Introduction in Internet

www.instigate-training-center.am

Ինստիգեյթ
Ուսումնական Կենտրոն

Topics

- Internet
- The uses of Internet
- What is Web?
- How to access the Internet
- Internet Service Provider
- How to access the Web
- Client/Server Structure of the Web
- IP Addressing
- Domain Name Adressing
- Uniform Resource Locators
- Structure of Uniform Resource Locators

Topics (cont.)

- HTTP
- HTTP 1.0 / 1.1
- GET request example
- Request Header Fields
- Other request Methods
- Caching
- Cookies

Internet

- It is the largest network in the world that connects hundreds of thousands of individual networks all over the world.
- The popular term for the Internet is the “information highway”.
- Rather than moving through geographical space, it moves your ideas and information through cyberspace – the space of electronic movement of ideas and information.

Internet (cont.)

- No one owns it
- It has no formal management organization.
- As it was originally developed by the Department of defense, this lack of centralization made it less vulnerable to wartime or terrorist attacks.
- To access the Internet, an existing network need to pay a small registration fee and agree to certain standards based on the TCP/IP (Transmission Control Protocol/Internet Protocol) .

The uses of the Internet

- Send e-mail messages.
- Send (upload) or receive (down load) files between computers.
- Participate in discussion groups, such as mailing lists and newsgroups.
- Surfing the web.

What is Web?

- The Web (World Wide Web) consists of information organized into Web pages containing text and graphic images.
- It contains hypertext links, or highlighted keywords and images that lead to related information.
- A collection of linked Web pages that has a common theme or focus is called a Web site.
- The main page that all of the pages on a particular Web site are organized around and link back to is called the site’s home page.

How to access the Internet?

- Many schools and businesses have direct access to the Internet using special high-speed communication lines and equipment.
- Students and employees can access through the organization’s local area networks (LAN) or through their own personal computers.
- Another way to access the Internet is through Internet Service Provider (ISP).

How to access the Internet?

- To access the Internet, an existing network need to pay a small registration fee and agree to certain standards based on the TCP/IP (Transmission Control Protocol/Internet Protocol) reference model.
- Each organization pays for its own networks and its own telephone bills, but those costs usually exist independent of the internet.
- The regional Internet companies route and forward all traffic, and the cost is still only that of a local telephone call.

Internet Service Provider (ISP)

- A commercial organization with permanent connection to the Internet that sells temporary connections to subscribers.
- Examples:
- Prodigy, America Online, Microsoft network, AT&T Networks, UCom

How to access the Web?

- Once you have your Internet connection, then you need special software called a browser to access the Web.
- Web browsers are used to connect you to remote computers, open and transfer files, display text and images.
- Web browsers are specialized programs.
- Examples of Web browser: Netscape Navigator (Navigator) and Internet Explorer.

Client/Server Structure of the Web

- Web is a collection of files that reside on computers, called Web servers, that are located all over the world and are connected to each other through the Internet.
- When you use your Internet connection to become part of the Web, your computer becomes a Web client in a worldwide client/server network.
- A Web browser is the software that you run on your computer to make it work as a web client.

IP Addressing

- The combination of the four IP address parts provides 4.2 billion possible addresses (256 x 256 x 256 x 256).
- This number seemed adequate until 1998.
- Members of various Internet task forces are working to develop an alternate addressing system that will accommodate the projected growth.
- However, all of their working solutions require extensive hardware and software changes throughout the Internet.

Domain Name Addressing

- Most web browsers do not use the IP address t locate Web sites and individual pages.
- They use domain name addressing.
- A domain name is a unique name associated with a specific IP address by a program that runs on an Internet host computer.
- This program, which coordinates the IP addresses and domain names for all computers attached to it, is called DNS (Domain Name System) software.
- The host computer that runs this software is called a domain name server.

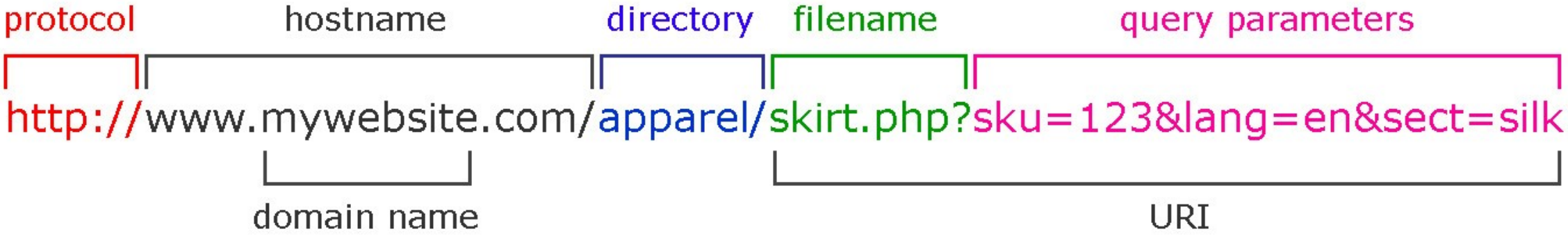
Domain Name Addressing

- Domain names can include any number of parts separated by periods, however most domain names currently in use have only three or four parts.
- Domain names follow hierarchical model that you can follow from top to bottom if you read the name from the right to the left.
- For example, the domain name gsb.uchicago.edu is the computer connected to the Internet at the Graduate School of Business (gsb), which is an academic unit of the University of Chicago (uchicago), which is an educational institution (edu).
- No other computer on the Internet has the same domain name.

Uniform Resource Locators

- The IP address and the domain name each identify a particular computer on the Internet.
- However, they do not indicate where a Web page's HTML document resides on that computer.
- To identify a Web pages exact location, Web browsers rely on Uniform Resource Locator (URL).
- URL is a four-part addressing scheme that tells the Web browser:
  - What transfer protocol to use for transporting the file
  - The domain name of the computer on which the file resides
  - The pathname of the folder or directory on the computer on which the file resides
  - The name of the file

Structure of a Uniform Resource Locators



HTTP

- The transfer protocol is the set of rules that the computers use to move files from one computer to another on the Internet.
- The most common transfer protocol used on the Internet is the Hypertext Transfer Protocol (HTTP).
- Two other protocols that you can use on the Internet are the File Transfer Protocol (FTP) and the Telnet Protocol

HTTP 1.0 and 1.1

- HTTP/1.0 allows only connectionless message passing
  - each request/response requires a new connection
  - to download a page with images requires multiple connections can overload the server, require lots of overhead

- HTTP/1.1 provides persistent connection by default
  - once client & server connect, remains open until told to close it (or timeout) reduces number of connections, saves overhead
  - client can send multiple requests without waiting for responses
    - e.g., can request all images in a page at once

GET Example

```
bash-3.1$ telnet www.csc.liv.ac.uk 80
Trying 10.128.0.3...
Connected to www.csc.liv.ac.uk
    (10.128.0.3).
Escape character is '^]'.
GET /~martin/index.html HTTP/1.1
Host: www.csc.liv.ac.uk
```

server response has assorted header information, followed by the page

```
HTTP/1.1 200 OK
Date: Mon, 08 Oct 2011 10:01:15 GMT
Server: Apache/2.0.58  HP-UX_Apache-based_Web_Server
    (Unix) mod_perl/1.99_16 Perl/v5.8.7 DAV/2 PHP/5.0.4
Last-Modified: Mon, 01 Oct 2011 14:55:16 GMT
ETag: "ec3f-1122-9fd83d00"
Accept-Ranges: bytes
Content-Length: 4386
Content-Type: text/html

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN“
    "http://www.w3.org/TR/xhtml11/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en"
    lang="en">
<head>
. . .
</head>

<body>
. . .
</body>
</html>

Connection closed by foreign host.
```

- The first line of the server’s response contains a status code
  - 200 OK - request was processed successfully
  - 301 Moved - permanently document has been moved
  - 304 Not modified- if cached version is up-to-date
  - 400 Bad request- syntax error in client’s request
  - 403 Forbidden - client is not allowed access (e.g., protected)
  - 404 Not found - file could not be found
  - 500 Internal server error - server failed
  - 503 Service unavailable - server is overloaded

Other response header fields

- In addition to the status code, the server’s response may include
  - Date response time (in GMT)
  - Server identification info on the server
  - Last-modified time document was last changed (in GMT)
  - Content-length size of document, in bytes
  - Content-type file format (e.g., html, gif, pdf)
  - Expires prevents browser from caching beyond date

Other Request Methods

- HEAD
  - similar to GET, but requests header information only, useful for checking to see if a document exists, how recent
- POST
  - similar to GET, but encodes inputs differently, useful for submitting form contents to a CGI program, and is also often used in PHP scripts for submitting information
- PUT
  - upload a document to the server (new in HTTP/1.1)
- DELETE
  - delete a document from the server (new in HTTP/1.1)

Caching

- Browsers cache pages to save downloading
  - maintain temporary storage (cache) for recent pages
  - when a page is requested, check to see if already in cache
  - if not in the cache, issue GET request
    - when response message arrives,
      - display page and store in cache (along with header info)
  - if already stored in the cache, send GET request with If-Modified-Since header set to the data of the cached page
    - when response message arrives,
      - if status code 200, then display and store in cache
      - if status code 304, then display cached version instead

Cookies

- HTTP message passing is transaction-based, stateless
  - many e-commerce apps require persistent memory of customer interactions
  - e.g., amazon.com

    remembers your name, credit card, past purchases, interests

    Other sites use “shopping carts” for purchases, logins for access, etc.

Cookies (cont.)

- Netscape’s solution: cookies
  - a cookie is a collection of information about the user, and a browser can store some information on your machine that it can later request
  - server can download a cookie to the client’s machine using the “Set-cookie” header in a response

```
Set-cookie: CUSTOMER=Alex_Thompson; PATH=/;
EXPIRES=Thursday, 29-Jan-2010 12:00:00
```

  - when user returns to URL on the specified path, the browser returns the cookie data as part of its request

```
Cookie: CUSTOMER=Alex_Thompson
```

Thank You

Ինստիգեյթ
Ուսումնական Կենտրոն