



SWE 363: WEB ENGINEERING & DEVELOPMENT

Spring Semester 2015 (2014-2)

Web Architecture – Part 1

Dr. Nasir Al-Darwish

Computer Science Department

King Fahd University of Petroleum and Minerals

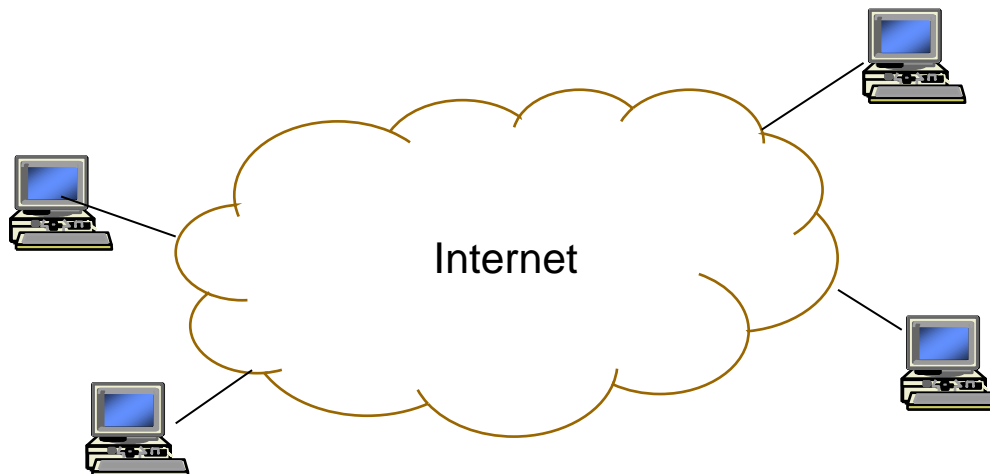
darwish@kfupm.edu.sa

Outline

- Introduction to the Internet
 - Definition
 - Hierarchical Structure
 - Internet Protocols
- Introduction to the Web
 - Definition
 - Web Clients and Servers
 - Web Architecture & Operation
 - Websites & Web Documents

What is the Internet?

- A global heterogeneous network that connects millions of computers world wide. **It uses TCP/IP protocols for the *network, transport and application* layers.**
- IP (Internet Protocol) is a network-layer protocol that makes it possible to connect (e.g. using a router) two or more networks into one internetwork.
- TCP/IP is designed to work with a multitude of data-link and physical layers (e.g. Ethernet, ATM, FDDI, PPP)



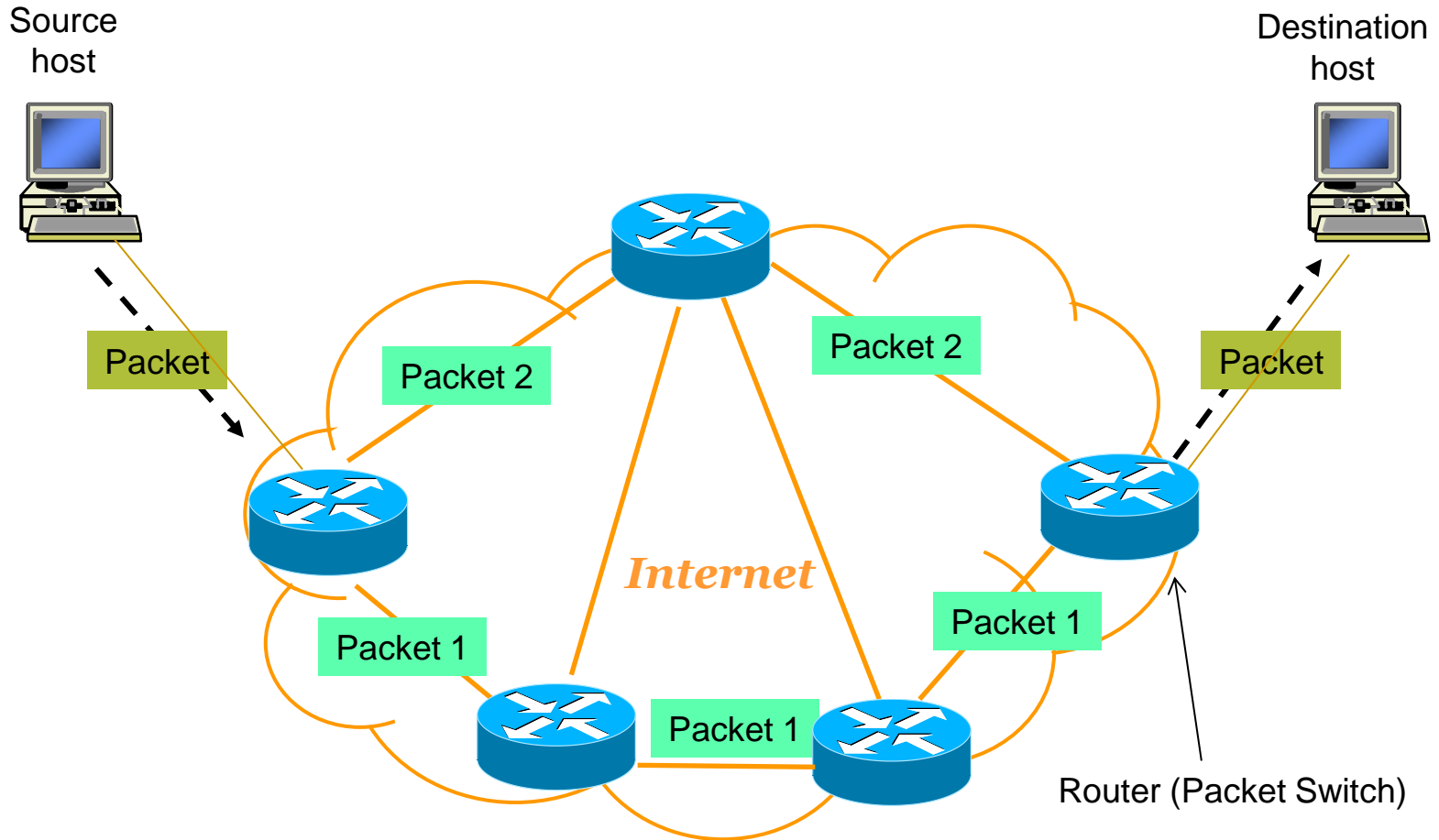
What is the Internet? ...

- The goal is to provide connectivity between computers (hosts) and their users to
 - ❑ Share resources
 - ❑ Increase reliability and availability
 - ❑ Collaborate (email, distributed computing, etc.)
 - ❑ Access remote information
- Thus, the Internet is a vehicle for transferring data from one computer to another.

Common Internet Applications

- Each Internet end-user application has its own (application-layer) protocol.
- The World-Wide Web (WWW); uses HTTP protocol
- Electronic mail (email); uses SMTP/POP/IMAP protocols
- File transfer (e.g. ftp); uses FTP protocol
- Remote login (e.g. telnet)
- Streaming multimedia (e.g. Internet feeds of live audio and video, video on demand (VoD))
- Internet telephony (VoIP), Chatting
- Many others

How the Internet works

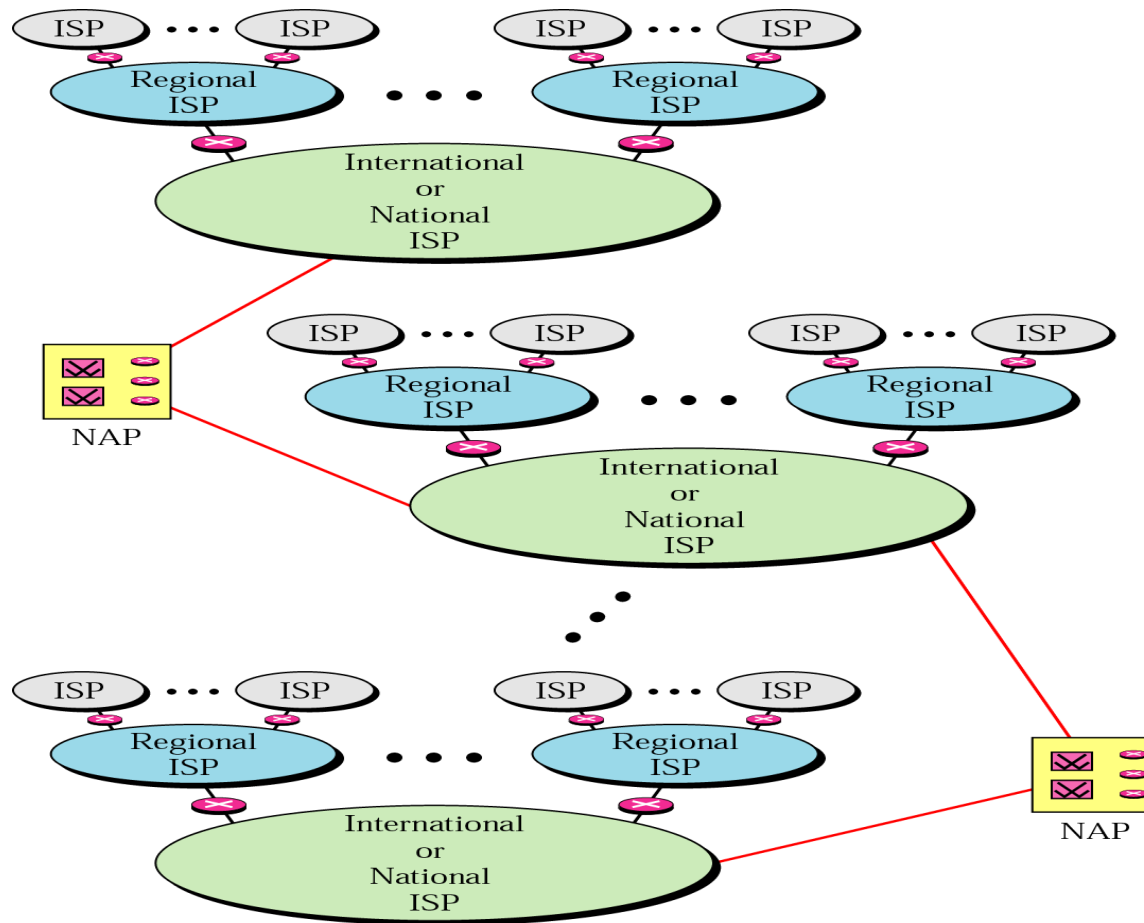


How the Internet works

- Source host generates a message & converts it into packets
- Source host or a router may fragment a packet into smaller packets (fragments)
- Packets transferred independently across network
- Destination router delivers packets to the destination host
- Destination host reassembles fragments to build the original packet
- TCP layer of destination host rearranges received packets to build the submitted message

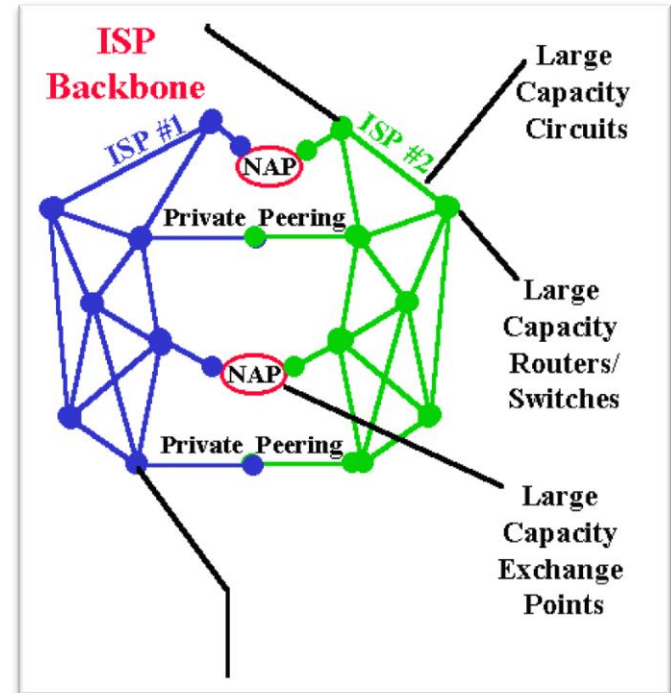
Internet Hierarchical Structure

- To manage the growth (scalability) of the Internet, it is roughly structured in a hierarchical manner



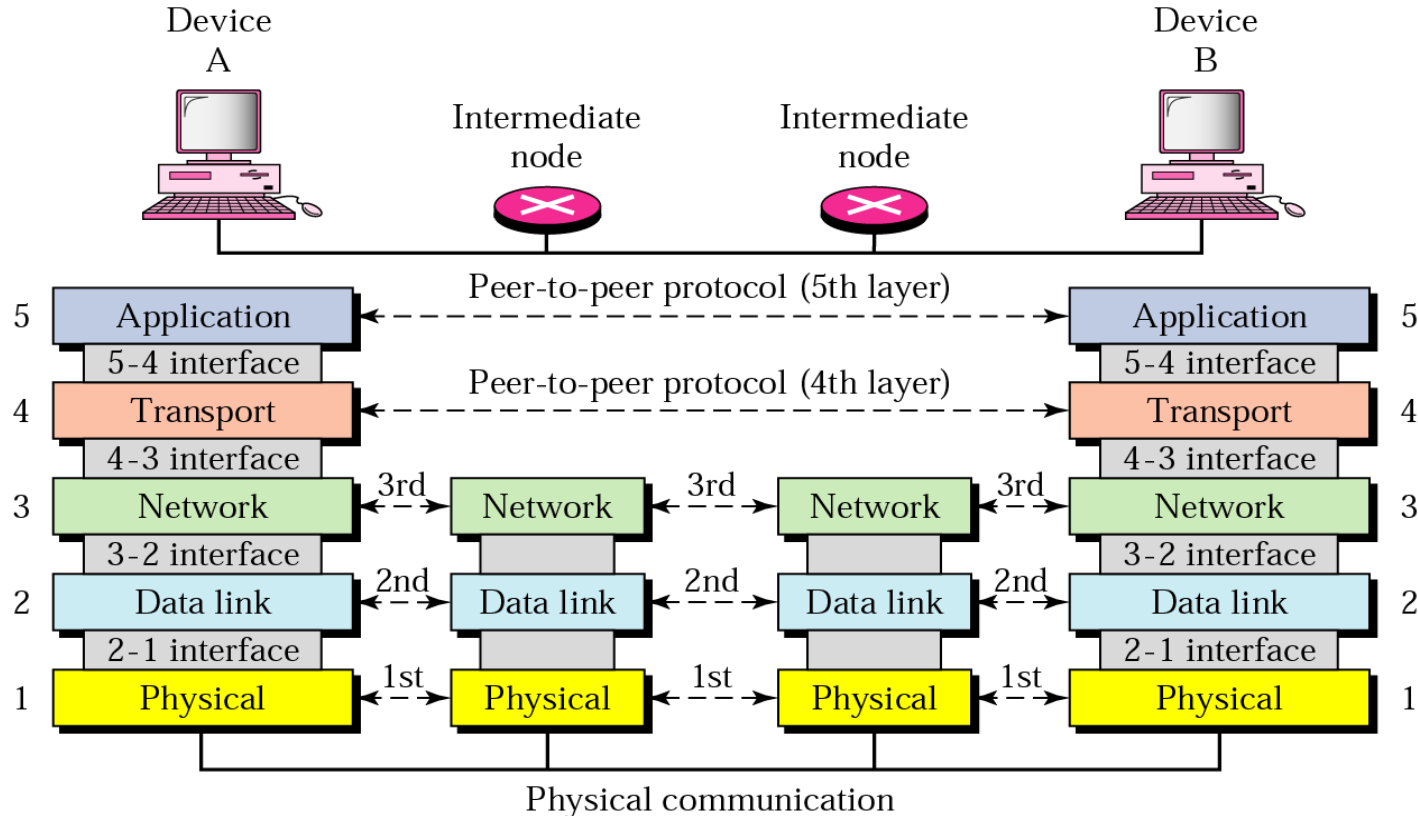
Internet Hierarchical Structure ...

- Internet Service Provider (ISP)
 - International, National, Regional, Local
- At the lowest level are the organizations networks
 - e.g. KFUPM network
 - Which can act as ISP for lower-level user networks as well (e.g. home networks or small office networks)
- Two ISP networks can be connected to each other through network access points (NAP) or private peering.
 - NAP: data communication facilities that provide access to higher-speed links
- Routers: networking devices that forward data packets across an internet network toward their destinations



Layered Architecture

- The operation of the Internet is controlled by software which is functionally structured into layers called TCP/IP protocol stack
- Each layer has a number of protocols to govern and facilitate communication between different devices

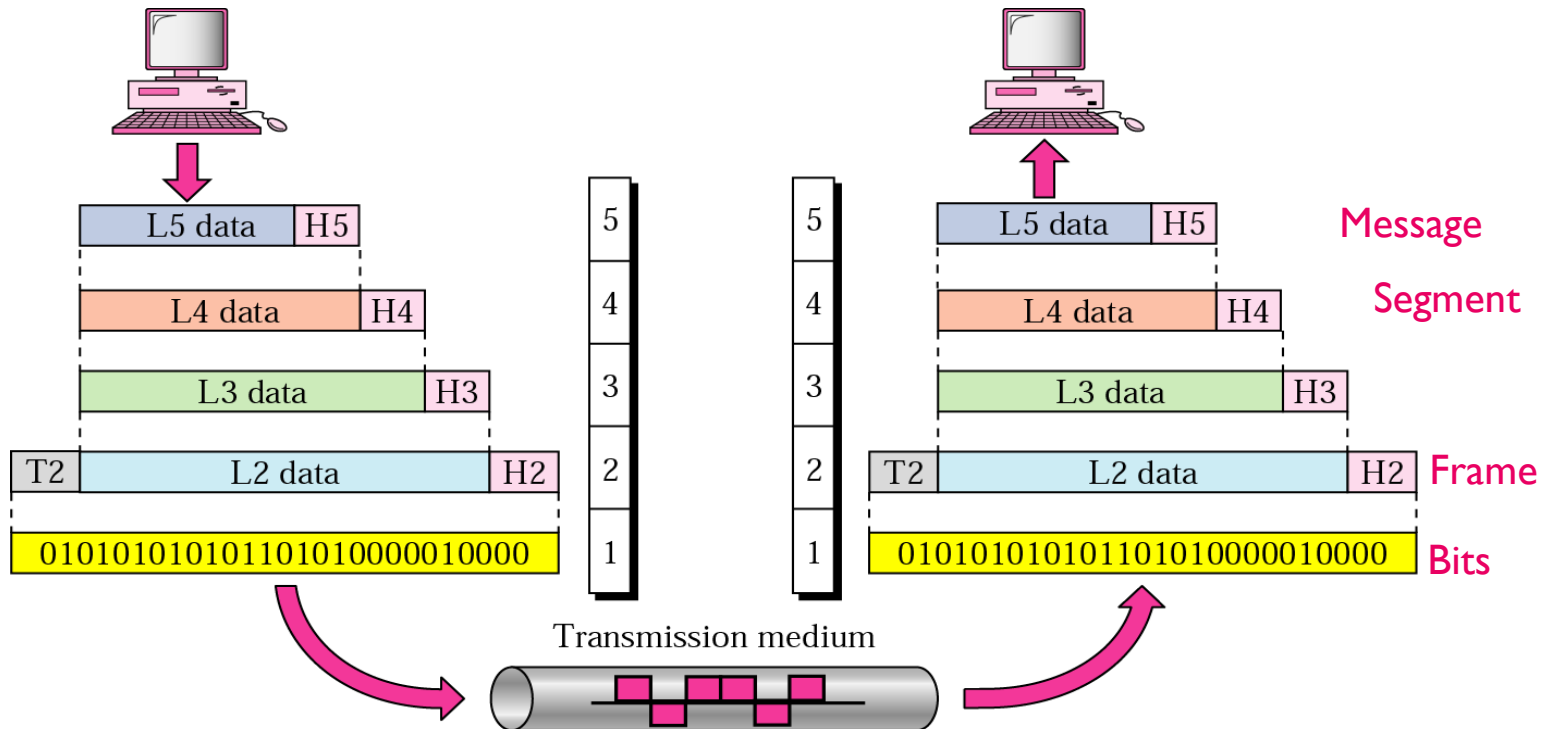


Layered Architecture ...

- *Application layer*: acts as an interface for user to allow them to access network resources. Each Internet application has its own application layer protocol.
- *Transport layer*: provides reliable channel for message delivery and error recovery between applications
- *Network layer*: moves packets from the source host to the destination host across the network
- *Data-link layer*: organizes data into frames and move frames from one machine to the next machine of the path to the destination (hop-to-hop delivery)
- *Physical layer*: transmits bits over a medium and provide mechanical and electrical specifications.

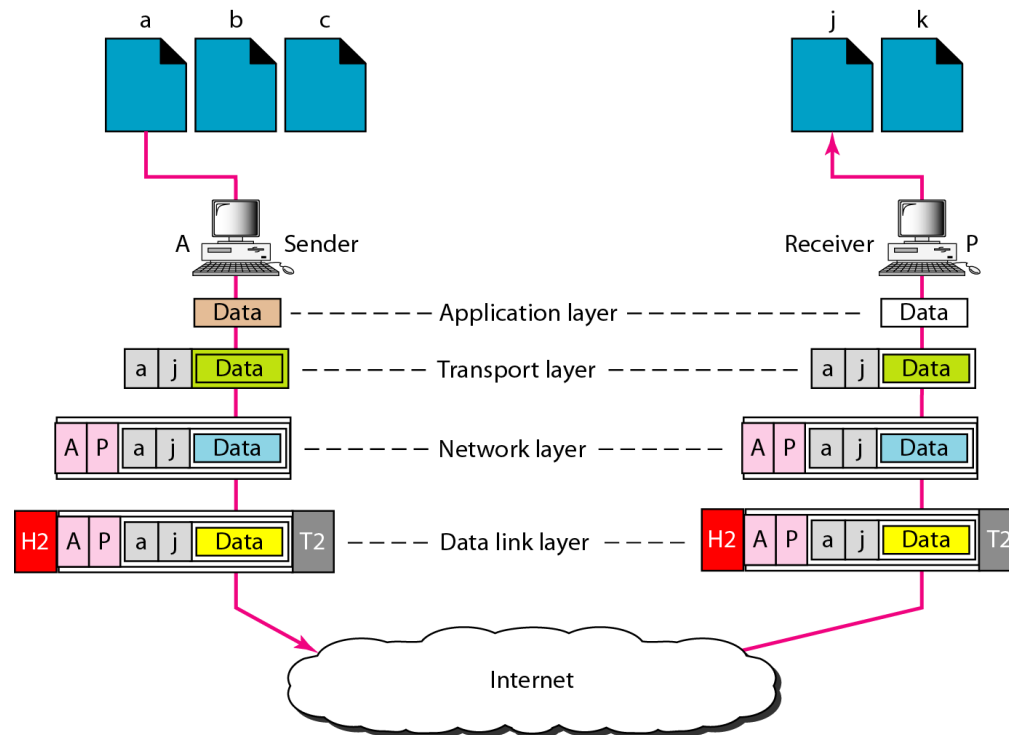
Layered Architecture ...

- Each layer adds meta-data (control information) to facilitate communication between corresponding layers at different hosts
- The control data includes address info. (source/destination address of hosts, port numbers (sending/receiving application), error control information, etc.



Addressing

- Each host is identified using a unique global address (called IP address) or hostnames (also a host can be assigned a symbolic name; DNS servers map hostnames to IP addresses)
- Processes on the same host are uniquely identified using the host address (IP address) + the transport layer protocol port number (e.g. a web server is running at TCP port # 80)



The Web

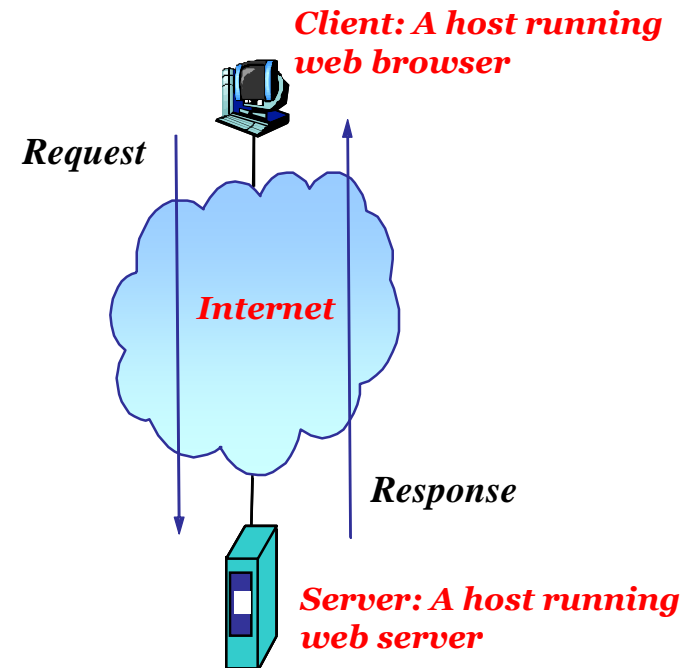
- Also called World Wide Web, WWW, W3; started in 1990s at CERN (European Center for Nuclear Research) by Tim Berners Lee
- Forms an overlay network over the Internet; massive repository of information
- A collection of distributed web documents and other web resources uniquely identified (using URLs), linked to each other, can be accessed via the Internet
- Web documents can have different types of information (multimedia): text, images, audio and video
- Provide a medium to share information; access and interact with remote applications and systems
- Main features: Portability, Scalability, User-Friendly Interface (Accessibility)

Primary Components of the Web

- Web Browser, Web Server, Internet
- Web server (i.e., HTTP server): is the application-layer software that
 - ❑ Stores a set of Web documents (web pages) and other resources
 - ❑ Responds to requests from the browser by sending a copy of the document
 - ❑ *Examples:* Apache, Microsoft IIS
- Web browser (also called user agent, web client or HTTP client)
 - ❑ End-user application program that represents the user interface to the Web
 - ❑ Fetches information from Web server and displays it to the user
 - ❑ *Examples:* Mosaic (1993), NN (1994), IE (1995), Mozilla (1998), Firefox (2004), Opera, Safari, many others
- Web standards
 - ❑ Transfer (delivery) protocol: HyperText Transfer Protocol (HTTP)
 - ❑ Hypermedia links: Uniform Resource Locator (URL) to identify web resources
 - ❑ Document encoding: HTML (HyperText Markup Language), XML (eXtensible Markup Language)

Web Architecture & Operation

- Web pages (documents) are hosted (stored) in computers running special software called web servers
- Web documents can be accessed and viewed using special programs called browsers (web clients or user agents)
- Browsers & Web servers are often running on different computers
- Browsers use HTTP protocol to communicate with the Web servers
- The HTTP protocol is a *command-response protocol*, with commands sent by the client and responses to them sent by the web server



The web is a client-server Internet application

Components of Web Infrastructure

- *Content*: Web Pages expressed in HTML
 - HTML contain references to other types of objects such as images
- *Clients*
 - Send requests to Servers / Receive responses
- *Servers*
 - Receive requests from clients / Send responses
 - Store or generate the responses
- *Proxies*
 - Placed between clients and servers
 - Act as a server for the client, and a client to the server
 - Provide extra functions
 - Caching, anonymization, logging, filtering access

Ingredients of Web Implementation

- **HTML (HyperText Markup Language):** uses markup or tags
 - The *anchor* tag is the most important tag (*Why?*)
 - In my web page:
I work at `KFUPM`
 - *It is a Web because html pages are nodes, the anchor tags are the links*
 - From functionality viewpoint, the anchor tag is the most important tag. It is the tag that makes the web a web.
- **URLs** (or more generally, *URIs*)
- **HTTP (HyperText Transfer Protocol):** Application-Layer protocol used between a web client and a web server

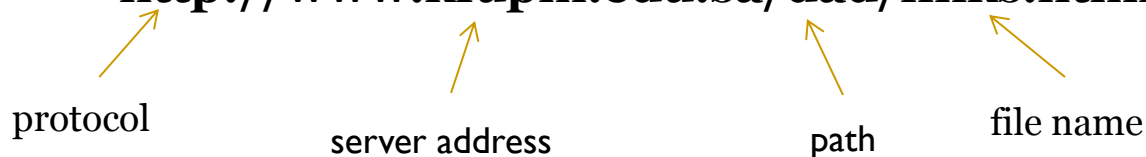
URL Structure

- Uniform Resource Locator (URL)
 - Represents the address of a resource on the Web
- A URL defines:
 - Protocol used to access/transfer the document (such as HTTP or FTP; the default is HTTP)
 - Server that hosts the document using its domain name (or IP address)
 - Protocol port number of the server (optional; the default is 80)
 - Path and document name
- General form of URL

protocol://server.domain-name:port/path-and-file-name

■ Example

http://www.kfupm.edu.sa/dad/links.html



Web Operation: Fetching Content

- A user starts a browser on his computer and request a web document by specifying its URL, e.g. <http://www.kfupm.edu.sa/>
- The browser uses DNS find the web server IP address.
- Then, the browser sends a message to the server requesting the required document
- The server finds the document in its file system and sends it back to the browser
- The browser interprets the content of the document and displays (renders) it for the user
- If the document contains images,
 - Images will be on separate files and only their URLs will be embedded in the base document
 - The browser will send a new request for each image file

Website Structure

- A website is made up of a set of related web pages linked to each other; these pages can be on one server or distributed on different servers

- Navigation order

- Linear

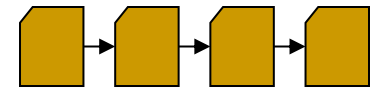
- Web pages are linked in a strict sequence
 - Useful for guided navigation, tutorials, multi-part articles, etc.

- Hierarchical (tree)

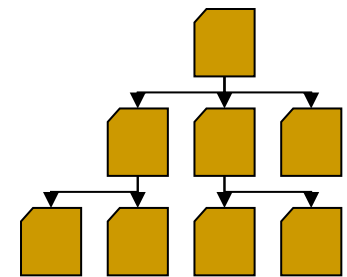
- Web pages are arranged hierarchically (also called tree)
 - The “root” is the website’s homepage

- Hybrid (non-linear)

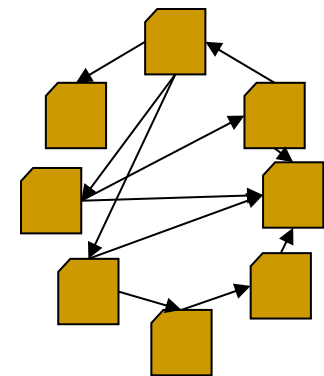
- Webpage links can form sequences, trees, loops, or whatever paths are needed
 - Might this be confusing to a site visitor?



Linear



Hierarchical



Hybrid

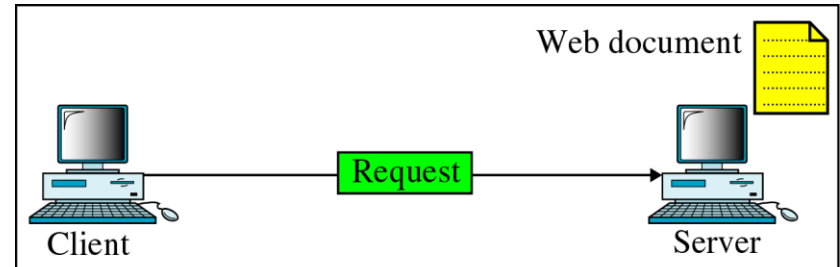
Categories of Web documents

- Three primary categories:
 - Static documents
 - Dynamic documents
 - Active documents

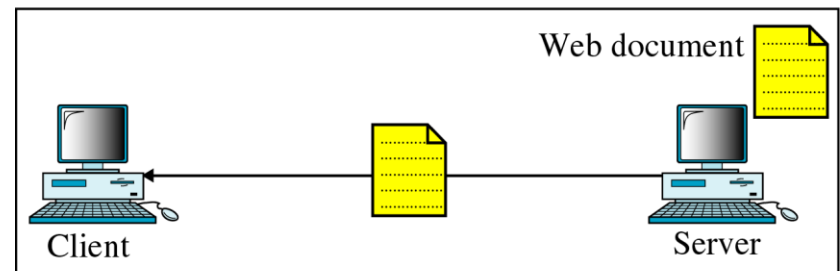
Categories of Web documents ...

■ Static documents

- ❑ Fixed-content document created and stored in a server
 - Content is determined when it is first created not when it is used
- ❑ The file contains text and formatting instructions
- ❑ The client cannot change the content of the document
- ❑ Use HTML technology



a. Request

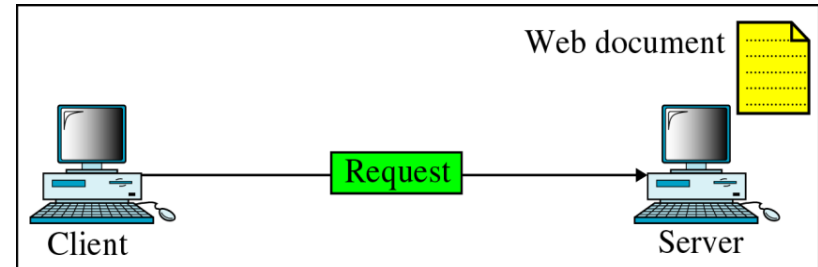


b. Response

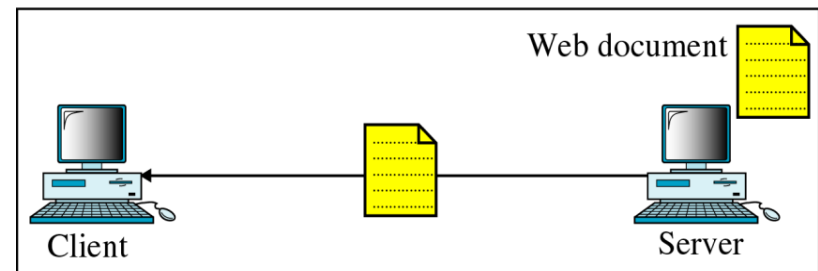
Categories of Web documents ...

■ Static documents

- ❑ Fixed-content document created and stored in a server
 - Content is determined when it is first created not when it is used
- ❑ The file contains text and formatting instructions
- ❑ The client cannot change the content of the document
- ❑ Use HTML technology



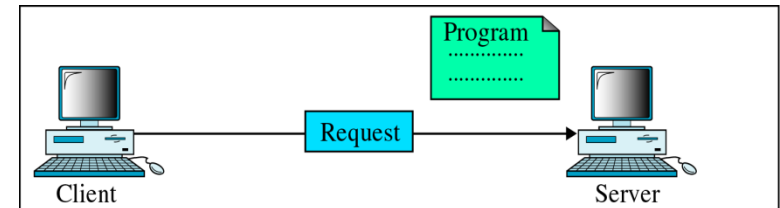
a. Request



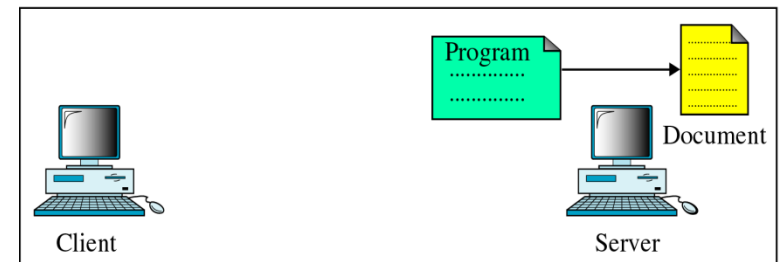
b. Response

Categories of Web documents ...

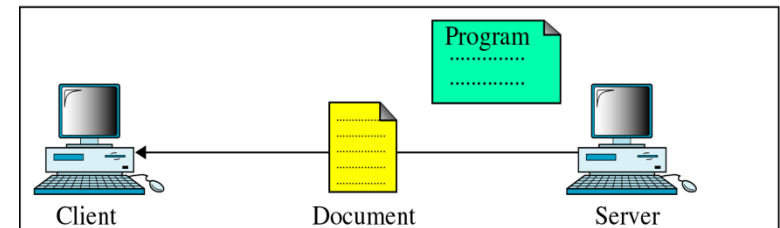
- Dynamic documents
 - ❑ The requested document does not exist in a predefined format but is created on demand
 - ❑ The accessed document contains code (application program) that the server executes to generate the document to be submitted to the browser
 - ❑ Personalized (customized) web pages
 - ❑ Can access restricted resources on the server such as databases
 - ❑ High load on the server
 - ❑ Use Common Gateway Interface (CGI) or ASP technologies for server side programming



a. Request for running a program



b. Running the program and creating the document

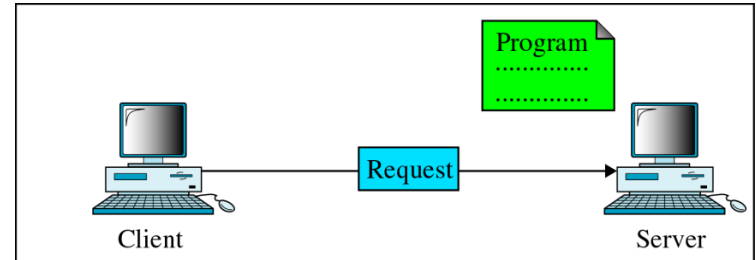


c. Response

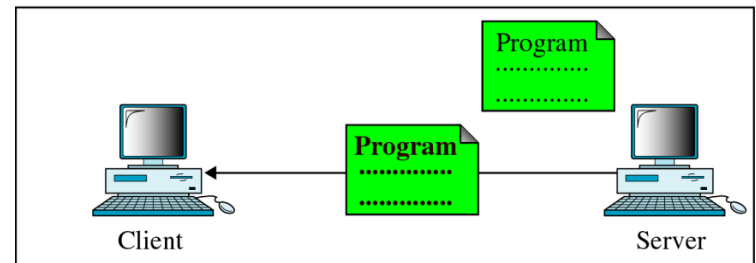
Categories of Web documents ...

■ Active documents

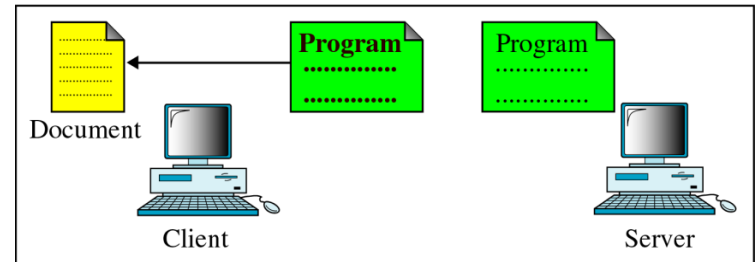
- ❑ The requested document contains a program; the server submits a copy of the program to the browser; the browser will run the program at the client side
- ❑ Can interact with user
- ❑ Does not create overhead for the server in the same way as dynamic documents
- ❑ The client can store the document and run it again and again without making another request
- ❑ Can save bandwidth and transmission time
- ❑ Use JavaScript language for client side programming



a. Request for a copy of a program

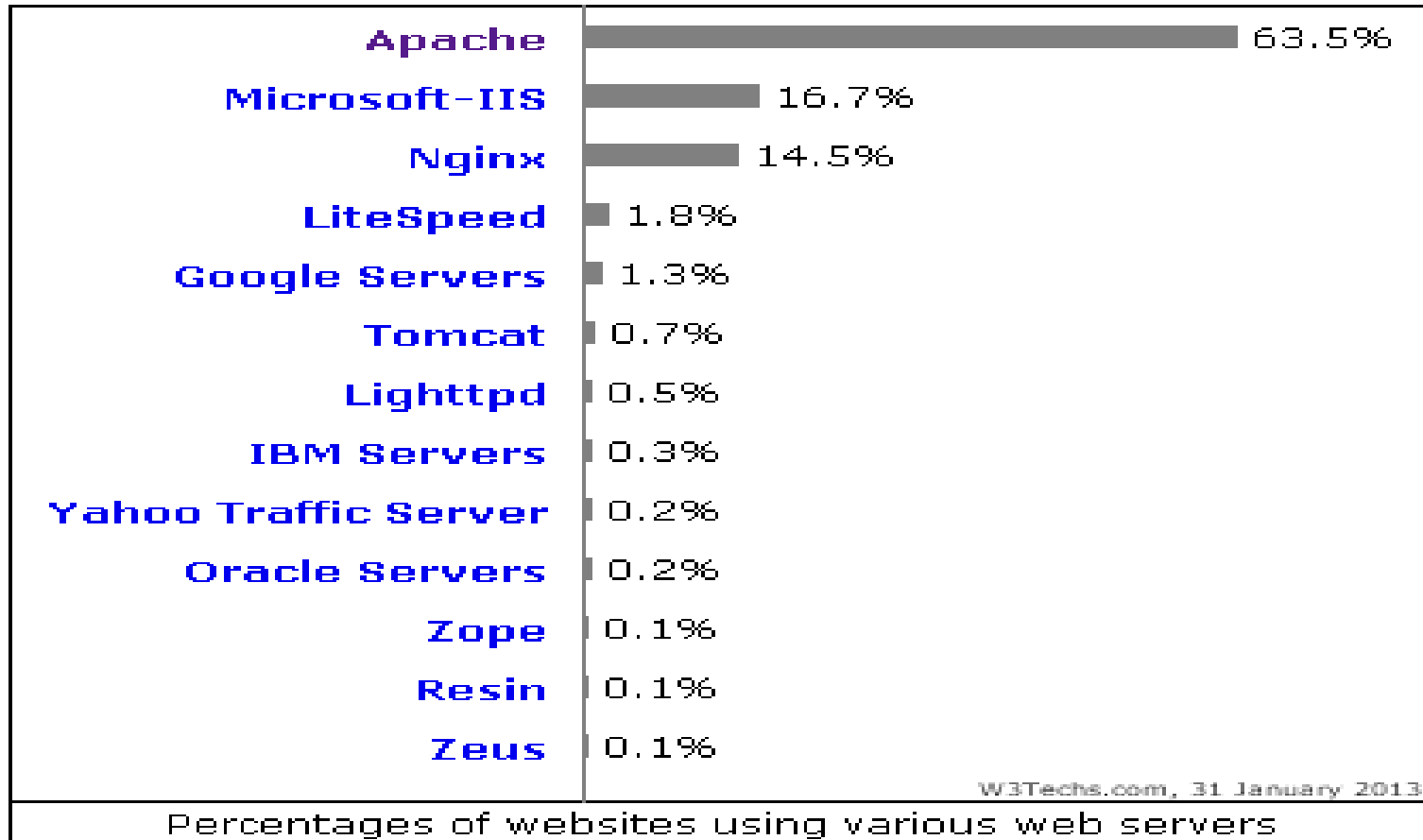


b. Sending a copy of the program



c. Running the program and creating the document

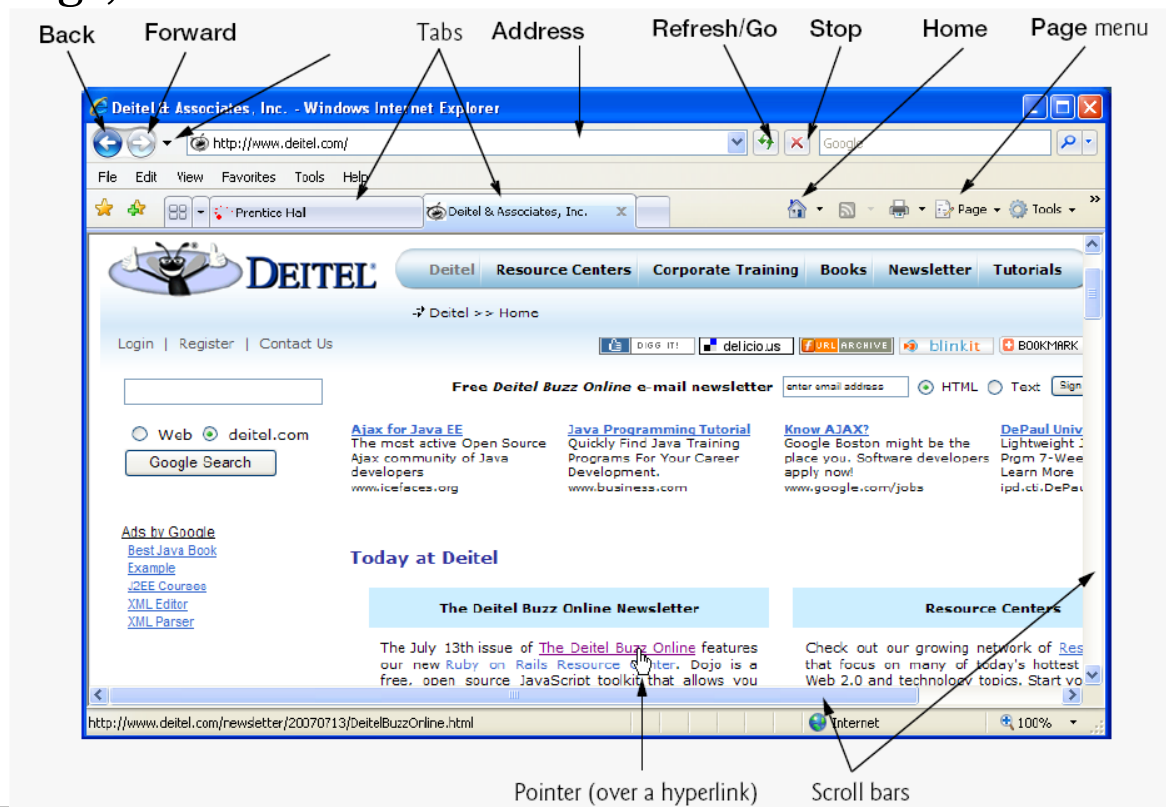
Web Servers Market Share



Source: http://w3techs.com/technologies/overview/web_server/all

Web Browser

- Software programs that fetch, interpret and display web pages
- Popular web browsers: [Mozilla Firefox](#), Microsoft [Internet Explorer](#) (IE), Apple [Safari](#), [Google Chrome](#), [Opera](#)
- Some features: Tabbed Browsing, History Feature, AutoComplete, Favorites, Off-Line Browsing, Viewing source code, Downloading files, Caching, Privacy settings, etc.

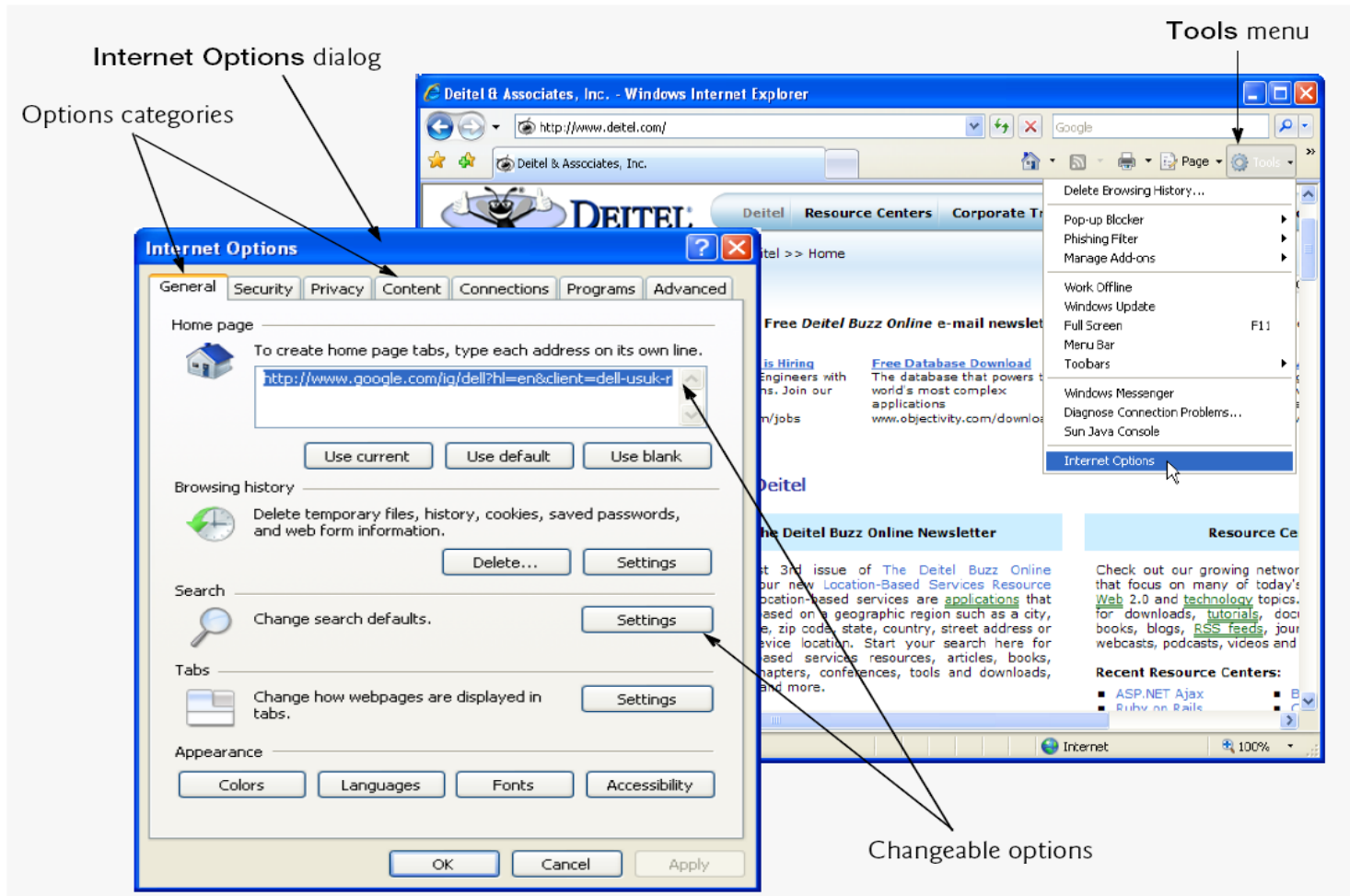


Web Browser ...

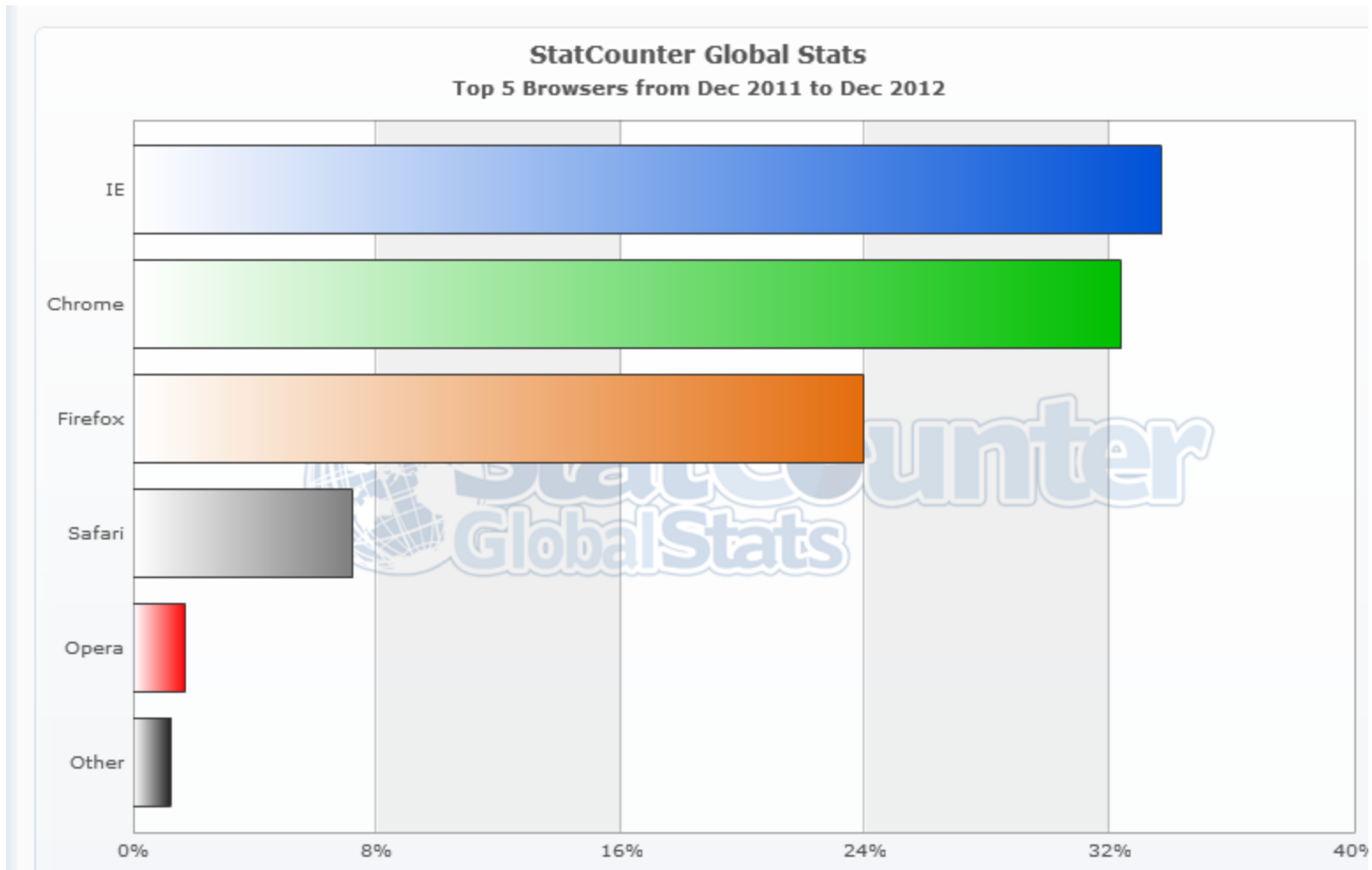
- Is the *client*
- Generates HTTP requests
 - User types URL, clicks a hyperlink or bookmark, clicks “reload” or “submit”
 - Automatically downloads images referenced in an HTML page
- Submits the requests (fetches content)
 - Via one or more HTTP requests
- Presents the response
 - Parses HTML and *renders* the Web page
 - Invokes helper applications (*e.g.*, Acrobat, RealPlayer)
- Maintains cache
 - Stores recently-viewed objects and ensures freshness

Web Browser ...

- Browser settings determine how output will be displayed, how security measures are applied, etc.



Web Browsers – Market Share



Source: <http://gs.statcounter.com/#browser-ww-monthly-201112-201212-map>

Components of a Web Browser

1. User and display interface
2. Browser engine, an embeddable component that provides a high-level interface for querying and manipulating the Rendering Engine;
3. Rendering engine, which performs parsing and layout for HTML documents, optionally styled with CSS
4. Networking subsystem
5. JavaScript interpreter
6. XML parser
7. Display backend, which provides drawing and windowing primitives, user interface widgets, and fonts
8. Data persistence subsystem, which stores various data associated with the browsing session on disk, including bookmarks, cookies, and cache

Cross-Browser Compatibility

- Great diversity of web browsers
 - ❑ There are hundreds of web browsers (and of varying versions) in use
 - ❑ Differ in functionality, performance and supported features for HTML, JavaScript, CSS, etc.
 - ❑ A cross-browser compatible Web page will look more or less the same in all of the existing Web browsers
- Cross-browser compatibility is difficult to achieve
 - ❑ May make Web pages look totally different
 - ❑ Achieving 100% compatibility with all potential browsers is impossible
- W3C is working toward the goal of a universal client-side platform
 - ❑ Write clean code that conforms to the www consortium (W3C) standards to get consistent results across different browser platforms
- HTML editors are, on their part, notorious for creating non-compliant and garbage code.
 - ❑ It is best if you write HTML code directly using a text-editor
 - ❑ If you must use an HTML front-end, use Adobe Dreamweaver or MS Visual Studio (2008 or later)