Connecting with Computer Science
Greg Anderson
David Ferro
Robert Hilton

Chapter

the Internet
Objectives

• Learn what the Internet really is
• Become familiar with the architecture of the Internet
• Become familiar with Internet-related protocols
• Understand how the TCP/IP protocols relate to the Internet
• Learn how IP addresses identify devices connected to the Internet
Objectives (continued)

- Learn how DHCP can be used to assign IP addresses
- Learn how routers are used throughout the Internet
- Learn how a DNS server translates a URL into an IP address
- Learn how port numbers are used with IP addresses to expand Internet capabilities
- Learn how NAT is used in networking
Objectives (continued)

- Learn how to determine your own TCP/IP configuration
- Learn how HTML and XML are used with the World Wide Web
- Learn how to develop a simple Web page using HTML
- Learn how search engines make the World Wide Web more usable
Why You Need to Know About... The Internet

- Internet as revolutionary information technology
- Impacts most spheres of human thought and action
  - E-commerce, information exchange, embedded devices
- Intersection with computer science
  - Networks
  - Protocols
  - Server and client programs
What is the Internet?

- The Internet: global collection of LANs and WANs
- Internet service provider (ISP) connects desktop to the Net
- No single entity owns the Internet
  - Some groups propose rules
  - Other groups provide maintenance
- Sum of Net activities serves the larger social good
The Architecture of the Internet

- Anatomy of a connection
  - Desktop linked via modem/transceiver to LAN of ISP
  - ISP switching center: called a point-of-presence (POP)
  - POP connected to larger ISP with larger POP
  - Larger ISP connects to national or international ISPs, called national backbone providers (NBPs)

- Network equipment and protocols critical to process
Figure 7-1 Internet data can pass through several levels of ISPs

- local ISP
- regional ISP
- national backbone provider
- national backbone provider
- regional ISP
- local ISP
Protocols

- Protocol: set of rules that facilitate communication
- Many protocols involved with the Internet
  - HTTP (Hypertext Transfer Protocol)
  - SMTP (Simple Mail Transfer Protocol)
  - FTP (File Transfer Protocol)
- Protocols especially vital for networking
TCP and IP

- TCP (Transmission Control Protocol)
  - Responsible for the reliable delivery of data
  - Separates data into manageable, fixed-size packets
  - Establishes virtual circuit for transmission
  - Manages packet sequencing
  - Re-transmits packets received in error
  - Header appended to data segment contains relevant information
### Table 7-1 TCP header fields

<table>
<thead>
<tr>
<th>header field</th>
<th>size in bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source port</td>
<td>16</td>
</tr>
<tr>
<td>Destination port</td>
<td>16</td>
</tr>
<tr>
<td>Sequence number</td>
<td>32</td>
</tr>
<tr>
<td>Acknowledgment (ACK) number</td>
<td>32</td>
</tr>
<tr>
<td>Data offset</td>
<td>4</td>
</tr>
<tr>
<td>Reserved</td>
<td>6</td>
</tr>
<tr>
<td>Flags</td>
<td>6</td>
</tr>
<tr>
<td>Window</td>
<td>16</td>
</tr>
<tr>
<td>Checksum</td>
<td>16</td>
</tr>
<tr>
<td>Urgent pointer</td>
<td>16</td>
</tr>
<tr>
<td>Options</td>
<td>32</td>
</tr>
</tbody>
</table>
### Table 7-2  IP version 4 datagram header

<table>
<thead>
<tr>
<th>header field</th>
<th>size in bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version</td>
<td>4</td>
</tr>
<tr>
<td>Header length</td>
<td>4</td>
</tr>
<tr>
<td>Type of service</td>
<td>8</td>
</tr>
<tr>
<td>Total length of datagram</td>
<td>16</td>
</tr>
<tr>
<td>Packet identification</td>
<td>16</td>
</tr>
<tr>
<td>Flags</td>
<td>4</td>
</tr>
<tr>
<td>Fragment offset</td>
<td>12</td>
</tr>
<tr>
<td>Time to live (TTL)</td>
<td>8</td>
</tr>
<tr>
<td>Protocol number</td>
<td>8</td>
</tr>
<tr>
<td>Header checksum</td>
<td>16</td>
</tr>
<tr>
<td>Source IP address</td>
<td>32</td>
</tr>
<tr>
<td>Destination IP address</td>
<td>32</td>
</tr>
<tr>
<td>IP options</td>
<td>32</td>
</tr>
</tbody>
</table>
TCP and IP (continued)

- IP layer adds address header to TCP packets
  - Most widespread version (IPv4) has 32 bit value
  - New version of IP (IPv6) has 128-bit addresses

- IPv4 address hierarchy
  - First part identifies network class (A, B, C, D, E)
  - Middle part identifies the host on the network
  - Final part identifies the node connected to the host
<table>
<thead>
<tr>
<th>bits</th>
<th>class type</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>class A</td>
</tr>
<tr>
<td>0 hosts (126 possible) first number 1–126</td>
<td>0 nodes (16,777,214 possible)</td>
</tr>
<tr>
<td>10</td>
<td>class B</td>
</tr>
<tr>
<td>10 hosts (16,382 possible) first number 128–191</td>
<td>10 nodes (65,534 possible)</td>
</tr>
<tr>
<td>110</td>
<td>class C</td>
</tr>
<tr>
<td>110 hosts (2,097,150 possible) first number 192–223</td>
<td>110 nodes (254 possible)</td>
</tr>
<tr>
<td>1110</td>
<td>class D</td>
</tr>
<tr>
<td>broadcast</td>
<td>11110</td>
</tr>
<tr>
<td>future use</td>
<td></td>
</tr>
</tbody>
</table>
TCP and IP (continued)

- Class indicates entity size and IP address allocation
- IANA (Internet Assigned Numbers Authority)
  - maintains global high-level registry of IP addresses
- ARIN (American Registry for Internet Numbers)
  - Agency allocates IP addresses to NBPs and ISPs
- ISPs allocate addresses to other ISPs and home users
- IP supports subnet
DHCP

- DHCP: Dynamic Host Configuration Protocol
  - Allows for automatic assignment of IP addresses
  - Computer uses DHCP to get IP address from router

- Flexibility for Database Administrator
  - Sets up server to allocate block of addresses
  - One time cost to configure computer for DHCP
Routers

- Router
  - Computer linked to different communication lines
  - Routes packets on to line closer to destination
  -Joins networks together, including the Internet
  -Monitors communication lines for congestion
  -May send message packets along different paths
- Routers work in a manner similar to post office
Figure 7-3 Routers provide many alternate routes for packets
High-Level Protocols

- Suite of protocols can be compared to OSI layers
  - TCP/IP spans Session, Transport, Network layers
  - SMTP, HTTP, FTP “above” TCP/IP in OSI model
- High-level protocols use TCP/IP to accomplish tasks
  - TCP splits messages into packets, if necessary
  - TCP layer adds header, forwards to IP for address
  - IP sends packets to Data Link and Physical layers
### Connecting with Computer Science

#### Figure 7-4 TCP/IP protocols compared to the OSI model

<table>
<thead>
<tr>
<th>OSI model</th>
<th>TCP/IP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> Physical layer</td>
<td>IP</td>
</tr>
<tr>
<td>electrical signals and cabling</td>
<td></td>
</tr>
<tr>
<td><strong>2</strong> Data Link (MAC) layer</td>
<td>TCP (delivery ensured) UDP (delivery not ensured)</td>
</tr>
<tr>
<td>transmits packets from node to node based on station address</td>
<td></td>
</tr>
<tr>
<td><strong>3</strong> Network layer</td>
<td></td>
</tr>
<tr>
<td>routes data to different LANs and WANs based on network address</td>
<td>Telnet</td>
</tr>
<tr>
<td><strong>4</strong> Transport layer</td>
<td></td>
</tr>
<tr>
<td>ensures delivery of entire file or message</td>
<td>SMTP</td>
</tr>
<tr>
<td><strong>5</strong> Session layer</td>
<td></td>
</tr>
<tr>
<td>starts, stops session; maintains order</td>
<td></td>
</tr>
<tr>
<td><strong>6</strong> Presentation layer</td>
<td></td>
</tr>
<tr>
<td>encryption, data conversion: ASCII to EBCDIC, BCD to binary, etc.</td>
<td></td>
</tr>
<tr>
<td><strong>7</strong> Application layer</td>
<td>FTP</td>
</tr>
<tr>
<td>type of communication; e-mail, file transfer, Web page</td>
<td></td>
</tr>
</tbody>
</table>

**network user**
SMTP

- SMTP: Simple Mail Transport Protocol (SMTP)
  - Used to send e-mail messages over the Internet
  - Establishes link from e-mail client to e-mail server
  - Handshaking creates parameters of communication
- Receipt of e-mail handled by another protocol
  - POP3 (Post Office Protocol version 3)
  - IMAP (Internet Message Access Protocol)
FTP

- FTP: File Transfer Protocol
  - Provides efficient transmission of data files
  - Requires client and server programs (like SMTP)
  - Most OSs include command-line FTP client
    - Windows: type FTP at the command prompt or
    - Specify server address in the IE address bar
Figure 7-5 Command-line FTP session

```
Command Prompt - ftp
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.
C:\Documents and Settings\rbhilton>ftp
ftp> open ftp.aol.com
Connected to ftp.aol.com.
220-
220- *********************************************
220- NNN NN EEEEE W W W W A AA 0000 LL
220- NNN NN EE W W W W A A A A 00 00 LL
220- NN N NN EEEE W W W W A A A A 00 00 LL
220- NN NNN EE WWWW WWWW AAAAAA 00 00 LL
220- NN NNN EEEEE W W W W A A A A 0000 LLLLLL
220- *********************************************
220- Please be advised, all activity on this server is logged.
220- 
220- *********************************************NOTICE**********NOTICE******************************************
220- ONLY GUEST ACCESS IS ACCEPTED AT THIS SITE
220- Guest logins are “ftp” or “anonymous”. Please use your full e-mail
220- address as the password when accessing the archive as a guest user.
220- 
220- ftp.newaol.com FTP server (SunOS 5.8) ready.
User (ftp.aol.com:(none)): 
```
Telnet

- Telnet: Internet standard protocol for remote login to a UNIX host
  - Telnet runs on top of TCP/IP
  - Allows client computer remote control over host
  - Most OSs include a command-line Telnet client
HTTP

- HTTP: Hypertext Transfer Protocol (HTTP)
  - Developed in 1990 by Tim Berners-Lee
  - Allows Web browsers and Web servers to communicate
  - Central to the idea of the World Wide Web
  - Example, http://www.course.com
    - “http” tells browser you are retrieving Web page with Hypertext Transfer Protocol
URLs and DNS

• Development of the Domain Name System (DNS)
  – Layers natural language name over IP address
  – Provides user friendly interface with Internet

• Uniform Resource Locator (URL)
  – Consists of the domain name followed by specific folder and/or filenames
  – DNS server resolves domain names from URLs into IP addresses
Figure 7-7 Structure of a URL

http://www.course.com/myfolder/myfile.html

protocol
host name
network name
folder
filename
• DNS server: computer maintained by ISP
  – Performs lookup on URL (Uniform Resource Locator)
  – Responsible for a portion of the world’s domains
  – Communicates with other DNS servers
• Domain levels
  – Top level domains (TLDs): .com, .edu, .gov, .net, .org
  – Server at each level has knowledge of lower-level
  – Example: faculty.weber.edu
### Table 7-3 Top-level domain names used on the Internet

<table>
<thead>
<tr>
<th>TLD</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>.aero</td>
<td>Air-transport industry</td>
</tr>
<tr>
<td>.arpa</td>
<td>Address and Routing Parameter Area</td>
</tr>
<tr>
<td>.biz</td>
<td>Business</td>
</tr>
<tr>
<td>.com</td>
<td>Commercial</td>
</tr>
<tr>
<td>.coop</td>
<td>Cooperative</td>
</tr>
<tr>
<td>.edu</td>
<td>U.S. educational</td>
</tr>
<tr>
<td>.gov</td>
<td>U.S. government</td>
</tr>
<tr>
<td>.info</td>
<td>Information</td>
</tr>
<tr>
<td>.int</td>
<td>International organization</td>
</tr>
<tr>
<td>.mil</td>
<td>U.S. military</td>
</tr>
<tr>
<td>.museum</td>
<td>Museum</td>
</tr>
<tr>
<td>.name</td>
<td>Individuals, by name</td>
</tr>
<tr>
<td>.net</td>
<td>Network</td>
</tr>
<tr>
<td>.org</td>
<td>Organization</td>
</tr>
<tr>
<td>.pro</td>
<td>Profession</td>
</tr>
<tr>
<td>.ca, .mx</td>
<td>Canada and Mexico and other countries are represented by two-letter codes</td>
</tr>
</tbody>
</table>
Port Numbers

• Port number: address specification below IP layer
• Port functions like apartment number in address
• Most protocols have a standard port number
  – A possible 65,636 port numbers for each IP address
  – Specify port by appending number to domain or IP address
    • 192.168.2.33 specifies the IP address
    • 8080 specifies port number
<table>
<thead>
<tr>
<th>port number</th>
<th>protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>21 FTP (File Transfer Protocol)</td>
</tr>
<tr>
<td>23</td>
<td>Telnet</td>
</tr>
<tr>
<td>25</td>
<td>SMTP (Simple Mail Transfer Protocol)</td>
</tr>
<tr>
<td>53</td>
<td>DNS (Domain Name System service)</td>
</tr>
<tr>
<td>68</td>
<td>DHCP (Dynamic Host Control Protocol)</td>
</tr>
<tr>
<td>80</td>
<td>HTTP (Hypertext Transfer Protocol)</td>
</tr>
<tr>
<td>110</td>
<td>POP3 (Post Office Protocol version 3)</td>
</tr>
<tr>
<td>139</td>
<td>NetBIOS</td>
</tr>
</tbody>
</table>
NAT

• NAT (Network Address Translation) protocol
  – Multiple computers share one Internet connection
  – Dependent on DHCP and port numbers

• NAT structure
  – Internal range: 192.168.0.0 to 192.168.255.255
  – Only IP address presented to the Internet: 192.168
  – Internal nodes use port number with IP address
  – TCP routes messages to node with matching port
Checking Your Configuration

• Type IPCONFIG command at in console window

• Screen reflects
  – Current IP address
  – Subnet mask
  – Address of your gateway to the Internet

• More information: use IPCONFIG /ALL command
Figure 7-8 Executing the IPCONFIG command at the command prompt

Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

C:\Windows> IPCONFIG

Windows IP Configuration

Ethernet adapter Belkin Connect Ethernet:

  Connection-specific DNS Suffix  . :
  IP Address.  .  .  .  .  .  .  .  .  .  .  : 192.168.0.33
  Subnet Mask  .  .  .  .  .  .  .  .  .  .  : 255.255.255.0
  Default Gateway  .  .  .  .  .  .  .  .  : 192.168.0.1

C:\Windows>
HTML

• Web page
  – Largest class: text files
  – Contains text information and HTML (Hypertext Markup Language) tags

• HTML tags
  – Formatting commands
  – Browser uses tags to display graphical content
  – Knowledge needed to use Web page design tools
<table>
<thead>
<tr>
<th>tag</th>
<th>purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;HTML&gt;</code> <code>&lt;/HTML&gt;</code></td>
<td>Used to provide a boundary for the HTML document; everything between the <code>&lt;HTML&gt;</code> and the  <code>&lt;/HTML&gt;</code> is considered to be part of the Web page</td>
</tr>
<tr>
<td><code>&lt;HEAD&gt;</code> <code>&lt;/HEAD&gt;</code></td>
<td>The <code>&lt;HEAD&gt;</code> tags are placed inside the <code>&lt;HTML&gt;</code> tags; the <code>&lt;HEAD&gt;</code> tags provide a boundary for items that are not part of the document, but are used to direct the browser to do certain things</td>
</tr>
<tr>
<td>tag</td>
<td>purpose</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><code>&lt;TITLE&gt;</code></td>
<td><code>&lt;TITLE&gt;</code> tags are used to surround the Web document's title. This is the title that appears in the title bar of the browsers when the page is displayed. The <code>&lt;TITLE&gt;</code> tags go inside of the <code>&lt;HEAD&gt;</code> tags.</td>
</tr>
<tr>
<td><code>&lt;BODY&gt;</code></td>
<td><code>&lt;BODY&gt;</code> tags enclose the part of the Web page document that is to be displayed in the browser; the <code>&lt;BODY&gt;</code> tags are placed inside of the <code>&lt;HTML&gt;</code> tags, but not inside the <code>&lt;HEAD&gt;</code> tags.</td>
</tr>
<tr>
<td><code>&lt;BR&gt;</code></td>
<td><code>&lt;BR&gt;</code> is used to force the browser display area to go to a new line. Note that the <code>&lt;BR&gt;</code> is a single tag. There is no closing tag for the <code>&lt;BR&gt;</code> tag.</td>
</tr>
<tr>
<td><code>&lt;P&gt;</code></td>
<td>The <code>&lt;P&gt;</code> tags are used to define a paragraph in the Web document and to cause a paragraph break to take place.</td>
</tr>
<tr>
<td><code>&lt;FONT&gt;</code></td>
<td>The <code>&lt;FONT&gt;</code> tags are used to specify the font for the text between the tags; the opening <code>&lt;FONT&gt;</code> tag has special arguments that are used to specify the type, size, and color of the font for the included text.</td>
</tr>
<tr>
<td><code>&lt;A&gt;</code></td>
<td>The <code>&lt;A&gt;</code> tag pair is used to specify a link to another Web page or a specific location on the current page; the opening <code>&lt;A&gt;</code> tag has arguments that reference the linked page or position.</td>
</tr>
<tr>
<td><code>&lt;IMG&gt;</code></td>
<td>The <code>&lt;IMG&gt;</code> tag is used to insert an image in the document; the <code>&lt;IMG&gt;</code> tag has arguments that are used to specify the location and size of the image to be displayed.</td>
</tr>
</tbody>
</table>
Table 7-5 (*Continued*)

<table>
<thead>
<tr>
<th>tag</th>
<th>purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;FORM&gt;</code></td>
<td>The <code>&lt;FORM&gt;</code> tags are used to provide the boundaries for an input form on the Web page; other tags are placed inside of the <code>&lt;FORM&gt;</code> tag to create items such as input boxes and buttons on the Web page</td>
</tr>
<tr>
<td><code>&lt;INPUT&gt;</code></td>
<td>The <code>&lt;INPUT&gt;</code> tag is used to specify data input objects inside of the <code>&lt;FORM&gt;</code> tags; these tags allow data input to take place on a Web page</td>
</tr>
<tr>
<td><code>&lt;TABLE&gt;</code></td>
<td>The <code>&lt;TABLE&gt;</code> tags are used to define an area on the Web page that displays tabular data in rows and columns</td>
</tr>
<tr>
<td><code>&lt;TR&gt;</code></td>
<td>The <code>&lt;TR&gt;</code> tags are placed inside of the <code>&lt;TABLE&gt;</code> tags to signify the start of a table row</td>
</tr>
<tr>
<td><code>&lt;TD&gt;</code></td>
<td>The <code>&lt;TD&gt;</code> tags are placed inside of the <code>&lt;TR&gt;</code> tags to define a column within the row of a table</td>
</tr>
</tbody>
</table>
Creating a Simple Web Page

- Open Notepad window
  - Type in the HTML document shown in Figure 7-9
  - Save file to disk with .htm or .html extension
  - Use Windows Explorer to locate file
  - Browser displays document formatted in HTML code
- HTML is free form and not case sensitive
- Hyperlinks: connection to another page
Figure 7.9 HTML tags for a simple Web page

```html
<html>
<head>
<title>My First Web Page</title>
</head>
<body>
<p>My First Web Text<br>
   <b><font size="3" face="Arial">My First Table</font></b></p>
<table width="30%" border="1">
  <tr>
    <td><b>Protocol</b></td><td><b>Purpose</b></td>
  </tr>
  <tr>
    <td>TCP</td><td>Reliable Delivery</td>
  </tr>
  <tr>
    <td>IP</td><td>Addressing</td>
  </tr>
  <tr>
    <td>HTTP</td><td>Web Pages</td>
  </tr>
</table>
</body>
</html>
```
Figure 7-10  Simple Web page as displayed on the browser screen

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCP</td>
<td>Reliable Delivery</td>
</tr>
<tr>
<td>IP</td>
<td>Addressing</td>
</tr>
<tr>
<td>HTTP</td>
<td>Web Pages</td>
</tr>
</tbody>
</table>
Creating a Simple Web Page (continued)

- **Web Server Programs:** provide dynamic Web pages
  - DHTML (Dynamic Hypertext Markup Language)
  - HTML/DHTML and scripting code
- **Completely dynamic Web pages**
  - CGI, ASP, JSP, PHP, and Python
- **Web services:** program contains data used by other programs or Web pages
XML

• SGML: specification source for HTML and XML
• HTML limitation: does not affect Web page content
• XML (Extensible Markup Language)
  – Similar to HTML in structure
  – Also provides data and metadata (information about data)
  – Can be used to display Web pages
  – Most important use: transfer data
<?xml version="1.0" encoding="ISO-8859-1"?>
<?xml-stylesheet type="text/xsl" href="sample.xsl"?>
<dvd_library>
  <dvd>
    <id>D0146</id>
    <title>The Lord of the Rings: The Return of the King</title>
    <rating>PG-13</rating>
    <price>24.95</price>
    <review>****</review>
  </dvd>
  <dvd>
    <id>D3218</id>
    <title>Dumb and Dumber</title>
    <rating>PG-13</rating>
    <price>14.95</price>
    <review>****</review>
  </dvd>
  <dvd>
    <id>D4482</id>
    <title>Mom and Dad Save The World</title>
    <rating>PG</rating>
    <price>8.95</price>
    <review>****</review>
  </dvd>
</dvd_library>
Using the Internet

- Internet as tremendous resource
- Invest time to develop searching skills
  - Professional, home, student life
- Search engines
  - Replace newsgroups as primary search vehicle
Search Engines

• Most common search method: “crawling”
  – Uses program called a bot (for robot) or a spider
  – Bot starts with a few pages submitted for indexing
  – Indexed pages scanned for links to other Web pages
  – Process continues for every retrieved page
  – Index to page made from relevant words and keywords of <META> tag

• Computer scientists design and maintain search engines
One Last Thought

• The Internet has grown into global forum
  – Information gathering and communication
  – Entertainment and E-commerce

• Some issues
  – Anonymity and reliability of information
  – Identity theft and virus replication

• Computer scientist and Internet co-evolve
Summary

- Internet is a collection of LANs and WANs
- No single entity controls the Internet
- Hierarchy of ISPs and NBPs manage connections
- Protocols are critical in Internet Operation
- TCP/IP lies at heart of protocol suite
- TCP manages data delivery
Summary (continued)

- IP provides for addressing
- Every node in the Internet has a unique address
- DHCP automatically assigns addresses
- Routers: key hardware component of Internet
- High-level protocols: HTTP, FTP, SMTP, POP3, and IMAP
- Uniform resource locator (URL): specifies address
Summary (continued)

- DNS server: resolves domain name to IP number
- Hypertext markup Language (HTML): language of World Wide Web
- Web page: content formatted with HTML tags
- Scripting Languages: generate dynamic pages
- Hyperlink: connection to another Web page
- Search engines: generate database of searchable Web sites