

i-NET + Review Notes

**Compiled by
Bryce Embry**

Preface

I compiled the following set of notes when teaching a course in i-Net+ certification during the 2000-2001 school year. The notes are based mostly on the text we used in the course, [i-Net+ Study Guide](#) by David Groth, David Wall and Michael DeBeer, published by Sybex.

I have not taught a course in i-NET+ since the 2000-2001 school year, and have not worked to keep them current. There have certainly been some changes in the Internet world, but in reviewing these notes it seems that there are not many changes to make yet. New standards, such as IPv6, have still not been fully implemented, and though XML probably deserves more attention, this work was really not intended to cover web programming. In the course I taught, we studied web programming in much more detail after covering much of this material.

If you are using this document to prepare for the i-NET+ exam, you will want to supplement it with something that covers HTML and web programming in more detail.

These notes, along with more web programming information, are available on my web site, www.bembry.org.

I hope you find this information useful.

Bryce Embry
June 6, 2002

Copyright (c) 2002 Bryce Embry

Permission is granted to copy, distribute and/or modify this document

under the terms of the GNU Free Documentation License, Version 1.1

or any later version published by the Free Software Foundation; with no Invariant Section,

with the Front-Cover Texts being “i-NET+ Review Notes Compiled by Bryce Embry”,

and with no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

Table of Contents

Network Hardware	4-5
Bandwidth Technologies	6
Internet Basics	7-8
TCP/IP Basics	9
Servers	14-16
Client Basics	17
Client Configurations	18-20
How to Use Client Utilities	21-23
Security Issues	24-26
Web Programming	27-30
Site Function and Design	31-33
Troubleshooting	34-36
Legal and Business Issues	37-38
GNU Free Documentation License	39-43

Network Hardware

Network Topologies

- A **LAN** (Local Area Network) is a collection of computers sharing resources in a confined area.
- LANs can be configured in a bus topology (all computers connected along a single line), a star topology (each device connected to a central hub), ring topology (each computer connected to two others in a circular layout), or a mesh topology (each computer physically connected to every other computer). Of these, the star topology is the most common.
- A LAN may be broken up into smaller workgroups, each connected via to the main backbone or hub
- LANs use either an ethernet or token ring method for transmitting data.
- In an **ethernet** network, each computer takes turns sending data along the network lines using a technology called Carrier Sense Multiple Access with Collision Detection (CSMA/CD). A computer waits until the line is free, and then sends its data. If the data collides with another set of data being sent simultaneously, the computer stops, waits a random period of time, and then tries to send again.
- In a **token ring** network, all computers are connected to a multi- station access unit (MSAU) or controlled- access unit (CAU). Within the unit, a token is passed sequentially from one computer to another. A computer wishing to send data must wait for the token, and then attach data to the token, which will in turn be passed around the ring of computers, being repeated by each computer and changed slightly by the client for whom the token was intended, until it returns back to the computer that sent it.
- A **WAN** (Wide Area Network) can cover an unlimited geographical range, typically runs slower than a LAN, and will likely use routers and public network links.
- A centralized WAN connects to a central computer, while a distributed WAN connects numerous computers.

Hardware Components

- On a network, a workstation is a computer connected to the network, while a client is any component that requests resources from the network (such as a printer).
- A **server** is a computer that provides resources to the network. These can include files (file server), print management (print server), applications (application server), web pages (web server), e- mail hosting and delivery (mail server), fax services (fax server), functions on behalf of other computer (proxy server), remote connections to the network (remote access server), or even telephone related services (telephony server).
- A **NIC** (Network Interface Card) is the physical component on the computer that allows the workstation to connect to the network. The NIC must be compatible with both your computer and the network interface being used.
- **Thin-net** coaxial cabling has a range limit of 185 meters and must have a special terminating device at the ends of the cable. It uses a **BNC** connection.
- **Thick-net** coaxial cabling has a range of 500 meters but is more cumbersome and expensive.
- Twisted pair cabling comes in either shielded (**STP**) or unshielded (**UTP**) varieties. The most common is Category 5 (**Cat 5**) UTP, which contains four twisted pairs, is rated for 100 Mbps, has a range of 100 meters, and is the standard cable used in a majority of LANs.
- Cat5 UTP cabling is connected to the NIC with an **RJ-45** connector that looks a lot like a phone plug.
- **Fiber-optic** cabling is the fastest cabling available, is immune to electro-magnetic interference (**EMI**) and radio-frequency interference (**RFI**), and has a range of up to 4 kilometers, but it is the most expensive of the cabling options.

- A **hub** is a box to which network clients connect. The hub serves as a central connection port, taking the signal from one port and sending it out over all the other ports.
- An **active hub** is a powered hub that amplifies and cleans the signal it receives before broadcasting it to the connected ports.
- A **passive hub** is not powered and serves only as a physical connection point. Cables connected to a passive hub may have reduced range.
- A **switch** (or switching hub) improves upon the design of the hub by sending out data only to the designated receiver instead of broadcasting the signal down all connections.
- A **bridge** connects two segments of a LAN, allowing people on each side of the LAN to communicate with a specific client on the other side. Bridges also help reduce traffic flow by keeping local traffic from crossing over to the other portion of the LAN.
- A **router** "intelligently" connects LANs by reading the address of the data being sent and selecting the best route to send that data based on the information stored in the router's memory.
- A **firewall** is any device (computer or specialized firewall device) residing between your LAN and the Internet that prevents unauthorized users from accessing your private network. Firewalls may offer protection through **packet filtering** (accepting or rejecting each packet based on your rules), **application gateways** (applies security measures to specific applications only), **circuit level gateways** (applies security measures when a connection is first made, then allows all packets to pass freely), **proxy servers** (intercepts all messages and hides the true network address), or some combination of these technologies.
- A **modem** (MODulator / DEModulator) converts digital data into analog signals, thus allowing a computer to communicate over telephone lines, cable lines, or other analog communication networks.
- A **gateway** is a combination of hardware and software that allows two different types of networks to communicate.

Network Operating Systems

- A network operating system (NOS) is in charge of running the server computer as well as managing services provided over the network.
- The three most popular network operating systems are Microsoft Windows NT / 2000, Novell NetWare, and various flavors of Unix / Linux.

Boxed Solutions

- An **Internet-in-a-box** component provides Internet access to a workgroup or very small LAN without the need for a true server computer. The box is connected to a LAN cable and also to a modem, DSL, or other type of Internet connection (depending on the type of box), thereby sharing a single Internet connection with a small workgroup.
- Internet- in-a-box solutions vary in the features they provide, but may serve as DHCP servers, firewalls, web caches, search engines, or variety of other features.
- A **cache-in-a-box** device stores portions of commonly accessed web pages, and then delivers those portions from within the local LAN instead of the Internet in order to speed up data delivery.
- Internet- in-a-box (Ibox) and cache- in-a-box devices are types of **firmware**, hardware with hard-wired software.

Bandwith Technologies

Speed Abbreviations

- **Kbps** = Kilo-bits per second
- **Mbps** = Mega-bits per second.
- **Gbps** = Giga-bits per second.

Bandwidth Technologies and Link Types

- **DDS**, or digital data service, is a dedicated digital connection providing up to 56Kbps throughput.
- A **T1** provides a 1.544Mbps digital connection through two pair of UTP wires or fiber-optic cabling. The cable is divided into 24 separate data channels, with each channel delivering 64Kbps. A full T1 connection uses all 24 channels, while a fractional T1 connection will use only a portion of the 24 channels.
- **T3** connections provide 44.736 Mbps transfer rates using 672 channels (equivalent to 28 T1 lines). T3 connections are used primarily for the backbone of the Internet and Internet service providers.
- **Asynchronous Transfer Mode (ATM)** is a transfer technology that can use a variety of cabling options that can support data rates from 25Mbps up to 622Mbps. With ATM, data is broken down into small, fixed-size packets called **cells**, and then sent along a fixed channel or route.
- **ISDN (Integrated Services Digital Network)** uses regular telephone wiring (Plain Old Telephone Service) with a digital signal to provide transfer rates of 64Kbps (with one line, or "B channel) to 128Kbps (with two lines). ISDN requires an ISDN terminal adapter on the host computer.
- **DSL (Digital Subscriber Lines)** also use regular telephone wiring to transmit data digitally. Using special modulation schemes, DSL lines are able to reach transfer rates downstream of 9Mbps and higher (some claim 32Mbps), and upstream transmissions from 64Kbps to more than 1Mbps. DSL lines can only run up to 20,000 feet from a central telephone station, so service areas are limited.
- DSL comes in a variety of formats, including HDSL (High Data-Rate Subscriber Lines), SDSL (Symmetric Digital Subscriber Lines), and ADSL (Asymmetric Digital Subscriber Lines), which are collectively referred to as xDSL.
- **SONET (Synchronous Optical Network)**, also known as **SDH (Synchronous Digital Hierarchy)**, is a standard for transmitting multiple transmission types over fiber-optic cables using multiplexing and fixed frame sizes. SONET uses from 1 to 768 channels, which are designated Optical Character (OC) lines. Transfer speeds go from 51.84Mbps (OC-1), up to multi-Gbps (OC-768).
- **Public Switched Telephone Network (PSTN)**, also known as **Plain Old Telephone System (POTS)**, is using the regular telephone lines to send data in analog format. This method requires a modem on the host computers and is capable of data transmission rates of 56Kbps.
- **Frame Relay** is a transfer method that uses **virtual circuits** and variable packet sizes to send data over a network. Frame relay does not require a direct connection between the two communicating computers; instead it can use whatever lines are available between computer A and B and establish a virtual circuit, thus making it a very cost-efficient technology.
- **X.25** is similar to frame relay but includes some built- in error correction and flow control.

Internet Basics

History

- Began as government project in 1973 to maintain computer communications in case portions of network were destroyed in a war. This effort produced ARPANet (Advanced Research Projects Agency network) and the TCP/IP protocol suite.
- In 1986 the National Science Foundation created the NSFnet, which connected ARPANet and a number of other computer networks together.
- In 1989, Tim Berners-Lee in Geneva, Switzerland proposes architecture for a multimedia hypertext information system. This design comes to be known as the World Wide Web.
- In 1993 - 1994 the first graphical web browsers are introduced, first Mosaic and the next year Netscape Navigator.
- The **World Wide Web** is a portion of the much larger Internet. The Internet encompasses numerous forms of communication and data transmission, including file transfers, news groups, email, and more. The World Wide Web refers to the collection of HTML and related multimedia hypertext documents available within the Internet.

Physical Layout

- The **Internet** is comprised of LANs and individual workstations connected to **Internet Service Providers (ISP)**, which are interconnected via a high-speed WAN.
- An **ISP** is a company with a connection to the Internet backbone (or another ISP) who sells Internet connections to consumers. An ISP may also offer web site hosting, e- mail services, or other services.
- The WANs used to connect ISPs are usually special phone lines leased from telephone companies.
- The Internet **backbone** is composed of ISPs who are interconnected in a via highspeed WAN.

Domain Names and URLs

- Domain Name Services (**DNS**) is a database service that links domain names with their corresponding IP addresses. This allows the computers to take a human friendly name and translate it into a computer- friendly numeric address.
- **URLs** (Uniform Resource Locators) are the addresses for information on the Internet.
- A URL is composed of up to six different parts:
 - ▶ **Protocol**: describes the protocol being used. Protocols include **HTTP** for web page transfer, **FTP** for file transfers, and **TELNET** for a terminal emulation (command line interface).
 - ▶ **Domain name** : Also called the **host name**, this is the name registered with the Internet domain name service and generally correlates to a specific web site.
 - ▶ **Port**: The port number is an optional element that designates which Internet port to connect to on the host computer.
 - ▶ **Path**: The directory path to the file or resource being requested.
 - ▶ **File**: The actual name of the file or resource being accessed.
 - ▶ **Query String**: A query string is information added to the end of a URL to submit information to a script or program being run on the web server. These are used frequently with forms and search engines on the Internet.
 - ▶ **Example**: **http://www.bembry.org /tech/inet/netbasics.html**

- The domain name of a URL is not case sensitive, but the path and file names are often case sensitive.
- The **Top Level Domain** for a domain name is the suffix or extension at the end of a domain name (the last two to four letters after the last dot on the host name). Top level domains include the following:
 - ▶ **com**: Commercial organization or business.
 - ▶ **edu**: Educational institution.
 - ▶ **org**: Non-profit organization.
 - ▶ **gov**: Government organization.
 - ▶ **mil**: US military branch.
 - ▶ **int**: International organization (i.e. the United Nations)
 - ▶ **net**: A network organization.
 - ▶ Other top level domains include country codes (DE for Denmark, JP for Japan, etc.), as well as the new domains **aero**, **biz**, **coop**, **info**, **museum**, **name**, and **pro**.
- The **gov**, **mil**, **edu**, and **int** top level domains are restricted in who may own or use them, as are most of the new domains. The .com, .org, .net and .info domains are available for anyone to use. For more info on the new top level domains, see www.icann.org/tlds.
- In order to save your own domain name, you must purchase that name from a company that is accredited by **ICANN**, the Internet Corporation for Assigned Names and Numbers which is responsible for ascertaining that there are no domain name conflicts. In the past there was only one company licensed to sell domains, but now the privilege has been opened up and prices are more competitive.

Internet Communication Process

When requesting a web page while surfing the Internet, the following steps occur:

- The browser sends the requested URL to a DNS server to be translated into an IP address.
- Using the IP address, the request is passed to routers, who transfer the request to the intended server.
- The server decodes the URL request and finds the resource or file requested.
- If the requested resource requires processing, as with server-side scripting, then the server will perform the necessary processing.
- The resulting resource is sent back to the computer that made the original request.
- The web browser on the host computer decodes the information it receives and displays it on the screen.

TCP / IP Basics

TCP/IP Overview

- A **protocol** is a set of rules on how data will be packaged and transferred between computers.
- TCP/IP (Transmission Control Protocol / Internet Protocol) is the main suite of protocols used for the Internet. This set of protocols includes TCP, IP, HTTP, FTP, PPP, and many others.
- TCP/IP was designed as an open standard, to be capable of implementation on all types of hardware and software systems.

Transmission Control Protocol

- TCP is responsible for taking the data to be transmitted, breaking it down into packages called **datagram**, encoding the datagram with special codes to ensure its safe delivery, keeping track of the datagram that have been successfully transferred, retransmitting datagram that have been lost, and reassembling all the datagram once they have safely reached their destination.
- TCP is a **connection-oriented protocol**, since it has methods of assuring the delivery of data.
- TCP **header data**, the data that TCP adds at the beginning of each datagram, includes the following information:
 - **Source port number** and **destination port number**. This information assures that the data sent connects to the correct process on each computer.
 - Since TCP breaks data into smaller units, the header includes a **sequence number** identifying which order the packets should be reassembled in.
 - To make sure the received data has not been corrupted, TCP places in the header a **checksum**, which is a value based on the number of bits in the data package. The receiving computer will calculate the checksum again, and if the checksum numbers do not match, the system will assume that the data has been corrupted.
 - The **acknowledgment number** is a number that the receiving computer sends back to the source computer to confirm that this particular datagram has been received correctly. If the source computer does not receive this acknowledgment number back within a specified period, then it will assume the datagram has been lost and retransmit it.

Internet Protocol

- The IP protocol is responsible for routing the packages created by TCP.
- IP is a **connectionless** protocol. It is not concerned with whether or not the data actually reaches the recipient, just with moving that data to its designated destination.
- IP adds a header to the datagram created by TCP, resulting in a total of two different headers added to the original source data.
- The IP header includes the following information:
 - ▶ A **checksum** to provide a means of checking data integrity at each stopover point.
 - ▶ A **hop count** or time to live, which determines the maximum number of hops a package can make.
 - ▶ Both source and destination addresses are also included in the IP header.
- The IP protocol is used to determine the route a data packet will take to its destination. If the destination IP address is not known by the local gateway, that gateway will pass the packet on to its default gateway. This process will continue until the desired destination is reached.

- Using IP, different datagram from a single data source may take different routes to their destination, thus causing some packets to arrive out of order. To avoid this randomness, it is also possible to prescribe a set route for the data to take.

IP Addressing

- In order to participate in a TCP/IP network, each computer (or **host**) must have a unique IP address. These addresses may be automatically assigned using **DHCP** (Dynamic Host Configuration Protocol) or manually entered into the host computer.
- An IP address is made up of a single 32-bit number (meaning it has 32 ones or zeros). This number is usually divided into four 8-bit segments separated by dots. Each 8-bit segment has a value between 0 and 255. *Example:* $01111111.00111111.00011111.00000111 = 127.63.31.7$.
- **Dotted decimal** notation refers to writing IP addresses using four decimal numbers (numbers between 0 and 255) separated by dots.
- The first portion of an IP address is usually used to identify the network, while the second portion identifies a particular machine within that network.
- An IP address composed of the network portion of the IP followed by all zeros identifies the network itself. *Example:* $192.168.0.0$ refers to the 192.168 network.
- An IP address composed of the network portion of the IP followed by all 255s is called a **broadcast address**. *Example:* A packet addressed to $192.168.255.255$ would be delivered to every machine on the 192.168 network.
- The IP address $192.168.x.x$ is reserved for private networks.
- The current version of IP addressing is IPv4 (version 4) and allows over 17 million address, which is proving insufficient. A new version, called **IPng** (IP next generation) or **IPv6** is currently being phased in and will provide more IP addresses (over 70 octillion).
- IP addresses are divided into the following classes:
 - ▶ **Class A:** Highest-order bit set to zero; IP address range from $1.x.x.x$ to $126.x.x.x$; first octet makes up the network portion of the IP address. There may be 127 class A networks, each having up to 16,777,214 connected hosts. All Class A networks are currently taken.
 - ▶ **Special:** The address $127.0.0.1$ is reserved for loop-back tests.
 - ▶ **Class B:** Highest order bit set to 10; IP address range from $128.0.x.x$ to $191.255.x.x$; first two octets make up the network portion of the IP address. There are no Class B addresses free.
 - ▶ **Class C:** Highest order bits set to 110; IP address range from $192.0.0.x$ to $223.255.255.x$; first three octets determine network portion of IP address.
 - ▶ **Class D:** Highest order bits set to 1110; used exclusively for **multicasting** (delivery to a group of host computers).
 - ▶ **Class E:** Highest order bits set to 1111; reserved for experimental use.
- A new addressing scheme called **CIDR** (Classless Inter-Domain Routing Scheme) breaks down IP addresses into segments smaller than class C to fit the needs of different companies.

Subnet Masks

- A **subnet mask** is a way of dividing a single network into multiple physical networks by reallocating the hosts portion of the IP addressing scheme. The new IP address scheme has a network portion, a subnet portion, and a host address that is shorter than under the original scheme.
- Subnets help reduce network traffic by keeping local traffic on one side of a router and isolating the information from the LAN on the other side of the router.
- A router must be used to implement a subnet scheme.
- To define a subnet mask, convert the network portion of the IP address into binary notation. Next, select the number of binary digits to use for the subnet mask. Finally, calculate the new dotted decimal ranges available under each subnet.
- Example:
 - ▶ **Key:** *Network; Subnet; Host*
 - ▶ **IP Network Address:** 172.25.16.x
 - ▶ **Binary IP Network Address:** 10101100 00011001 00010000 xxxxxxxx
 - ▶ **Add Subnet Mask:** 10101100 00011001 0001000 11xxxxxx
 - ▶ **Four New Subnets Available:**
 - A. 10101100 00011001 0001000 00xxxxxx
 - B. 10101100 00011001 0001000 01xxxxxx
 - C. 10101100 00011001 0001000 10xxxxxx
 - D. 10101100 00011001 0001000 11xxxxxx
 - ▶ **Dotted Decimals of New Subnets:**
 - A. 172.25.16.0 to 172.25.16.63
 - B. 172.25.16.64 to 172.25.16.127
 - C. 172.25.16.128 to 172.25.16.191
 - D. 172.25.16.192 to 172.25.16.255
- On a subnet, the first available address in the subnet class is the new network number and the last available address is the new broadcast number. Example: In subnet A above, 172.25.16.0 is the network number and 172.25.16.63 is the subnet broadcast number.

DHCP

- DHCP (Dynamic Host Configuration Protocol) dynamically assigns IP addresses to hosts as they log onto a network, then revokes those addresses when the hosts log off. This method of allocating IP addresses makes administration easier and allows a network that has more computers than available IP addresses to continue to function.
- DHCP servers are configured with a range of addresses they are permitted to lease out to client computers. The server then keeps track of which addresses have been leased and which are still available.
- When a network client boots up, it sends out a broadcast message to discover the DHCP server. The DHCP server responds with an offer of an available IP address, the appropriate subnet mask, and data defining how long the address lease is good for. The client host accepts the IP address and returns a request to the DHCP server for the lease of the IP address offered and any other needed information. The DHCP finalizes the assignment by returning an acknowledgment packet containing all the pertinent network information.
- DHCP leases can be configured to expire within a certain period of time or to not expire at all.
- When half of a DHCP lease period has expired, the client sends a message requesting a lease renewal. The server again responds with an acknowledgment.

Port Numbers

- A **port** is a TCP/IP identifier that indicates what application or process a request is associated with. On the server, an application like Telnet monitors its assigned port number for activity, and then communicates with the remote computer using this port.
- Well Known Port Numbers include the following:
 - ▶ **20**: FTP Data (File Transfer Protocol)
 - ▶ **21**: FTP Control (File Transfer Protocol)
 - ▶ **23**: Telnet
 - ▶ **25**: SMTP (Simple Mail Transfer Protocol)
 - ▶ **53**: DNS (Domain Name Server)
 - ▶ **70**: Gopher
 - ▶ **79**: Finger
 - ▶ **80**: HTTP (Hypertext Transfer Protocol)
 - ▶ **110**: POP3 (Post Office Protocol 3)
 - ▶ **119**: NNTP (Network News Transfer Protocol)
- Some systems are capable of **dynamically allocating** port numbers. If a port is currently busy when a new request is received, the system can assign a different port number to handle the request.
- An IP address combined with a port number creates a **socket**. Two sockets (one sending, one receiving) are required to establish a TCP connection.

Other Protocols

- **SLIP** (Serial Line Internet Protocol) is a protocol used to transmit over serial lines, such as with a modem over phone lines. SLIP is a simpler protocol that has a low overhead, but its lack of some desired features (such as password encryption and error checking) has caused it to be largely replaced by PPP.
- **PPP** (Point-to-Point Protocol) is commonly used to establish remote connections to Internet service providers or LANs. PPP can run over various types of connections, provides error correction, supports automatic TCP/IP configuration, and provides a number of other benefits above SLIP, although it does demand higher overhead.
- With **PPTP** (Point-to-Point Tunneling Protocol) is a Microsoft-created protocol that uses PPP to create a **Virtual Private Network** (VPN). To use PPTP, a user establishes a PPP connection to the desired server, and then launches a PPTP connection. In effect, the user is then connected to the server via PPP, but is able to transfer information securely from within the PPP connection thanks to the PPTP session. PPTP is currently not standard and not supported by all operating systems.
- **HTTP**, or Hypertext Transfer Protocol, is used to communicate between a web browser and a web server. Web pages are transmitted over HTTP.
- **FTP** (File Transfer Protocol) is used to transfer files between two computers. FTP requires users to authenticate with a user name and password, while a similar protocol, **TFTP** (Trivial File Transfer Protocol) works the same without requiring user authentication.
- **SMTP** (Simple Mail Transfer Protocol) moves mail from one mail server to another or from an email client to an email server. Either way, SMTP is the way to get email onto a mail server.
- **POP3** (Post Office Protocol 3), on the other hand, is used to get mail off of a mail server. When used by an email client, POP3 downloads all the client messages available from the server. The **IMAP** (Internet Message Access Protocol) is a different protocol for retrieving mail off an email server; however, IMAP supports downloading selected messages only, and leaving the rest on the server.
- The **NNTP** (Network News Transfer Protocol) provides the facilities for transferring information on newsgroups (Usenet news). The protocol allows posting, distribution, and retrieval of the messages among both clients and servers.
- **LDAP** (Lightweight Directory Access Protocol) is an open protocol for accessing information directories, which supply such data as email addresses and names.
- **Gopher** was a method for organizing and displaying files on an Internet server before the advent of the World Wide Web. This system has largely been replaced by the web.
- **TELNET** is a protocol used mostly on Unix servers which allows users to log onto a remote computer and use it as they were sitting at the console themselves.
- **LPR** (Line Printer Remote) allows a user to send a print file to a remote server for printing.

Servers

HTTP Servers

- HTTP servers are responsible for serving text and graphical content using Hypertext Transfer Protocol.
- An HTTP or web server runs an **HTTP daemon**, which is a program responsible for responding to requests from web browsers. The server may also be configured to process different types of **scripts**.
- An **Internet web server** is designed and configured to provide web content to the general public.
- An **intranet server** is configured to provide web content to authorized users only. Intranet servers are usually located behind a firewall within a local LAN and intended for use by only the members of the organization or company.
- An **extranet server** is an intranet server that also permits some limited access by the public or other authorized users outside the local area network. These servers are popular for business partners who wish to share information.
- A web server designed to provide e-commerce services may be called an **ecommerce server**. These services, which are supported through special programs or scripts, may include inventory searches, a shopping cart program, credit card verification, and an automatic email confirming your order.
- There are three main HTTP daemons in use today. These are HTTPd, Microsoft IIS, and Apache. Currently the most popular is Apache, which is on over 60% of all servers.
- The web server software **HTTPd** (HTTP daemon), was the original HTTP server software. HTTPd was developed by the National Center for Supercomputing Applications for the Unix operating system and is free, open source software.
- Microsoft's **IIS** (Internet Information Server) is packaged with Windows NT and Windows 2000. It provides a user-friendly graphical interface but works only on Windows NT or 2000 servers.
- **Apache** is a free, open source HTTP daemon built on HTTPd. Although originally designed for use on Linux / Unix systems, it is also available for use on Windows NT and OS/2. Apache is currently the most widely used HTTP daemon on the Internet.
- If the traffic on a web server becomes too heavy, the site may be **mirrored**. Mirroring a site means placing a copy of the site on another web server.

FTP Servers

- An **FTP server** runs an **FTP daemon** in order provide file services (download, upload, delete) using the FTP protocol.
- An FTP server requires a name and password from the user before allowing file transfers. For public downloads, "anonymous" transfers are permitted with the user name "anonymous" or "ftp" and any text for the password.
- Two popular FTP server programs are **FTPd** (FTP daemon) and Microsoft's **IIS**. FTPd is a free Unix-based program in wide use, while IIS is designed for specifically for use on Windows NT or 2000 server systems.

SMTP Servers

- SMTP (Simple Mail Transfer Protocol) servers allow users to send and receive email through the Internet.
- An SMTP server runs a daemon responsible for locating the IP address of the recipient email account, verifying that mail has been delivered, displaying an appropriate error message if mail is undeliverable, retrying to send mail if the recipient server is busy, and generally getting the mail through.
- Mail servers may also provide **SMTP relaying** services, which forward email coming from one SMTP to a different SMTP.
- Mail servers may also be used as **list servers**, allowing email messages to be sent to all the recipients who have signed up on a list, and allowing people on that list to in turn reply to the list.
- The most commonly used SMTP daemon is **Sendmail**, another product of open source software.

Proxy Servers

- Proxy servers provide a means for workstations on a secure intranet or private LAN to access the Internet while preventing Internet traffic from accessing the private LAN.
- A proxy server sits between the Internet server and the client workstation, processing the requests made by the client and determining whether to pass those requests on to the Internet server or deny the request. Similarly, the proxy server can manage some of the requests sent from the Internet server. Proxy servers can be configured as to what types of services they permit.
- Since proxy servers can provide **packet filtering**, they can be used to restrict the types of files that network users' access over the Internet. For example, a proxy server may be set to prohibit any MP3 files from being downloaded.
- Proxy servers may also provide **caching** services. With **active caching**, the server uses low activity periods to download and store documents that may be requested by network users, while **passive caching** determines whether or not to store documents as they are requested by users on the network.

NNTP Servers and Newsgroups

- NNTP (Network News Transfer Protocol) servers are used to provide access to newsgroups.
- Newsgroup servers store a variety of newsgroups, and can be configured to communicate with other NNTP servers so that each server has a copy of all the messages available within the newsgroup.
- If a newsgroup is **moderated**, any new messages posted to the newsgroup are first reviewed by a moderator, who determines whether to publicly post the message or discard it.
- Newsgroups may be configured with an expiration date, thereby automatically erasing messages after they reach a designated age.

Other Servers

- A server running a telnet daemon is a **telnet server**. The telnet daemon, used mostly on Unix / Linux systems, allows a remote user to log onto the server and type in commands as if he were sitting directly at that computer.
- A **directory server** provides access to directories of information, such as phone books, email address lists, and various other directory-type information. Directory servers are also known as **LDAP servers**, since most conform to the Lightweight Directory Access Protocol.
- **Mirror servers** are backup servers that duplicate everything that takes place on the main server so that they can immediately replace the main server if it fails.
- A single computer may run a variety of daemons and thus be more than one type of server. For example, an HTTP server may also be a telnet and FTP server.

Client Basics

Hardware

- You computer's hardware must meet the software manufacturer's minimum specifications for processor, RAM, and hard disk space in order to load and run the required client software.
- Internet software may be run on a traditional computer, an Internet appliance (such as WebTV), or other device designed with Internet capabilities (i.e. Cellular phone, handheld PC, etc.).
- Any hardware device connecting to the Internet must have hardware designed specifically for such connections.
- A **modem** is required to connect to the Internet through conventional phone lines. The modem will be used to dial into an Internet service provider (ISP).
- A **NIC** (Network Interface Card) is required to connect to the Internet through a LAN.

Software

- On a traditional PC, you must have an operating system installed and running in order to use Internet software. Furthermore, the software you choose must be compatible with the operating system you are using.
- Along with the operating system, a computer must have the TCP/IP **protocol stack** installed. The protocol stack is the set of software required to support the TCP/IP protocol suite.
- In Microsoft Windows, the **WINSOCK.DLL** file supplies TCP/IP support.
- A **web browser**, such as Netscape Navigator, Microsoft Internet Explorer, or Opera, is required to view HTML files and their associated graphics.
- HTML files may also be viewed without the associated graphics from text-only web browsers, such as Lynx.
- An **FTP utility** is required to upload and download files using the File Transfer Protocol. These utilities may be standard command line interfaces (text only) or graphical utilities that provide more user friendly interfaces. Both Microsoft Internet Explorer and Netscape Navigator provide basic FTP download capabilities.
- A **telnet** utility is required to interact using the telnet protocol. Microsoft Windows 95 and higher operating systems come with the telnet feature preinstalled.
- A **news** client program is required for subscribing to, reading, and posting to Usenet news groups. Netscape Communicator and Microsoft Outlook Newsreader (a feature of Microsoft's Internet Explorer) both provide this capability.
- An **e-mail** program is required to send and receive e-mail over the Internet. Both Microsoft Internet Explorer and Netscape Communicator have integrated e-mail programs. Microsoft Internet Explorer supplies e-mail services through the Microsoft Outlook program, which has become a popular target for viruses. Stand-alone e-mail programs, such as Eudora, are also available.

Client Configuration

IP Configuration

- To connect a computer to the Internet, you must configure the TCP/IP address and name resolution properties in the Network portion of the Control Panel. If connecting through a modem, the dial-up properties must also be set, while connecting through a LAN will require further configuration in the TCP/IP properties.
- In TCP/IP properties, the computer must be configured either to obtain an IP address automatically or assigned a specific IP address.
- A computer may be assigned an IP address and subnet mask automatically if it is connecting to a server running DHCP (Dynamic Host Configuration Protocol). On a DHCP network, the server has a set of IP addresses that it hands out to host computers as they log on. Most dial-up ISPs run DHCP.
- A computer must have an IP address and subnet mask manually set if it is in a network where IP addresses are **static**. In this type of network, the administrator chooses what IP addresses each computer receives and sets them manually.
- A **default gateway** refers to the address of the router or routing computer where the host will send all unfamiliar IP address requests. The default gateway can be automatically set on a DHCP network, but must be set manually on a network using static IP addresses.
- Running `windowsipcfg` on a Windows computer will display the computer's current IP configuration.

Name Resolution

- **Name resolution** is the method a computer uses to translate text addresses into IP addresses. A client PC may use **HOSTS**, **DNS**, or **WINS** to accomplish this task.
- A **HOSTS** file is a manually edited text file that maps host names to their IP address. If using a HOSTS file for name resolution, you must change the text file every time you make any changes to the IP addresses available on the network.
- Using **DNS** (Domain Name Service) requires enabling the service in the TCP/IP properties menu, the host name (of your computer), the name of the network domain, IP addresses of the DNS servers on the network, and optionally a domain suffix that the computer will automatically add at the end of unrecognized names in an attempt to resolve the name you type with a known IP address.
- The DNS Configuration on Windows requires the host, domain, and the IP address of at least one DNS server to be entered.
- DNS must be enabled and configured (either automatically or manually) in order for a host computer to access the Internet.
- **WINS** (Windows Internet Name Service) is a dynamic directory designed to keep track of resources, mapping their NetBIOS name with their assigned IP addresses on a Windows network. Unlike DNS, WINS is updated automatically. Also, WINS is limited to Windows networks, while DNS is supported on a variety of platforms.

Dial-Up Configuration

- A computer must have a modem installed to communicate over telephone lines.
- The modem may be configured using the following Hayes AT Command Set:
 - ▶ **DT nnnnnnn**: Dial phone number using tone dialing.
Example: ATDT5551212
 - ▶ **n,**: Comma calls for a pause. Used especially when a number must be dialed before accessing an outside line.
Example: ATDT9,5551212
 - ▶ ***70**: Turn off call waiting
*Example: ATDT*70,5551212*
 - ▶ **A**: Answer the phone.
Example: ATA
 - ▶ **H0 (or ++++)**: Hang up immediately
Example: ATH0
 - ▶ **S0-n**: Wait 'n' number of rings before answering the phone.
Example: ATSO-5
 - ▶ **Mn**: Speaker controls.
 - M0 = Speaker always off
 - M1 = Speaker off during carrier detect only
 - M2 = Speaker always on*Example: ATM1*
- In Windows, the Dial-Up Networking (DUN) software must be installed before you can use your modem to connect to an ISP.
- In order to set up Dial-Up Networking, you must have an account with an Internet service provider.
- DUN configuration involves entering the phone number for the ISP, the user name and password for your account, the DNS names used for mail service, as well as the TCP/IP information for the network.
- Windows includes a Dial-Up Networking connection configuration wizard called "Make New Connection" in the Dial-Up Networking section of the My Computer folder (or the Control Panel). Use this wizard to set up your dial-up connection.

Browser Configuration

- A browser can be configured to open either blank or to a specific web page.
- The **home page** on a browser can be set to any web site you choose. This page can be configured to show every time the web browser is launched, or simply be the web site the browser goes to when you click the *home* button.
- **MIME** (Multipurpose Internet Mail Extension) types originally allowed e-mail clients to send and receive files other than plain ASCII text files. In browsers, they allow non-HTML documents and files to be processed, such as PDF and RealAudio files. Most of these files require browser add-ons, which will automatically configure their MIME type settings.
- Netscape Navigator has a section for setting the browser's MIME types, while Internet Explorer relies on the Windows operating system to maintain the MIME relationships.
- MIME types may need to be manually configured if two applications are both set up to handle the same type of file,

thus resulting in a conflict.

- **Cookies** are text files sent by a web server to be stored by the web browser. Cookies enable web servers to keep track of your personal preferences and can be created, updated, and read without the knowledge of the user, although each cookie can be read only by the domain that created it.
- Cookies permit such features as personalized web sites, virtual shopping carts at an online store, automated member login, and targeted web site advertising.
- Browsers can be configured to accept all cookies, warn the user before accepting cookies, or disable all cookies. Netscape may also be configured to accept only cookies sent back to the original web server, to avoid accepting cookies maintained by a third party.
- Netscape stores all cookie information in the file `cookies.txt` in each user's subfolder and can hold a maximum of 300 different cookies. Internet Explorer saves cookies individually in the cookies sub- folder of the Windows folder and can use up to two percent of your hard drive to store cookies. Cookies may be erased by editing the `cookies.txt` file in Netscape or by deleting the individual cookie files in IE.
- The **local cache** stores copies of HTML pages and graphics visited by the browser. The size of the cache is adjustable.
- Both Netscape and IE store a history of all sites visited by the browser. This history is useful in returning to web sites whose URLs you no longer remember. The browser can be configured for how many days are stored in history, and may be cleared manually.
- If a host is using a proxy server, then the browser must be configured to use the **proxy cache server**.

E-Mail Client Configurations

- Email clients must be configured with the DNS name of the SMTP mail server to send mail and the DNS of the POP3 or IMAP server for receiving mail. A server is either POP3 or IMAP, with most servers currently using POP3.
- A **signature file** is a text file that can be added automatically to the end of outgoing emails. To set up a signature file, create a text file with the text you would like at the end of your emails, then, in the appropriate place, provide the email software with the location of the signature file.
- Email clients may be configured with a reply-to address that is different from the address being used to send the email. This is useful if you are sending from a restricted account and wish to receive your email elsewhere, such as a personal account.

Using Client Utilities

Connecting to Dial-Up ISP

- To connect through a Dial-Up modem, double click the correct Dial-Up icon.
- To disconnect, double-click the DUN icon in the icon tray and select disconnect.
- A connection to an ISP can be maintained even if when a web browser is closed.

Web Browsing

- When browsing the web, a **hyperlink** is any clickable text or object that links to another portion of the page or another web site.
- Although links are designed to be displayed in blue, underlined text, this format is easily changed. The easiest way to determine if a graphic or section of text is a link is to point to it. If it is a link, the cursor will change from an arrow to a pointing hand.
- To stop loading a page, click the stop button.
- To reload a page, click the reload or refresh button.
- Most browsers can be configured to work without displaying pictures.

Using Email

- If configured properly, the email utility will automatically check for new messages at pre-set intervals. To check more frequently, you may reconfigure the email utility or click the "check messages" button.
- To reply to an email click "Reply To" and the email utility will automatically address the response and fill in the subject line.
- To send a received message on to someone else, click the forward button.
- A single message may be sent to more than one person at a time by filling in the "CC" field.
- Email may be stored on the server or on the local computer, or even both. This parameter is configurable. Be certain to know where the mail is and delete as necessary.

Using FTP

- In Windows, the FTP command line interface is started by typing `FTP` at a DOS prompt.
- The following commands are used at the FTP prompt:
 - ▶ **open**: Makes connection with server
Example: open ftp.website.com
 - ▶ **ls**: Lists files in the directory
 - ▶ **cd**: Change directory
 - ▶ **get**: "Gets" or downloads files.
Example: get file.exe
 - ▶ **mget**: Gets multiple files.
*Example: mget *.exe*
 - ▶ **put**: "Puts", or uploads, a file onto the server.
Example: put file.exe newfile.exe
 - ▶ **mput**: Uploads multiple files to the server.
*Example: mput *.exe*
 - ▶ **binary**: Sets the FTP for transferring binary files. Used for most files.
 - ▶ **ascii**: Sets the FTP for transferring ASCII files.
 - ▶ **quit**: Quits the FTP program.
- When working on a Unix or Linux server, all file names and commands are case sensitive.
- When logging in as an anonymous user, type the user name `anonymous` and an email address for the password. The email address does not have to be real as long as it is formatted like an email address (`sample@lookatme.com`).

Using Telnet

- In order to use Telnet, you must first connect to the server then login with a valid username and password.
- On a Unix server, a forward slash is used to denote directory paths.
- The following commands are available on a Unix / Linux server running the bash shell:
 - ▶ **ls**: Lists files in the current directory.
 - ▶ **cd**: Changes to a different directory.
 - ▶ **cp**: Copies a file from one place to another.
Example: cp filea /mywork/filea
 - ▶ **mv**: Moves a file from one location to another. May also be used to rename a file.
Example: mv yourfile.txt myfile.txt
 - ▶ **rm**: Removes or deletes a file.
 - ▶ **pwd**: Print working directory. Tells the path for the directory you are currently working in.
 - ▶ **exit**: Exits the telnet session.

Using Newsgroups

- In order to participate in a newsgroup, you must first subscribe to that particular news service. Subscribing involves providing a username and email address to the server.
- When providing your email address to register for a news service, it is a good idea to obviously alter the address so that humans can know what your email address is but computers harvesting email for spam purposes will not be able to add your real email to their spam list.
 - ▶ *Example:* johndoeDON'T-SEND-SPAM@aol.com
- There are over 53,000 different newsgroups covering almost every conceivable topic.
- To view new items posted on a newsgroup, you must download the listings of new topics from the server, then click on the subject lines to download the messages that look interesting to you. News service is not automatically updated for the client.
- A plus sign + next to a subject header indicates that there are a number of other postings or replies associated with this particular posting.

Security Issues

Authentication

- **Authentication** is the process of verifying that a user is the same person he claims to be.
- **Password authentication** simply requires that the user enter a password to verify their identity. This measure is only as secure as the password itself.
- A good password will have a mix of numbers, lower case letters, capital letters, and symbols. Further, a good password will not be a dictionary word or proper name associated with the user.
- **Key or card authentication** requires that the user have a physical object, such as a key or card, to further authenticate their identification to the computer.
- **Biometric authentication** allows the computer to scan a person's unique physical features as proof of their identity. Biometric authentications include fingerprints, voiceprints, face recognition, and retinal scanning.
- Biometric authentications are currently limited by cost as well as by the possibility of changes and fluctuations in a person's physical appearance.
- A **digital signature** is a value obtained by performing unique mathematical algorithms on data. This special value, and the algorithm used to obtain it, is encrypted using the recipient's public key and then sent to the recipient along with the original data, which is not encrypted. This signature assures the user that the data has not been altered in transit.
- A **digital certificate** is issued by a third party, such as www.verisign.com, to authenticate the identity of a server or an individual. This third party, called the **certificate authority**, verifies that the public key being used is valid and associated with a particular individual or organization.

Access Control

- **Firewalls** are devices (computers with specialized software or a stand-alone specialized hardware component) that prevent unwanted traffic from accessing a network.
- Firewalls using **Access Control Lists** analyze a host's IP address and refer to a special list to determine which processes the host is permitted to access. This type of access restriction is subject to **IP spoofing** (hosts using a false IP address to gain unauthorized access).
- **Dynamic packet filtering** allows a firewall to keep track of the data packets it is transferring and determine whether or not a packet actually belongs in the sequence of packages.
- The **protocol switching** technique translates TCP/IP data into a different network protocol (such as IPX/SPX) to limit the effectiveness of certain TCP/IP specific attacks.
- A **demilitarized zone** is a section of network made accessible to the public and to the internal LAN, thus providing public access to a portion of the network while securing the remainder of the network from public access.
- **Proxy servers** make Internet requests on behalf of internal hosts, and then forward the received data on to that host. This setup protects the hosts on the LAN because they are never directly exposed to the Internet.

Encryption

- **Encryption** involves translating clear-text into cipher-text using mathematical algorithms.
- Encryption works by one computer applying an algorithm to encode a message, sending the coded message, and then decoding the message using a key.
- A **private key** is a single, secret key shared by two individuals. Both individuals have the same key and use it for encoding and decoding their messages.
- **Public keys** use one key to encrypt a message and separate key to decode the message. In this format, the code needed to encrypt a message is publicly available, but the decryption key is private. This way, anyone can encrypt a message to the recipient, but only the recipient can read it.
- Public keys use a **one-way encryption scheme**
- A key's security is measured by the bit length of the key. A 40-bit key is less secure than a 128-bit key.
- **Pretty Good Privacy (PGP)** is a free encryption scheme using public-keys of variable lengths. PGP is one of the most common encryptions used on the Internet.
- **Secure Sockets Layer (SSL)** is a method of establishing a secure connection between a server and a client. With SSL, the server uses a digital certificate to identify itself to the client, and then the two computers collaborate on a private key to be used for the remainder of the transmission.
- **S/MIME**(Secure Multipurpose Internet Mail Extension) is an encryption technique for email. This technique uses the recipient's public key to encode the email, then attaches a digital signature to the data to ensure that the data has not been tampered with during transmission.

Auditing

- **Auditing** is the process of keeping track of the events that occur on a system, including successful logins, failed login attempts, changes in user privileges, remote logins, and system shut-downs and restarts.
- Auditing information is typically stored in **log files** on the computer.
- Log files may be analyzed automatically by a software application to aid in administration and help alert administrators of possible problems.

Attacks and Suspicious Activities

- People may attempt to break into a network for a variety of reasons, including the desire to do harm to the organization, a yen for personal profit, or merely for fun.
- There are two main types of attacks: **denial of service** (DOS) attacks and information theft or destruction.
- In a **denial of service** attack, a server is overwhelmed with bogus requests, to the point that it is unable to service the legitimate requests it receives.
- Attackers may gain access to a network through **social engineering**, which is simply persuading legitimate users to give out their user name and password on a network. Such attackers often pose as an administrator, though a true system administrator may access any account without having to know the user's password.
- The **brute force** method of gaining access uses a computer program to try every possible letter combination until it finds a working user name and password, or until it is able to duplicate the key used for decryption.
- Flaws in software design can also open up gateways for attackers to enter or harm a system. These "bugs" in software are often fixed in patches released by the software maker.
- Some attacks do not require a user name and password; instead, they simply exploit the very services the server was designed for. Such attacks include **mail flooding** (signing a system user up for hundreds of mailing lists or sending them very large messages), **ping floods** (issuing a multitude of pings, or requests for connection verification, to a server), a **ping of death** (a ping package larger than 65,536 bytes), and **SYN floods** (filling the TCP/IP buffer with SYN requests for connections that are never answered by the requesting host).

Network Security Requirements

- To maintain security with an **Internet** connection, use some type of firewall to keep out unauthorized traffic, and send sensitive information only within a secure environment using proper encryptions.
- To secure an **intranet**, virtual private network, or LAN which is accessible only to employees (some of which may become upset at the organization), restrict each user's access to only the information they need to access, back up data regularly, use an anti-virus program to prevent virus infections, require periodic password changes, and educate users on basic security issues and precautions.

Web Programming

Network Software

- In a **client-server** design, one process (the server) serves information to another process (the client). Typically the server process is on one computer and the client process is on another computer, but it is possible to have both on the same computer.
- On the web, the web browser is the client process that requests and process the information from the server.
- Web browsers are capable of interpreting and displaying HTML and other markup language content, special graphic formats, executing scripting language codes, and Java applets or ActiveX Controls.
- A **plug-in** is a program which allows a browser to display types of data it was not originally designed to interpret.
- Web server software is responsible for serving data files requested of it. However, **server extensions** may be used to expand the capabilities of the web server software.
- **Server-side scripting languages** are codes associated with HTML data that the server will process before sending out a web page. PHP3 (and now PHP4) and ASP are two of the most popular types. These scripts are often used to integrate a database with a web site.
- **Forms** are used to allow web users to enter data on a web page and send it to the web server for processing. Such data transmission relies on **CGI** (Common Gateway Interface), which is a standard for defining how data is packaged and sent over the network.
- **Enterprise computing** deals with sharing information among various organizations to improve the way they do business.
- **Distributed computing** is a method of spreading the modules of an application or task across several computers, thereby sharing the work load and effectively increasing the speed at which the task is accomplished. This distribution may be accomplished using **Java Beans** or **Component Object Models (COM)**.

Programming Languages

- Programming languages may be full featured development languages (such as Java, C, C++, and Visual Basic), scripting languages (like Perl, PHP, and ASP), or a combination of the two (like Python).
- Full featured development languages are used to write stand-alone programs. These are usually compiled languages, meaning the code you type must be processed through a compiler that will translate the human-readable code into machine language. Java and Python are not technically compiled languages, since they are translated into an intermediate byte code and require another interpreter to be run in order to execute the code.
- The **Java** programming language is different from **JavaScript**. The Java language is intended to be platform neutral, thus allowing a program to be written in Java and run on any operating system without changing the program. This is performed using the **Java Virtual Machine**, a program that mediates between the Java code and the underlying system.
- **Python** is a development language noted for both its power and its simplicity. Like Java, Python code is not directly compiled but is translated into byte code before being executed, thus making the code more portable. The simplicity of Python makes it attractive as a first language for those interested in learning to program.
- **C** and **C++** are two of the most powerful and complex programming languages. These languages are used to create programs that execute quickly and are platform dependent (will run on only the type of machine it's code was written for). C++ is generally considered the successor to C, providing object-oriented features not available with C.
- **Visual Basic** is a language used to quickly create programs for Microsoft Windows, and some Web functions. Visual Basic is typically easier to use, but much less powerful and slower than other full featured programming languages.

- A **scripting language** is a language whose code is not compiled. Instead, the code is interpreted by a program running on the host computer.
- **Shell scripting** languages are the basic languages built into operating systems for automating routine tasks. These languages have limited usefulness in web development.
- **Perl** is one of the most popular web scripting languages. It is a powerful open-source language that is free to download and use. Perl is sometimes referred to as the 'duct tape' of the Internet.
- **Server-side scripting languages** are scripts embedded in text documents that are intended to be processed by the server before being sent to the client computer. These scripts may be used to insert a date, a portion of a web site, or provide a wide variety of other functions. Generally server-side scripts are much faster than client-side scripts. Since the code may be executed at the server instead of having to be downloaded by the client before execution. In order to use a server-side script, the server must support the language you wish to use.
- **PHP** is a popular and powerful open-source server-side scripting language. It is often used in conjunction with databases (usually MySQL) to integrate database connectivity features with a website.
- **ASP** (Active Server Pages) is Microsoft's version of a server-side scripting language.
- **Client-side Markup Languages** are mark-up languages interpreted by the client browser. These include typical HTML; **XML** (extensible markup language), which allows programmers to add their own tag definitions similar to HTML; and **VRML** (Virtual Reality Modeling Language), which allows a programmer to use tags to define three-dimensional space layouts.
- **Client-side scripting languages** are scripts embedded in an HTML document which must be processed by the browser on the client computer. **JavaScript**, a derivative of the Java language, and **VBScript**, a derivative of Microsoft's Visual Basic, are the two most popular client side scripts.

Multimedia

- Multimedia components, such as music and videos, may be incorporated as **streaming** or **non-streaming** files. Non-streaming files, such as MPEG and AVI files, must be fully downloaded before they are processed and displayed. Streaming media, such as RealPlayer, begins to play the media file after only a portion has been downloaded. While it is playing this portion, the remainder of the file continues to be downloaded in the background, allowing the user to experience the media with less wait time. Streaming media uses a buffer to store a few seconds worth of media ahead of what is being played in order to compensate for fluctuations in data downloads.
- **Macromedia Shockwave** is a browser plug-in used to display animations and interactive games. It has been succeeded by **Macromedia Flash**, which uses vector graphics to create smoother animations and smaller file sizes.
- **RealPlayer** is a browser plug-in designed especially for playing streaming music and video files.
- **Windows Media Player** and **Apple Quick Time** are two more plug-ins used for sound, video, and animation files. Quick Time files have a **.MOV** file extension and Media Player files may have a **.AVI** extension.
- **MPEG** (Motion Picture Experts Group) is a file standard for full motion video and sound. The popular **MP3** is a file extension associated with MPEG audio layer 3, a coding and compression scheme that produces small, high-quality sound files.

Graphic Formats

- In order to make picture files small enough to fit over the Internet, the files must be compressed. Different compression schemes offer a variety of benefits and drawbacks.
- A graphic compression scheme that does not change the image or cause it to lose any quality is called **lossless**, while a scheme that results in a change in the image data or a loss of quality is called **lossy**.
- The **GIF** (Graphics Interchange Format) is a proprietary graphic format developed by CompuServe that has been around for years. GIF uses a lossless compression scheme, supports up to 256 (8 bit) colors, allows for simple animations, and supports both interlacing and transparency.
- **JPEG** (Joint Photographic Experts Group) is a lossy compression scheme that supports 16.7 million (24 bit) colors. Standard JPEGs do not support transparency or interlacing. However, the progressive JPEG format does allow interlacing of JPEGs. JPEGs are best used with photos that have a lot of color.
- **PNG** (Portable Network Graphics) was developed in response to legal problems over the GIF format. The PNG format is a lossless compression scheme that allows either an indexed 256 color image or a 16.7 million color image. PNG also supports interlacing, 256 levels of transparency, gamma information, and a variety of other options. PNGs may be used with any graphic.
- Other graphic formats include **BMP** (Windows Bitmap), which is not a compressed image and is supported only by Internet Explorer; and **TIFF** (Tagged Image File Format), which is not supported by browsers much either.

Other File Formats

- **ZIP** for Windows, **BinHex** (.HQX) for Macs, and **GZIP** (.gz) for Unix / Linux are popular file compression formats.
- **TAR** is a Unix / Linux utility often associated with zipped files. The TAR utility takes a number of different files and combines them into a single file. Technically, the utility does not do any compression.
- **PDF** files are document files designed to be displayed using Adobe Acrobat Reader. PDF has become a popular way of sending manuals, texts, and other "books" over the Internet, since the Acrobat reader allows for browsing of the document and printing of all or selected pages.
- **RTF** (Rich Text Format) is a basic text formatting file extension. Most word processors should be able to read an RTF file and correctly display the document.
- **PostScript** files are files formatted for printing on a PostScript compatible printer. PDF is largely replacing PostScript files on the Internet.

Databases

- A **non-relational database** is very simple, flat databases with only one table or list.
- A **relational database** is a database that has more than one table, or data list, and has defined relationships between one or more fields of the tables.
- Database Management Systems (**DBMS**) are programs that manage adding, deleting, and organizing database info.
- A **database server** is a program that responds to queries and supplies matching data from the requested database. Brand-name database servers include Oracle, one of the most powerful and respected database servers; Microsoft SQL Server; MySQL, a basically free database server popular on the Linux / Unix platform; and PostgreSQL, a powerful open source database system.
- Information is obtained from databases using **SQL** (Structured Query Language), a standard descriptive language designed especially for databases.
- Integrating a database with a web site can create dynamic page content, thereby saving a great deal of programming time. For example, a sales web site may have all product data in a database and have a basic web page template created with a scripting language such as PHP or Perl. When a user searches for a particular item, the server searches the database for the correct data and then dynamically merges that data with the web page template.
- Databases may be also be integrated with a web site to accept user input through forms, then either store the data for future use or act upon the data as part of a query.

Site Function and Design

Server / Client Interactions

- Studies indicate that a typical Internet user will become frustrated and consider leaving a site if it takes longer than 10 seconds to load.
- When a user types in a URL, the following events occur before they see the web page:
 - ▶ URL looked up in DNS and connection created between client and server
 - ▶ Request waits in queue on server
 - ▶ Server responds to request
 - ▶ Data transmitted back to client
 - ▶ Client software decodes and then displays data
- The time it takes for a web page to be downloaded by a user may be estimated by dividing the size of the page in bits by the bandwidth in bits. This will yield an ideal data transmission time; network lags, DNS lookup and browser translation time must also be factored in to provide a better estimate of the time needed to download a page.
- Possible reasons for a web site request failing include the following:
 - ▶ The URL is not valid
 - ▶ The client software is not configured correctly
 - ▶ The client's Internet connection is down
 - ▶ The server computer is down
 - ▶ The server's Internet connection is down
 - ▶ The server is currently too busy
 - ▶ The server's DNS files are corrupt
- • **MIME-Types** identify to a web browser what type of application a file is associated with. The MIME types are stored in a database on the client web browser. If a server sends out a file that has a MIME type unknown by the browser, the browser will either prompt to download the proper plug- in software, or display an error.

Content Planning

- A web site should be designed for its intended audience, keeping in mind their skill level, available technology, and interests.
- Advanced features, such as animation or multimedia, should be used on a page only if they serve the page's purpose. Adding advanced features to a crummy page will only make it worse.
- Avoid using plug- ins that is available for only a certain operating system.
- For maximum compatibility, design the web page to fit within the 640x480 resolution used by a number of monitors (keeping the page at a maximum width of 640 pixels). Otherwise, if the page is designed to display larger than the screen resolution of the client monitor, the user will be forced to scroll in order to view the entire page.
- Determine a policy to apply when designing your web site. Policies may include **captive audience**, which designs for a specific browser and plug- ins determined by the designers; **lowest common denominator**, which designs so that the page will work in any browser, including text-only browsers; **85% policy**, where the site is designed so that most people can see it, but does not worry about reaching users without the technology for the site; or **adaptive content**, which is designing the page so that it can both take advantage of advanced features on clients that support them, and dismiss these features for clients unable to use them.

- Two ways of making a web site suitable for both high and low end browsers are **differential content** and **graceful degradation**. Differential content involves creating two web sites, one with advanced functions and another with only basic functions, and diverting users to the site that suits their needs. Graceful degradation is designing a single site in such a way that advanced features will be implemented if the browser supports them, but will be ignored if the browser does not support them.
- Policy should be written and have specific details on page sizes, browser support, multimedia, and any other basics of the page design.

Server Planning

- A **queue** is a set of objects waiting in line. In relation to servers, a queue is the number of requests waiting to be serviced.
- The **utilization rate** of a server is the rate of new requests divided by the maximum rate at which the requests can be fulfilled. A server that can handle a hundred requests per second and is receiving 50 requests per second has a utilization rate of 0.5.
- If the utilization rate approaches or exceeds 1, the computer will fall behind, causing delays for clients trying to access the server. To maintain a high utilization rate, you must reduce the number of requests coming in or increase the number of requests the server can handle.
- The rate of fulfilled requests is equal to the number of **active children** divided by the time each request takes. Active children are processes (or threads) being acted upon by the computer.
- Network performance is affected by the network bandwidth, amount of RAM, CPU speed, disk input/output speed, and database connections. Any one of these areas may impede performance and produce a bottleneck. The most common bottleneck for a web server is the network bandwidth.
- Redundancy is one the best back- up supplies to have in case of an emergency. Having a second server, Internet line, power supply, or other component always available can mean the difference between having no down time and being down for days.

Caching

- **Caching** is storing information in a quick-access area (such as RAM or a local server)to make it easier to access in the near future.
- In order to be useful, the files in a cache must be the most recent, up-to-date version. To keep information current, the cache may send out a query to the original document, checking if the cached copy is still the same version as the most recent original. If it is out-dated, the cache will download the newest version.
- Most web browsers perform caching by saving all accessed web files on the hard drive. If the user returns to a cached web site, the browser will then determine whether to use the cached copy or reload the original web site. The rules a browser uses to determine when to use the cached file can be set within the browser configuration system.
- A proxy server may provide caching for the network it serves, thus increasing speed for the users on the network. Proxy caching may be established on a LAN, on a dial-up ISP, or even among other proxy servers to create a regional proxy system.
- **Reverse proxy servers** (or caching Web servers) sit between a web server and the rest of the network, providing cached copies of the files on the main web server. Using a reverse proxy helps to lower the demands placed on the main server and may speed system response time. However, reverse proxies are not very common.

Searching and Indexing

- A **site map**, or site index, is like a table of contents for a web site. The site map lists the main sections of the site and the topics or pages available for each section.
- A **simple search** searches a database of web sites for all occurrences of a specified word. A simple search can yield thousands of web sites, which is more than most users want.
- Different search engines use different rules to develop more complex searches. Commonly, a + before a word indicates that it must exist in the results, a - indicates it must not exist in the results, and the word AND signifies that both the first and second word must appear in the results.
- A **reverse index** is a list of all the unique words that appear in a set of documents linked to the documents which contain those words. Reverse indexes make it much easier to locate the documents that contain a certain set of words.
- Searches may be refined and improved by focusing only on the information in META tags, finding sites that use the desired word most frequently; by delivering the most popular web sites at the top of the list, or by using various language concepts to better refine the search.
- Language concepts used by search engines include omitting unnecessary words (a, an, the), **stemming** words (recognizing the various forms of a word as relating to the same concept, such as bake, baked, and baking), classifying the meaning of a word based on its context (so that a search for "fly fishing" does not yield results about button fly jeans, planes that fly, and the common house fly).
- Different search engines employ different techniques to yield their results. Some rely on various data searching techniques, while others incorporate human editors and some sell their search results to the highest bidder.

Troubleshooting

Pre-Testing Web Site

- Before a web site is made public, it should be tested in a private space to assure that everything works.
- Private testing may be done on a non-public portion of the regular web server. This option allows the developer to incorporate and test server functions (such as server-side scripting, CGI, and style-sheets), as well as make the site available to others for critique before going public.
- A **staging server**, or separate web server for privately testing a web site, may be used if site testing would conflict with the active site if done on a private portion of the active web server.
- When moving tested files from the private testing area to the public server, be certain to maintain the integrity of all links.
- **Storyboards** aid in site development and testing by recording how each page on the site should look and connect to other pages within the site. A tester can then use the storyboard to compare the actual site to the intended plan.
- Test all links on every page of a site to make sure they function correctly. Software is available to run these tests automatically and return a list of all broken links.
- Before making a website public, be certain to test it in a variety of browsers. Since browsers are notorious for not supporting the same standards, a site that works well in one browser may be un-readable in another. Some minor code changes can often correct some of these problems.
- If a page uses CGI scripts or forms, test it by entering valid information, and then test with a wide variety of invalid information (blank entries, overly long entries, garbage text, and so forth) to assure the script handles errors well.
- Load testing software can help test how well equipped a server is to handle a large volume of traffic to a web site.
- If the expected volume of traffic on a server is more than the server can handle, a **mirror site** would be helpful. A mirror site is a separate web server that holds the same web site files as the original. Mirror sites may be used to ease the load on a main server, or to speed transactions by providing geographically closer access to the files.

Troubleshooting

- In determining and solving a problem, it helps to follow a procedure such as the following:
 - ▶ 1. Identify the exact issue.
 - ▶ 2. Re-create the problem.
 - ▶ 3. Isolate the cause
 - ▶ 4. Formulate a correction
 - ▶ 5. Try the corrective measure
 - ▶ 6. Test the solution
 - ▶ 7. Document the problem and its solution
 - ▶ 8. Give feedback
- If a user complains about a problem, attempting to recreate the problem can help determine if it was a random and isolated incident, a true problem, or simply a user error.
- Asking specific questions can help isolate the cause of a problem. If a user is having troubles, determine if they were ever able to perform the procedure, if so what has happened since they were last able to do it, and if the problem is effecting other users as well. Questions along these lines can help to isolate the cause of an error.
- After a problem has been isolated, determine and write down possible fixes. If possible, rank the fixes in the order of their likelihood of success and try the most likely candidate first. Be sure to keep track of what fixes you have tried and how each attempt turned out, including the fix that actually worked.
- If a browser encounters a file it is unable to process, it will give an error message.
- Error messages such as *No DNS Entry*, *Connection Reset by Peer*, *Server Not Responding*, and *File Contains No Data* indicate a problem communicating with the server. These may be indications of high network traffic, mistyped address, or a more serious problem.
- HTTP response codes come in four basic varieties to help diagnose the source of a problem:
 - ▶ **200s** -- A successful transmission.
 - ▶ **300s** -- The file has been moved.
 - ▶ **400s** -- An error on the client's part.
 - ▶ **500s** -- An error on the server's part
- If a site is not accessible to anyone, then there is a problem with the connection to the web server.
- If the site is inaccessible by a group of people, but accessible by others, then there may be a problem with their Internet connection, even if they are able to access other sites. It is possible they have been disconnected from a portion of the Internet.
- If a user cannot access anything on the Internet, then the problem is with his connection only.
- Web servers use file permission settings to limit user access. These file permissions determine whether a "general" user is permitted to read, write, or execute a file. If the permissions are set improperly, they may render portions of the website inaccessible.
- Log files keep track of all of a server's transactions. These log files can be quite useful in identifying the source of a problem.

Network Utilities

- **ARP** (Address Resolution Protocol) resolves IP address with hardware (MAC) addresses. The ARP utility displays a table showing which IP addresses correspond to which MAC addresses. This information is very useful for determining if more than one host has been assigned the same address, thus causing conflict.
- **netstat** displays the statistics on all TCP/IP transmissions currently running on your machine.
- The **Ping** utility simply sends a test message to a given IP address to see if the addressed host is connected and responding.
- **winipcfg** and **ipconfig** are TCP/IP configuration utilities that allow a user to see his IP address configuration and make basic changes.
- The **tracert** (trace route) utility actually traces the path taken by a packet to reach its destination. The utility reveals the IP address of every device it passes through and the time it took to be sent on to the next device.
- **Network analyzers** are software tools that allow a user to monitor and manipulate the traffic on a local network.

Viruses

- **Viruses** are malicious programs that cause damage on a computer system and then spread to other systems.
- A virus **in the wild** is one that is currently spreading around a network or the Internet.
- Viruses come in two main types: **macro viruses**, which are typically limited to specific applications, and **boot sector viruses**, which rewrite the boot sector on a hard drive making it appear as if the hard drive itself has been erased.
- Antivirus programs use definition files to identify known viruses and an engine to scan files, compare them with the known viruses, and clean the files as necessary.
- Virus definitions should be updated at least monthly.
- Anti-virus scans may be either **on-demand**, meaning they are run only when the user specifically requests them (which may mean as part of an automatic maintenance schedule), or **on-access**, meaning the virus scan is run whenever a file is accessed.

Legal and Business Issues

Legal Issues

- **Intellectual property** refers to "products" created by a person or company. These products are usually intangibles, such as stories, images, software, and music.
- Intellectual property is protected through **copyright law**, which insures that the original creator maintains control of his creation and receives compensation from those who wish to use it.
- Creations that are not copyrighted are in the **public domain** and may be used by any one for any purpose.
- Copyrights may be obtained by common law or by official registration. Simply writing "Copyright" or ©, the date, and the author's name on a work creates a common law copyright. For an official copyright, a work must be filed with the US Copyright Office.
- In general, a copyright is international. If a work is copyrighted in one country, it is automatically copyrighted in all other countries that adhere to the Berne Convention treaty.
- **Fair use** of copyright material allows the material to be used without consulting the copyright owner as long as the use is considered "fair". Fairness typically means that the owner is not deprived of an opportunity to make money.
- To use copyrighted material beyond "fair use", the user must obtain permission from the copyright owner. The owner may deny permission, allow use free of charge, or require a **licensing fee** to use the copyrighted material.
- A copyright may be transferred to a different owner with a written statement of assignment.
- Copyright violations are typically a matter of civil, rather than federal, law.
- A **trademark** is anything that identifies or distinguishes a product or service. Trademarks may be declared by simply using the TM symbol or may be formally registered with the government. A registered trademark is denoted by ®.
- Physical devices and processes are protected with **patents**, which grant the creator exclusive control over the use of his creation. All patents must be formally registered to be truly patented. A device or process will merit a patent only if the product is useful, new, and not an obvious solution.
- **Patent pending** status simply means that the creator has applied for a patent but it has not been granted yet.

The Open Source Movement

- Most commercial software is sold in executable form only, meaning there is no way to look at the original code. **Open Source Software** is software that includes the original code, allowing users with programming skills to make changes to the program.
- Some Open Source Software is simply public domain software, meaning anyone may use it for free. Other Open Source Software may be protected by different types of copyrights, such as **copyleft** or the **GPL**.
- Under copyleft and GPL restrictions, software is free to use, distribute, and modify. However, these licenses require that anyone who modifies the original program releases their work under the same license as the original work, thus assuring that the software will remain free and open source.
- Open source software is built on the idea that allowing programmers world-wide to modify a program will result in better software than can be produced by only a few programmers working in isolation.

Business Issues

- Be aware of differences in language and culture when developing a site or communicating with international users.
- When dealing with international payments, credit cards supply the easiest currency transfers.
- In international business, be aware of the need to conform to the law in both countries, including export or import restrictions.
- **Push technologies** send information from a server to a client without the client having to directly request the data. This technology is used for automatic news updates, mailing lists, user-specific advertising, and everyone's favorite, SPAM.
- **Pull technologies** rely on a user or client browser to request information. These requests can be automated by the HTML code so that a user's screen is refreshed automatically at set intervals. Since the client browser is making the request, this is a pull technology.
- A **banner ad** is a graphic link placed on one web site advertising a different web site. Site owners may charge advertisers for space on their web page, or may work out some other deal for showing the advertisement.
- **Webrings** are collections of sites focused on a central theme. Site creators subscribe to the webring, usually providing a link to the webring home page index and possibly to other sites on the ring.
- In business, be aware that the Internet can be used by your company as a positive promotion tool, as well as by customers who have a complaint against your company. Monitor the consumer sites, new groups, and mailing lists that others might use to complain publicly about your company so that you have an opportunity to respond.
- Commerce sites on the Internet focus on either business-to-consumer (B2C) sales or business-to-business (B2B) sales.
- **EDI** (Electronic Data Interchange) is a technology standard for exchanging data across networks. EDI is used to make electronic fund transfers, including transfers made using ATMs (Automatic Teller Machines).
- **Online cataloging** involves making a business' product catalog available online. This merchant system improves upon standard paper catalogs by allowing for easier catalog updates, customizing sales to the individual as he browses, and automatically tracking user interactions.
- Business web sites should be designed to provide customer self-service, allowing the customer to search through owner's manuals, walk through trouble-shooting guides, contact customer service through email, and in general relieving some of the load from customer service personnel.

GNU Free Documentation License

Version 1.1, March 2000

Copyright (C) 2000 Free Software Foundation, Inc.
 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA
 Everyone is permitted to copy and distribute verbatim copies
 of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other written document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondly, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you".

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (For example, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, whose contents can be viewed and edited directly and straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input

to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup has been designed to thwart or discourage subsequent modification by readers is not Transparent. A copy that is not "Transparent" is called "Opaque".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML designed for human modification. Opaque formats include PostScript, PDF, proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies of the Document numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a publicly-accessible computer-network location containing a complete Transparent copy of the Document, free of added material, which the general network-using public has access to download anonymously at no charge using public-standard network protocols. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through

your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A.** Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- B.** List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has less than five).
- C.** State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D.** Preserve all the copyright notices of the Document.
- E.** Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F.** Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G.** Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H.** Include an unaltered copy of this License.
- I.** Preserve the section entitled "History", and its title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J.** Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K.** In any section entitled "Acknowledgements" or "Dedications", preserve the section's title, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L.** Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M.** Delete any section entitled "Endorsements". Such a section may not be included in the Modified Version.
- N.** Do not retitle any existing section as "Endorsements" or to conflict in title with any

Invariant Section.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties--for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections entitled "History" in the various original documents, forming one section entitled "History"; likewise combine any sections entitled "Acknowledgements", and any sections entitled "Dedications". You must delete all sections entitled "Endorsements."

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, does not as a whole count as a Modified Version of the Document, provided no compilation copyright is claimed for the compilation. Such a compilation is called an "aggregate", and this License does not apply to the other self-contained works thus compiled with the Document, on account of their being thus compiled, if they are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one quarter of the entire aggregate, the Document's Cover Texts may be placed on covers that surround only the Document within the aggregate. Otherwise they must appear on covers around the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License provided that you also include the original English version of this License. In case of a disagreement between the translation and the original English version of this License, the original English version will prevail.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See <http://www.gnu.org/copyleft/>.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.