

## The Internet Module

### *Pre-reading Material for Candidates*

Please read this booklet carefully before attending the course; you will be tested on its contents.

#### *Anti-virus Software*

- A virus is a piece of computer programming code that makes copies of itself without any human intervention. Viruses are deliberately written to behave as they do. Some viruses do more than simply replicate themselves, they might display messages, install other software or files, delete software or files, etc.
- A virus requires the presence of some other program to replicate itself. Typically viruses spread by attaching themselves to programs and in some cases files, typically e-mails or files downloaded from the internet. Music download sites are particularly involved in this too.
- Anti-virus software is designed to detect viruses before they are copied onto a computer and prevent them from working. This can only detect files that are known to the software distributor so it is essential that the anti-virus software is kept up to date.
- Many viruses have been constructed so that if they can install themselves on a computer they can modify the anti-virus software and prevent it working correctly. Once this happens it can be extremely difficult to detect and remove the virus.

#### *Anti-spyware Software*

- Spyware is software, usually unwanted, that installs itself on a computer and records what the user is doing. It may copy keystrokes, such as passwords, or other activities and then send them via the Internet to another computer.
- Spyware usually infects the computer when items are downloaded from the Internet; particularly music files or pornographic pictures.
- Anti-spyware software checks for such files and either prevents them from installing or reports them to the user. Most anti-spyware can also be used to scan a computer and search for any Spyware that may already be installed. It can only detect spyware that is known to the software distributor so it is essential that the anti-spyware software is kept up to date.

#### *Firewalls*

- A firewall is a combination of hardware and software that separates a Network, or perhaps an individual computer link, into two parts for security purposes. One part is connected to the outside world and the other to the internal hardware.
- Only software items approved by the user can cross the firewall thus preventing damaging material entering or confidential material leaving the system

#### *Passwords and Security*

- Access to many computer systems is controlled by a combination of user identities and passwords. The user identity may be widely known but the password is secret and known only to the individual user. This is to prevent unauthorised users from using the system and provides a way for individual users' work to be kept apart from others.
- The password should be simple enough for the user to remember but difficult for anyone else to guess. It should be a mixture of upper & lower case letters and digits.
- Passwords are not usually stored on the system. When they are created, the software securely encrypts them and it is this that is stored. When the user enters their password to log onto the system it is again encrypted and checked against the stored entry. In this way, even if someone could gain access to the stored data they could not find out the actual password to be able to get at the user's files.

## *Internet*

- The Internet is a world-wide network of tens of thousands of networks connected together. It grew out of the ARPANET created by the US military to link research centres together.
- The Internet permits the transfer of data in many different ways of which the World-Wide Web is only one.

## *Intranet*

- An Intranet is a system on a private network that offers many facilities similar to the Internet but only to those users on that network.

## *IP Addresses*

- Every computer connected to a network, particularly the Internet, has its own unique identity code called the Internet Protocol (IP) Address. This is four numbers, each between 0 and 255, separated by full-stops. It would look something like 198.34.0.129 and may change each time the computer is switched on.
- The number of possible IP Addresses is thus  $256 \times 256 \times 256 \times 256$  or 4,294,967,296 and is quite limited. To get round this most networks are allocated an IP Address for the public facing side of their firewall and their internal network uses an entirely different and much bigger set of numbers.
- There are moves afoot to extend IP Addresses to five sets of numbers which would give 1,099,511,627,776 different values and thus avoid the shortages.
- Note: you don't need to remember the actual numbers above; just the underlying ideas!

## *Domain Name*

- It is very difficult to remember the IP Address of every computer that you may want to connect to on the Internet. Instead, each computer is given a unique domain name such as **lancing.org.uk** which is much easier to remember.

## *DNS*

- When a domain name is used on the Internet it is sent to a Domain Name Server (DNS) somewhere. The DNS converts the domain name into the IP Address which is then used to contact the computer.
- Each network has its own DNS to look after internal data transfers and each Internet Provider will have a DNS with all their customers' IP Addresses recorded. Worldwide there are a small number of DNS centres that handle global addresses.

## *Routers*

- When computers are connected to the Internet there has to be some way to direct the data packets to the right destination. This is done by routers.
- In individual broadband connections the router can be a very simple device that simply transfers data. In bigger installations the router will have much more traffic to deal with and many different types. The router will sort the data and may well act as part of a firewall.

## **HTTP**

- The Hypertext Transfer Protocol (HTTP) is a set of rules and methods used to transfer web pages between computers on the Internet. The computer supplying the file uses an HTTP Server and the computer receiving the file uses an HTTP client.

## **HTTPS**

- HTTPS is a version of HTTP that permits secure transfers.

## **FTP**

- The File Transfer Protocol is a set of rules and methods used to transfer data files between computers on the Internet.

## **ADSL**

- Asymmetric Digital Subscriber Line; a DSL line where the upload speed is different from the download speed. Usually the download speed is much greater. Commonly used as broadband links in homes and small businesses.

## **Apache**

- This is the most common web server (or HTTP server) software on the Internet. Apache is an open-source application originally created at the National Centre for Supercomputing Applications, the same place the Mosaic web browser was created.

## **Applet**

- An applet is a small Java program that can be embedded in an HTML page. Applets differ from full-fledged Java applications in that they are not usually allowed to access certain resources on the local computer such as files and devices such as modems, printers etc and are prohibited from communicating with most other computers across a network.

## **ARPANet**

- The Advanced Research Projects Agency Network was the precursor to the Internet. It was developed in the late 60's and early 70's by the US Department of Defence as an experiment in wide-area-networking to connect together computers that were each running different system so that people at one location could use computing resources from another location.

## **ASCII**

- The American Standard Code for Information Interchange is the world-wide standard for the code numbers used by computers to represent all the upper and lower-case Latin letters, numbers, punctuation, etc. There are 128 standard ASCII codes each of which can be represented by a 7 digit binary number: 0000000 through 1111111.

## **Backbone**

- The backbone is a high-speed line or series of connections that forms a major pathway within a network usually used for linking the major critical components

## **Bandwidth**

- This measures the amount of data you can send through a connection in a particular time; usually measured in bits-per-second (bps). A full page of English text is about 16,000 bits. A fast modem can move about 57,000 bits in one second, an ADSL line can handle around 8 Mbits. Full-motion full-screen video needs about 10 Mbits-per-second, depending on compression.

## **Binary**

- Binary data consists entirely of ones and zeros; sometimes also used to refer to files that are not simply text files e.g. images.

## **Bit**

- A Binary digIT is a single digit number in base-2; in other words either a one or a zero. It is the smallest unit of computerized data. Bandwidth is usually measured in bits-per-second.

## **Blog**

- A weB LOG is basically a journal that is available on the web. The activity of updating a blog is "blogging" and someone who keeps a blog is a "blogger". Blogs are typically updated daily using software that allows people with little or no technical background to update and maintain the blog.

## **Broadband**

- This generally refers to connections to the Internet with much greater bandwidth than you can get with a modem. There is no specific definition of the speed of a "broadband" connection but in general any Internet connection using DSL, ADSL or a link via Cable-TV may be considered a broadband connection.

## **Browser**

- A browser is a program that is used to look at various kinds of Internet resources; usually the World Wide Web.

## **Byte**

- A set of bits that represent a single character. There are 8 Bits in a Byte.

## **CATP**

- The Caffeine Access Transport Protocol is a common method of moving caffeine across Wide Area Networks such as the Internet. CATP was first used at the Binary Cafe in Cybertown and quickly spread world-wide. There are reported problems with short-circuits and rust and de-caffeinated beverages were not supported until version 1-5-3. Sometimes used to refer to weak drinks.

## **CGI**

- The Common Gateway Interface is a set of rules that describe how a Web Server communicates with another piece of software on the same machine, and how the other piece of software (the CGI program) talks to the web server.

## **Client**

- A client is just a program that is used to contact and obtain data from a software server on another computer, often across a great distance. Each client program is designed to work with one or more specific kinds of server programs, and each server requires a specific kind of client. A Web Browser is a type of client.

## **Cookie**

- The most common meaning of "Cookie" on the Internet refers to a piece of information sent by a Web Server to a Web Browser that the Browser software is expected to save and to send back to the Server whenever the browser makes further requests from the Server. Depending on the type of Cookie used, and the Browsers' settings, the Browser may accept or not accept the Cookie, and may save the Cookie for either a short time or a long time. Cookies might contain information such as login or registration information, online "shopping cart" information, user preferences etc. When a Server receives a request from a Browser that includes a Cookie, the Server is able to use the information stored in the Cookie. For example, the Server might customize what is sent back to the user, or keep a log of particular users' requests. Cookies are usually set to expire after a predetermined amount of time and are usually saved in memory until the Browser software is closed down, at which time they may be saved to disk if their "expire time" has not been reached. Cookies do not read your hard drive and send your life story to the CIA, but they can be used to gather more information about a user than would be possible without them.

## **CSS**

- A Cascading Style Sheet is a way of specifying the appearance of text and other elements in a web page. CSS was developed for use with HTML in web pages but is also used in other situations. CSS is typically used to provide a single "library" of styles that are used over and over throughout a large number of related documents, as in a web site. A CSS file might specify that all numbered lists are to appear in italics. By changing that single specification the look of a large number of documents can be easily changed.

## **DHCP**

- Dynamic Host Configuration Protocol is way by which a machine can obtain an IP number (and other network configuration information) from a server on the local network.

## **DHTML**

- Dynamic HyperText Markup Language refers to web pages that use a combination of HTML, JavaScript, and CSS to create features such as letting the user drag items around on the web page and some simple kinds of animation etc.

## *DNS*

- The Domain Name System is the system that translates Internet domain names into IP numbers. A "DNS Server" is a server that performs this kind of translation.

## *Domain Name*

- The unique name that identifies an Internet site is its Domain Name. Domain Names always have 2 or more parts, separated by dots. The part on the left is the most specific, and the part on the right is the most general. A given machine may have more than one Domain Name but a given Domain Name points to only one machine. For example, the domain names:

matisse.net  
mail.matisse.net  
workshop.matisse.net

can all refer to the same machine, but each domain name can refer to no more than one machine. Usually, all of the machines on a given Network will have the same thing as the right-hand portion of their Domain Names (matisse.net in the examples above). It is also possible for a Domain Name to exist but not be connected to an actual machine. This is often done so that a group or business can have an Internet e-mail address without having to establish a real Internet site. In these cases, some real Internet machine must handle the mail on behalf of the listed Domain Name.

## *Download*

- Transferring data (usually a file) from another computer to the computer you are using is called downloading. In the opposite direction it is called an upload.

## *DSL*

- A Digital Subscriber Line is a method for moving data over regular phone lines. A DSL circuit is much faster than a regular phone connection but the wires coming into the subscriber's premises are the same copper wires used for regular phone service. A DSL circuit must be configured to connect two specific locations, similar to a leased line however a DSL circuit is not a leased line and is much cheaper. If the upload and download speeds aren't the same it is called ADSL and in theory ADSL allows download speeds of up to 9 megabits per second and upload speeds of up to 640 kilobits per second.

## *Email*

- Email or Electronic Mail is messages, usually text, sent from one person to another via computer. E-mail can also be sent automatically to a large number of addresses in which case it may become Spam.

## *Ethernet*

- Ethernet is a very common method of networking computers in a LAN.

## *Firewall*

- A firewall is a combination of hardware and software that separates a Network into two or more parts for security purposes.

## *FTP*

- The File Transfer Protocol is a common method of moving files between two Internet sites. There are many Internet sites that have established large amounts of publicly accessible material that can be obtained using FTP, by logging in using the account name "anonymous", thus these sites are called "anonymous ftp servers". FTP was invented and in use long before the advent of the World Wide Web and originally was always used from a text-only interface.

## *Gateway*

- The technical meaning is a hardware or software set-up that translates between two dissimilar protocols; for example America Online has a gateway that translates between its internal, proprietary e-mail format and Internet e-mail format. Another, sloppier meaning of gateway is to describe any mechanism for providing access to another system e.g. AOL might be called a gateway to the Internet.

## *GIF*

- The Graphic Interchange Format is a commonly used method for storing image files. It is especially suitable for images containing large areas of the same colour. GIF format files of simple images are often smaller than the same file would be if stored in JPEG format but GIF format does not store photographic images as well as JPEG.

## *Gigabyte*

- One Gigabyte is 1,024 Megabytes or 1,073,741,824 bytes; sometimes used to mean 1,000,000,000 depending on who is doing the measuring.

## *Hit*

- As used in reference to the World Wide Web, a hit means a single request from a web browser for a single item from a web server; thus in order for a web browser to display a page that contains 3 graphics, 4 hits would occur at the server: 1 for the HTML page, and one for each of the 3 graphics.

## *Home Page*

- The web page that your browser is set to use when it starts or simply the main page out of a collection of web pages is the Home Page.

## *HTML*

- The HyperText Markup Language is the coding language used to create Hypertext documents for use on the World Wide Web. HTML looks a lot like old-fashioned typesetting code, where you surround a block of text with codes that indicate how it should appear. The "hyper" in Hypertext comes from the fact that in HTML you can specify that a block of text, or an image, is linked to another file on the Internet. HTML files are meant to be viewed using a "Web Browser". HTML is loosely based on a more comprehensive system for markup called SGML, and is expected to eventually be replaced by XML-based XHTML standards.

## *HTTP*

- HyperText Transfer Protocol is used for moving hypertext files across the Internet. It requires an HTTP client program on one end and an HTTP server program (such as Apache) on the other end. HTTP is the most important protocol used in the World Wide Web (WWW).

## *Hypertext*

- Generally, any text that contains links to other documents is called Hypertext.

## *Internet*

- The vast collection of inter-connected networks that are connected using the TCP/IP protocols and that evolved from the ARPANET of the late 60's and early 70's is now called the Internet. The Internet connects millions of independent networks into a vast global internet and is the largest Wide Area Network in the world.

## *Intranet*

- A private network inside a company or organization that uses the same kinds of software that you would find on the public Internet, but that is only for internal use.

## *IP Number*

- The Internet Protocol Number is sometimes called a dotted quad. A unique number consisting of 4 parts separated by dots e.g.

165.113.245.2

Every machine that is on the Internet has a unique IP number - if a machine does not have an IP number, it is not really on the Internet. Many machines (especially servers) also have one or more Domain Names that are easier for people to remember.

## *IRC*

- Internet Relay Chat is basically a huge multi-user live chat facility. There are a number of major IRC servers around the world which are linked to each other. Anyone can create a channel and anything that anyone types in a given channel is seen by all others in the channel. Private channels can be and are created for multi-person conference calls.

## *ISDN*

- Integrated Services Digital Network is just a way to move more data over existing phone lines. It can provide speeds of roughly 128,000 bits-per-second over regular phone lines. In practice, most people will be limited to 56,000 or 64,000 bits-per-second. Unlike DSL, ISDN can be used to connect to many different locations, one at a time, just like a regular telephone call, as long the other location also has ISDN.

## *ISP*

- The Internet Service Provider is an organisation that provides access to the Internet in some form, usually for money.

## *Java*

- Java is a network-friendly programming language invented by Sun Microsystems. Using small Java programs called "Applets", Web pages can include functions such as animations, calculators, and other fancy tricks.

## *JavaScript*

- JavaScript is a programming language that is mostly used in web pages, usually to add features that make the web page more interactive. When JavaScript is included in an HTML file it relies upon the browser to interpret the JavaScript. When JavaScript is combined with Cascading Style Sheets (CSS), and later versions of HTML (4.0 and later) the result is often called DHTML.

## *JPEG or JPG*

- The Joint Photographic Experts Group or JPEG is most commonly mentioned as a format for image files. JPEG format is preferred to the GIF format for photographic images as opposed to line art or simple logo art.

## *Kilobyte*

- Approximately one thousand bytes; actually 1024 ( $2^{10}$ ) bytes.

## *LAN*

- A LAN is a local area network. A network where all the computers are in the same location such as one building.

## *Leased Line*

- Refers to a line such as a telephone line or fibre-optic cable that is rented for exclusive 24-hour, 7-days-a-week use from your location to another location. The highest speed data connections require a leased line but it is quite expensive.

## *Login*

- Noun: The account name used to gain access to a computer system. Not a secret (contrast with Password).
- Verb: the act of connecting to a computer system by giving your credentials; usually both your "username" and "password".

## *Megabyte*

- Approximately one million bytes; actually it's 1024 kilobytes or 1,048,576 bytes, which is more than a million.

## *Modem*

- A MODulator, DEModulator is a device that connects a computer to a phone line. A modem allows a computer to talk to other computers through the phone system. The maximum practical bandwidth using a modem over regular telephone lines is currently around 57,000 bps.

## *Netiquette*

- The etiquette to be used on the Internet; it's just good manners.

## *Network*

- Any time you connect 2 or more computers together so that they can share resources, you have a computer network.

## *NIC*

- The Network Interface Card is the card in a computer that you plug a network cable into.

## *Node*

- Any single computer connected to a network.

## *Password*

- A code used to gain access (login) to a locked system. Good passwords contain letters and non-letters and are not simple combinations such as "virtue7". A good password might be "5%df(29)" but it is more difficult to remember.

## *PDF*

- Portable Document Format is a file format designed by the Adobe Corporation to enable printing and viewing of documents with all their formatting (typefaces, images, layout, etc.) appearing the same regardless of what operating system is used, so a PDF document should look the same on Windows, Macintosh, Linux, OS/2, etc. The PDF format is based on the widely used Postscript document-description language.

## *Ping*

- To check if a server is running. From the sound that a sonar system makes, like in movies, when they are searching for a submarine.

## *Port*

- Generally, a place where information goes into or out of a computer or both is called a port. E.g. the serial port on a personal computer is where a modem would be connected.

## *Protocol*

- A protocol is a set of rules that control communication between systems. For example the HTTP protocol defines the format for communication between web browsers and web servers.

## *Proxy Server*

- A Proxy Server sits in between a Client and the "real" Server that a Client is trying to use. The client makes all of its requests to the Proxy Server, which then makes requests from the "real" server and passes the result back to the Client. Sometimes the Proxy server will store the results and give a stored result instead of making a new one (to reduce use of a Network). Proxy servers are commonly established on Local Area Networks to assist in controlling the traffic.

## *Router*

- A special-purpose computer (or software package) that handles the connection between 2 or more Packet-Switched networks is called a router. Routers spend all their time looking at the source and destination addresses of the packets passing through them and deciding which route to send them on.

## *Search Engine*

- A search engine is a, usually web-based, system for searching the information available on the internet. Some search engines work by automatically searching the contents of other systems and creating a database of the results. Other search engines contain only material manually approved for inclusion in a database, and some combine the two approaches.

## *Server*

- A computer or piece of software that provides a specific kind of service to client software running on other computers is called a server. The term can refer to a particular piece of software, such as a WWW server, or to the machine on which the software is running, e.g. "Our mail server is down today, that's why e-mail isn't getting out". A single server machine can and often does have several different server software packages running on it, thus providing many different services to clients on the network.

## *Spam (or Spamming)*

- An inappropriate attempt to use a mailing list, or USENET or other networked communications facility as if it was a broadcast medium (which it is not) by sending the same message to a large number of people who didn't ask for it. The term probably comes from a famous Monty Python skit which featured the word spam repeated over and over. The term may also have come from someone's low opinion of the food product with the same name, which is generally perceived as a generic content-free waste of resources.

## *Spyware*

- Spyware is a rather vague term generally referring to software that is secretly installed on a user's computer and that monitors use of the computer in some way without the user's knowledge or consent. Most spyware tries to get the user to view advertising and/or particular web pages. Some spyware also sends information about the user to another machine over the Internet. Spyware is usually installed without a user's knowledge as part of the installation of other software, especially software such as music sharing software or pornography obtained via download.

## *SQL*

- Structured Query Language is a specialized language for sending queries to databases.

## *Tag*

- The term "tag" is a basic element of the languages used to create web pages (HTML) and similar languages such as XML.

## *TCP/IP*

- Transmission Control Protocol/Internet Protocol is the set of rules that defines how the internet works. Originally designed for the UNIX operating system, TCP/IP software is now included with every major kind of computer operating system. To be truly on the Internet, your computer must have TCP/IP software.

## *Terabyte*

- A terabyte is 1024 Gigabytes or 1,099,511,627,776 bytes.

## *Terminal*

- A device that allows you to send commands to a computer somewhere else is called a terminal. At a minimum, this usually means a keyboard and a display screen and some simple circuitry. Usually you will use terminal software in a personal computer that allows you to type commands to a computer somewhere else.

## *Trojan (Horse)*

- A computer program is that is either hidden inside another program or that masquerades as something it is not in order to trick potential users into running it is called a Trojan Horse. For example a program that appears to be a game or image file but in reality performs some other function. The term "Trojan Horse" comes from a mythical trick of war used by the Greeks sometime between 1500 and 1200 B.C.

## *URI or URL*

- A Uniform Resource Identifier or Uniform Resource Locator is an address for a resource available on the Internet. The first part of a URI is called the "scheme". The best known scheme is http, but there are many others. Each URI scheme has its own format for how a URI should appear.

Here are examples of URIs using the http, telnet, and news schemes:

`http://www.matisse.net/files/glossary.html`

`telnet://well.sf.ca.us`

`news://new.newusers.questions`

## *Virus*

- A virus is a piece of programming code that makes copies of itself without any conscious human intervention. Some viruses do more than simply replicate themselves, they might display messages, install other software or files, delete software or files, etc. A virus requires the presence of some other program to replicate itself. Typically viruses spread by attaching themselves to programs and in some cases files, for example the file formats for Microsoft word processor and spreadsheet programs allow the inclusion of programs called "macros" which can in some cases be a breeding ground for viruses.

## *VOIP*

- Voice over IP is a system used to make telephone calls over IP networks, especially the Internet. Just as modems allow computers to connect to the Internet over regular telephone lines, VOIP technology allows humans to talk over Internet connections. Costs for VOIP calls can be a lot lower than for traditional telephone calls.

## *VPN*

- A Virtual Private Network is one in which some of the parts are connected using the public Internet, but the data sent across the Internet is encrypted, so the entire network is "virtually" private.

## *WAN*

- A Wide Area Network is one that covers an area larger than a single building or campus.

## *Web page*

- A document designed for viewing in a web browser is called a web page.

## *Website*

- The entire collection of web pages and other information (such as images, sound, and video files, etc.) that are made available through what appears to users as a single web server is called a website. Typically all the pages in a web site share the same basic URL, for example the following URLs are all for

pages within the same web site:

<http://www.microsoft.com/>  
<http://www.microsoft.com/whatis/>  
<http://www.microsoft.com/teenagers/>

## **Wi-Fi**

- Wireless Fidelity is a popular term for a form of wireless data communication; basically Wi-Fi is "Wireless Ethernet".

## **Worm**

- A worm is a virus that does not infect other programs. It makes copies of itself, and infects additional computers (typically by making use of network connections) but does not attach itself to additional programs; however a worm might alter, install, or destroy files and programs.

## **WWW**

- The World Wide Web (or simply Web for short) is not "The Internet"; which is a much larger thing. The WWW is the collection of hypertext servers (HTTP servers), more commonly called "web servers", which are the servers that serve web pages to web browsers.