INTERNET & WEB APPLICATION DEVELOPMENT SWE 444

Fall Semester 2008-2009 (081)

Module 1: Internet Basics for Web Development (III)

Dr. El-Sayed El-Alfy
Computer Science Department
King Fahd University of Petroleum and Minerals
alfy@kfupm.edu.sa

Objectives/Outline

- Module Objectives
  - Learn the basics of the Internet and the Web
  - Identify and describe the key elements of the Internet and the Web

- Lecture Outline
  - Web Caching
  - Firewalls & Web Security
  - Web Cookies
  - Web Development Tools & Technologies
  - Standards
  - Standardization Bodies
  - Brief History
  - Questions & Answers

Web Caching Technology

➤ Caching = temporary storage
➤ Web caching: temporary storage of recently accessed web objects (such as HTML documents, web graphics, etc) for later retrieval without involving the origin server
➤ Web cache sits between web servers and web clients
➤ Advantages of web caching:
  ◦ reduced bandwidth consumption (fewer requests/responses going over the network),
  ◦ reduced server load (fewer requests for a server to handle),
  ◦ reduced latency (since responses for cached requests are available immediately, and closer to the client being served)
  ◦ the web becomes less expensive and better performing

Origin web servers

Internet

LAN

Clients   Proxy Cache
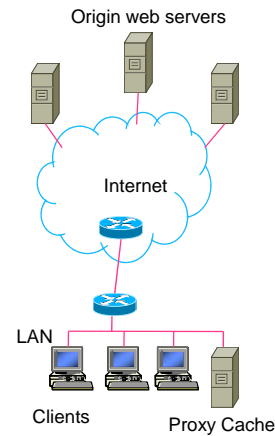
Caching Techniques

➤ Built-in caches (browser caches)
  ◦ Performed by the client application (browser); limited in capability
➤ Personal proxies – Products that extend or replace built-in caches
  ◦ Allows larger storage, more features, or better performance
  ◦ Cache web objects from many servers but all for a single user
  ◦ Examples: NetSonic, Webcellerator, and MicroSurfer
➤ Institutional proxy caches
  ◦ Caching is performed in the middle, between the client and the server as part of a proxy located near network gateways
  ◦ Cached objects requested by one client can retrieved later by another client; serve many clients with cached objects from many servers
  ◦ Examples: Cisco Cache Engine, Lucent IPWorX, WinProxy, WinGate

Caching Techniques …

- ➢ Reverse proxy caches (Gateway caches)
  - ◦ Located directly in front of a particular server, to reduce the number of requests that the server must handle (also known as inverse cache, or sometimes httpd accelerator)
  - ◦ Caches objects for many clients but from (usually) only one server
  - ◦ Examples: IBM Websphere Edge Server, Netplicator

- ➢ Hierarchical proxy caches
  - ◦ A set of proxy caches (organized in a tree) that can interact with each other to check for the requested object to reduce the need to fetch the object directly.

Cache or not to cache …

- ➢ A cache can 'hide' users and make it difficult to see who's using the site
  - ◦ But there are ways that can be used to manage that
- ➢ Caches can serve content that is out of date, or stale.
  - ◦ have a set of rules to determine when to serve cached objects from the cache
  - ◦ web caches can use expiration time or periodically poll the web server to see if a document is modified
- ➢ Sites containing dynamic scripts will be hard to cache
  - ◦ But writing cache-aware scripts can resolve this problem
- ➢ The fact is that caches will be used whether you like it or not
  - ◦ Good planning of your site, makes it load faster, and save load on your server and Internet link;
  - ◦ Users will appreciate a fast-loading site, and will visit it more often
  - ◦ If you don't configure your site to be cached correctly, it will be cached using whatever defaults the cache's administrator decides upon.

Firewalls
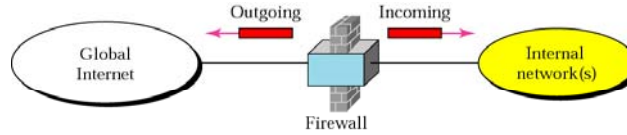
- A firewall is a device (usually a router or computer) installed between an organization's network and the rest of the Internet to control access to the organization's network by filtering exchanged packets.
- Examples of what a firewall can do
  - A firewall may filter all incoming packets destined for a specific host or a specific server such as HTTP
  - It can deny access to a specific service
- Firewalls can be classified as:
  - packet-filter firewall
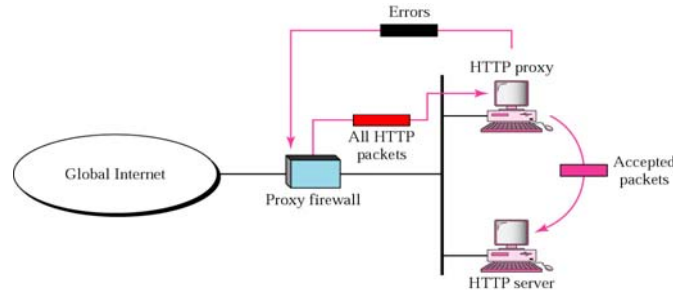 - proxy-based firewall

Packet-Filter Firewall

- It is a router that uses a filtering table to decide which packet must be blocked based on the information in the network layer and transport layer headers (source and destination IP addresses, source and destination ports, type of protocols TCP or UDP)

According to this table:

Incoming pkts from network 131.34.0.0 are blocked   (* means any)

Incoming pkts destined for TELNET service are blocked

Incoming pkts destined for host 194.78.20.8 are blocked

Outgoing pkts destined for HTTP server are blocked

| Interface	Source IP	Source Port	Destination IP	Destination Port
1	131.34.0.0	*	*	*
1	*	*	*	23
1	*	*	194.78.20.8	*
2	*	80	*	*

Proxy-Based Firewall

- ➢ Used to filter a message based on its content at the application layer
- ➢ The firewall first submit the message to a proxy (also called application gateway) to be checked then if accepted it will be forwarded to its destination
- ➢ Example
  - ◦ An organization may want to implement the following policies regarding its web pages: "only Internet users who have previously established business relations with the company can have access"

Web Cookies

- ➢ A mechanism that enables the Web server to store a small amount of text information (up to 4KB) on the client and is sent back from the client to the server with each HTTP request
  - ◦ Enables a Web server distinguish between clients
  - ◦ Used to track user browsing and to customize/personalize pages based on user preferences
- ➢ A cookie is associated with a specific web site
- ➢ A cookie is sent in HTTP header
- ➢ There is a concern of privacy
- ➢ Can expire some time in the future

[Source: wikipedia]
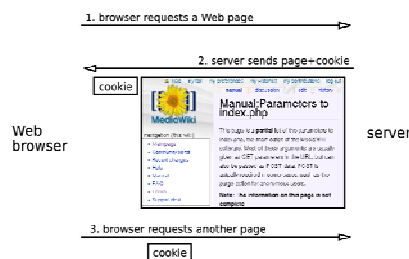
Web Development Tools & Technologies

- ➤ Web development technologies
  - ◦ HTML, XHTML , Cascading Style Sheets (CSS), JavaScript, Document Object Model (DOM), XML, Ajax
- ➤ Rich Internet Applications (RIAs): web applications that offer the responsiveness and rich GUI features of desktop applications
  - ◦ Key tools for building RIAs
    - • Adobe's Flex, Microsoft's Silverlight, ASP.NET Ajax, Sun's JavaServer Faces
- ➤ Other Web development tools and technologies
  - ◦ Adobe's Dreamweaver, JSON, web servers IIS and Apache, MySQL, PHP and ASP.NET
- ➤ Web services allow you to incorporate functionality from existing applications into your own applications quickly and easily, e.g.
  - ◦ Amazon Web Services
  - ◦ Google Maps web services
  - ◦ eBay web services

Standards

- ➤ What?
  - ◦ Guidelines, agreements, rules and working models established by authority, custom or general consent to manufacturers, vendors, government agencies and other service providers
- ➤ Why?
  - ◦ Ensure the kind of *interoperability* (compatibility) necessary in today's marketplace among multi-vendor/multi-service products
- ➤ Standards Categories
  - ◦ **De facto:** (by fact) have not been approved by an organized body but have been adopted as standards because of their widespread (dominant) use
  - ◦ **De jure:** (by law) standards have been legislated by an officially recognized body

Standardization Bodies

- Government regulatory agencies such as
  - FCC (Federal Communications Commission)
  - NIST (National Institute of Standards and Technology)
  - NTIA (National Telecom Information Admin)
- Standards creation committees (non-profit organizations)
  - ISO (International Standards Organization)
  - ITU-T (International Telecommunications Union- Telecommunication Standards Sector)
  - ANSI (American National Standards Institute)
  - IEEE (Institute of Electrical and Electronics Engineers)
  - EIA (Electronic Industries Association)
  - IAB (Internet Activity Board) has internet policies through task forces e.g. IETF - Internet Engineering Task Force
- Forums/Groups/Consortium (to speed up regulations)
  - advancing the deployment of multi-vendor/multi-service products
  - Examples: ATM Forum, MPLS Forum, DSL Forum, OMG Group, W3C

Standardization Bodies

- Internet & Web Standards Organizations
  - The Internet Society
    - A professional organization that provides leadership in addressing issues related to the future of the Internet
    - IETF-- Internet Engineering Task Force
      - RFC – Requests for Comments
    - IAB – Internet Architecture Board
  - ICANN - The Internet Corporation for Assigned Numbers & Names
    - Non-profit organization
    - Main function is to coordinate the assignment of:
      - Internet domain names and IP addresses
      - Protocol parameters and port numbers

Standardization Bodies …

- W3C – World Wide Web Consortium [http://www.w3c.org]
  - Develops recommendations and prototype technologies related to the Web
  - Produces specifications, called Recommendations, in an effort to standardize web technologies
  - WAI – Web Accessibility Initiative
    - Develops recommendations for
      - web content developers,
      - web authoring tool developers,
      - developers of web browsers, and
      - developers of other user agents to facilitate use of the web by those with special needs.
  - WCAG – Web Content Accessibility Guidelines
    - http://www.w3.org/WAI/WCAG20/quickref/

Brief History

- ARPANET – *The grandparent of today's Internet*
  - Implemented in late 1960's by ARPA (Advanced Research Projects Agency of the US Department of Defense (DOD)) for research into networking.
    - The first node was at UCLA, closely followed by nodes at Stanford Research Institute, the University of California at Santa Barbara, and the University of Utah.
    - Computer systems of a dozen universities and institutions were linked with 56KB communications lines
    - The objective was allowing fast communication between researchers using emails, sharing computing resources
  - In 1973, ARPANET linked 40 machines and had international connections to England and Norway.
  - In 1984, over 1,000 host computers were part of ARPANET, DNS was introduced
- In Jan 1983, the first TCP/IP wide area network was operational when the US National Science Foundation (NSF) constructed a university network backbone that would later become the NSFNet.
  - This date is held by some to be technically that of the birth of the Internet.
- 1986 – IETF was founded
- In 1989 , the first proposal for the WWW was made by Tim Berners-Lee, a scientist at CERN (European centre for High Energy Physics – Geneva)
  - provide a global information sharing system between scientists all over the world
  - merge technologies of PC, networking and hypertext

Brief History …

- In 1990, the first browser/editor program appeared ; there were over 300,000 host computers.
- 1991 – an early WWW system released to the high energy physics community via the CERN program library including a browser, Web server and a library
- In Dec. 1991, the first Web server in the US came online at Stanford Linear Accelerator Center (SLAC) in California
- In 1993, the first version of the Mosaic browser (running on X Window System environment) released at the National Center for Supercomputing Applications (NCSA), university of Illinois
- Late 1993, over 500 known Web servers, WWW accounted for 1% of Internet traffic (the rest was remote access, e-mail and file transfer)
- 1994 – Netscape Navigator
- 1995 – MS Internet Explorer
- Jan 1995, the International WWW Consortium (W3C) was founded
  - To develop interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential.
  - Other developments JAVA source code was released in 1995
- 1998 – Google was founded
- 1998 & 2002 – Mozilla
- 2004 – Firefox

Q & A

?

Other Resources

- *Data Communications and Networking*, 4/e. B.A. Forouzan, McGraw-Hill Higher Education 2007. http://www.mhhe.com/forouzan
- The World Wide Web Consortium (W3C)
- The Anatomy of a Large-Scale Hypertextual Web Search Engine, by Sergey Brin & Lawrence Page at Stanford University
- Dive Into Web 2.0, http://www.deitel.com/freeWeb20ebook/