

Internet Web Site Design

HISTORY OF INTERNET

The Internet, the network of networks, is a worldwide system of interconnected computer networks that use the TCP/IP protocols to reach billions of users. The Internet began as a U.S Department of Defense network to link scientists and university professors around the world.

The Internet had its roots during the 1960's as a project of the United States Government's Department of Defense, to create a non-centralized network. This project was called **ARPANET (Advanced Research Projects Agency Network)**, created by the Pentagon's Advanced Research Projects Agency in 1969 to provide a secure and survivable communications network for organizations engaged in defense-related research.

In order to make the network more global, a new sophisticated and standard protocol was needed. They developed **IP (Internet Protocol)** technology which defined how electronic messages were packaged, addressed, and sent over the network. The standard protocol was invented in 1977 and was called **TCP/IP (Transmission Control Protocol/Internet Protocol)**. TCP/IP allowed users to link various branches of other complex networks directly to the ARPANET, which soon came to be called the Internet. Researchers and academics in other fields began to make use of the network, and eventually the **National Science Foundation (NSF)** took over much of the TCP/IP technology from ARPANET and established a distributed network of networks capable of handling far greater traffic. In 1985, NSF began a program to establish Internet access across the United States. They created a backbone called the **NSFNet** and opened their doors to all educational facilities, academic researchers, government agencies, and international research organizations.

In 1991, the U.S. government decided that NSFNET was not capable of supporting the rapidly increasing Internet traffic. Three companies, IBM, Merit, and Verizon, filled the void by forming a nonprofit organization called **Advanced Network & Services (ANS)** to build a new, high-speed Internet backbone called **Advanced Network Services Network (ANSNET)**.

The term Internet, originally associated with government-funded connected networks, now referred to the connected networks using TCP/IP protocols.

Today, we witness a rapid growth both in the infrastructure and new applications. The Internet today is a set of networks that provide services to the whole world. What has made the Internet so popular is the invention of new applications.

The terms "Internet" and "World Wide Web" are often used interchangeably; however, the Internet and World Wide Web are not one and the same.

- The Internet is a vast hardware and software infrastructure that enables computer interconnectivity.

Internet Web Site Design

- The Web, on the other hand, is a massive hypermedia database – a myriad collection of documents and other resources interconnected by hyperlinks. Imagine the World Wide Web as the platform which allows one to navigate the Internet with the use of a browser such as Google Chrome or Mozilla Firefox.

WORLD WIDE WEB (ABBREVIATED WWW OR WEB)

The idea of 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 other's researches. The commercial Web started in the early 1990s.

The Web today is a repository of information in which the documents, called web pages, are distributed all over the world and related documents are linked together. The popularity and growth of the Web can be related to two terms in the above statement: **distributed and linked**.

Distribution allows the growth of the Web. Each web server in the world can add a new web page to the repository and announce it to all Internet users without overloading a few servers.

Linking allows one web page to refer to another web page stored in another server somewhere else in the world. The linking of web pages was achieved using a concept called **hypertext**, which was introduced many years before the advent of the Internet. The idea was to use a machine that automatically retrieved another document stored in the system when a link to it appeared in the document.

The Web implemented this idea electronically to allow the linked document to be retrieved when the link was clicked by the user. Today, the term hypertext, coined to mean linked text documents, has been changed to hypermedia, to show that a web page can be a text document, an image, an audio file, or a video file.

The purpose of the Web has gone beyond the simple retrieving of linked documents. Today, the Web is used to provide electronic shopping and gaming. One can use the Web to listen to radio programs or view television programs whenever one desires without being forced to listen to or view these programs when they are broadcast.

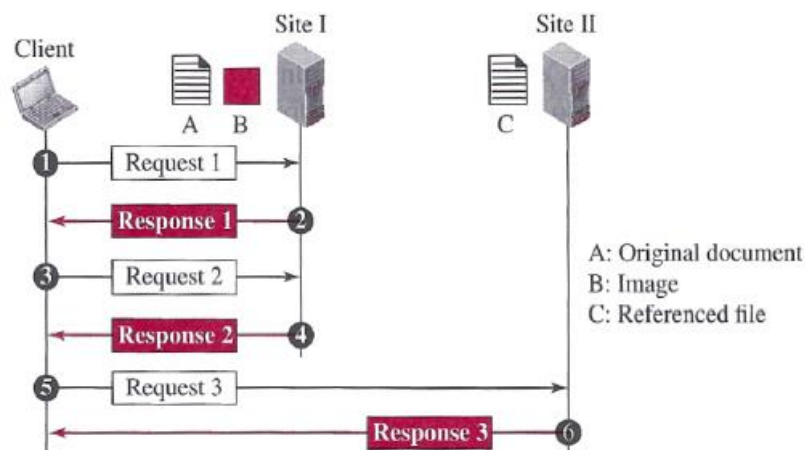
Internet Web Site Design

Architecture

The WWW today is a distributed client-server service, in which a client using a browser can access a service using a server. However, the service provided is distributed over many locations called sites. Each site holds one or more web pages. Each web page, however, can contain some links to other web pages in the same or other sites. In other words, 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.

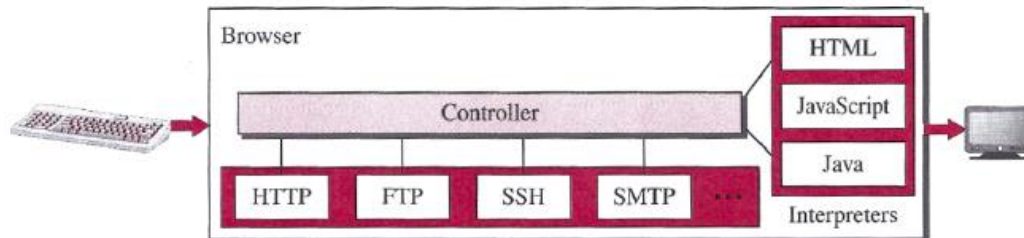
Example: Assume we need to retrieve a scientific document that contains one reference to another text file and one reference to a large image. The figure below shows the situation. The main document and the image are stored in two separate files (file A and file B) in the same site; the referenced text file (file C) is stored in another site. Since we are dealing with three different files, we need three transactions if we want to see the whole document. 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. Note that although files A and B both are stored in site 1, they are independent files with different names and addresses. Two transactions are needed to retrieve them. A very important point we need to remember is that file A, file B, and file C in the example are independent web pages, each with independent names and addresses. Although references to file B or C are included in file A, it does not mean that each of these files cannot be retrieved independently. A second user can retrieve file B with one transaction. A third user can retrieve file C with one transaction.



Internet Web Site Design

Web Client (Browser)

A variety of vendors offer commercial browsers that interpret and display a web page, and all of them use nearly the same architecture. Each browser usually consists of three parts: a controller, client protocols, and interpreters. (see the figure below).



The controller receives input from the keyboard or the mouse and uses the client programs (protocols) 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, Google Chrome, Opera, 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. A server can also become more efficient through multithreading or multiprocessing. In this case, a server can answer more than one request at a time. Some popular web servers include Apache and Microsoft Internet Information Server (IIS).

Uniform Resource Locator (URL)

A web page, as a file, needs to have a unique identifier to distinguish it from other web pages. To define a web page, we need three identifiers: **host, port, and path**.

However, before defining the web page, we need to tell the browser what client-server application we want to use, which is called the **protocol**. This means we need four identifiers to define the web page. The first is the type of vehicle to be used to fetch the web page; the last three make up the combination that defines the destination object (web page).

Internet Web Site Design

1–**Protocol**: The first identifier is the abbreviation for the client–server program that we need in order to access the web page. Although most of the time the protocol is HTTP (Hyper Text Transfer Protocol) we can also use other protocols such as FTP (File Transfer Protocol).

2–**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 such as 64.23.56.17; the name is normally the domain name that uniquely defines the host, such as forouran.com, which we discuss in Domain Name System (DNS) later.

3–**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.

4–**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 myfile, stored in the directory last, which itself is part of the directory next, which itself is under the directory top. In other words, the path lists the directories from the top to the bottom, followed by the file name.

To combine these four pieces together, the uniform resource locator (URL) has been designed; it uses three different separators between the four pieces as shown below:

protocol://host/path	Used most of the time
protocol://host:port/path	Used when port number is needed

Example: The URL <http://www.mhhe.com/compsci/forouzan/> defines the web page related to one of the authors of Data Communications and Networking book.

The string www.mhhe.com is the name of the computer in the McGraw–Hill company (the three letters www are part of the host name and are added to the commercial host).

The path is compsci/forouzan/, which defines Forouzan's web page under the directory compsci (computer science).

Internet Web Site Design

Web Documents

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

1– 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. In other words, the contents of the file are determined when the file is created, not when it is used. Of course, 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: HyperText Markup Language (HTML), Extensible Markup Language (XML), Extensible Style Language (XSL), and Extensible Hypertext Markup Language (XHTML).

2– 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 that creates 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 very 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. Dynamic documents are prepared using one of the scripting languages such as Java Server Pages (JSP), which uses the Java language for scripting, or Active Server Pages (ASP).

3– 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, suppose 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. Another way is to use JavaScript but download and run the script at the client site.

Internet Web Site Design

TYPES OF SCRIPTS

- A script is generally a series of instruction or program, which has to be executed on other program or application.
- The scripts can be written in two forms, at the client end (front end – client-side scripting) or at the server end (back end – server-side scripting).
- The main difference between server-side scripting and client-side scripting is that:
 - The server side scripting involves server for its processing.
 - On the other hand, client-side scripting requires browsers to run the scripts on the client machine but does not interact with the server while processing the client-side scripts.
- As we know that the web works in a client-server environment.
 - The client-side script executes the code to the client side which is visible to the users.
 - While a server-side script is executed in the server end which users cannot see.
- Client-side scripting languages: HTML, CSS, JavaScript.
- Server-side scripting languages: PHP, Python, Ruby, ColdFusion, C#, Java, C++ and etc.

