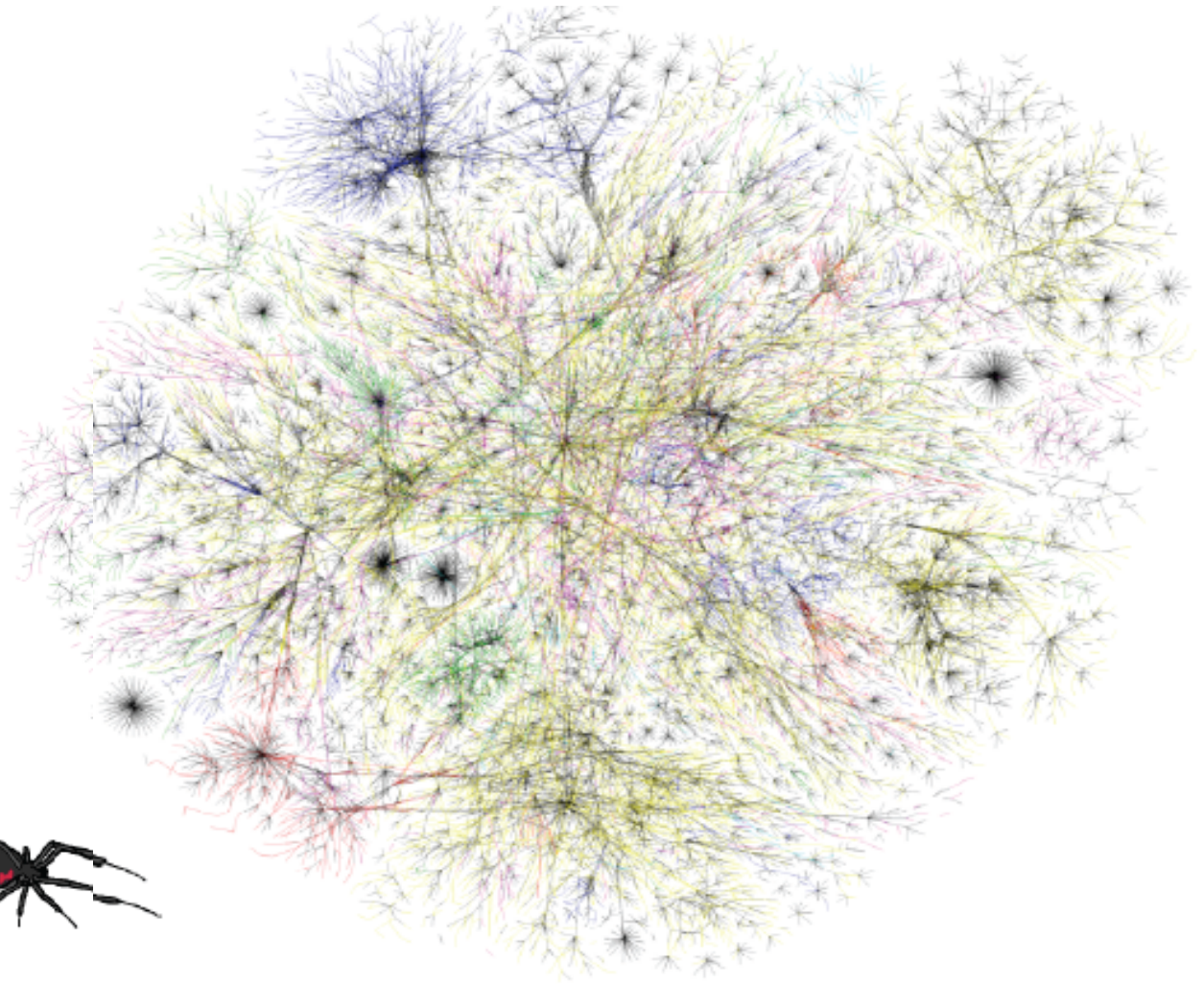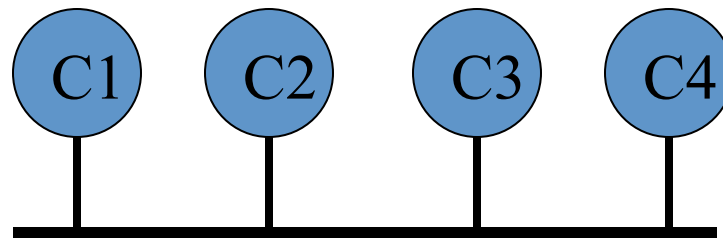# Networking and the Web

World-Wide Web

- Wide use of computers by the general public is directly attributable to the creation of the World Wide Web
- Key components of the web
  - Computer Communication Networks
  - Browsers
  - Content

"On the Internet, nobody knows you're a dog."

What's a network?

- A collection of computers connected together.

Networks: History

- Humble beginnings

- In 1972 the US government funded the development of the first network designed for communication between computers

- ARPA-net

- Connected a few
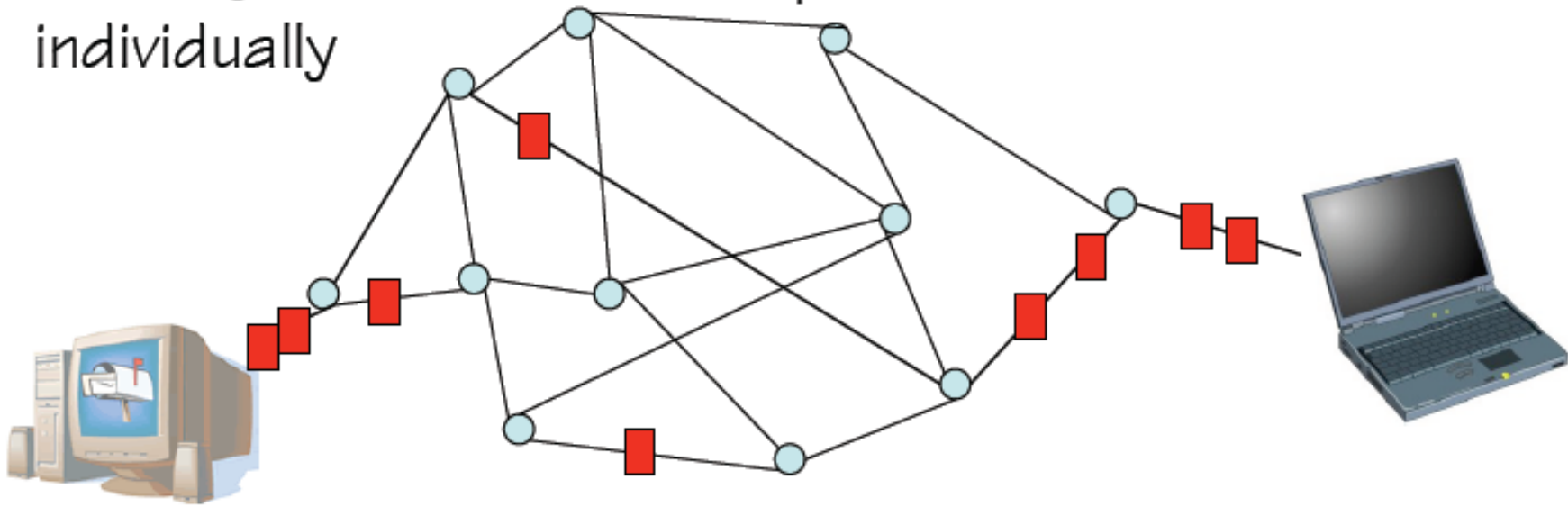 - large computer
  - systems nationwide

1974
Arpanet

Earlier Communication Networks

- Telephone Network

- ***Circuit-switched vs Packet-switched***

- A dedicated path is reserved from source to destination
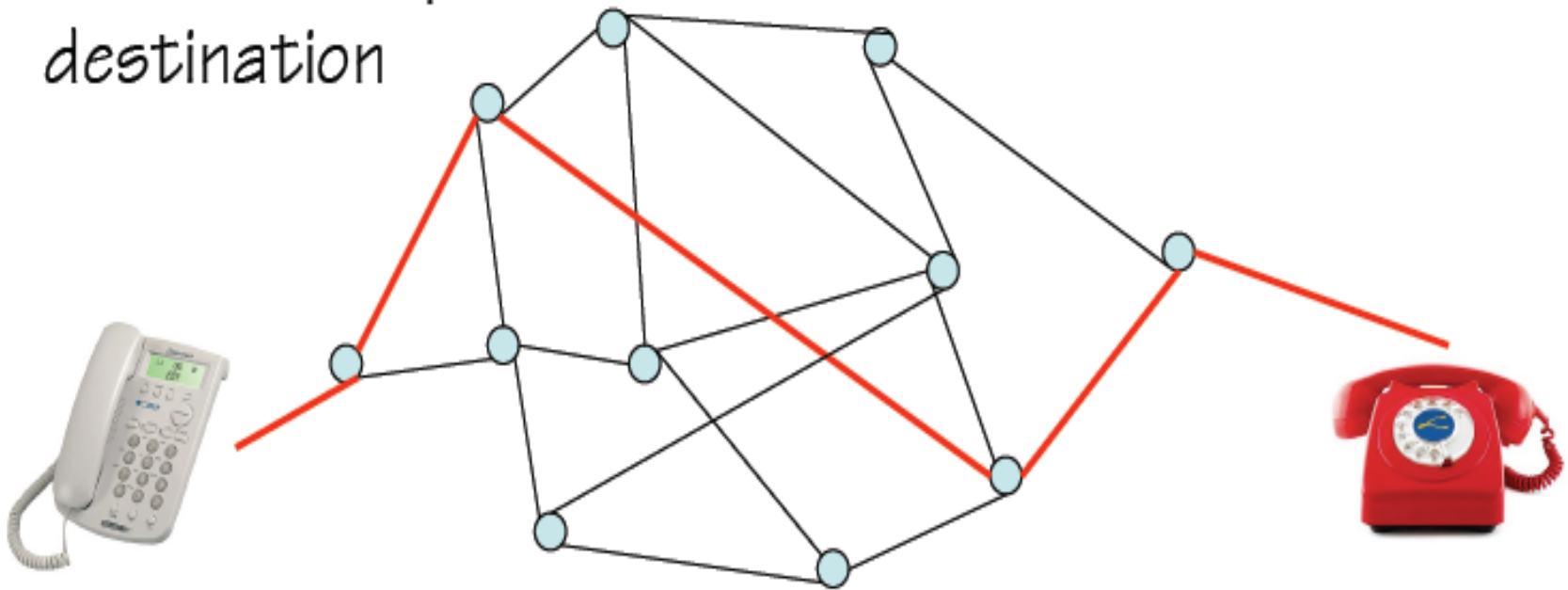
Earlier Communication Networks

- Telephone Network

- Circuit-switched vs **Packet-switched**

- Messages are divided into packets and routed individually

Routing and Forwarding

- So, how does a message get from computer A, to computer B?
  - Think of a letter.
  - What do you need to get a letter from point A to point B?
    - Address
  - What makes up an address?
    - Street and number
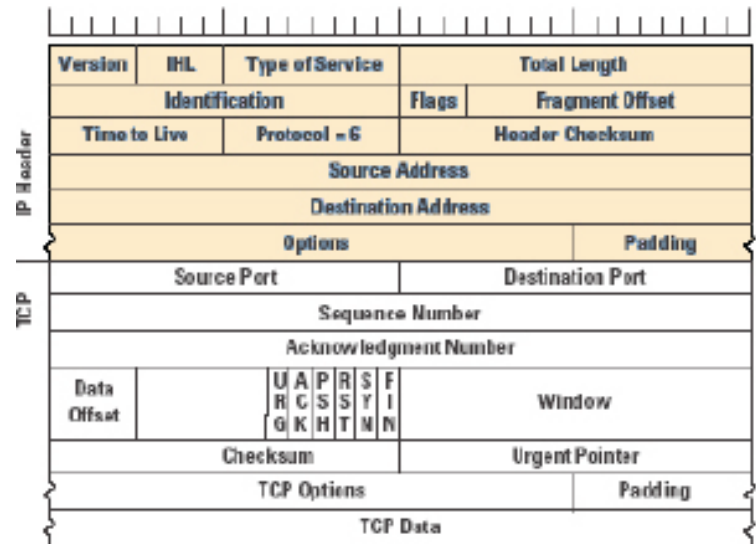   - City, State, and ZipCode.

Routing and Forwarding

- We can think of the letter being delivered in two parts:
  - First, we get it to the right zipcode.
  - Then, we get it to the right house.
- This happens for computer networks as well.
  - First we get it to the right network.
  - Then we get it to the right machine.

Packet Format / Protocol

- A network ***Protocol is how the information is actually packaged and sent***
- *TCP/IP (Transmission Control Protocol / Internet Protocol)*
  - Information is broken into a sequence of small fixed-size units called *IP packets. Each with an index or sequence number*
  - Each packet has space for the unit of data, the destination IP address, and a sequence number
  - Packets are sent over the Internet one at a time using whatever route is available
  - Because each packet can take a different route
  - Missed packet are resent
  - Congestion and service interruptions merely delay transmissions

IP Header					
Version	IHL	Type of Service	Total Length		
Identification			Flags	Fragment Offset	
Time to Live		Protocol = 6	Header Checksum		
Source Address					
Destination Address					
Options				Padding	

TCP					
Source Port			Destination Port		
Sequence Number					
Acknowledgment Number					
Data Offset		U R G / A C K / P S H / R S T / S Y N / F I N	Window		
Checksum			Urgent Pointer		
TCP Options				Padding	
TCP Data					

Moving Packets: Wires and More

- Internet uses electrical, electronic, and optical communication means
- Telephone lines, dedicated fiber optic lines, etc.
- The technology used to move the packet is independent from the protocol; transmission of a single file may use multiple technologies

Computer Phone Numbers

- Computers on the network have IP addresses
- 4 byte (or 32 bit) sequence 152. 2. 1.217 with two parts:
  - Network address
  - Host address
- Each number group is between 0-255 (8-bits)
- Mapping from a "Names" to "Numbers"
- Try using the command "ping" to find out the IP numbers of your favorite network addresses
  - unc.edu
  - amazon.com
- Now try "tracert" to trace a networking route

IP Addresses

- To understand an IP address completely, need to know:
  - 32-bit address
  - How many bits are for the network.
- Example: 152.2.131.156
- 10011000000000101000001110011100

Network Number   Host Number

Forwarding

- In the middle of the Internet, only the network part of an address matters.

- Message comes in some incoming link.

- Need to decide on which outgoing link to send the message on.

- This is called hop-by-hop forwarding.

Routing Table

- To do forwarding:
  - Build a table for all network numbers.
    - Table associated net number with outgoing link.
  - When message arrives, look up network number, find associated link, send message.
- Building the routing table is a different problem.

The Network Phonebook

- The *Domain Name System (DNS) translates the* human-readable names into IP addresses
- Internet host knows the IP address of its nearest DNS server, a computer that keeps a list of domain names and corresponding IP addresses
- When you use a domain name to send information, your computer asks the DNS server to look up the IP address
- If the DNS server doesn't know the IP address, it asks a *Root name server, which keeps the master* list of name-to-address relationships

Network Domains

- A *Domain is a related group of networked* computers
- Top-level domains appear in the last part of domain name:

  .edu   educational institutions

  .org   organizations

  .net   networks

  .mil   military

  .gov   government agencies

  Mnemonic two-letter country designators such as .ca (Canada)

Table 3.1. *Top-level Country Domain Abbreviations*

AF	Afghanistan	EG	Egypt	LT	Lithuania	SN	Senegal
AL	Albania	SV	El Salvador	LU	Luxembourg	SC	Seychelles
DZ	Algeria	GQ	Equatorial Guinea	MO	Macau	SL	Sierra Leone
AS	American Samoa	ER	Eritrea	MK	Macedonia, the for-	SG	Singapore
AD	Andorra	EE	Estonia		mer Yugoslav	SK	Slovakia (Slovak
AO	Angola	ET	Ethiopia		Republic of		Republic)
AI	Anguilla	FK	Falkland Islands	MG	Madagascar	SI	Slovenia
AQ	Antarctica		(Malvinas)	MW	Malawi	SB	Solomon Islands
AG	Antigua & Barbuda	FO	Faeroe Islands	MY	Malaysia	SO	Somalia
AR	Argentina	FJ	Fiji	MV	Maldives	ZA	South Africa
AM	Armenia	FI	Finland	ML	Mali	GS	South Georgia &
AW	Aruba	FR	France	MT	Malta		South Sandwiches
AU	Australia	FX	France, Metropolitan	MH	Marshall Islands	ES	Spain
AT	Austria	GF	French Guiana	MQ	Martinique	LK	Sri Lanka
AZ	Azerbaijan	PF	French Polynesia	MR	Mauritania	SH	St. Helena
BS	Bahamas	TF	French Southern	MU	Mauritius	PM	St. Pierre &
BH	Bahrain		Territories	YT	Mayotte		Miquelon
BD	Bangladesh	GA	Gabon	MX	Mexico	SD	Sudan
BB	Barbados	GM	Gambia	FM	Micronesia	SR	Suriname
BY	Belarus	GE	Georgia	MD	Moldova,	SJ	Svalbard & Jan
BE	Belgium	DE	Germany		Republic of		Mayen Islands
BZ	Belize	GH	Ghana	MC	Monaco	SZ	Swaziland
BJ	Benin	GI	Gibraltar	MN	Mongolia	SE	Sweden
BM	Bermuda	GR	Greece	MS	Montserrat	CH	Switzerland
BT	Bhutan	GL	Greenland	MA	Morocco	SY	Syrian Arab Republic
BO	Bolivia	GD	Grenada	MZ	Mozambique	TW	Taiwan, Province of
BA	Bosnia &	GP	Guadeloupe	MM	Myanmar		China
	Herzegovina	GU	Guam	NA	Namibia	TJ	Tajikistan
BW	Botswana	GT	Guatemala	NR	Nauru	TZ	Tanzania, United
BV	Bouvet Island	GN	Guinea	NP	Nepal		Republic of
BR	Brazil	GW	Guinea-Bissau	NL	Netherlands	TH	Thailand

Table 3.1. *Top-level Country Domain Abbreviations* **(continued).**

|							
IO	British Indian Ocean Territory	GY	Guyana	NC	New Caledonia	TG	Togo
BN	Brunei Darussalam	HT	Haiti	NZ	New Zealand	TK	Tokelau
BG	Bulgaria	HM	Heard & McDonald Islands	NI	Nicaragua	TO	Tonga
BF	Burkina Faso			NE	Niger	TT	Trinidad & Tobago
BI	Burundi	HN	Honduras	NG	Nigeria	TN	Tunisia
KH	Cambodia	HK	Hong Kong	NU	Niue	TR	Turkey
CM	Cameroon	HU	Hungary	NF	Norfolk Island	TM	Turkmenistan
CA	Canada	IS	Iceland	MP	Northern Mariana Islands	TC	Turks & Caicos Islands
CV	Cape Verde	IN	India				
KY	Cayman Islands	ID	Indonesia	NO	Norway	TV	Tuvalu
CF	Central African Republic	IR	Iran (Islamic Republic of)	OM	Oman	UG	Uganda
				PK	Pakistan	UA	Ukraine
TD	Chad	IQ	Iraq	PW	Palau	AE	United Arab Emirates
CL	Chile	IE	Ireland	PA	Panama		
CN	China	IL	Israel	PG	Papua New Guinea	GB	United Kingdom
CX	Christmas Island	IT	Italy	PY	Paraguay	US	United States
CC	Cocos (Keeling) Islands	JM	Jamaica	PE	Peru	UM	United States Minor Outlying Islands
		JP	Japan	PH	Philippines		
CO	Colombia	JO	Jordan	PN	Pitcairn	UY	Uruguay
KM	Comoros	KZ	Kazakhstan	PL	Poland	UZ	Uzbekistan
CG	Congo	KE	Kenya	PT	Portugal	VU	Vanuatu
CK	Cook Islands	KI	Kiribati	PR	Puerto Rico	VA	Vatican City State
CR	Costa Rica	KP	Korea, Democratic People's Republic of	QA	Qatar	VE	Venezuela
CI	Cote d'Ivoire			RE	Reunion	VN	Vietnam
HR	Croatia (local name: Hrvatska)	KR	Korea, Republic of	RO	Romania	VG	Virgin Islands
		KW	Kuwait	RU	Russian Federation	VI	Virgin Islands (US)
CU	Cuba	KG	Kyrgyzstan	RW	Rwanda	WF	Wallis & Futuna Islands
CY	Cyprus	LA	Lao People's Democratic Republic	KN	Saint Kitts & Nevis		
CZ	Czech Republic			LC	Saint Lucia	EH	Western Sahara
DK	Denmark	LV	Latvia	VC	Saint Vincent & the Grenadines	YE	Yemen
DJ	Djibouti	LB	Lebanon			YU	Yugoslavia
DM	Dominica	LS	Lesotho	WS	Samoa	ZR	Zaire
DO	Dominican Republic	LR	Liberia	SM	San Marino	ZM	Zambia
TP	East Timor	LY	Libyan Arab Jamahiriya	ST	Sao Tome & Principe	ZW	Zimbabwe
EC	Ecuador	LI	Liechtenstein	SA	Saudi Arabia		

Network Ports

- Each computer has > 65000 network ports (like phone extensions)
- Many offer a service
  - Sharing files
  - Serving Web pages
  - Email
  - Database

Far and Near: WAN and LAN

• Internet is a collection of *Wide Area Networks (WAN),* designed to send information between widely separated locations

• *Local Area Networks (LAN)* connect computers close enough to be linked by a single cable or wire pair
 – *Ethernet is the main technology for LAN*
 – Private LANs have dedicated IP Numbers 192.168.*.*, 10.*.*.*, 172.16-31.*.*
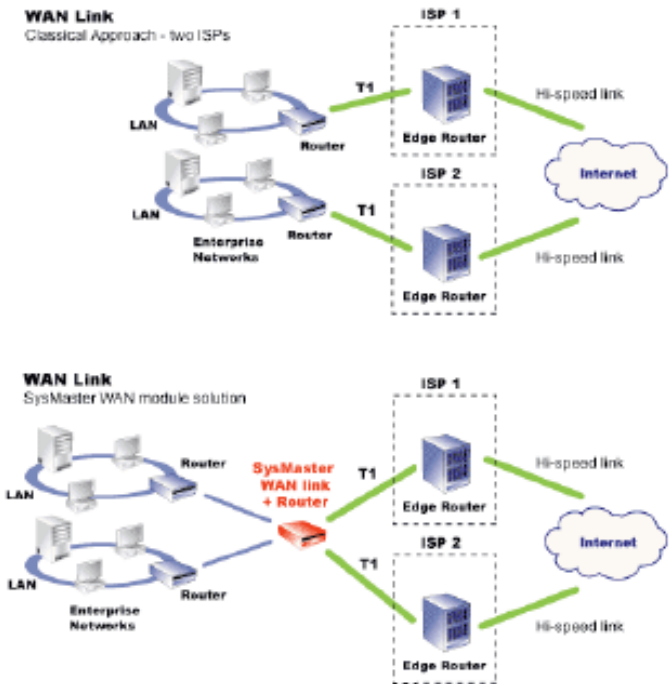
Ethernet

- Channel (wire, wire pair, or optical fiber) that winds past a set of computers
- Each computer is connected to the channel, allowing it to send a signal that can be detected by all computers connected to the channel

- Decentralized scheme: Each computer listens to the channel, and if it's quiet, it's grabs it. The computer transmits unless another computer starts at the same time. In that case, both stop for a random time and then try again.
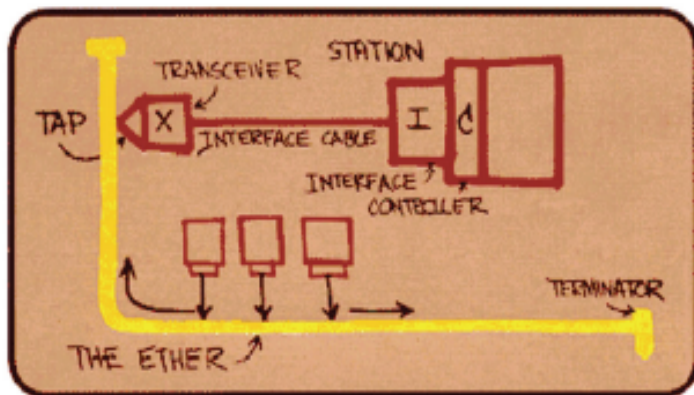

Figure 3.10. Robert Metcalfe's original drawing of the Ethernet design; the unlabeled boxes (computers), "tap" onto the wire that Metcalfe labeled "The Ether."

Connecting a Computer to the Internet

- By ISP:
  - Internet Service Providers sell connections to Internet (like AOL and Earthlink)
  - User plugs into telephone system, cable television, provider, or dedicated connection to ISP
  - Home computer talks to ISP's computer
  - ISP's computer is connected to Internet, and relays information for its customers

Connecting a Computer to the Internet (cont'd)

- By Enterprise Network Connections (LAN):
  - Large networked organizations such as schools, businesses, or governmental units
  - The organization creates a LAN or *intranet*
  - The intranet connects to the Internet by a gateway
  - Information from a Web computer is sent across Internet, through gateway, across LAN to user's computer

Wireless Networks

- A variation on a LAN connection
- A computer or router (called the *hub) is physically connected to* the Internet
- The hub also broadcasts and receives radio frequency (rf) signals
- Mobile computers also send and receive signals
- The hub relays Internet requests for the networked computers
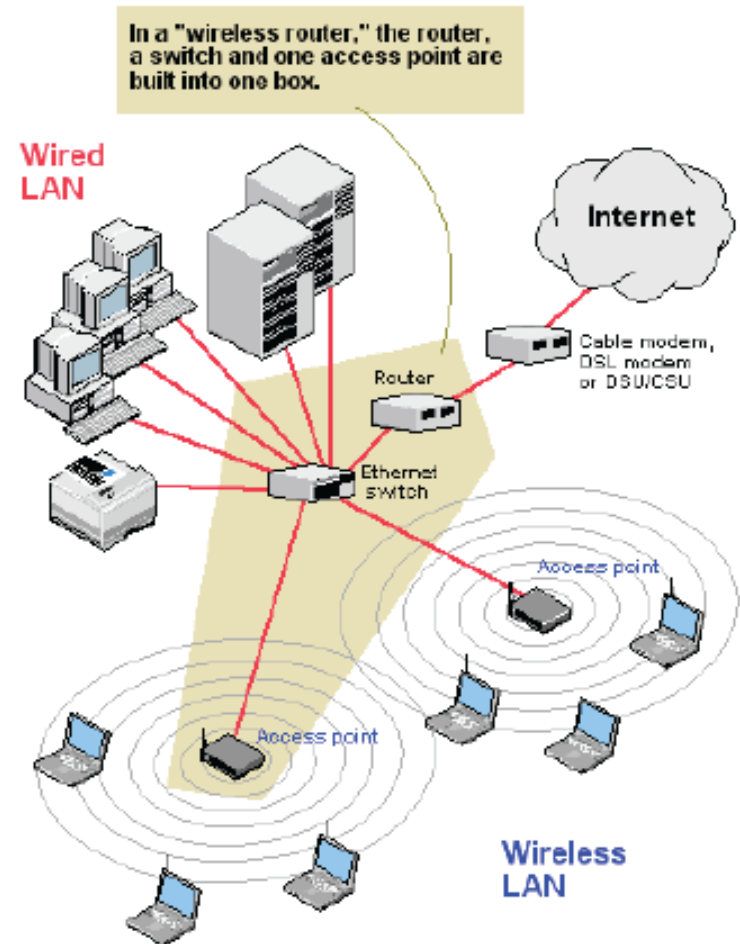
The World Wide Web

- The *World Wide Web is merely an internet service that runs on a port* of some computers (port 80 is the default)
- External computers request specially formatted files from this port, and the server delivers content
- Was designed in 1991 to share scientific communications (papers)
- *Web servers: Computers* programmed to send files to browsers running on other computers connected to the Internet
- Web servers and their files make up the World Wide Web

Requesting a Web Page

- Web request creates a client/server interaction
- Universal Resource Locator (URL) has three main parts
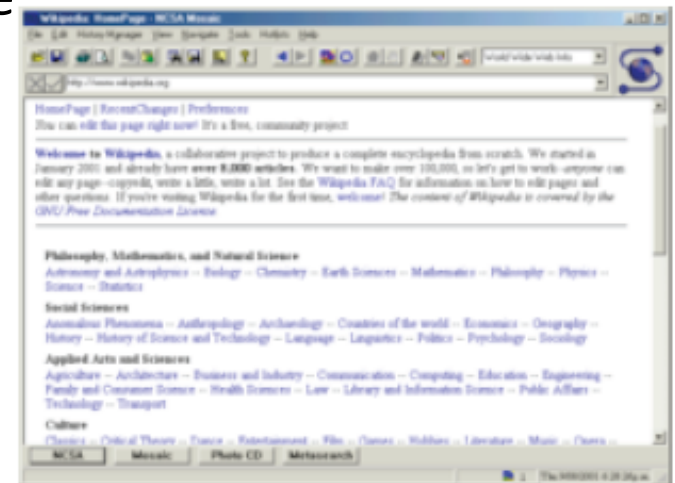 1. Protocol:
     - http://
     - Hypertext Transfer Protocol
     - Tells the computer how to handle the file
  2. Server computer's name:
     - Server's IP address given by the domain hierarchy
  3. Page's pathname:
     - Tells the server which file (page) is requested and where to find it.

Describing a Web Page

- Original WWW content was stored with only hints of how it should be rendered, the end user could specify font sizes, emphasis, etc.
- Eventually pages began to stored a description of how they should appear on the screen
- Web browser "renders" an image from this description file
  - Browser can adapt the source image more easily

Hypertext Markup Language

- *HyperText Transfer Protocal (HTTP) was designed for* transmitting web page descriptions
- *Markup languages describe the layout of a document*
  - Margin width
  - Font
  - Text style
  - Image placement
  - Etc.
- Hypertext provides a way to jump from point to point in documents (non-linear)
- Combination of hypertext with markup language lets us build nonlinear documents for the dynamic and interconnected Net and Web

The Internet and the Web

- When is the "www" required and when is it optional?
- WWW is just a name; web pages do not have to use it
- In order for DNS to work, user must give the exact domain name
- To help users reach them, organizations do two things:
  1. Redirection: browser inserts the "www"
  2. Registering multiple domain names
     - Museum of Modern Art has registered both "moma.org" and "www.moma.org" to the same IP address

URL Structure

- Web files are specified by *"Universal Resource Locater" (URL) which tells the server where to find the requested file*
- It assumes a directory hierarchy model, much like that of a file system. However, it may or may not actually reflect a file system
- *Directory Hierarchy: Directories can contain other* directories, which can contain other directories, etc.–
  -Down, or lower in the hierarchy, means moving into subdirectories
  – Up, or higher in the hierarchy, means into enclosing directories

URL Structure (cont'd)

- Part of the directory hierarchy is shown in the pathnames of URL's.
  http://www.nasm.si.edu/galleries/ga1100/pioneer.html
- Page is the terminal part of the pathname:
  /galleries/ga1100/pioneer.html
- If no file is specified then "index.html" or "index.htm" is generally assumed
- Each time we pass a slash (/), we move into a subdirectory or into the file (lower in the hierarchy)

An Old-School Web Page

- A simple static HTML web page

```
<html>
<head> <title> Alto Computer </title> </head>
<body bgcolor="white"><font face="Helvetica">
   <img align="right" src="alto.jpg">
   <h1>Alto, <br>A Computer of Note</h1>

       <p>The Alto was the first networked personal computer.
       It was invented at the Xerox Palo Alto Research Center
       (PARC) by the team of Ed McCreight, Chuck Thacker,
       Butler Lampson, Bob Sproull and Dave Boggs to explore
       office automation. Altos were the first production
       computers to have a bit-mapped display, windows and a
       mouse. Ethernet technology, also invented at PARC, was
       first used to connect Altos. </p>

       <p>Though Xerox was unable to market the Alto — they cost
       $32,000 in 1979 — the computer impressed many others who
       did push the technologies. For  example, Apple Computer
       co-founder Steve Jobs was so impressed when he saw the
       Alto, he created the revolutionary Apple Macintosh in
       its image.</p>

</body>
</html>
```

Figure 3.11. A Web page and the HTML source that produced it. Notice that an additional image file, alto.jpg, is also required to display the page. (continued).