CHAPTER 5: TELECOMMUNICATION TECHNOLOGY

➤ TELECOMMUNICATION IN DIGITAL LIBRARIES

➤ WEB SERVER TECHNOLOGY
5.1 MODES OF TELECOMMUNICATION IN DIGITAL LIBRARIES:

Digital libraries or digital collections are becoming ubiquitous in the information arena. Certainly the most fascinating and challenging with regard to access are those, which are predominantly or solely image based. A prime example of such collections is the American Heritage at The Library of Congress. Improvements in computer and telecommunications technologies have enabled information professionals to include them in their offerings either internally within the organization or as links to external sites. The need to access such collections is not only vital to basic research, but also invaluable to human communication in the digital age. The adage "a picture is better than a thousand words" has never been more appropriate as when it is used to refer to digital resources on the Web. The Internet, particularly the Web, has made it possible to access such collections and has in fact accentuated the creation of remotely accessible image-based digital collections. But unlike text-based information access, image intensive digital libraries are fraught with downloading and uploading bottlenecks. Careful design of distribution and receiving information systems is needed. Various alternatives have been used to alleviate the bandwidth bottleneck, including: cable modems, frame relay, ISDN, digital subscriber line (DSL), satellites, and high-speed analog modems. Comparative analysis of various alternatives of Digital library access is presented along with the information of their development expenditure and incoming/outgoing performances;

<table>
<thead>
<tr>
<th>Technology</th>
<th>Deployment</th>
<th>Monthly cost ($)</th>
<th>Performance incoming</th>
<th>Performance outgoing</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-modems</td>
<td>Universal</td>
<td>20</td>
<td>56kbps</td>
<td>33.6kbps</td>
</tr>
<tr>
<td>ISDN</td>
<td>Widespread</td>
<td>50-130</td>
<td>128kbps</td>
<td>128kbps</td>
</tr>
<tr>
<td>C-modems</td>
<td>Limited</td>
<td>30-65</td>
<td>1-5mbps</td>
<td>33.6kbps-2.5mbps</td>
</tr>
<tr>
<td>DSL</td>
<td>Limited</td>
<td>49-1,200</td>
<td>144kbps-9mbps</td>
<td>64kbps-9mbps</td>
</tr>
<tr>
<td>Satellite</td>
<td>Widespread</td>
<td>40-130</td>
<td>400kbps</td>
<td>33.6kbps</td>
</tr>
<tr>
<td>F-relay and T... series</td>
<td>Widespread</td>
<td>300-3,000</td>
<td>56kbps-45mbps</td>
<td>56kbps-45mbps</td>
</tr>
</tbody>
</table>

Alternatives for digital library access
It summarizes the latest developments in these technologies and how they can be used by various types of professionals and end-users in accessing digital libraries. While in a given information systems environment, the mode of implementation of the transmission technologies in the physical networks significantly affects the ease with which digital libraries are accessed, the basic access terminal has also to be reckoned with. Thus three major elements interplay to determine the final rate at which multimedia in the digital library are effectively transmitted to the end-user. First, an effective national telecommunications network, that includes signaling and switching techniques must be in place and operating efficiently. Second, an institutional distribution network which delivers information to the end-user must be available. Third, the access terminal used, which for many end-users is currently an intelligent personal computer (PC) must have the capacity to handle a variety of images and sound. Assuming that the intelligent PC is readily available to the end-user, albeit in a variety of flavors and degrees of regional penetration per capita, the telecommunications networks become the more significant elements.

5.11 The Communication Technologies:

5.11.1 High-End Analog Modems:

Starting with the lowest on the totem pole, Plain Old Telephone Service (POTS) as the delivery network for digital images, we have an analog modem based access system with maximum delivery speed of 56kbps. This is slow for multimedia and is certainly inadequate for teleconferencing, video conferencing, and animation, which may tap into digital libraries for source images and sound. However, the majority of Internet (Web browsers, especially in homes, still use analog modems to access digital libraries. There are several advantages of using analog modems. First, with a typical lowest recurring cost, they are affordable for small library and users. Second, they are part of a universally available POTS network and thus accessible to most users. Third, they are easy to install and in many instances come already installed in new deliveries of computers. Finally, they are easy to maintain and operate. The greatest disadvantage of analog modems is definitely speed for downloading and uploading of digital images,
which eliminates them as viable alternatives for optimum data delivery in medium and large organizations.

5.112 T... Series And Frame Relay:

Since the technologies in this section are very complex and expensive they are dealt with here to dispose of them as viable alternatives for most end-users or small to medium organizations. Some of the critics question on their cost benefit advantages even for large organizations, especially in their small branches for telecommuting employees that may have access to cable modems, satellite, or DSL. T... series (i.e. T-1, T-2c T-2, T-3 T-4) are telecommunications standards that define long distance digital lines used mainly for data communications. AT&T, and several North American communications carriers own T... series lines for leasing, many of which are fiber optic-based and digital. Their transmission speeds range from 1.544-274.176Mb per second (mps). Many large organizations lease dedicated or shared fiber optic-based digital T... series trunk lines for their own proprietary information networks such as metropolitan area networks (MANs) or intranets. They are very fast and ideal for accessing distant digital libraries - if only they were affordable. The main disadvantage for T... series are the charges, which may range from $300-$3,0000 or more per month. Frame relay is another method of data transmission with transmission rates ranging from 56kbps-1.5Mbps, but within the same price tag as the lower end of the T... series.

5.113 ISDN In Digital Transmission:

To augment the POTS, ISDN was one of the earliest broadband signaling systems developed in the 1980s, by the telecommunications carriers. It does indeed transmit images at a faster rate than the analog modem-based networks. The ISDN signaling algorithm works on the regular telephone network and requires ISDN switches at the telephone company's central office and an ISDN capable terminal at the user end. As it establishes a virtual digital network, it achieves high efficiency, for there is no signal conversion, comparable to the analog POTS, at either end - carrier or subscriber. Its two main flavors, the basic rate interface (BRI) and the primary rate interface (PRI) carry
signals at maximum rates of 128kbps and 1.544Mbps respectively. Such rates are a great improvement on the POTS analog modem-based digital image transmission.

Digital library access via ISDN

Although it has been on the market for almost 20 years as an alternative technology for transmitting digital images, ISDN has been slow to develop. Among the reasons often given are cost of equipment and installation, uneven deployment, and lack of "trigger" user application at its critical developmental period during the 1980s. Even in countries like the USA, with highly sophisticated telecommunications infrastructures, ISDN has not taken off as anticipated by the telecommunications carriers. Internet user demand for high bandwidth to accommodate digital images, as well as the threat of other technologies such as cable modems, accounted for the heightened telecommunications company’s interest in ISDN in the 1990s. But by the time ISDN became affordable, alternative technologies like DSL and cable modems were on the market and offered a better multimedia operational environment with higher bandwidth and faster throughput.

5.114 Cable Modems In Digital Transmission:

Cable modems are one of the most promising technologies for accessing digital libraries for the twenty-first century. The monthly cost for the technology is lesser than others. Most of the deployed systems are based on the existing CATV networks, which are plagued by two intrinsic design problems when modified to accommodate bi-
directional data communications. First, they were never designed with a symmetrical two-way point-to-point communication mode in mind. Second, they have been deployed on a regional or local basis with virtually no common standards and little quality control. Ordinarily, television programs are broadcast down the cable simultaneously in what is called a one-to-many mode, normally called the downstream. Because of the differential in the volume of signals upstream and downstream, most CATV companies use a split frequency spectrum for communication in either direction.

![Diagram of cable modem setup]

**Digital library access by cable modem**

The downstream is allocated 50-750MHz of the spectrum, which is the regular analog television broadcast band. This band is traditionally a one-way, or technically a simplex, method, which over the years of technical experience has been perfected in terms of error checking, trouble-shooting, and signal amplification requirements. Depending on need, digital data or video may be sent downstream to users (in homes or offices) using one or more of the unused channels within the broadcast band.

The upstream band used for data communications between the user terminals (PCs) via cable modems to the head-end is allocated the 5-40MHz part of the spectrum. Thus, the upstream portion of the network uses a many-to-one data communications mode, whereby data is sent to the head-end controller by the individual nodes on the network. It has a lot of noise, usually known as ingress. Such noise is external to the
network and comes from home equipment such as dryers, mixers, thermostats, as well as ham radios and is caused by the low frequency on the upstream channels. Cable network providers are developing filters to minimize such noise.

At the users level, cable modems tune to relevant channels, demodulate the signals transmitted from the head-end and sends them to the users' personal computers. It reverses the process on the return circuit. This simplified picture of the cable modem needs some elaboration, for it has more functionality than a traditional telephone modem. The cable modem also acts as a transceiver, i.e. it receives and transmits data to and from the head-end. In addition, depending on sophistication, it may contain routers, and diagnostics management software. For Internet access, each cable modem has an Ethernet port, which facilitates connection to the computer on one side and the cable connection on the other. The user must install an Ethernet adapter inside the PC, and connect it to the cable's Ethernet port by an RJ-45 connector. Appropriate software is used to configure the PC to operate the TCP/IP protocol and make a direct connection to the Internet. Depending on the cable network connected to, maximum downstream speeds may be 500kbps-30Mbps, while upstream may have 96kbps-10mbps.

Deployment of cable modems excites many people who are struggling with slow uploads and downloads, although cable modems do have inherent problems. First, many of the cable networks are not two-way capable and some analysts suggest that only 5 per cent of cable networks can deliver broadband without major upgrades. Second, one's neighborhood may not be among those served by broadband bi-directional cable networks. Third, standards are lacking and if a customer moves from one region served by one cable provider to one by a different company, there is no guarantee of compatibility. However, IEEE P802.14 Cable TV MAC and PHY Protocol Working Group, and ATM Forum's Residential Broadband Working Group are working collaboratively to alleviate the lack of standards problem. Finally, cable modems are still quite expensive compared to telephone modems.
5.115 DSL In Digital Transmission:

Digital subscriber line (DSL) is a relatively newcomer to the telecommunications market, but perhaps the most promising to transmit digital images to the end-user in corporations as well as homes. Asymmetric digital subscriber line (ADSL) is one of the flavors of a group of several digital signaling techniques that have been developed in the last decade or so to utilize the existing telephone network to carry high bandwidth. These systems go by digital subscriber line as the generic name and are sometimes collectively treated together as xDSL, where "x" is a variable replaceable by any specific character for the particular type of DSL. Other examples from the family of DSL signaling systems include: Symmetric DSL (SDSL) and Very high rate DSL (VDSL), which differ in both the mode of transmission and bandwidth. The renewed interest in DSL is attributable to the Internet demand particularly for the Web-based large files, especially digital images.

Since it is based on a signaling algorithm that uses the regular telephone twisted copper wire network, the very foundation for POTS, DSL has a nation-wide appeal for transmitting multimedia. Some analysts have asserted that it holds the greatest potential for mass deployment as it introduces the broadband characteristics needed for high volume image-based digital data transmission on a network, which is associated with a conventionally narrow band signal transmission. ADSL's asymmetric mode of transmission is well suited for Web multimedia access. Most users request digital images from remote servers using few textual commands, thus requiring minimal use of bandwidth upstream. At the server end, the downstream, massive multimedia is often requested requiring heavy use of the available bandwidth. ADSL is designed to serve such environments. The excitement with ADSL is justified, because the copper-based telephone network is the most ubiquitous in the industrialized world. Assuming that a reasonably stable POTS infrastructure exists, ADSL is poised to be one of the gems of the telecommunications industry for the twenty-first century. The often-quoted maximum transmission speed of 9Mbps is faster than analog modem speeds as well as ISDN. The use of satellites discussed in the following section improves access to digital collections especially in regions not well served by telephone or cable networks.
5.116 Satellites In Digital Transmission:

Communications satellites are an alternative medium for transmitting multimedia. They are ideal for sparsely populated areas or areas which have not been adequately covered by the regular telephone or cabling networks due to adverse terrains. They are also the main technology of choice for linking less-developed countries to advanced countries databanks for digital image access. Comparable to microwaves, their mode of transmission is based on high-frequency radio waves with very high bandwidth. Their mechanism includes a space satellite and two or more ground stations. The earth stations used for multimedia communications are similar to dish antennas commonly used by individuals or organizations to receive television signals. Two typical terminals have characterized satellite digital data access at the end-user end. Very small aperture terminals (VSAT) are mainly for text, while T-carrier small aperture terminals (TSAT) can carry multimedia as they achieve a 1.544Mbps data rate. Most communications satellites are placed in a geostationary orbit - an orbit timed to the earth's rotation approximately 23,000 miles above the earth's surface. Within such an orbit, the satellite stays in a fixed position with regard to the earth's antennas. This obviates the need for constant re-orientation of the earth's stations in order to remain in touch with the communications satellite.

Hughes Network Systems DirectPC is one of the pioneer systems developed with telecommuters, the general home end-user, and small to medium corporations in mind. One advantage of satellite digital image communications is the high bandwidth, which is ideal for digital image transmission. Satellite digital transmission mode is essentially broadcast in nature. This implies that messages beamed to the earth may be picked up by any station tuned to a given radio frequency channel and pointed to the communications space satellite. Although this allows the satellite to send signals to many earth stations simultaneously, within its footprint, privacy of data is hard to maintain. For corporate or otherwise confidential multimedia data, scrambling or encryption is normally used. At the receiving station, such data must be deciphered using special conversion algorithms. Yet another serious problem with satellites is propagation delay caused by transmitting
signals through space. All satellite signals using a relay station in ordinary geosynchronous orbit are subject to a quarter of a second delay in both directions. While the delay may be vital to some real time interactive data applications, file transfers can be done with relative convenience.

Given the exponential growth of the Internet and other information networks especially the World Wide Web, access to digital collections containing graphics, sound and moving pictures has become imperative. Information systems designers and end-users have to grapple with downloading and uploading bottlenecks, which are gradually being solved by using emerging telecommunications technologies.

While comparatively analyzing the costs of the relevant technologies using "lowest mean" costs as the baseline. While frame relay and the T... series (T1 ... T4) are used and afforded by giant corporations, small businesses, telecommuters and or other home users, and users in isolated areas may select one of the newly emerging technologies: DSL, cable modems, or satellite, depending on location. Competition in the last few years has definitely lowered the costs. Uneven deployment is also true in other industrialized countries where some of these technologies are marketed.

5.2 WEB SERVERS:

5.21 APACHE WEB SERVER:

5.211 Introduction:

The Apache Web server is the crown jewel of the open source software movement. It costs nothing to obtain, performs better than the competition, and is thus more widely used than all other Web servers combined. The propagation of open source software is tightly analogous to biological natural selection-the Linuxes and sendmails of the world eventually end up on the cover of Time magazine and are swallowed by the hype machine, while legions of DOS utilities slide slowly but inexorably to the /dev/null of history. Apache would not be popular if it didn't work well.
Apache has another virtue not quite so common in the open source world: It is simple enough that any reasonably competent computer user can master it. This is no slur on Linux, by the way; operating systems, particularly multiuser operating systems, are hugely complex. The only way to make them accessible to the average user is to dumb them down.

The collection of tasks delegated to Apache is thankfully not quite so vast. If somebody approaches Apache with little more than self-confidence and a sense of adventure, will be relieved to know that the configuration and care of the server itself really isn't a particularly complex task. The trick, depending on on the level of experience, will probably be to grasp the fundamental concepts of the operating system, learn the commands to make the machine do what you want it to do, and absorb the jargon.

The Apache server is descended from the httpd server created by Rob McCool at the National Center for Supercomputing Applications (NCSA). In 1995, httpd was the most popular Web server in existence, but when McCool left NCSA in 1994, development of the program was stalled. A small group of Web administrators formed the core of what came to be known as the Apache Group. The members included: Brian Behlendorf, Roy T. Fielding, Rob Hartill, David Robinson, Cliff Skolnick, Randy Terbush, Robert S. Thau.

Together with contributions from Eric Hagberg, Frank Peters, and Nicolas Pioch, the Apache Group incorporated published bug fixes for httpd 1.3, added some new features, and released Apache 0.6.2 in April 1995. Since then, the Apache group, as they came to be known, has been fine tuning and enhancing the base software. Software ports are now available for virtually all the major operating systems, though the Unix platform remains the forerunner. The Apache Web server is the end result of an enormous coordinated effort by some extremely skilled programmers.
Apache exists to provide a robust and commercial-grade reference implementation of the HTTP protocol. It must remain a platform upon which individuals and institutions can build reliable systems, both for experimental purposes and for mission-critical purposes. It is believed that the tools of online publishing should be in the hands of everyone, and software companies should make their money providing value-added services such as specialized modules and support, amongst other things. It is often seen as an economic advantage for one company to "own" a market-in the software industry, that means to control tightly a particular conduit such that all others must pay. This is typically done by "owning" the protocols through which companies conduct business, at the expense of all those other companies. To the extent that the protocols of the World Wide Web remain "unowned" by a single company, the Web will remain a level playing field for companies large and small. Thus, "ownership" of the protocol must be prevented, and the existence of a robust reference implementation of the protocol, available absolutely for free to all companies, is a tremendously good thing.

5.2.12 Open Source Software:

Apache is an open source product. Traditional shrink-wrapped software typically includes only the executable object code, not the human-readable source code from which it is compiled. Apache and the other open source products include with their distributions not only the executable object code, but also the source code files from which it was created.

From the end user's standpoint, this makes a lot of sense. For example, it is a common feature that the owner of a software may have a problem in his office. A large commercial software package running on a large commercial operating system may get into a state where it stopped responding to input and may be, in fact, unkillable. He may try a stack trace and a few other things, but without the source code, there really nothing to do. And the final solution is to dump out everything and shipped it off to the software vendor for analysis. Presumably, he'll get back to us in a week or two.
Apache and the other open source software products benefit from their constant exposure to the developer community. Because there are more developers working on each open source project than even the wealthiest corporation could afford to hire, flawed source code is located and fixed more quickly. The initial quality of open source code tends to be higher than that which was commercially developed. Because open source developers are motivated by the simple love of programming, you tend to get the best of the best working on open source software. Contrast this with traditional software shops, where much of the day is spent in meetings, on the phone, and trading stocks.

5.22 W3C'S JAVA SERVER (JIGSAW):

5.221 Introduction:

Jigsaw is W3C's leading edge Web server platform, providing a sample HTTP 1.1 implementation and a variety of other features on top of an advanced architecture implemented in Java. Jigsaw is a W3C Open Source Project, started in May 1996.

5.222 Different Jigsaw versions:

- Jigsaw 2.2.2 (January 8th, 2003)

  This new version fixes several bugs, and adds performance optimizations. It also provides HTTP compliance fixes. The only new feature is SSL support.

  - SSL Support for HTTP and WebDAV
  - HTTP/1.1 compliance
  - WebDAV support
  - Many bug fixes

  This version fixes several bugs, including a security problem. It also provides new features: support for WebDAV in JigEdit, a PushCache package, and a validating filter

  - HTTP/1.1 compliance
  - WebDAV support
  - PushCache package
• (X)HTML validation on PUT
• Apache mod_asis
• Many bug fixes

• **Jigsaw 2.2.1 (April 8th, 2002)**

• **Winie 1.0.8 (March 9th, 2001)**

  Winie is a network utility to put files on the web using HTTP/1.1. The main feature of Winie is to solve the "lost update problem". Winie uses the client side api of Jigsaw. Major changes are:
  • Content-Language support
  • Bugs fixed

Common features of Winie are:
  • PUT, GET and DELETE files on the web
  • Version conflict detection
  • Retries when connection closed (like wget does)
  • Upload all files located in a directory (recursively or not)
  • Support for proxies
  • Support for metadata configuration (language, charset)

• **Jigsaw WebDAV Package (November 24th 2000)**

  WebDAV stands for "Web-based Distributed Authoring and Versioning". It is a set of extensions to the HTTP protocol, which allows users to collaboratively edit and manage files on remote web servers. This package is preconfigured as a WebDAV server.

• **Jigsaw 2.0.5 (June 5th 2000)**

  • Servlet API Support, JSDK/2.2 support
  • JSP Support
  • Image metadata extraction
Many bug fixes.

Jigsaw 2.0: It was developed by the World Wide Web Consortium (W3C), is designed to be a technology demonstration rather than a full-fledged release. It's purposely intended as a project to showcase new technologies, but in the case of Jigsaw 2.0, this Web server also ends up being more robust than the average Web server. Most importantly, though, Jigsaw serves as a useful blueprint to the future of the HTTP protocol and object-oriented Web servers.

5.223 Common Platforms:
The server will run on any platform supporting Java. At this time, it has been tested on Win95, WinNT and Solaris 2.x. Other people have reported successful use of Jigsaw on OS/2, MacOS, BeOS, Linux, AS-400 and AIX.

5.23 TOMCAT SERVER:

5.231 Introduction:
TOMCAT IS A Servlet Engine that operates the Java Server Pages (JSP) technique. The server side tech of Java becomes useful with Tomcat. The example infrastructure shown in the figure is based on the Tomcat Web Server, Java Server Pages (JSP), Java Servlets, Open Database connectivity (ODBC) and the MS-Access database. Tomcat web server includes a "build" mechanism, which separates the internal configuration details of the web server directories from the development area. The team can work on the various files for an application in a completely independent directory, and then "build" the application into the web server for testing and development. This approach also allows the developers to port the application from platform to platform, e.g. applications develop under Windows and deploy under Unix. The MS-Access database is also a good choice database for small and medium sized libraries although it lacks the sophistication of commercial products like Oracle. It is easily available with Microsoft Office and installed quickly; also, ODBC drivers supplied with Windows for
MS-Access. The best way to get familiar with the infrastructure is to work with the simple example provided with the Tomcat bundle.

5.232 Installing Tomcat Server:

Tomcat can be used as an add-on to an existing web server (currently Apache, IIS and Netscape servers are supported). A web application is a collection of resources such as jsps, servlets, html files, images, etc. which are mapped to a specific “URI” prefix. For example, all the resources related to OPAC database access are assembled into a “opac” folder and correspondingly all the requests that start with “/opac” can be mapped to this application.

The installation of Tomcat requires installing Tomcat web server and Java Development tools. Following versions of software are better to use for maintaining web access to OPAC database on LAN.

<table>
<thead>
<tr>
<th>Tomcat</th>
<th>Tomcat 3.2.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Java</td>
<td>Java 2 SDK 1.3.1</td>
</tr>
<tr>
<td>Windows</td>
<td>Windows-2000</td>
</tr>
<tr>
<td>Database</td>
<td>MS-Access</td>
</tr>
</tbody>
</table>

Tomcat and Java are installed, preferably on same drive at the specified location. The first and foremost task is to install JDK and Tomcat and get ensure that these are ready to run in a better mode. After successful installation the Tomcat server it will be essential to start Tomcat manually. To escape oneself for its manual starting, it is essential to configure as a Windows service and register for Auto option.

5.2321 Configuring the web application:

Before configuring it is essential to develop the web application in Java Server Page (JSP) to search OPAC database. It is essential to place all HTML pages and JSP applications in a new folder and images used in HTML pages are stored in the other folder. Before launching the application it require some configuration at server as well as on client site.
5.23211 Server Site:

While configuring the server site it is essential to create system Data Source Names (DSNs). These data sources are local to a computer but not user dedicated, any user with privileges can access a system DSN. One DSN entry is required for each database used for search. Each DSN entry includes database and its driver for providing the Open Database Connectivity (ODBC) to the source database.

5.23212 Client Site:

No configuration is required on client site.

5.24 INTRNET INFORMATION SERVER (IIS):

5.241 Introduction:

IIS 4.0 allows you to have multiple web sites on one machine. Though IIS 3.0 has this capability, IIS 4.0 expands the functionality of multiple web sites by adding additional characteristics to sub-directories, and allow for multiple applications. Special considerations need to be made when designing and administrating multiple web sites on a single machine, including when to use sub-directories, when to use virtual directories, how to handle security, and the handling of multiple applications.

5.242 Web Site Design:

Web sites should be singular entries that are self-supporting. Each web site should be able to be moved to a different machine for load balancing, or just transportation purposes. In order to do this, they should be self-supporting, have their own security, and their own application scope. If you are an Internet Service provider, you will want to be able to design, move, and have the user update their web site without interfering with other sites on the same machine.

The HTTP protocol uses URLs to request files from the web server. Since most of these files are contained on the file system, IIS needs to translate the URL to the full path
name of the file. The Internet Information Server does this translation on every request. However it is up to the administrator to configure the server so that the right URLs are mapped to the right directories.

To properly design the file system structure on a machine that hosts multiple websites, it is necessary to understand the difference between a home directory, a virtual root, and a sub-directory. It is also essential to understand when to use virtual directories and when to use sub-directories.

5.2421 Home Directory:

A URL that just contains a domain name is requesting the home directory, sometimes called the root directory. For instance, the URL below is requesting the default file in the home directory.

    The minimal amount of work that the administrator needs to do to assist the web server in mapping URLs to directories is to map the home directory. For example the home directory of this web site could be mapped to:

    c:\inetpub\wwwroot

    Now it is clear that what a home directory is, let us see how to create one home directory in IIS 4.0. The home directory is the starting location of the web site in IIS 4.0 and is created when you create a web site. IIS 3.0 didn't require a home directory, but IIS 4.0 does. Here is how to create a new web site and specify the home directory:

    From MMC:

    1. Select the server that you want to create the web site on.
    2. Right Click and choose Create New | Web Site.
    3. The New Web Site Wizard appears and you are asked to enter a web site description.
    4. Enter a description and press Next.
5. The next page of the Wizard asks for the TCP/IP information, leave the default setting for now and press **Next**. You can always change these settings later.

6. The third page asks for your home directory, enter in the directory and press **Next**.

7. The forth page queries you about the access permission. Select the proper entries and press **Finish**.

**5.2422 Sub-Directories:**

Sub-Directories are directories that inherit the URL mapping from the file system structure. For example, if this directory existed:

c:\inetpub\wwwroot\sales

Then this URL would also exist:

http://www.myserver.com/sales

Sub-directories do not need to be defined to the web server by the system administrator. Because of this, just creating the sub-directory with Explorer will create the directory. There is no need to make any modification in the IIS 4.0 configuration.

**5.2423 Virtual Roots:**

Virtual directories are sub-directories of a URL that are mapped to file system directories that might not inherently exist on the file system. For example if you wanted your site to contain the following URL:

http://www.myserver.com/marketing

And this directory didn’t exist:

c:\inetpub\wwwroot\marketing

One can create a virtual directory that mapped the URL to:

c:\inetpub\marketing\website\external
Virtual directories make the web site appear as if it has a different directory structure than it actually has on the file system. Here is how to create a virtual directory in IIS 4.0:

From MMC:

1. Select the web site that you want to create the virtual directory in.
2. Right Click and choose **Create New | Virtual Directory**.
3. The New Virtual Directory Wizard appears and you are asked to enter an alias to the virtual directory.
4. Enter an alias and press **Next**.
5. The next page of the Wizard ask the physical directory location of the virtual directory, enter the physical directory information and press **Next**.
6. The third page queries you about the access permission. Select the proper entries and press **Finish**.

In IIS 3.0 the difference between virtual directories and sub-directories was significant. In IIS 3.0, sub-directories inherited the properties of the parent directories and virtual directories could have different properties. For instance if you made the home directory read-only and you created a sub-directory called scripts - that sub-directory would be read-only also. If you wanted the scripts directory to have read and execute permissions so that you could run ASP files, you would need to make it a virtual directory.

In IIS 4.0, sub-directories inherit the properties of the parent directory upon creation, but these properties can later be changed. In IIS 4.0 you can create a sub-directory called scripts and change its properties so that it has scripting permission without creating a virtual directory. Here is how to change the permissions of a sub-directory.

From MMC:

7. **Select the sub directory whose permissions you want to change.**
8. **Right Click and choose Properties** from the drop down menu.

10. Choose the Directories tab.

11. Select the proper permissions and press OK.

Virtual directories should only be used when sub-directories can not be used. Here is where we get into personal opinion. Because sub-directories take no web server, and they have all the functionality of virtual directories in IIS 4.0, they should be used whenever possible. Plus sub-directories organize all files into a central location for the web site.

Virtual directories should be used when all the files in the virtual directory does not fit on the physical disk. For instance, if you have a web site that is bigger then 2 Gigs, you might not be able to fit all of it on one disk. In this case, you will need to separate the web site into multiple virtual directories on the directory on each disk. For performance you can also divide your web site up onto multiple disks. In theory, random access across multiple disk drive should be faster then the same number of accesses on the same disk.

If you have multiple web sites and you are sharing information, virtual directories can be used to accomplish this task. For instance if you are sharing graphics, both web sites could have a virtual directory called graphics that is mapped to the same physical disk location. This would be impossible to do with sub directories. Updates to the files in the graphics directory would affect both sites. There is also a performance consideration here, two sites sharing the same files would allow NT to do more memory caching of those files than if they where in separate directories.

One of the main differences between IIS 3.0 and IIS 4.0 is Application Scope. In IIS 3.0, the scope of the application covered the whole machine. In IIS 3.0, if you had two web sites running on the machine, they both shared the same application. In IIS 4.0, you can have more then one application in each web site, and many applications scopes
on the whole machine. In order for each web site to be a singular entity you need to understand how to assign each web site its own application scope.

For instance, in IIS 3.0 if you have the web site http://www.myserver.com and http://www.myofferserver.com and the user linked from one of the servers to the other, he would be within the same application scope. Which means that if you had two global.asa files, one for each web site, only the first global.asa would be called and the second would not. The one called would be the global.asa that corresponded to the first web site that you entered on that machine.

IIS 4.0 gives you the ability to have an application scope start anywhere that you have a directory. The scope then extends to all files in that directory and all files in the subdirectory below. The subdirectory rule however only pertains if there isn't another application scope defined in any of the subdirectories themselves.

One of the problems with the word security is that it means different things to different people. In this issue we will be referring to security as the ability to restrict access to pages on the web server. In the IIS context, security can also refer to SSL encryption, which we will not be addressing.

Most of us run anonymous security configurations and do not think much about web security. However, if you are going to secure your web site, you will want to design the file structure to make administering security easy. You also need to take into consideration multiple web sites on the same machine, each might have different security requirements that need your attention.