Introduction to Computer Science
Lesson 10

BSc in Computer Science
University of New York, Tirana

Assoc. Prof. Marenglen Biba

Networking and the Internet

- Last lesson
  - Network Fundamentals
  - The Internet
- Today
  - The World Wide Web
  - Internet Protocols
  - Security

World Wide Web

- Hypertext and HTTP
- Browser gets documents from Web server
- Documents identified by URLs

The World Wide Web

- Multimedia information disseminated over the Internet, is based on the concept of **hypertext**
  - a term that originally referred to text documents that contained links, called **hyperlinks,** to other documents.

- Today, hypertext has been expanded to encompass images, audio, and video, and because of this expanded scope it is sometimes referred to as **hypermedia.**

WWW: the networkwide web

- A reader of hypertext documents can explore related documents or follow a train of thought from document to document.
- As portions of various documents are linked to other documents, an intertwined web of related information is formed.
- When implemented on a computer network, the documents within such a web can reside on different machines, forming a network-wide web.
- The web that has evolved on the Internet spans the entire globe and is known as the **World Wide Web** (also referred to as **WWW, W3,** or the **Web**).
- A hypertext document on the World Wide Web is often called a **Web page.**
- A collection of closely related Web pages is called a **Web site.**

WWW: Tim Berners-Lee

- The World Wide Web had its origins in the work of Tim Berners-Lee who realized the potential of combining the linked-document concept with internet technology and produced the first software for implementing the WWW in December of 1990.

WWW: Tim Berners-Lee

- Using concepts from his earlier hypertext systems like ENQUIRE, British engineer, computer scientist and at that time employee of the CERN, Sir Tim Berners-Lee, now Director of the World Wide Web Consortium (W3C), wrote a proposal in March 1989 for what would eventually become the World Wide Web.
- At CERN, a European research organisation near Geneva straddling the border between France and Switzerland, Berners-Lee and Belgian computer scientist Robert Cailliau proposed in 1990 to use hypertext "to link and access information of various kinds as a web of nodes in which the user can browse at will", and they publicly introduced the project in December of the same year.

The World Wide Web Consortium

- The World Wide Web Consortium (W3C) was formed in 1994 to promote the World Wide Web by developing protocol standards (known as W3C standards).
- W3C is headquartered at CERN, the high-energy particle physics laboratory in Geneva, Switzerland.
- CERN is where the original HTML markup language was developed as well as the HTTP protocol for transferring HTML documents over the Internet.
- Today W3C is the source of many standards (including standards for XML and numerous multimedia applications) that lead to compatibility over a wide range of Internet products.
- You can learn more about W3C via its Web site at http://www.w3c.org.

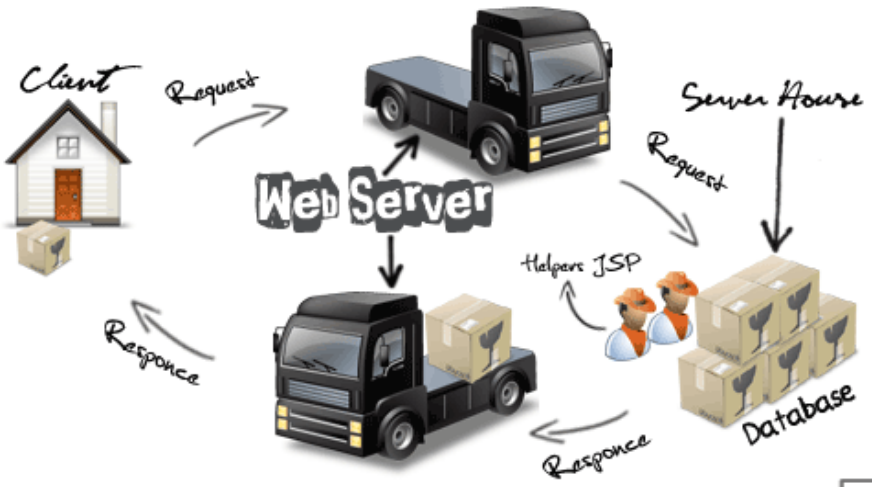
Web Implementation: the client

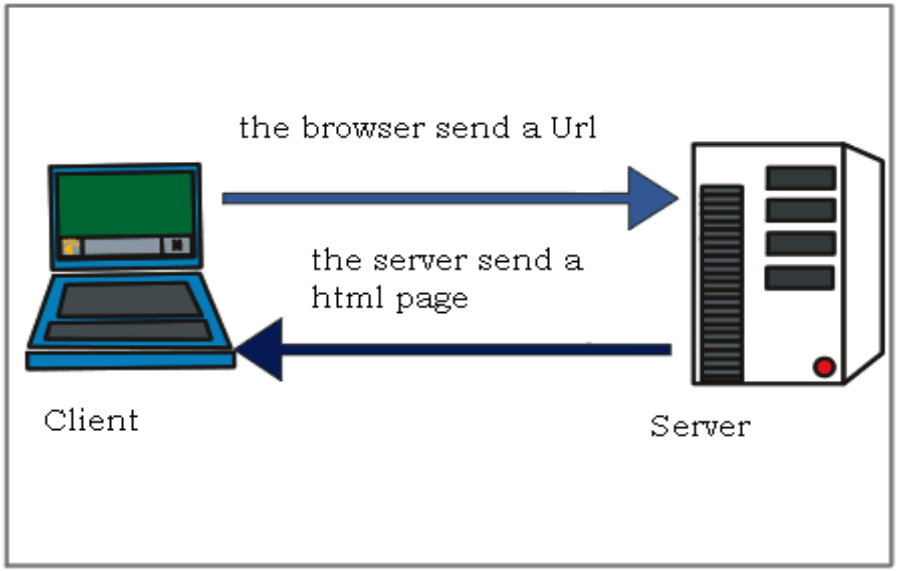
- Software packages that allow users to access hypertext on the Internet fall into one of two categories:
  - packages that play the role of clients, and packages that play the role of servers.
- A client package resides on the user's computer and is charged with the tasks of obtaining materials requested by the user and presenting these materials to the user in an organized manner.
- It is the client that provides the user interface that allows a user to browse within the Web.
- Hence the client is often referred to as a **browser,** or sometimes as a Web browser.

Web Implementation: the client

- The server package (often called a **Web server**) resides on a computer containing hypertext documents to be accessed.
- Its task is to provide access to the documents under its control as requested by clients.
- In summary, a user gains access to hypertext documents by means of a browser residing on the user's computer.
- This browser, playing the role of a client, obtains the documents by soliciting the services of the Web servers scattered throughout the Internet.
- Hypertext documents are normally transferred between browsers and Web servers using a protocol known as the **Hypertext Transfer Protocol (HTTP).**

Web server





Picture taken from: http://www.9lessons.info/2008/10/webserver.html

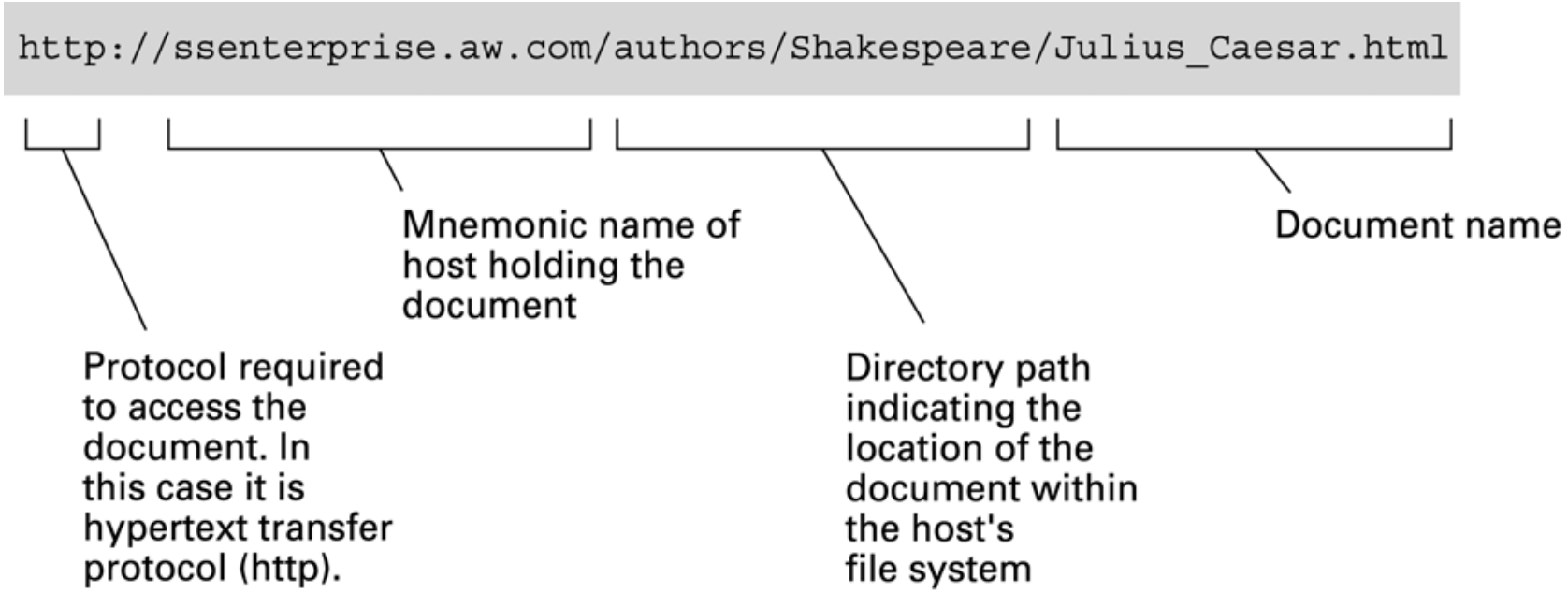
Location of documents: URL

- In order to locate and retrieve documents on the World Wide Web, each document is given a unique address called a **Uniform Resource Locator (URL).**

- Each URL contains the information needed by a browser to contact the proper server and request the desired document.

- Thus to view a Web page, a person first provides his or her browser with the URL of the desired document and then instructs the browser to retrieve and display the document.

Figure 4.8 **A typical URL**



HTML

- A traditional hypertext document is similar to a text file because its text is encoded character by character using a system such as ASCII or Unicode.
- The distinction is that a hypertext document also contains special symbols, called **tags,** that describe how the document should appear on a display screen, what multimedia resources (such as images) should accompany the document, and which items within the document are linked to other documents.
- This system of tags is known as **Hypertext Markup Language (HTML).**

Hypertext Document Format

- Encoded as text file
- Contains tags to communicate with browser
  - Appearance
    - <h1> to start a level one heading
    - <p> to start a new paragraph
  - Links to other documents and content
    - <a href = . . . >
  - Insert images
    - <img src = . . . >

Figure 4.9 **A simple Web page**

a. The page encoded using HTML.

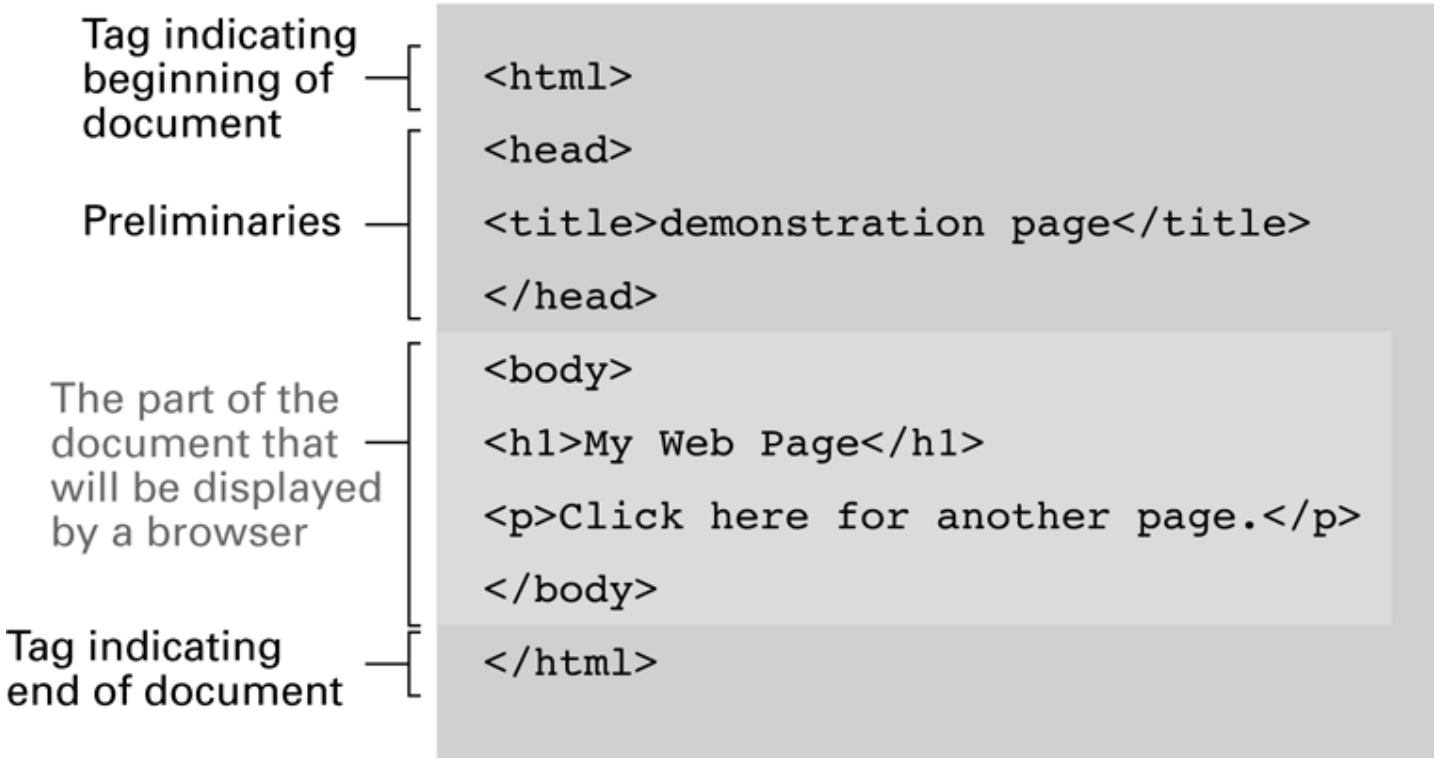


Figure 4.9 A simple Web page (continued)

b. The page as it would appear on a computer screen.

Figure 4.10 An enhanced simple Web page

a. The page encoded using HTML.


```
<html>
<head>
<title>demonstration page</title>
</head>
<body>
<h1>My Web Page</h1>
<p>Click
   <a href="http://crafty.com/demo.html">
   here
   </a>
   for another page.</p>
</body>
</html>
```

Figure 4.10 An enhanced simple Web page (continued)

b. The page as it would appear on a computer screen.

Extensible Markup Language (XML)

- XML: A language for constructing markup languages similar to HTML
  - A descendant of SGML
  - Opens door to a World Wide *Semantic* Web

XML

- HTML is essentially a notational system by which a text document along with the document's appearance can be encoded as a simple text file.
- In a similar manner we can also encode nontextual material as text files — an example being sheet music.

XML

- The **eXtensible Markup Language (XML)** is a standardized style (similar to that of our music example) for designing notational systems for representing data as text files.
- Actually, XML is a simplified derivative of an older set of standards called the Standard Generalized Markup Language, better known as SGML.
- Following the XML standard, notational systems called **markup languages** have been developed for representing mathematics, multimedia presentations, and music.
- In fact, XHTML is the markup language based on the XML standard that was developed for representing Web pages.
  - Actually, the original version of HTML was developed before the XML standard was solidified, and therefore some features of HTML do not strictly conform to XML.
- That is why you might see references to XHTML, which is the version of HTML that rigorously adheres to XML.

Using XML

<staff clef = “treble”> <key>C minor</key>

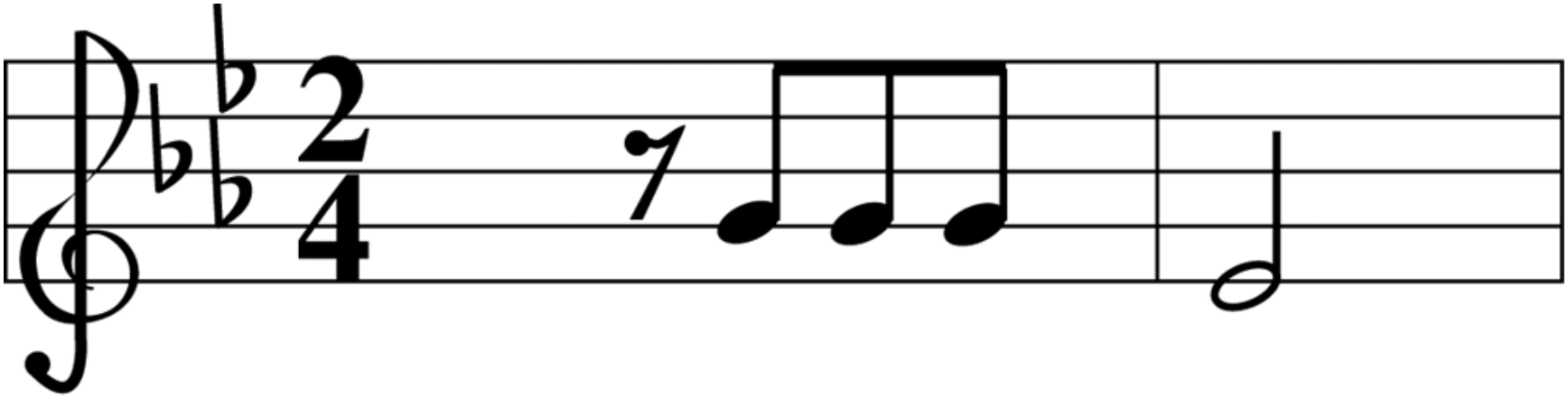
<time> 2/4 </time>

<measure> < rest> egth </rest> <notes>
egth G, egth G, egth G
</notes></measure>

<measure> <notes> hlf E
</notes></measure>

</staff>

Figure 4.11 **The first two bars of Beethoven's Fifth Symphony**



Client Side Versus Server Side

- Client-side activities
  - Examples: java applets, javascript, Macromedia Flash
- Server-side activities
  - Common Gateway Interface (CGI)
  - Servlets
  - PHP

Client Side and Server Side scripts

- If we want a Web page involving animation or one that allows a customer to fill out an order form and submit the order,
  - These needs would require additional activity by either the browser or the Web server.
  - Such activities are called **client-side** activities if they are performed by a client (such as a browser) or **server-side** activities if they are performed by a server (such as a Web server).

Example: Travel agent web page

- As an example, suppose a travel agent wanted customers to be able to identify desired destinations and dates of travel, at which time the agent would present the customer with a customized Web page containing only the information pertinent to that customer's needs.
- In this case the travel agent's Web site would first provide a Web page that presents a customer with the available destinations.
- On the basis of this information, the customer would specify the destinations of interest and desired dates of travel (a client-side activity).
- This information would then be transferred back to the agent's server where it would be used to construct the appropriate customized Web page (a server-side activity) which would then be sent to the customer's browser.

Example: Search engine

- Another example occurs when using the services of a search engine.
- In this case a user at the client specifies a topic of interest (a client-side activity) which is then transferred to the search engine.
- The search engine constructs a customized Web page identifying documents of possible interest (a server-side activity) and sent back to the client.

Example: Web Mail

- Still another example occurs in the case of **Web mail** — an increasingly popular means by which computer users are able to access their email by means of Web browsers.
- In this case, the Web server is an intermediary between the client and the client's mail server.
- Essentially, the Web server builds Web pages that contain information from the mail server (a server-side activity) and sends those pages to the client where the client's browser displays them (a client-side activity).
- Conversely, the browser allows the user to create messages (a client-side activity) and sends that information to the Web server, which then forwards the messages to the mail server (a server-side activity) for mailing.

Client-side scripts

- There are numerous systems for performing client- and server-side activities, each competing with the others for prominence.
- An early and still popular means of controlling client-side activities is to include programs written in the language JavaScript (developed by Netscape Communications, Inc.) within the HTML source document for the Web page.
- From there a browser can extract the programs and follow them as needed.
- Another approach (developed by Sun Microsystems) is to first transfer a Web page to a browser and then transfer additional program units called applets (written in the language Java) to the browser as requested within the HTML source document.
- Still another approach is the system Flash (developed by Macromedia) by which extensive multimedia client-side presentations can be implemented.

Server-side scripts

- An early means of controlling server-side activities was to use a set of standards called CGI (Common Gateway Interface) by which clients could request the execution of programs stored at a server.
- A variation of this approach (developed by Sun Microsystems) is to allow clients to cause program units called servlets to be executed at the server side.
- A simplified version of the servlet approach is applicable when the requested server-side activity is the construction of a customized Web page, as in our travel agent example.
- In this case Web page templates called JavaServer Pages (JSP) are stored at the Web server and completed using information received from a client.

ASP and PHP

- A similar approach is used by Microsoft, where the templates from which customized Web pages are constructed are called Active Server Pages (ASP).

- In contrast to these proprietary systems, PHP (originally standing for Personal Home Page but now considered to mean PHP Hypertext Processor) is an open source system for implementing server-side functionality.

Scripts: Security issues

- The fact that Web servers routinely transfer programs to clients where they are executed leads to ethical questions on the server side and security questions on the client side.
  - If the client blindly executes any program sent to it by a Web server, it opens itself to malicious activities by the server.
- Likewise, the fact that clients can cause programs to be executed at the server leads to ethical questions on the client side and security questions on the server side.
  - If the server blindly executes any program sent to it by a client, security breaches and potential damage at the server could result.

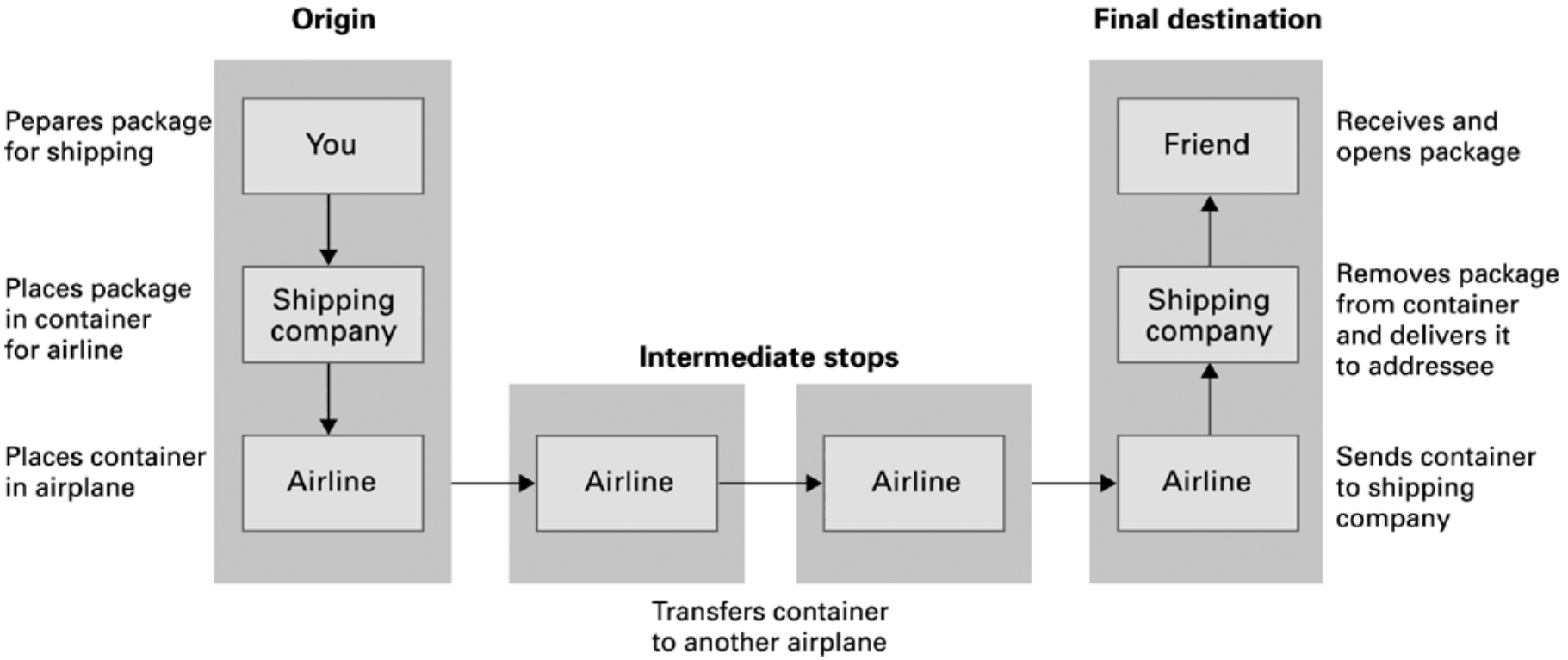
Internet Protocols

The Layered Approach to Internet Software

- A principal task of networking software is to provide the infrastructure required for transferring messages from one machine to another.

- In the Internet, this message-passing activity is accomplished by means of a hierarchy of software units, which perform tasks analogous to those that would be performed if you were to send a gift in a package from the West Coast of the United States to a friend on the East Coast

Figure 4.12 **Package-shipping example**



Internet Software Layers

- **Application:** Constructs message with address
- **Transport:** Chops message into packets
- **Network:** Handles routing through the Internet
- **Link:** Handles actual transmission of packets

Figure 4.13 **The Internet software layers**

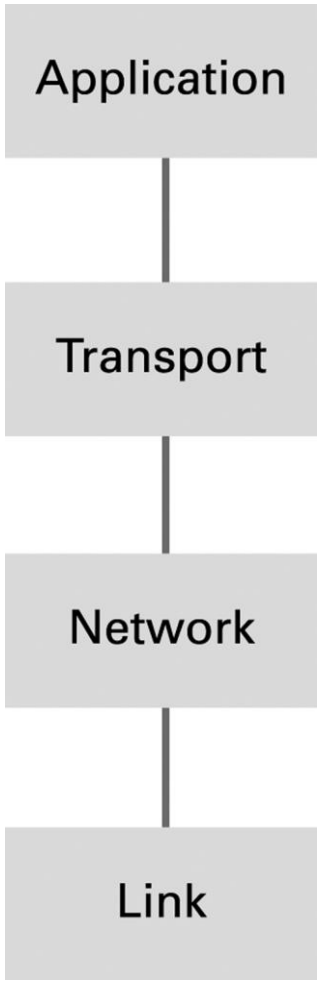
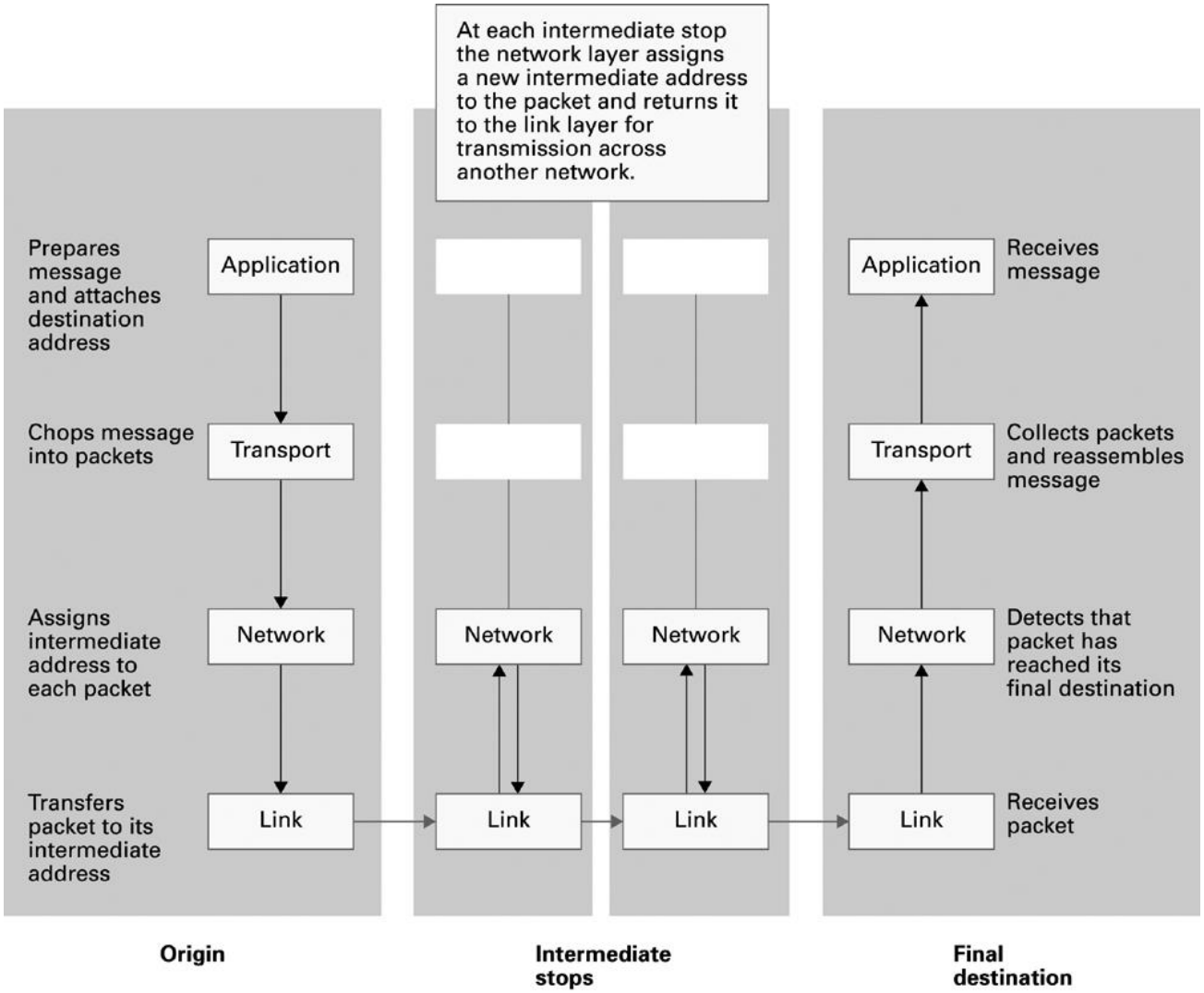


Figure 4.14 Following a message through the Internet



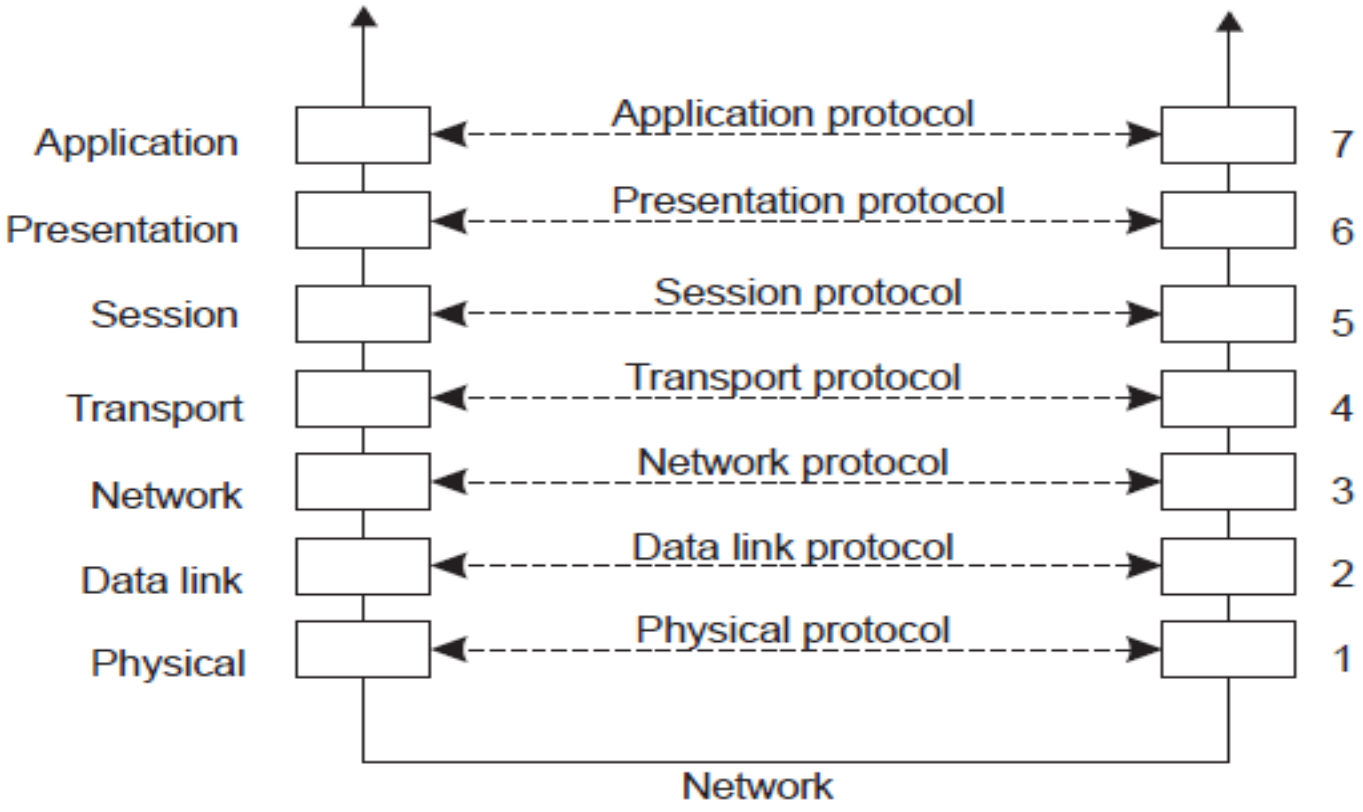
TCP/IP Protocol Suite

- Transport Layer
  - TCP
  - UDP
- Network Layer
  - IP (IPv4 and IPv6)

OSI reference model

- The demand for open networks has generated a need for published standards by which manufacturers can supply equipment and software that function properly with products from other vendors.
- One standard that has resulted is the **Open System Interconnection (OSI)** reference model, produced by the International Organization for Standardization.
- This standard is based on a seven-level hierarchy as opposed to the four-level hierarchy we have just described.

OSI



TCP and IP

- The TCP/IP protocol suite is a collection of protocol standards used by the Internet to implement the four-level communication hierarchy implemented in the Internet.
- Actually, the **Transmission Control Protocol (TCP)** and the **Internet Protocol (IP)** are the names of only two of the protocols in this vast collection — so the fact that the entire collection is referred to as the TCP/IP protocol suite is rather misleading.
- More precisely, TCP defines a version of the transport layer.
- We say a *version* because the TCP/IP protocol suite provides for more than one way of implementing the transport layer; one of the other options is defined by the **User Datagram Protocol (UDP).**

Differences between TCP and UDP

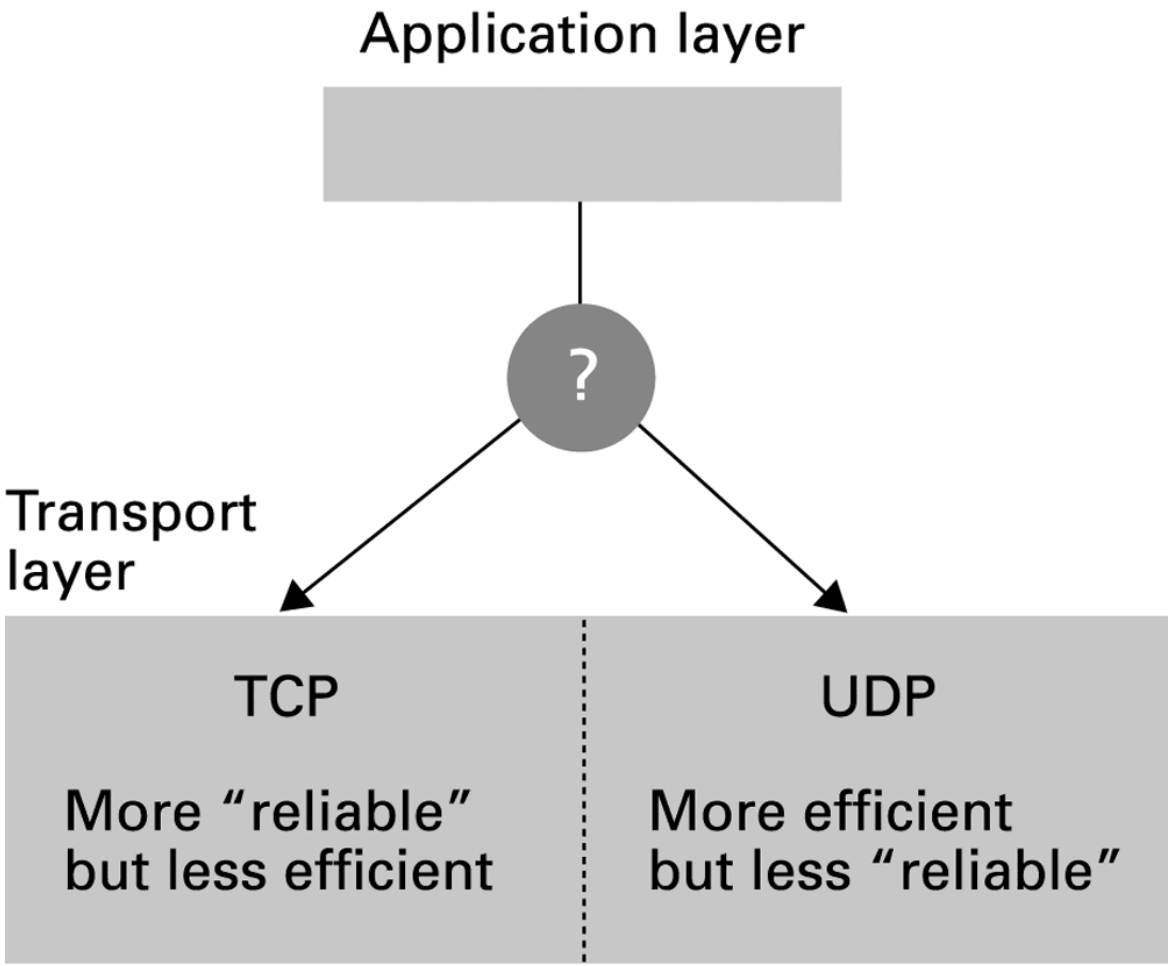
- There are several differences between TCP and UDP.
- One is that before sending a message as requested by the application layer, a transport layer based on TCP sends its own message to the transport layer at the destination telling it that a message is about to be sent.
- It then waits for this message to be acknowledged before starting to send the application layer's message.
- In this manner, a TCP transport layer is said to establish a connection before sending a message.
- A transport layer based on UDP does not establish such a connection prior to sending a message. It merely sends the message to the address it was given and forgets about it.
- For all it knows, the destination computer might not even be operational.
- For this reason, UDP is called a connectionless protocol.

Differences between TCP and UDP

- Another difference between TCP and UDP is that TCP transport layers at the origin and destination work together by means of acknowledgments and packet retransmissions to assure that all segments of a message are successfully transferred to the destination.

- For this reason TCP is called a reliable protocol, whereas UDP, which does not offer such retransmission services, is said to be an unreliable protocol.

Figure 4.15 Choosing between TCP and UDP



Choice of protocol

- Considering the differences, it does not mean that UDP is a poor choice.
- If an application is prepared to handle the potential consequences of UDP, that option might be the better choice.
- For example, the efficiency of UDP makes it the protocol of choice for DNS lookups and VoIP.
- However, because email is less time sensitive, mail servers use TCP to transfer email.

IP

- IP is the Internet's standard for implementing the tasks assigned to the network layer.
- We have already observed that this task consists of forwarding, which involves relaying packets through the Internet, and routing, which involves updating the layer's forwarding table to reflect changing conditions.
- For instance, a router may malfunction, meaning that traffic should no longer be forwarded in its direction, or a section of the Internet may become congested, meaning that traffic should be routed around the blockage.
- Much of the IP standard associated with routing deals with the protocols used for communication among neighboring network layers as they interchange routing information.

Hop count

- An interesting feature associated with forwarding is that each time an IP network layer at a message's origin prepares a packet, it appends a value called a **hop count,** or time to live, to that packet.
- This value is a limit to the number of times the packet should be forwarded as it tries to find its way through the Internet.
- Each time an IP network layer forwards a packet, it decrements that packet's hop count by one.
- With this information, the network layer can protect the Internet from packets circling endlessly within the system.
- Although the Internet continues to grow on a daily basis, an initial hop count of 64 remains more than sufficient to allow a packet to find its way through the maze of routers within today's ISPs.

IPv4 and IPv6

- For years a version of IP known as IPv4 (IP version four) has been used for implementing the network layer within the Internet.
- However, the Internet is rapidly outgrowing the 32-bit internet addressing system dictated by IPv4.
- To solve this problem as well as to implement other improvements such as multicast, a new version of IP known as IPv6, which uses internet addresses consisting of 128 bits, has been established.
- The process of converting from IPv4 to IPv6 is currently underway — and it is expected that the use of 32-bit addresses within the Internet will be extinct by 2025.

Security

Encryption

- FTPS, HTTPS, SSL
- Public-key Encryption
  - Public key: Used to encrypt messages
  - Private key: Used to decrypt messages
- Certificates and Digital Signatures

Encryption

- Encryption can be used so that even if the data fall into unscrupulous hands, the encoded information will remain confidential.
- Today, many traditional Internet applications have been altered to incorporate encryption techniques, producing what are called "secure versions" of the applications.
- Examples include **FTPS,** which is a secure version of FTP, and SSH, as a secure replacement for telnet.

HTTPS

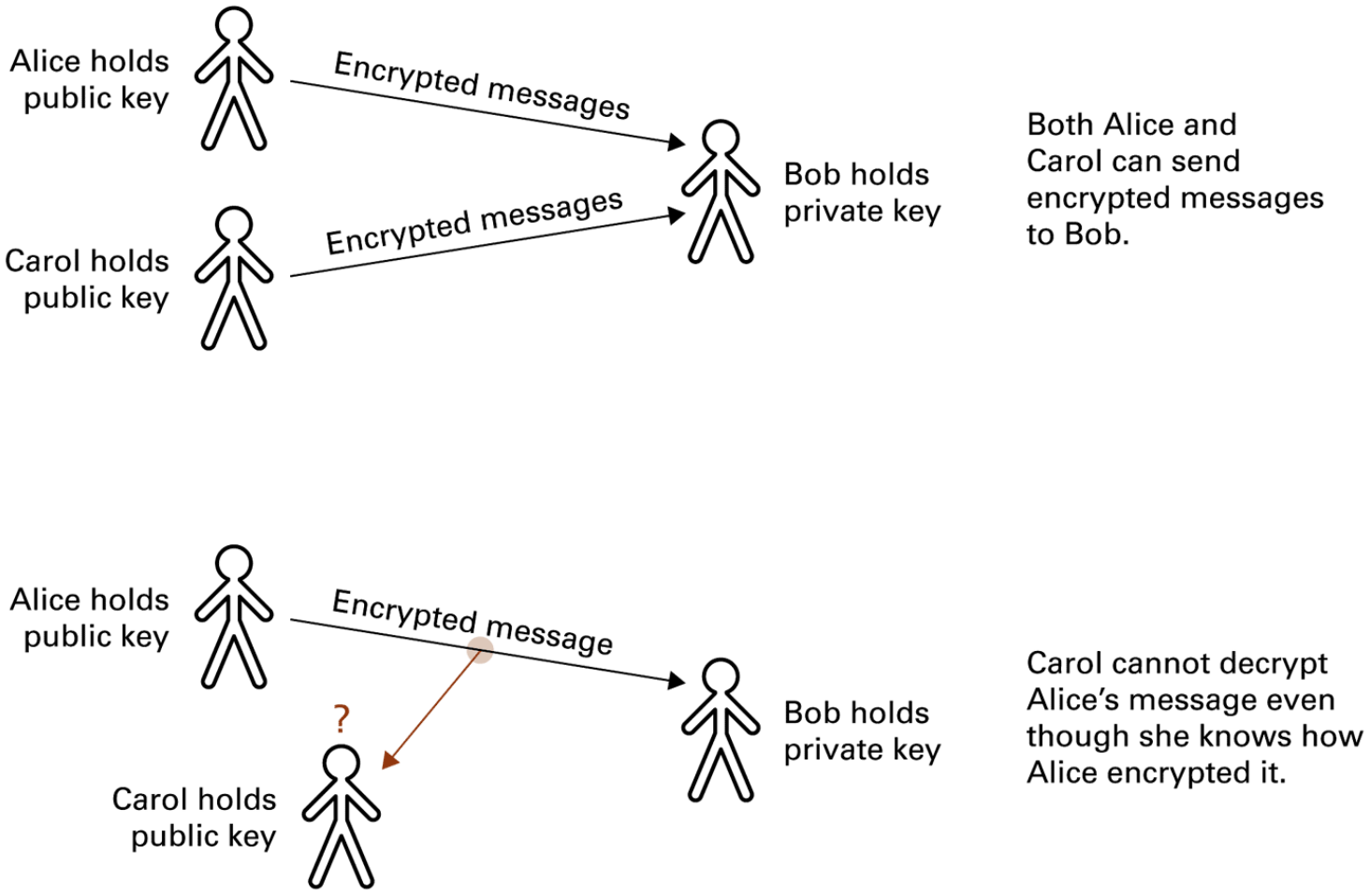
- Still another example is the secure version of HTTP, known as **HTTPS,** which is used by most financial institutions to provide customers with secure Internet access to their accounts.
- The backbone of HTTPS is the protocol system known as **Secure Sockets Layer (SSL)** which was originally developed by Netscape to provide secure communication links between Web clients and servers.
- Most browsers indicate the use of SSL by displaying a tiny padlock icon on the computer screen.

Public-key encryption

- The pair public-private key is unique.
- The public key is made public: everybody can encrypt messages with it.
- The private key is held only by the entity that needs to decrypt the messages.

- Important: it is not possible to find the private key in reasonable computation time.

Figure 4.16 **Public-key encryption**



Issues in public-key systems

- There are subtle problems within public-key systems.
- One is to ensure that the public key being used is, in fact, the proper key for the destination party.
- For example, if you are communicating with your bank, you want to be sure that the public key you are using for encryption is the one for the bank and not an impostor.
- If an impostor presents itself as the bank (an example of spoofing) and gives you its public key, the messages you encrypt and send to the "bank" would be meaningful to the impostor and not your bank.
- Thus, the task of associating public keys with correct parties is significant.

Certificate authorities

- One approach to resolving this problem is to establish trusted Internet sites, called **certificate authorities,** whose task is to maintain accurate lists of parties and their public keys.
- These authorities, acting as servers, then provide reliable public-key information to their clients in packages known as certificates.
- A **certificate** is a package containing a party's name and that party's public key.
- Many commercial certificate authorities are now available on the Internet, although it is also common for organizations to maintain their own certificate authorities in order to maintain tighter control over the security of the organization's communication.

Authentication

- Finally, we should comment on the role public-key encryption systems play in solving problems of **authentication** — making sure that the author of a message is, in fact, the party it claims to be.
- The critical point here is that, in some public-key encryption systems, the roles of the encryption and decryption keys can be reversed.
- That is, text can be encrypted with the private key, and because only one party has access to that key, any text that is so encrypted must have originated from that party.
- In this manner, the holder of the private key can produce a bit pattern, called a **digital signature,** that only that party knows how to produce.

Legal Approaches to Network Security

- Another way of enhancing the security of computer networking systems is to apply legal remedies.
- There are, however, two obstacles to this approach.
- The first is that making an action illegal does not preclude the action.
- All it does is provide a legal recourse.
- The second is that the international nature of networking means that obtaining recourse is often very difficult.
- What is illegal in one country might be legal in another.
- Ultimately, enhancing network security by legal means is an international project, and thus must be handled by international legal bodies—a potential player would be the International Court of Justice in The Hague.

US laws: Computer Fraud and Abuse Act

- Passed in 1984, although it has been amended several times.
- It is under this act that most cases involving the introduction of worms and viruses have been prosecuted.
- In short, the act requires proof that the defendant knowingly caused the transmission of a program or data that intentionally caused damage.

US laws: Computer Fraud and Abuse Act

- The Computer Fraud and Abuse Act also covers cases involving the theft of information.
- In particular, the act outlaws obtaining anything of value via the unauthorized access of a computer.
- Courts have tended to assign a broad interpretation to the phrase “anything of value,” and thus the Computer Fraud and Abuse Act has been applied to more than the theft of information.
- For instance, courts have ruled that the mere use of a computer might constitute “anything of value.”

Electronic Communication Privacy Act

- The right of privacy is another, and perhaps the most controversial, networking issue facing the legal community.
- Questions involving an employer's right to monitor the communications of employees and the extent to which an Internet service provider is authorized to access the information being communicated by its clients have been given considerable thought.
- In the United States, many of these questions are addressed by the Electronic Communication Privacy Act (ECPA) of 1986, which has its origins in legislation to control wiretapping.

ECPA

- Although the act is lengthy, its intent is captured in a few short excerpts. In particular, it states that

- *Except as otherwise specifically provided in this chapter any person who intentionally intercepts, endeavors to intercept, or procures any other person to intercept or endeavor to intercept, any wire, oral, or electronic communication . . . shall be punished as provided in subsection (4) or shall be subject to suit as provided in subsection (5).*

ECPA

- *…any person or entity providing an electronic communication service to the public shall not intentionally divulge the contents of any communication . . . on that service to any person or entity other than an addressee or intended recipient of such communication or an agent of such addressee or intended recipient.*

ECPA and government agencies

- Moreover, the act goes on to give some government agencies authority to monitor electronic communications under certain restrictions.
- These provisions have been the source of much debate.
- For example, in 2000 the FBI revealed the existence of its system, called Carnivore, that reports on the communication of all subscribers of an Internet service provider rather than just a court-designated target.
- In 2001 in response to the terrorist attack on the World Trade Center, congress passed the controversial USA PATRIOT (Uniting and Strengthening America by Providing Appropriate Tools Required to Intercep and Obstruct Terrorism) Act that modified the restrictions under which government agencies must operate.

End of lesson 10

- Readings
  - Chapter 4