5.2 WORLD WIDE WEB

The Web was first proposed by Tim Berners-Lee in 1989 at *CERN*†, the European Organization for Nuclear Research, to allow several researchers at different locations throughout Europe to access each others researches.The commercial Web started in 1990s.The web pages,are distributed all over the world and related documents are linked together.Today, the Web is used to provide electronic shopping and gaming.

Architecture of WWW

The WWW is a distributed client-server service, in which a client using a browser can access a service using a server. The service provided is distributed over many locations called sites as in figure 5.2.1.

A web page can be simple or composite. A simple web page has no links to other web pages; a composite web page has one or more links to other web pages. Each web page is a file with a name and address.

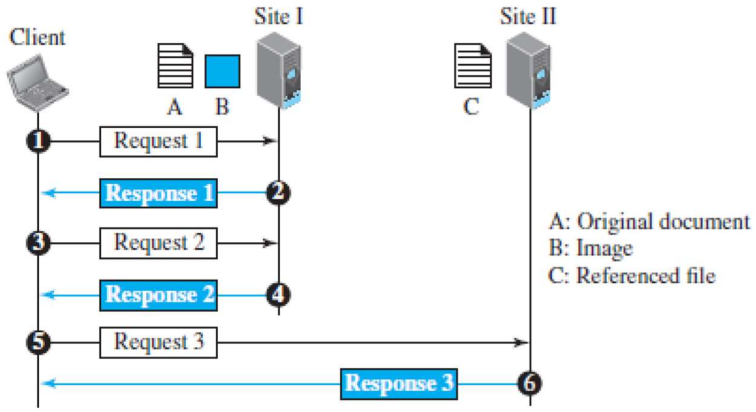

Fig5.2.1: The web architecture.

[Source :"Data Communications and Networking" by Behrouz A. Forouzan,Page-821]

The first transaction (request/response) retrieves a copy of the main document (file A), which has references (pointers) to the second and third files. When a copy of the main document is retrieved and browsed, the user can click on the reference to the image to invoke the second transaction and retrieve a copy of the image (file B). If the user needs to see the contents of the referenced text file, she can click on its reference (pointer) invoking the third transaction and retrieving a copy of file C.

Web Client (Browser)

A variety of vendors offer commercial browsers that interpret and display a web page, and all of them use the same architecture.Each browser has three parts: a controller, client protocols, and interpreters. The controller receives input from the keyboard or the mouse and uses the client programs to access the document.

After the document has been accessed, the controller uses one of the interpreters to display the document on the screen.The client protocol can be one of the protocols described later, such as HTTP or FTP.The interpreter can be HTML, Java, or JavaScript, depending on the type of document. Some commercial browsers include Internet Explorer, Netscape Navigator, and Firefox.

Web Server

The web page is stored at the server. Each time a request arrives, the corresponding document is sent to the client. To improve efficiency, servers normally store requested files in a cache in memory; memory is faster to access than a disk.

Uniform Resource Locator (URL)

A web page, as a file, needs to have a unique identifier to distinguish it from other web pages.

URL is a standard for specifying any kind of information on the internet.URL defines four things: protocol, host computer, port and path.

Protocol. It is the client-server program that we need to access the web page.Example protocol is HTTP (HyperText Transfer Protocol) and FTP (File Transfer Protocol).

Host. The host identifier can be the IP address of the server or the unique name given to the server. IP addresses can be defined in dotted decimal notation.

Port. The port, a 16-bit integer, is normally predefined for the client-server application. For example, if the HTTP protocol is used for accessing the web page, the well-known port number is 80. However, if a different port is used, the number can be explicitly given.

Path. The path identifies the location and the name of the file in the underlying operating system.

The format of this identifier normally depends on the operating system. In UNIX, a path is a set of directory names followed by the file name, all separated by a slash.

For example, /top/next/last/myfile is a path that uniquely defines a file named my file, stored in the directory last, which itself is part of the directory next, which itself is under the directory top.

To combine these four pieces together, the uniform resource locator (URL) is used. It uses three different separators between the four pieces.

Web Documents

The documents in the WWW can be grouped into three broad categories: static, dynamic, and active.

Static Documents

Static documents are fixed-content documents that are created and stored in a server. The client can get a copy of the document only. The contents in the server can be changed, but the user cannot change them.When a client accesses the document, a copy of the document is sent. The user can then use a browser to see the document.Static documents are prepared using one of several languages: Hyper Text Markup Language (HTML), Extensible Markup Language (XML), Extensible Style Language (XSL), and Extensible Hypertext Markup Language (XHTML).

Dynamic Documents

A dynamic document is created by a web server whenever a browser requests the document.

When a request arrives, the web server runs an application program or a script thatcreates the dynamic document.The server returns the result of the program or script as a response to the browser that requested the document. Because a fresh document is created for each request, the contents of a dynamic document may vary from one request to another.

A simple example of a dynamic document is the retrieval of the time and date from a server.

Time and date are kinds of information that are dynamic in that they change from moment to moment. The client can ask the server to run a program such as the date program in UNIX and send the result of the program to the client.

Active Documents

For many applications, we need a program or a script to be run at the client site. These are called active documents.For example, if we want to run a program that creates animated graphics on the screen or a program that interacts with the user. The program definitely needs to be run at the client site where the animation or interaction takes place.

When a browser requests an active document, the server sends a copy of the document or a script.

The document is then run at the client (browser) site. One way to create an active document is to use Java applets, a program written in Java on the server. It is compiled and ready to be run. The document is in byte code (binary) format.