

CSC 121 Computers and Scientific Thinking

Fall 2005

The Internet and the Web

1

History of Internet



recall: the Internet is a vast, international network of computers

the Internet traces its roots back to the early 1960s

- MIT professor J.C.R. Licklider published a series of articles describing a "Galactic Network" of communicating computers
- in 1962, Licklider became head of computer research at the U.S. Department of Defense's *Advanced Research Project Agency (ARPA)*
- in 1967, Licklider hired Larry Roberts to design and implement his vision of a Galactic Network

the ARPANet (precursor to the Internet) became a reality in 1969

- it connected computers at four universities: UCLA, UCSB, SRI, and Utah
- it employed dedicated cables, buried underground
 - the data transfer rate was 56K bits/sec, roughly same as dial-up services today
- the ARPANet demonstrated that researchers at different sites could communicate, share data, and run software remotely

2

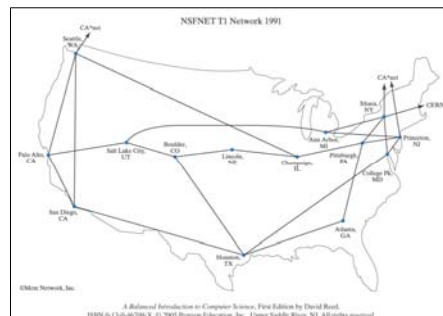
NSFNet

in the early 1980s, the ARPANet experienced an astounding growth spurt

- applications such as email, newsgroups, and remote logins were attractive to all colleges and universities
- by 1984, the ARPANet encompassed more than 1,000 sites

to accommodate further growth, the National Science Foundation (NSF) became involved with the ARPANet in 1984

- NSF funded the construction of high-speed transmission lines that would form the backbone of the expanding network



5

"Internet"

the term "Internet" was coined in recognition of the similarities between the NSFNet and the interstate highway system

- backbone connections provided fast communications between principal destinations, *analogous to interstate highways*
- connected to the backbone were slower transmission lines that linked secondary destinations, *analogous to state highways*
- local connections were required to reach individual computers, *analogous to city and neighborhood roads*

note: Al Gore did not INVENT the Internet, *nor did he ever claim to*

- *he sponsored legislation in the late 1980s to support growth and expand access*

recognizing that continued growth would require significant funding and research, the government decided in the mid 90s to privatize the Internet

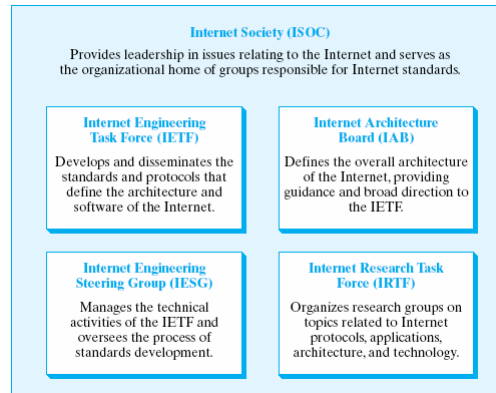
- control of the network's hardware was turned over to telecommunications companies and research organizations (e.g., MCI WorldCom, GTE, Sprint)
- research and design are administered by the *Internet Society*

6

Internet Society

Internet Society is an international nonprofit organization (founded in 1992)

- it maintains and enforces standards, ensuring that all computers on the Internet are able to communicate with each other
- it also organizes committees that propose and approve new Internet-related technologies and software



7

Internet Growth

according to the Internet Software Consortium, the Internet has more than doubled in size every 1 or 2 years

- will this trend continue?

Year	Computers on the Internet ¹
2002	162,128,493
2000	93,047,785
1998	36,739,000
1996	12,881,000
1994	3,212,000
1992	992,000
1990	313,000
1988	56,000
1986	5,089
1984	1,024
1982	235

8

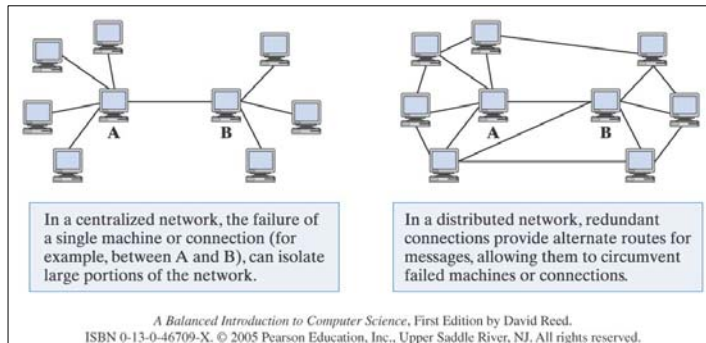
Distributed Networks

the design of the ARPANet was influenced by the ideas of Paul Baran, a researcher at the RAND Institute

- Baran proposed 2 key ideas: *distributed network* and *packet-switching*

recall: the ARPANet was funded by the Dept of Defense for communications

- as such, it needed to be resistant to attack or mechanical failure

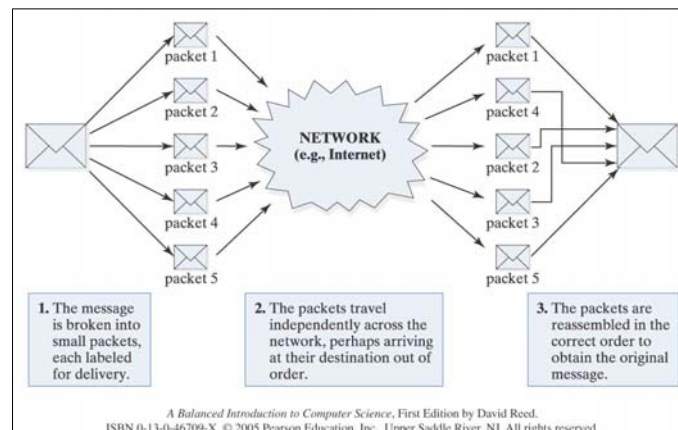


9

Packet Switching

in a packet-switching network, messages to be sent over the network are first broken into small pieces known as *packets*

- these packets are sent independently to their final destination



10

Advantages of Packets



1. sending information in smaller units increases the efficient use of connections
 - large messages can't monopolize the connection
 - *analogy: limiting call lengths at a pay phone to limit waiting*
2. transmitting packets independently allows the network to react to failures or network congestion
 - routers (special-purpose computers that direct the flow of messages) can recognize failures or congestion and reroute the packet around trouble areas
3. breaking the message into packets can improve reliability
 - since the packets are transmitted independently, it is likely that at least part of the message will arrive (even if some failures occur within the network)
 - software at the destination can recognize which packets are missing and request retransmission

11

Protocols and Addresses



the Internet allows different types of computers from around the world to communicate

- this is possible because the computing community agreed upon common *protocols* (sets of rules that describe how communication takes place)
- the two central protocols that control Internet communication are:
 1. *Transmission Control Protocol (TCP)*
 2. *Internet Protocol (IP)*

these protocols rely on each computer having a unique identifier (known as an *IP address*)

- *analogy: street address + zip code provide unique address for your house/dorm using this address, anyone in the world can send you a letter*
- an IP address is a number, written as a dotted sequence such as 147.134.2.20
- each computer is assigned an IP address by its Internet Service Provider (ISP)
- some ISPs (e.g., AOL, most colleges) maintain a pool of IP addresses and assign them dynamically to computers each time they connect

12

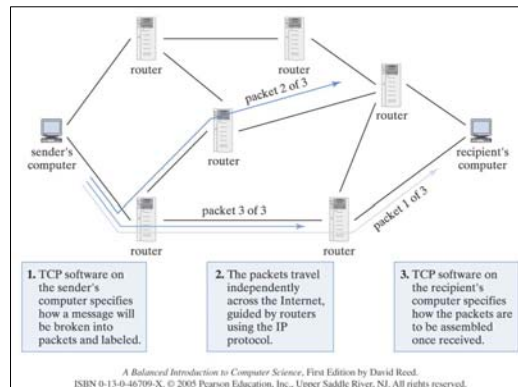
TCP/IP

Transmission Control Protocol (TCP)

- controls the method by which messages are broken down into packets and then reassembled when they reach their final destination

Internet Protocol (IP)

- concerned with labeling the packets for delivery and controlling the packets' paths from sender to recipient

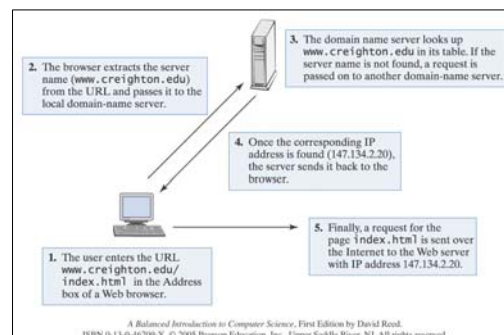


13

Routers and DNS

the Internet relies on special purpose computers in the network

- routers* are computers that receive packets, access the routing information, and pass the packets on toward their destination
- domain name servers* are computers that store mappings between domain names and IP addresses
 - domain names* are hierarchical names for computers (e.g., bluejay.creighton.edu) they are much easier to remember and type than IP addresses
 - domain name servers translate the names into their corresponding IP addresses



14

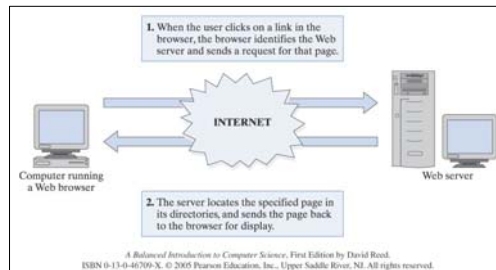
History of the Web

the World Wide Web is a multimedia environment in which documents can be seamlessly linked over the Internet

- proposed by Tim Berners-Lee at the European Laboratory for Particle Physics (CERN) in 1989
- designed to facilitate sharing information among researchers located all over Europe and using different types of computers and software

Berners-Lee's design of the Web integrated two key ideas

1. hypertext (documents with interlinked text and media)
 - Web pages can contain images and links to other pages
2. the distributed nature of the Internet
 - pages can be stored on machines all across the Internet, known as *Web servers*
 - logical connections between pages are independent of physical locations



15

Web Timeline

- 1990: Berners-Lee produced working prototypes of a Web server and browser
- 1991: Berners-Lee made his software available for free over the Internet
- 1993: Marc Andreessen and Eric Bina of the University of Illinois' National Center for Supercomputing Association (NCSA), wrote the first graphical browser: Mosaic
 - Mosaic integrated text, image & links, made browsing more intuitive
- 1994: Andreessen founded Netscape, which marketed the Netscape Navigator
- 1995: Microsoft released Internet Explorer → the browser wars begin!
- 1999: Internet Explorer becomes the most popular browser (~90% of market in 2002)

Year	Computers on the Internet ⁵	Web Servers on the Internet ⁶
2002	162,128,493	33,082,657
2000	93,047,785	18,169,498
1998	36,739,000	4,279,000
1996	12,881,000	300,000
1994	3,212,000	3,000
1992	992,000	50

in 2002, Google indexed more than 3 billion Web pages

in 2005, Google claims more than 8 billion pages

16

How the Web Works



like Internet communications, the Web relies on protocols to ensure that pages are accessible to any computer

- HyperText Markup Language (HTML) defines the form of Web page content
- HyperText Transfer Protocol (HTTP) defines how messages exchanged between browsers and servers are formatted

- the prefix `http://` in a URL specifies that the HTTP protocol is to be used in communicating with the server
- the prefix is NOT used for local file access since no server communication is necessary

for efficiency reasons, browsers will sometimes *cache* pages/images

- to avoid redundant downloads, the browser will store a copy of a page/image on the hard drive (along with a time stamp)
- the next time the page/image is requested, it will first check the cache
 - if a copy is found, it sends a *conditional* request to the server
 - essentially: "send this page/image only if it has been changed since the timestamp"
 - if the server copy has not changed, the server sends back a brief message and the browser simply uses the cached copy