Chapter 4: Networking and the Internet

- 4.1 Network Fundamentals
- **4.2 The Internet**
- **4.3 The World Wide Web**
- **4.4 Internet Protocols**
- **4.5 Security**

The Internet

- The Internet: one internet spanning the world, involving millions of machines
  - Started by DARPA in 1973: ARPANET
  - Development of the Internet shifted from a government-sponsored project to an academic research project.
  - 1990's - the Internet expanded into commercial areas
- Initially text based, today any digital format

- The Internet is an internet (network of networks) operating with different hardware and software platforms but still communicating with each other
- It is run by the individual networks co-operating through communications protocols which facilitate the inter-platform communication.
- Familiar protocols are;
  - TCP/IP
  - HTTP
  - FTP

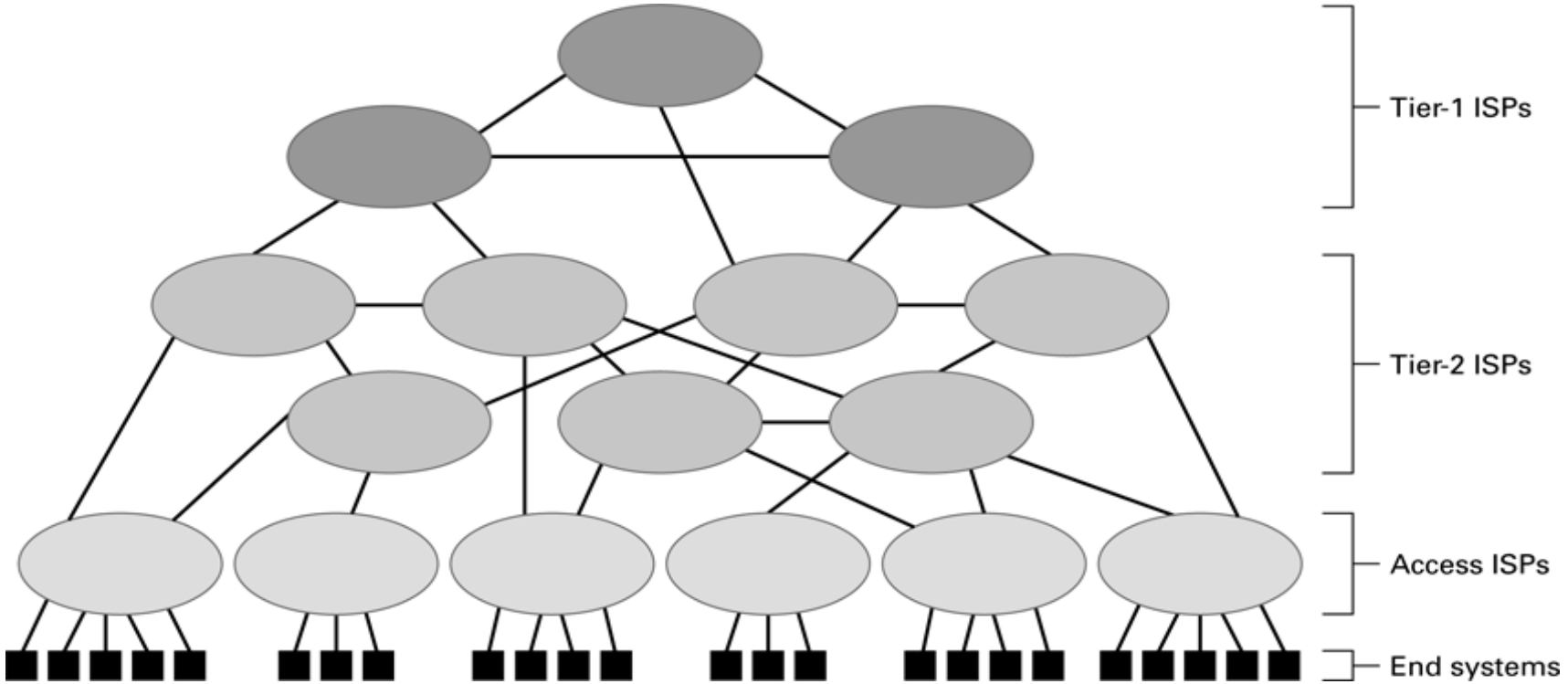
The Internet allows a user to:

- Communicate with people on the other side of the world
- Tap into the resources of other computers
- Search libraries and visit virtual places (museums, exhibitions,…)
- Watch videos, listen to music, read multimedia magazines
- Shop (goods, tickets), compare, sell, …
- Value of online sales in UK (2023) around 2.22 billion pounds

Internet Architecture

- Internet Service Provider (ISP)
  - Tier-1
  - Tier-2
- Access ISP: Provides connectivity to the Internet
  - Traditional telephone (dial up connection)
  - Cable connections
  - DSL
  - Wireless

Figure 4.7 Internet Composition

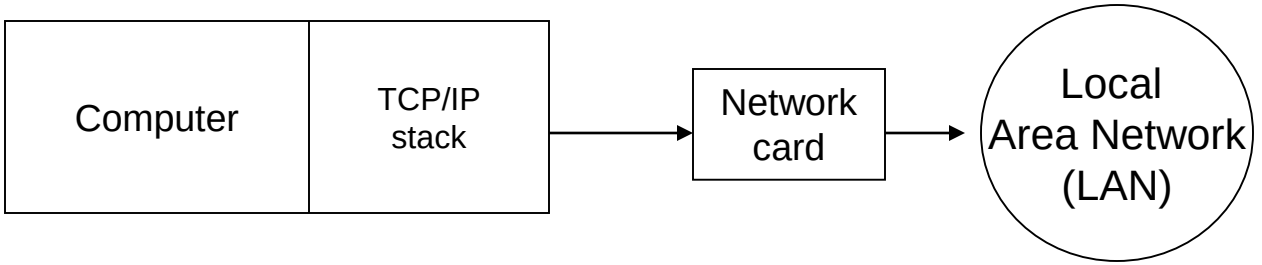


Connecting to the Internet

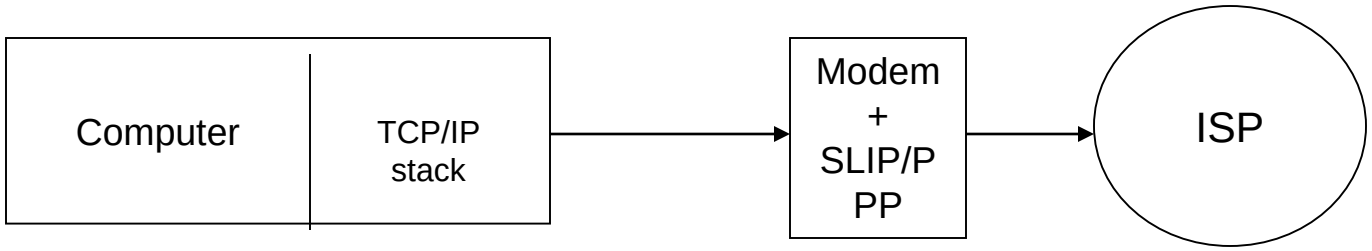
- Requires software that can understand and interpret the TCP/IP protocols
  - i.e. a socket or TCP/IP stack
  - e.g. Winsock used on PC
- Full internet access (usually) achieved by:
  - Computer on a network using a network card *or*
  - Dialling to an ISP with a modem

Connecting to the Internet

- A network card and
- A local area network connection



Connecting to the Internet (cont.)



- A modem,
- A communication protocol
  - Serial Line Internet Protocol (SLIP)
  - Point-to-point Protocol (PPP)
- And an Internet service provider

Connecting to the internet

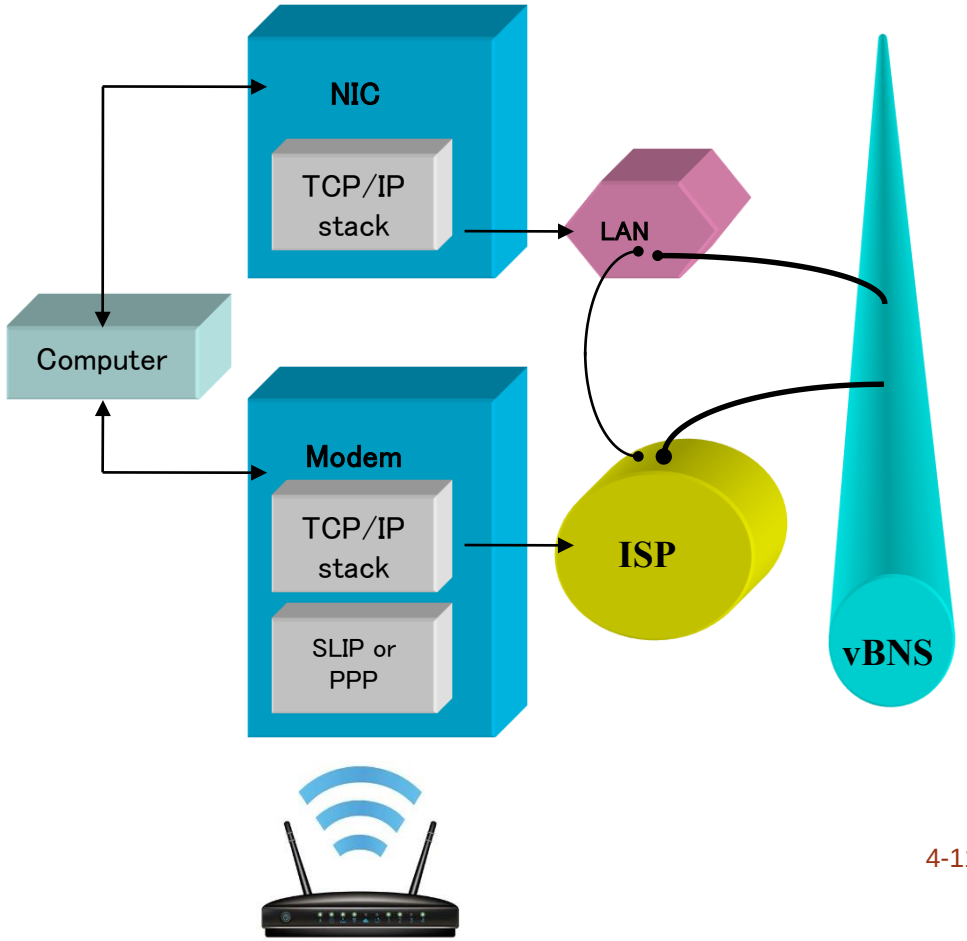
- Network Interface Card (NIC)
  - Transmission Control Protocol (TCP) and Internet Protocol (IP)

or

- Modem
  - Serial Line Internet Protocol (SLIP)

or

  - Point-to-point Protocol (PPP)



TCP/IP

To send information on the internet

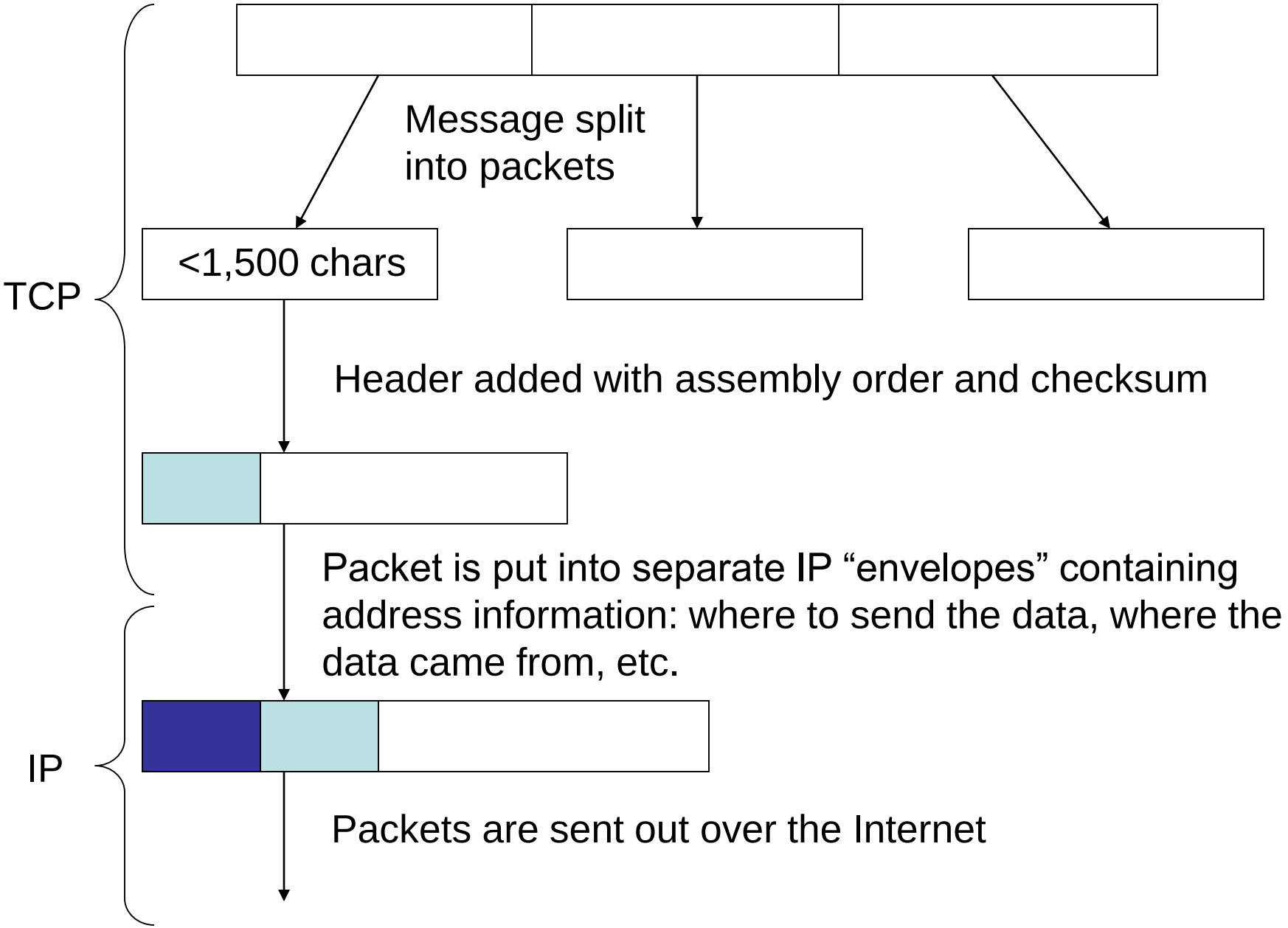
- Break up every piece of information and messages into packets
- Deliver the packets to the appropriate destination
- Reassemble the packets into their original form

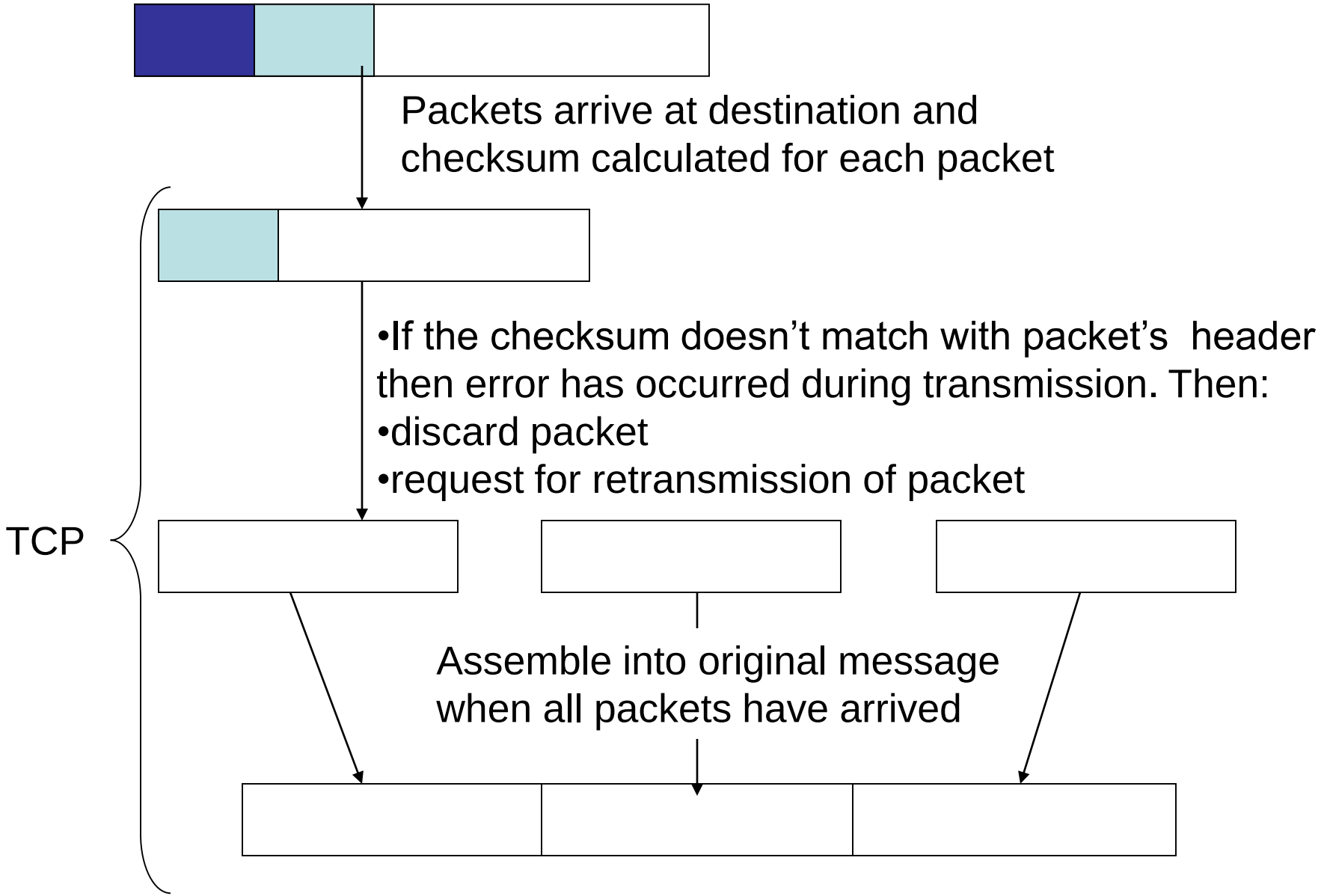
TCP/IP

- Transmission control protocol (TCP)
  - Is responsible for the break down of the message into packets and the re-assembly of the packets into the original message
- Internet protocol (IP)
  - Is responsible for making sure the packets are sent to the correct destination

Internet is a *packet-switched network*

- There is no single, unbroken connection between sender and receiver
- information is broken into small chunks called packets
- packets are sent over many different routes at the same time
- packets are reassembled at the receiving end into the original information


Message split into packets
<1,500 chars
TCP
Header added with assembly order and checksum
Packet is put into separate IP “envelopes” containing address information: where to send the data, where the data came from, etc.
IP
Packets are sent out over the Internet


Packets arrive at destination and
checksum calculated for each packet
•If the checksum doesn't match with packet's header
then error has occurred during transmission. Then:
•discard packet
•request for retransmission of packet
TCP
Assemble into original message
when all packets have arrived

Internet Addressing: IP Addresses

- Any internet needs a system for assigning identifying addresses to each computer in the system. In the Internet these are the "Internet Protocol" addresses
- IP address = 32 bit identifier for a machine

  = network identifier + host address

  Network identifier = part assigned by ICANN (https://www.icann.org/)

  Host address = part assigned by domain owner

Dotted decimal notation

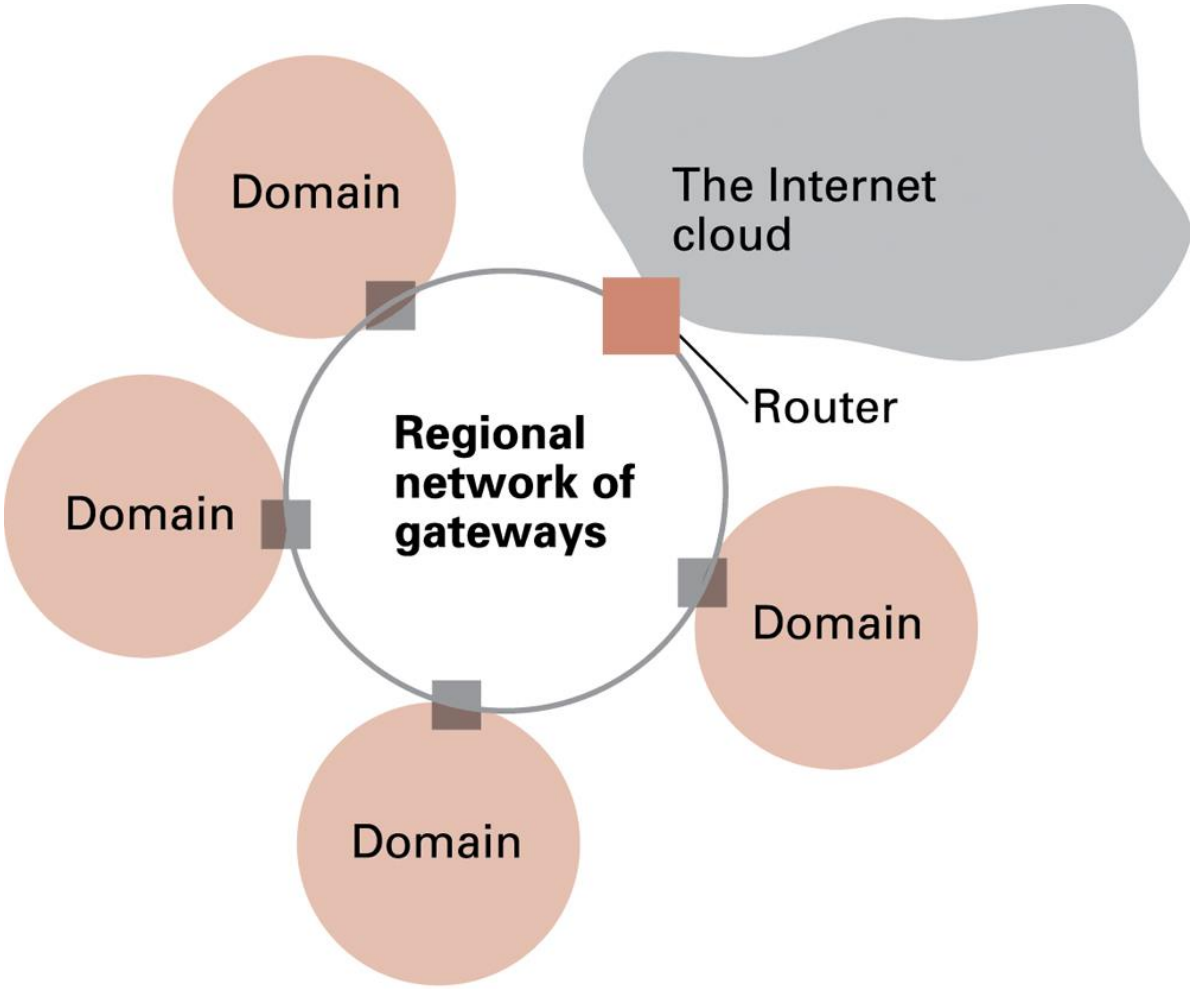
- An IP address is represented in dotted decimal notation, e.g. 192.207.177.133
- Each decimal number is in the range 0-255 and specifies 8 bits
  - the above IP address is

    11000000.11001111.10110001.10000101

Internet Architecture

- **Domain:** network or internet controlled by one organisation
- **Gateway:** router connecting a domain to the rest of the Internet (the Internet cloud')
- Domains must be registered by their owners
  - Internet Corporation for Assigned Names & Numbers (ICANN) serves as **registrar**

Figure 4.7 A typical approach to connecting to the Internet



Internet addressing: host names

- Host name = mnemonic name
  - Example: mymachine.aw.com
  - Domain name = part assigned by a registrar
    - Example: aw.com
    - Top level domain = classification of domain owner
      - By country – Example: .au = Australia
      - By usage – Examples
        - » .com = commercial
        - » .edu = educational
        - » .org = nonprofit organizations
        - » .gov = governmental

.

Internet addressing: host names

- Sub-domains and individual machine names
  - Assigned by domain owner
  - Domain owner must run a **name server**, maintaining a directory of mnemonic addresses and corresponding IP address for computers in the domain
  - The domain name system (DNS) is an Internet-wide directory system converting between mnemonic and IP addresses: “DNS lookup”
  - *Note*: Dots in host names are not related to the dots in the dotted decimal notation used for the host IP address

.

Internet applications

“traditional” applications

- Electronic mail (e-mail)
- File Transfer Protocol (FTP)
- Remote login: Telnet, Secure Shell (SSH)
- World Wide Web

Electronic mail

- Mail server: set up by domain owner
  - Mail sent from domain members goes through mail server
  - Mail sent to domain members is collected by mail server
  - Mail delivered to clients on demand
    - POP3 (Post Office Protocol - version 3)
    - IMAP (Internet Mail Access Protocol)

E-Mail

- Mail server: computer which handles email within a domain.
- Email address formats:

  individual@mailservermnenomic

  individual@domain name

Telnet and Secure Shell

- Telnet and SSH (Secure Shell) are both network protocols used to remotely access and manage devices (such as servers, routers, and computers) over a network.
- Telnet has several weaknesses. The main problem is communication is not encrypted, including username and password information: these could be intercepted and misused.
- Secure Shell (SSH) is a communication system that solves this problem, by providing *encryption* of data being transferred, and *authentication* - making sure the two parties communicating are who they claim to be.

File Transfer Protocol (FTP)

- client-server protocol transfer files between a client and a server
- accesses to files can be guarded by password
- allows users to upload, download, and manage files on remote systems.

World Wide Web (WWW)

- WWW is just one of many applications supported by the basic inter-network message transfer service of the internet
- Originated in the European Centre for Nuclear Research (CERN). Developed as a method to simplify the exchange of information between physics researchers in different countries
- The World Wide Web or the Web is the part of the Internet associated with HTML documents.

World Wide Web

- The data is the collection of documents (pages) while the structure comes from the links between the documents (hypertext)
- Server disseminates hypertext (or hypermedia) documents
  - Web site: all hypertext documents controlled by one organization or individual
    - Usually all at same internet address
  - HTML: language of hypertext documents

World Wide Web implementation

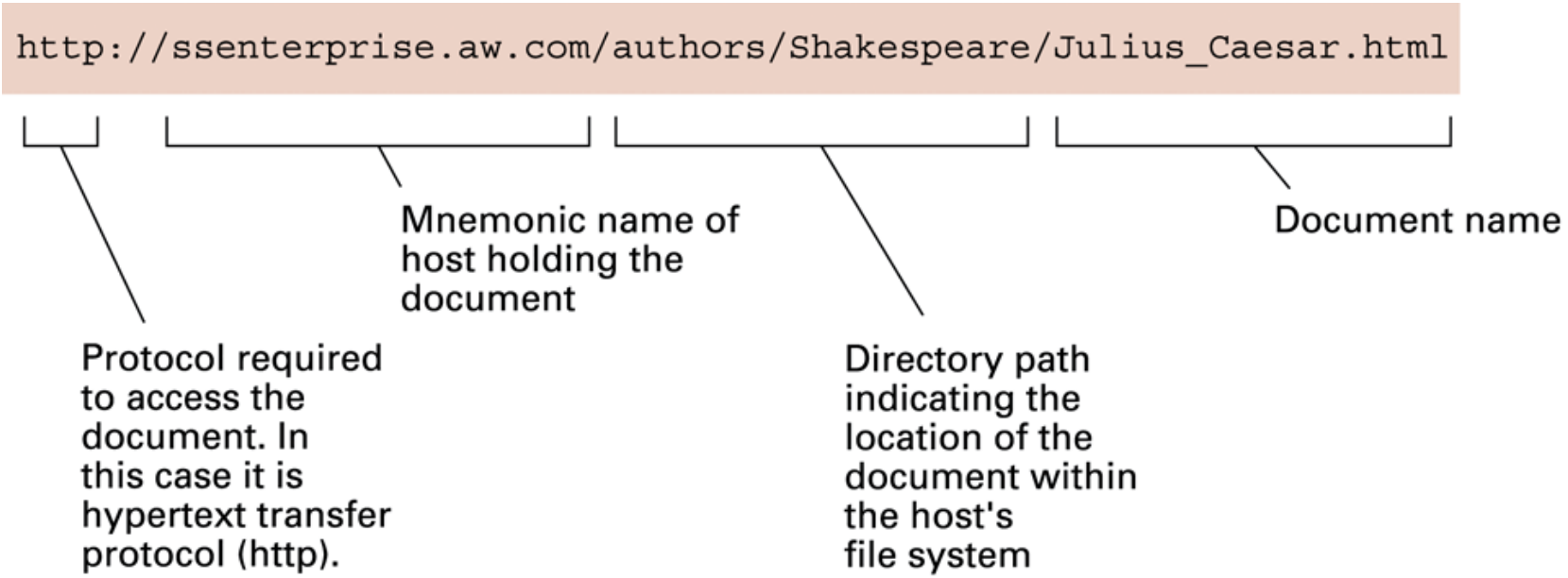
- Web server: provides access to documents on its machine as requested
- Browser: allows user to access web pages
- Hypertext Transfer Protocol (HTTP): communication protocol used by browsers and web servers
- Uniform Resource Locator (URL): unique address of a document on the web

Original HTTP Protocol (1991)

- Client: connect to server.
- Server: accept connection.
- Client: request document (GET document address).
- Server: Send message (= stream of ASCII characters).
- Server: close connection.

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

Figure 4.8 A typical URL



Client Server Communication

Client Server Communication is a two-stage ‘handshake’.

1. a request containing a URL is sent to the server from the client.

2. a response containing the required WWW page is sent to the client from the server.

HTTP's handshake involves an exchange of textual messages based on the format of electronic mail messages

Client request

e.g.

GET http://www.crawl.non/searcher/front-page.html HTTP/1.0
If-Modified-Since: Friday, 10-May-96 14:27:43 GMT
User-Agent: Tapestry/11.96

Server reply

e.g.

HTTP/1.0 200 Document follows
Date: Mon, 20 May 1996 07:47:41 GMT
Server: Webwide/7.2.18
Content-type: text/html
Last-modified: Tue, 14 May 1996 23:32:16 GMT
Content-length: 1523

<html>
<head><title>Web Crawl Inc. Home Page </title></head>
<body bgcolor ="#ffffff">

<h1> Web Crawl Inc. - Surfing the Internet for you!<h1/>

.
.
</body>
</html>

Hypertext document format

- Entire document is printable characters
- Contains tags to control display
  - Display appearance
  - Links to other documents and content
  - Dynamic functions

- This system of tags is Hypertext Markup Language (HTML)
- An HTML document contains the text, etc. and the instructions for its display

Figure 4.7 A simple Web page

a. The page encoded using HTML.

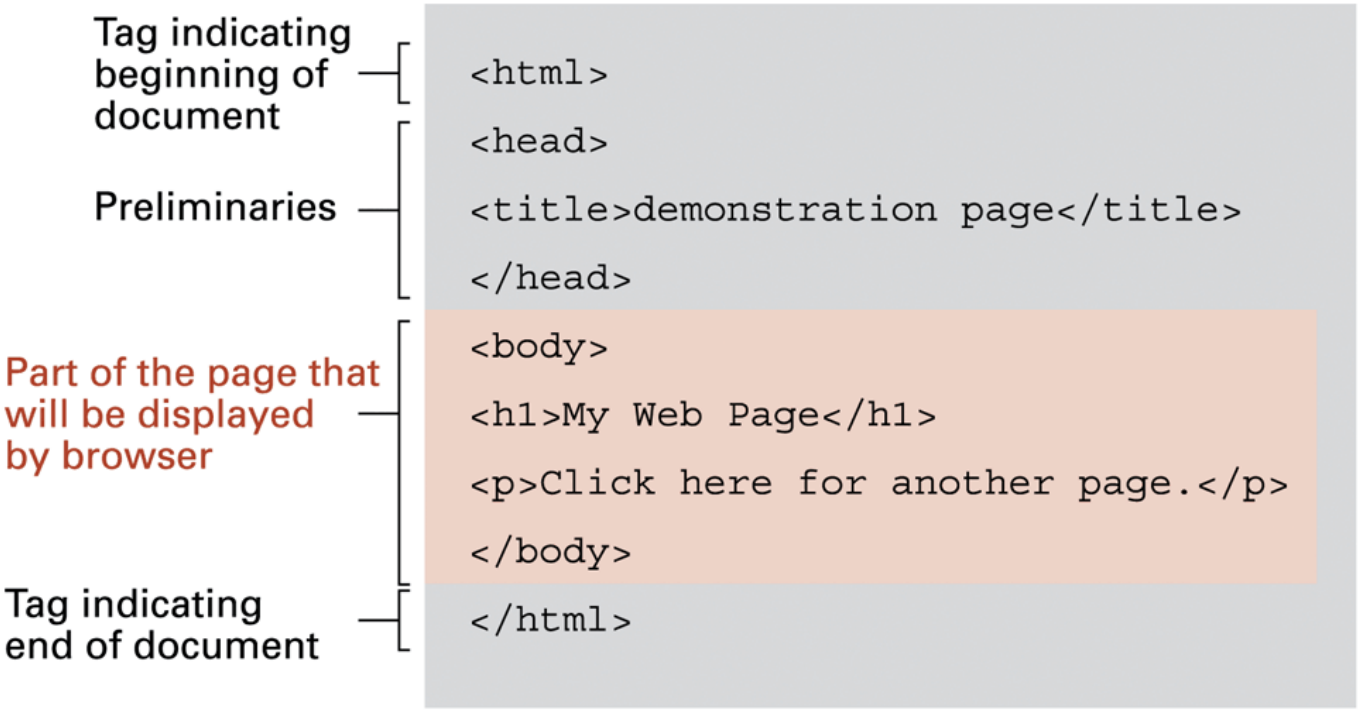


Figure 4.7 A simple Web page (cont'd)

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

Figure 4.8 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>
```

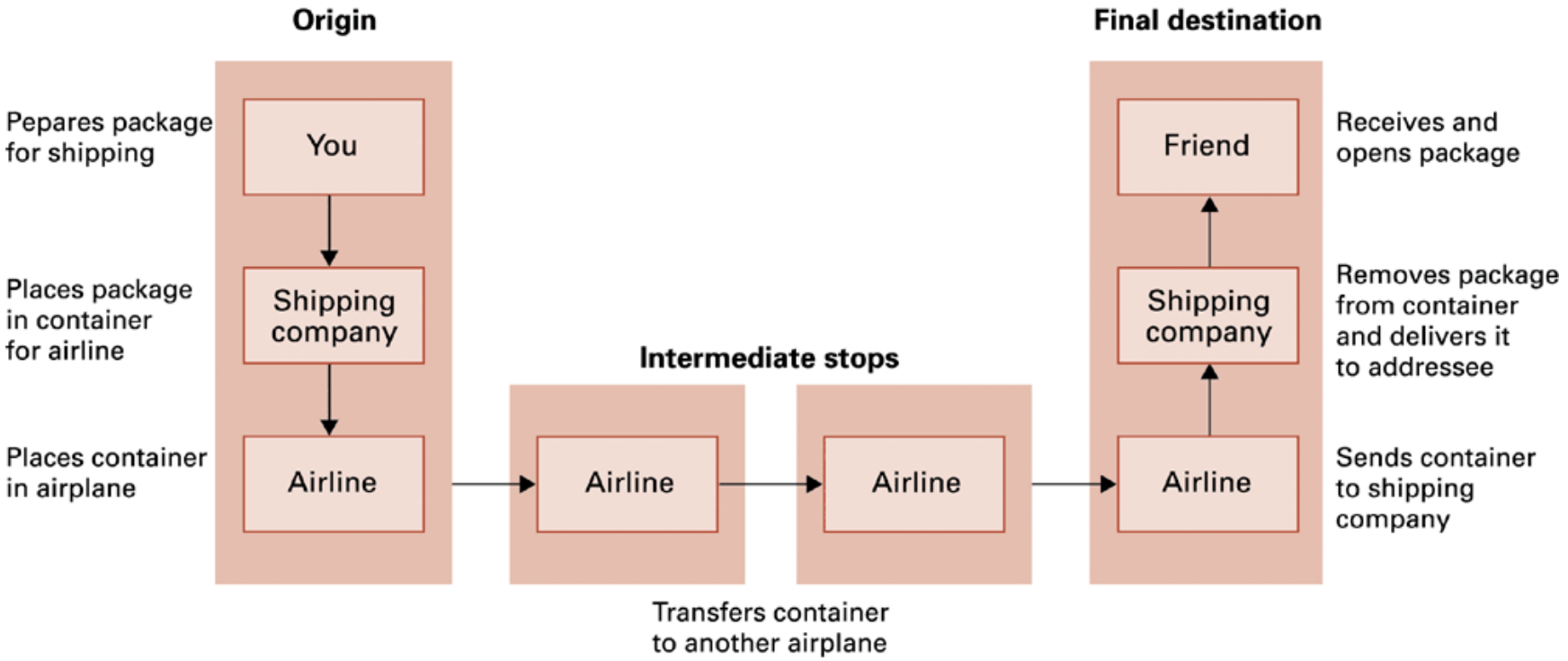
Figure 4.8 An enhanced simple Web page (cont'd)

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

4.4 Internet Protocols
Internet software layers

- Application layer:
  - Example: browser
- Transport layer: TCP/IP
- Network layer: handles routing through the internet
- Link layer: handles actual transmission of packets
  - Token ring or Ethernet

Figure 4.12 Package-shipping analogy



Internet message passing is analogous to sending a package across USA: 3-level hierarchy of (1) user level, (2) shipping company, (3) airline.

Figure 4.13 The Internet software layers

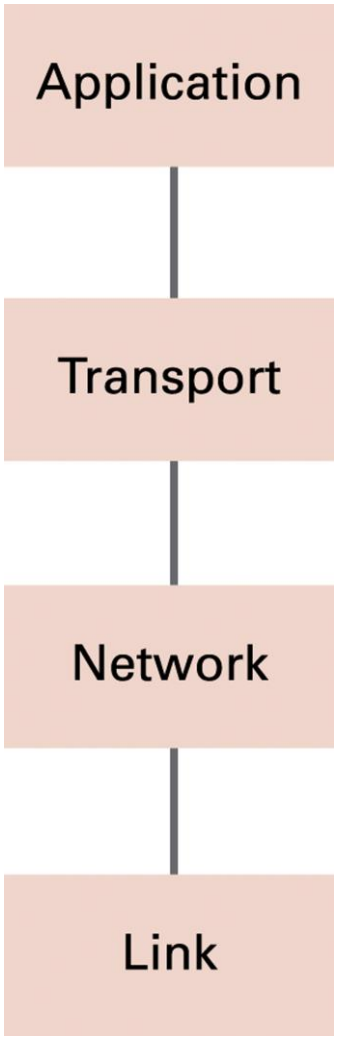
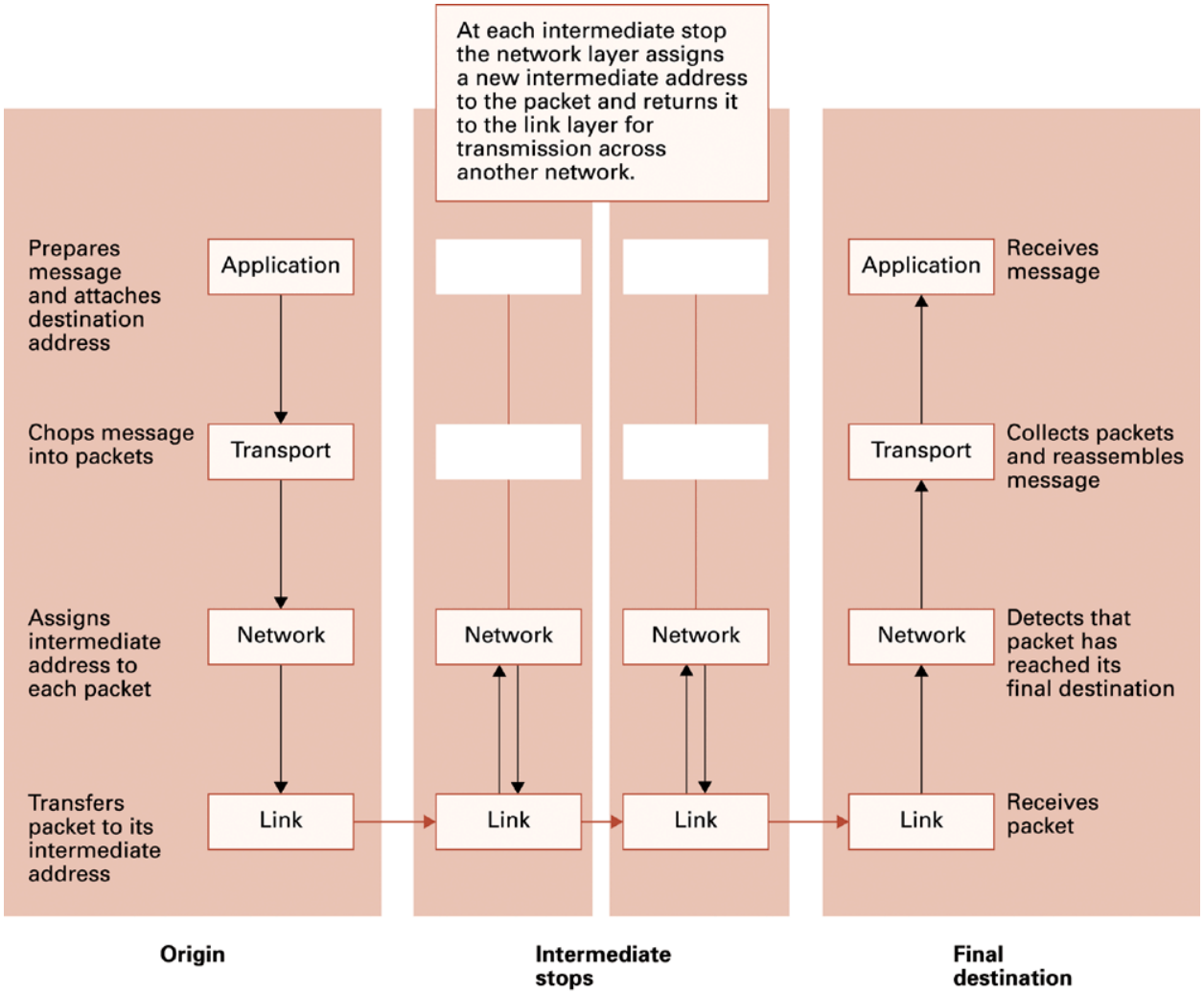


Figure 4.14 Following a message through the Internet



4.5 Network security

- Integrity of machine exposed to internet
  - Attacks: (malicious software, “malware”)
    - virus
    - worm
    - Trojan horse
    - spyware
    - denial of service attack
  - Defense

Software that may be transferred to and executed on the computer itself

A **virus** infects a computer by inserting itself into programs that already reside in the machine.

- When the 'host' program is executed, so is the virus
  - some viruses just try to transfer themselves to other programs in the computer
  - others also corrupt data and programs with devastating effects:
    - damaging the operating system
    - erasing blocks of mass storage

A **worm** is an autonomous program -- it doesn't need a host program

It transfers itself through networks by forwarding copies of itself to other computers

May be designed just to replicate or perform extreme vandalism

A characteristic result of a worm is an explosion of worm copies that interfere with / block legitimate applications and can overload an entire network or internet

A **Trojan horse** program enters a computer system disguised as a desirable program, e.g. a game or utility that is imported by the computer user. Often they arrive as attachments to e-mails, and are activated when the attachment is opened. The effects of the Trojan horse can be very harmful.

!! Never open attachments to mail from unknown sources

The Trojan horse may activate immediately, or wait for a *trigger*, such as a specific date

Spyware (or **sniffing** software) collects information about activities on the computer, and sends the information back to the investigator of the attack.

May be used by companies to build customer profiles. May be used to record keyboard input in search of passwords and credit card numbers.

Phishing is trying to get such information by asking for it, with the criminal sending emails pretending to be from a bank, law enforcement agency, government bureau, etc.

Attacking a computer from a remote computer

Denial of service attack: overload a computer with requests, disrupting the company's business and in some cases stopping them.

Attack requires many requests over a brief period of time, usually achieved by illicitly planting software on numerous computers that can be triggered by a signal to swamp the target computer(s)

Destructive Attacks

"How can I cause most damage by attacking this system?"

- Work of terrorists, malicious employees or 'black-hat' hackers.
- Usually criminal! Examples include the distributed denial-of-service attacks against Yahoo!, Amazon.com, E*Trade, Buy.com, CNN, and eBay

Protection

Firewall

- Filtering traffic attempting to pass through a point in the network, e.g. at a domain's gateway
  - Block outgoing messages with certain addresses
  - Block incoming messages from certain addresses
  - Block incoming messages with addresses inside the domain: they must be from an outsider pretending to be an insider (i.e. someone is **spoofing**)

- Also used to protect individual computers, e.g. Any web server, email server, name server

Protection

Proxy servers

- To shield a client from adverse actions of a server
  - The server could collect lot of information about the client's domain, and possibly use it in attacks
- Instead of dealing directly with a server the client deals with a proxy; the proxy server plays the role of the client in dealing with the actual server
  - Server cannot gather information about the client or its domain
  - Proxy server can check/filter all messages from the server to the client

Protection

- **Auditing software**
  - Detect sudden increases in message traffic
  - Monitor firewalls
  - Analyse patterns of requests from individual computers
  - *Identify problems before they get out of control*

- **Antivirus software**
  - detect and remove known viruses
  - But viruses are constantly ‘evolving’, so antivirus software needs regular updating

- **Caution**
  - Never open email attachments from unfamiliar sources
  - Do not download software without first confirming its reliability
  - Do not respond to pop-up ads
  - Do not leave a PC connected to the internet when the connection isn’t needed

Privacy of communication
Cryptography – 'hidden writing'

Encrypt (encode) data for protection in transfer over networks and internets

Many Internet applications have been altered to incorporate encryption:

FTPS, secure version of FTP

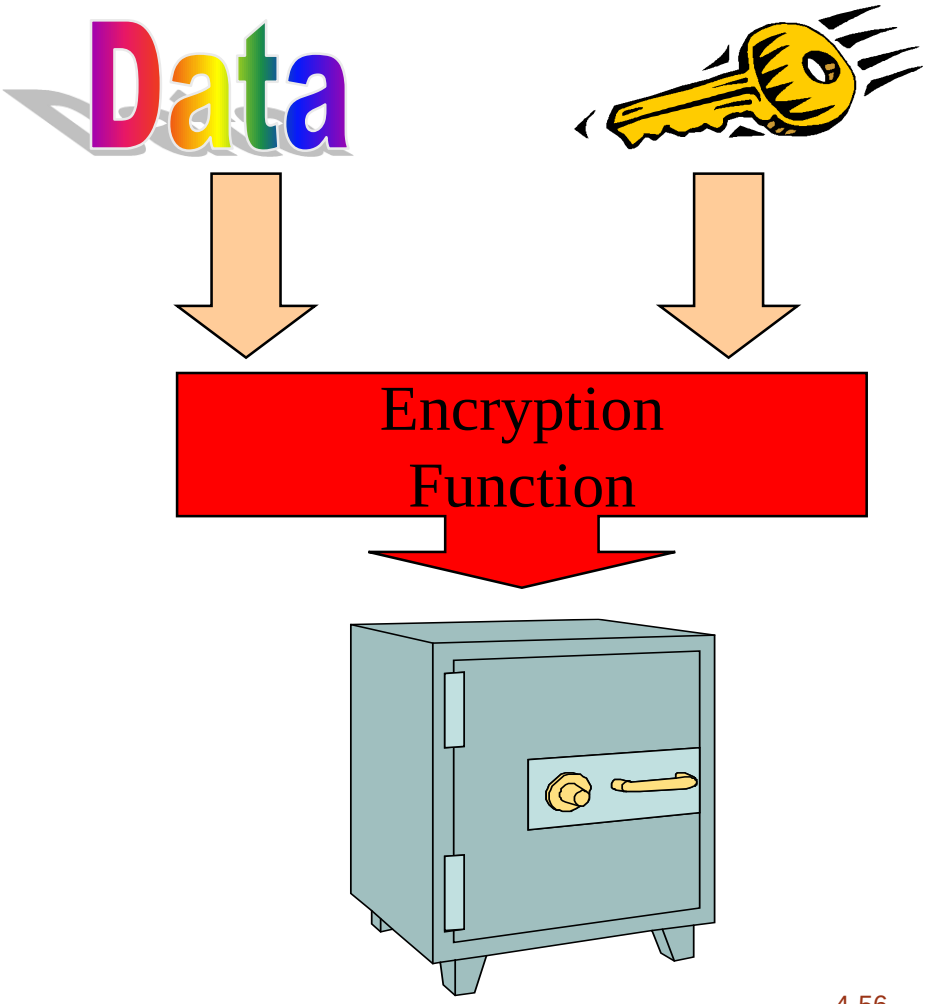
SSH, secure replacement for telnet

HTTPS, secure version of HTTP

- uses the Secure Sockets Layer (SSL) protocol

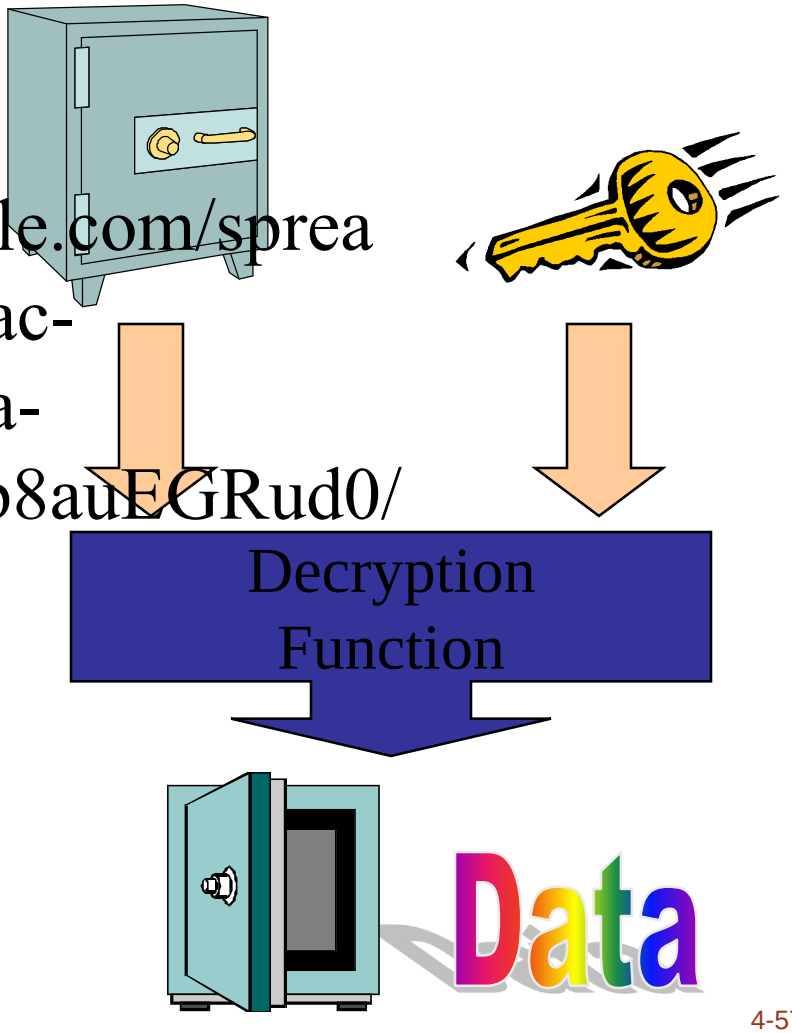
Encryption

- Encryption is the process of taking some data and a key and feeding it into a function and getting encrypted data out
- Encrypted data is, in principal, unreadable unless decrypted



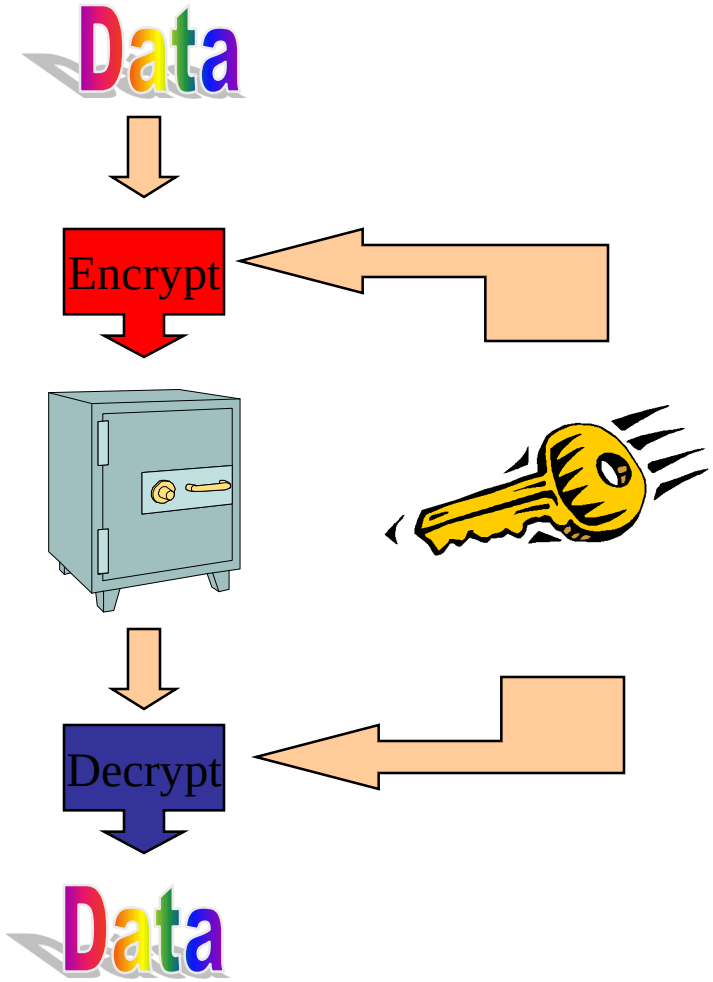
Decryption

- Decryption is the process of taking encrypted data and a key and feeding it into a function and getting out the original data
  - Encryption and decryption functions are linked



Symmetric Encryption

- Encryption and decryption functions use same key

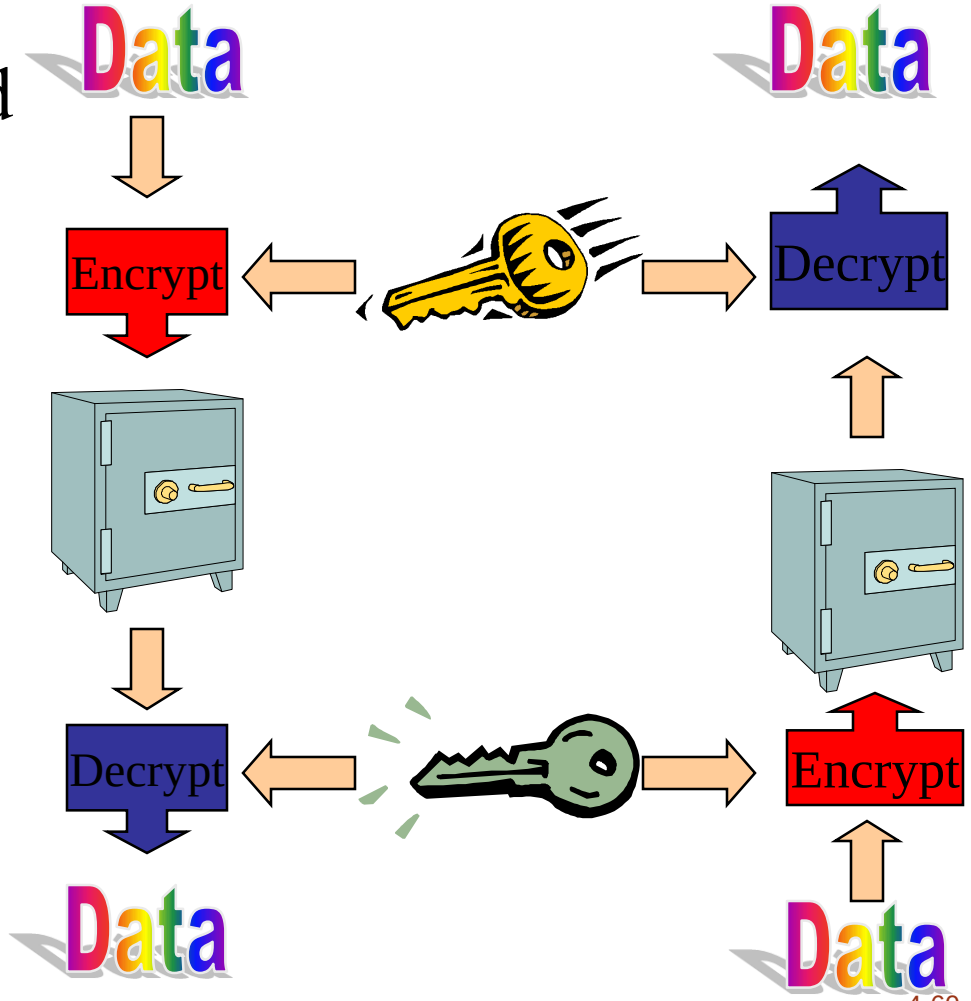


Asymmetric Encryption

- Encryption and decryption functions that use a key pair are called asymmetric
  - Keys are mathematically linked

Asymmetric Encryption

- When data is encrypted with one key, the other key must be used to decrypt the data
  - And vice versa



Public and Private Keys

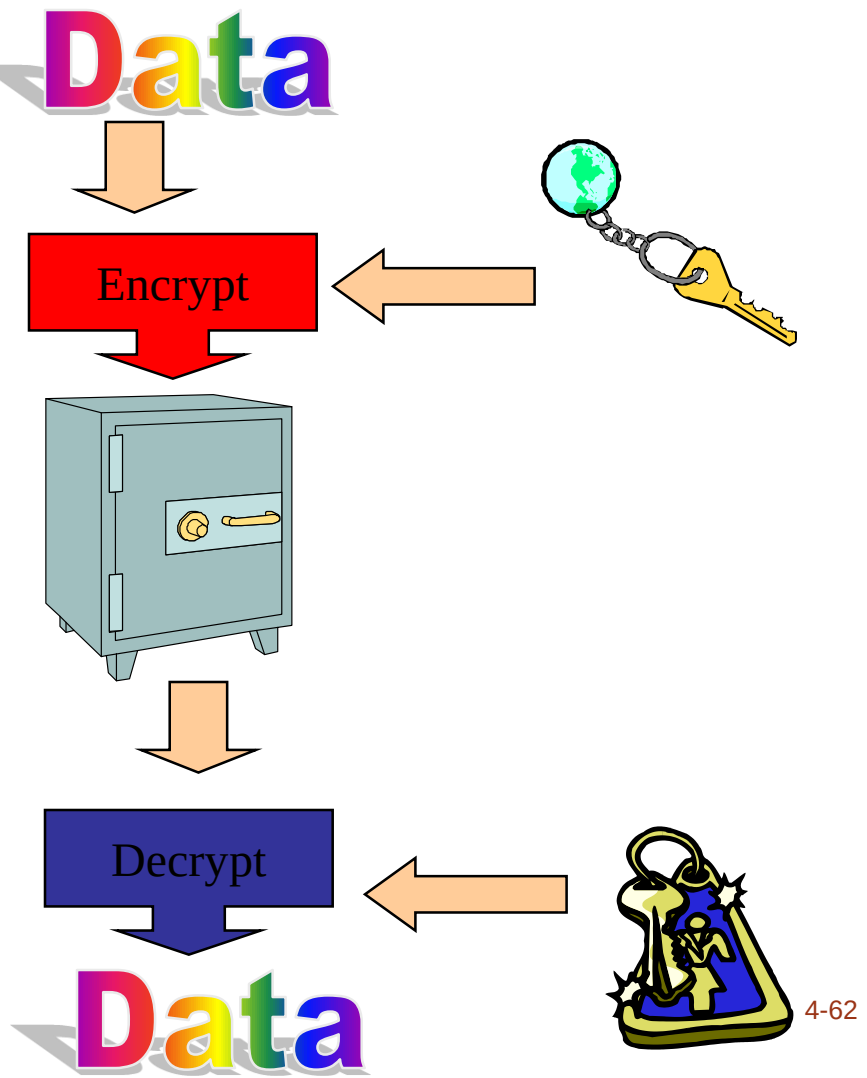
- With asymmetric encryption each user can be assigned a key pair: a private and public key

Private key is known only to owner

Public key is given away to the world

Public and Private keys

- Anything encrypted with the public key can only be decrypted with the private key
- And vice versa
- Since the private key is known only to the owner, this is very powerful.
- Message Privacy!



Public-key cryptography

Bob lets his public key be known to everyone, but keeps the private key secret.

Alice may send a confidential message to Bob like this:

A gets B's public key (she can get it from his web page, or he can just send it to her).

A encrypts the message with B's public key, and sends it to B.

B decrypts the message with his private key.

And even if Carol knows the public key and intercepts the message, she cannot decrypt it

For this method to work, the system must guarantee that it is (effectively) impossible to decrypt the message without knowledge of the private key.

In particular, it must be impossible to decrypt using the public key, or to derive the private key from the public key.

Public key cryptography is secure provided it is computationally unfeasible to reverse the encryption.

(see Chapter 11 section 6 of Brookshear 9th edition)