.