Internet Technologies
1- Introduction

F. Ricci

2010/2011

Contact Details

- Francesco Ricci
  - Room 204 (POS)
  - fricci@unibz.it
- Availability Hours:
  - Thursday 16:00 – 18:00
  - **by prior arrangement via e-mail**
- **Course web site**
  - http://www.inf.unibz.it/~ricci/IT/

Course Structure

- **Lectures:** 24 **hours**
- **Labs:** 12 **hours**
- **Timetable:**
  - Lectures: Thursday 10:30 – 12:30, Room C4.01
  - Labs:
  - Dario Cavada : Thu 15:00 - 16:00 Room E531
  - Mehdi Elahi: Thu 15:00 - 16:00 Room E431
  - Today (Feb 24th) all students in room E531
- **Assessment:**
  - final exam, written, 50% of the grade
  - project (1 student per project !) 50%.

Motivations

- **Internet** and **World Wide Web** is modifying in a radical way how individuals and organizations interacts, for **business**, **learning** or **leisure**

- Millions of people around the world have access to an **extraordinary amount of information**, they can **search** it, exchange **email**, make **phone calls**, **buy and sell** goods and services

- All of this **is changing and will keep changing the world we live**.

Goals

- Introduction - both **methodological** and **practical** - to the **most basic** Internet:
  - **Languages**
  - **Protocols**
  - **Standards**
  - **Application Architectures**
  - **Tools**
- But also illustrate some of the most challenging and **innovative techniques** on the fore
- Self contained introduction to motivate further study and provide **prerequisite material** for more advanced courses on internet and www ("Advanced Internet Technologies" and "Internet and Mobile Services").

What you should learn

- A catalogues of **languages (API) and protocols**

- The **basic elements required for building a dynamic, database supported, web application**

- To reason about the **benefits** of a language or protocol

- The capability to **decide** when (in which context, where in your application) a technique can be **useful** or not recommendable

- *How many things you have seen does actually work?*

Course Format

- **12 Lectures** on various topics in Internet Technologies
- **12 Labs** where we shall
  - Run yourself the examples (software) shown during the lectures
  - Solve some new exercises
  - Build your own example applications
  - Work on your final exam project
- **Books**
  - Andrew S. Tanenbaum, Computer Networks, Fourth Edition, Prentice Hall PTR, 2002
  - Marty Hall and Larry Brown, Core Servlets and JavaServer Pages, Vol. 1: Core Technologies, Second Edition, Prentice Hall PTR, 2004. *(PDF available online)*

Syllabus

- Architecture of the web
- Networking fundamentals
- HTML and HTTP
- Dynamic web sites:
  - Client Side: Java Script
  - Server Side: CGI, Perl, Java Servlets and Server Pages
- Web application model
- Java servlets: generating dynamic content, session management, connecting to a data repository
- Java server pages
- XML
- Web 2.0

Challenges

- Internet technologies are **evolving** very **fast**
- To build a Web-based application you should have a very **wide knowledge** of many **software and communication technologies**
- There are dozens of **competing approaches** for building web applications
- You must learn the most updated information from Internet
- *We cannot cover all possible approaches and languages in this course*
- BUT you have a lot of space to build something innovative and useful!

What we shall **not** cover

- Ajax-enabled rich Internet applications
- Adobe Flash
- Adobe Dreamweaver
- PHP
- Ruby on Rails
- ASP.NET
- C#
- JavaServer Faces
- Java FX
- Objective C
- Web services
- Google Web Toolkit
- …

Project

- The project is conducted **individually**
- The objective is to develop **your dynamic**, **database supported**, web site:
  - Choose an **application domain**: music, trekking, soccer, photography, etc.
  - Manage **items** (music tracks, trekking paths, soccer matches, cameras, …) and users of the application
  - Identify the functionality (extending the base functionality describe later)
  - Enable **users** to access items (search, select, comment) and provide new items
- All the techniques illustrated in the lectures must be properly applied (*not a simple, static HTML-based web site*)
- The project results are a **running system** and a written **report.**

Structure of the Project

- The **application** must run on the application server that we shall indicate in the labs
- The **report** must describe clearly in **min 2000**, **max 3000** words (plus images):
  - The **functions** of the web application and their **motivation**
  - The **architecture** of the application (modules and their roles) – use figures
  - Main **classes** and main **methods**
  - Major **technical problems** found during the work
- The project will be **evaluated** according to: coverage and complexity of the implemented functions, user interface usability and completeness, organization of the code, **coverage of the required technologies**.

What a student must do to pass

- *Read* the book chapters or articles that will be suggested for each lecture
- The slides should be enough **only for a general understanding of the topic**
- If something is not clear during a lecture you must take a note and **rise a question** (especially in the labs)
- Develop and test the web application - if there are bugs and it will not run on both Firefox and IE you will not pass
- Upload the project and send me the report on time.

Exam

- The final grade is obtained evaluating the project result and the knowledge acquired about the lectures' topics in an written exam

- **Written exam:** questions on the topic illustrated – you find on the web site the previous ones

- The final written report must be sent to me ten days before the written exam (*exact timing will be provided*)

- **You cannot attend the written exam if you have not passed the project part**

- You will have two grades: P (project), max 15 points, and W (written exam), max 15 points

- The final grade is F = P + W

- Both P and W must be greater or equal to 9.

Internet Technologies
1 – Internet and other Networks

Content

- What is Internet and the World Wide Web
- Internet usage and statistics
- Introduction to computer networks
- Distributed systems
- Client-Server Architecture
- Usage of computer networks
- LAN, MAN and WAN
- Internetworks
- ARPANET
- NSFNET
- Internet Architecture

What is the Internet?

- WWW
- Video conferencing
- ftp
- telnet
- Email
- Instant messaging
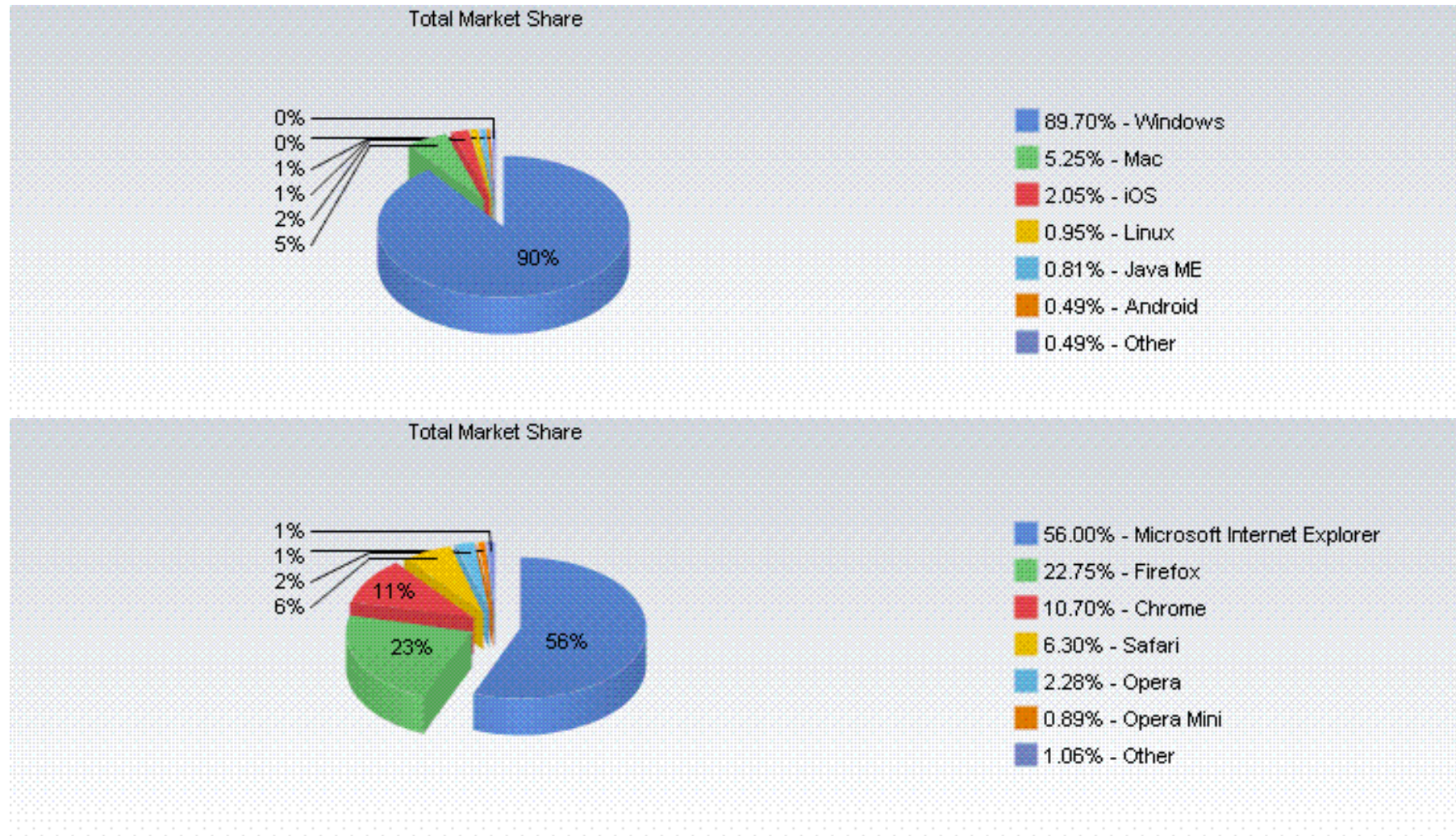- …

A communication infrastructure
Usefulness is in exchanging information

Internet Usage and Population Statistics

WORLD INTERNET USAGE AND POPULATION STATISTICS					
World Regions	**Population (2010 Est.)**	**Internet Users Dec. 31, 2000**	**Internet Users Latest Data**	**Penetration (% Population)**	**Growth 2000-2010**	**Users % of Table**
Africa	1,013,779,050	4,514,400	110,931,700	10.9 %	2,357.3 %	5.6 %
Asia	3,834,792,852	114,304,000	825,094,396	21.5 %	621.8 %	42.0 %
Europe	813,319,511	105,096,093	475,069,448	58.4 %	352.0 %	24.2 %
Middle East	212,336,924	3,284,800	63,240,946	29.8 %	1,825.3 %	3.2 %
North America	344,124,450	108,096,800	266,224,500	77.4 %	146.3 %	13.5 %
Latin America/Caribbean	592,556,972	18,068,919	204,689,836	34.5 %	1,032.8 %	10.4 %
Oceania / Australia	34,700,201	7,620,480	21,263,990	61.3 %	179.0 %	1.1 %
WORLD TOTAL	6,845,609,960	360,985,492	1,966,514,816	28.7 %	444.8 %	100.0 %

www.internetworldstats.com. June 2010

Internet Users in the European Union

EUROPEAN UNION	Population (2010 Est.)	Internet Users, Latest Data	Penetration (% Population)	User Growth (2000-2010)	Users % Table
Austria	8,214,160	6,143,600	74.8 %	192.6 %	1.3 %
Belgium	10,423,493	8,113,200	77.8 %	305.7 %	1.7 %
Bulgaria	7,148,785	3,395,000	47.5 %	689.5 %	0.7 %
Cyprus	1,102,677	433,800	39.3 %	261.5 %	0.1 %
Czech Republic	10,201,707	6,680,800	65.5 %	568.1 %	1.4 %
Denmark	5,515,575	4,750,500	86.1 %	143.6 %	1.0 %
Estonia	1,291,170	969,700	75.1 %	164.5 %	0.2 %
Finland	5,255,695	4,480,900	85.3 %	132.5 %	0.9 %
France	64,768,389	44,625,300	68.9 %	425.0 %	9.4 %
Germany	82,282,988	65,123,800	79.1 %	171.3 %	13.7 %
Greece	10,749,943	4,970,700	46.2 %	397.1 %	1.0 %
Hungary	9,992,339	6,176,400	61.8 %	763.8 %	1.3 %
Ireland	4,622,917	3,042,600	65.8 %	288.1 %	0.6 %
Italy	58,090,681	30,026,400	51.7 %	127.5 %	6.3 %
Latvia	2,217,969	1,503,400	67.8 %	902.3 %	0.3 %
Lithuania	3,545,319	2,103,471	59.3 %	834.9 %	0.4 %
Luxembourg	497,538	424,500	85.3 %	324.5 %	0.1 %
Malta	406,771	240,600	59.1 %	501.5 %	0.1 %
Netherlands	16,783,092	14,872,200	88.6 %	281.3 %	3.1 %
Poland	38,463,689	22,450,600	58.4 %	701.8 %	4.7 %
Portugal	10,735,765	5,168,800	48.1 %	106.8 %	1.1 %
Romania	21,959,278	7,786,700	35.5 %	873.3 %	1.6 %
Slovakia	5,470,306	4,063,600	74.3 %	525.2 %	0.9 %
Slovenia	2,005,692	2,003,136	64.8 %	332.8 %	0.3 %
Spain	46,505,963	29,093,984	62.6 %	440.0 %	6.1 %
Sweden	9,074,055	8,397,900	92.5 %	107.5 %	1.8 %
United Kingdom	62,348,447	51,442,100	82.5 %	234.0 %	10.8 %
European Union	499,671,847	337,779,055	67.6 %	257.8 %	100.0 %

Web access by OS and Browser

Total Market Share

0% / 0% / 1% / 1% / 2% / 5% — 90%

- 89.70% - Windows
- 5.25% - Mac
- 2.05% - iOS
- 0.95% - Linux
- 0.81% - Java ME
- 0.49% - Android
- 0.49% - Other

Total Market Share

1% / 1% / 2% / 6% — 11% — 23% — 56%

- 56.00% - Microsoft Internet Explorer
- 22.75% - Firefox
- 10.70% - Chrome
- 6.30% - Safari
- 2.28% - Opera
- 0.89% - Opera Mini
- 1.06% - Other

http://marketshare.hitslink.com

Generation Y, X and Baby Boomers

Generations Explained			
Generation Name*	Birth Years, Ages in 2009	% of total adult population	% of internet-using population
Gen Y (Millennials)	Born 1977-1990, Ages 18-32	26%	30%
Gen X	Born 1965-1976, Ages 33-44	20%	23%
Younger Boomers	Born 1955-1964, Ages 45-54	20%	22%
Older Boomers	Born 1946-1954, Ages 55-63	13%	13%
Silent Generation	Born 1937-1945, Ages 64-72	9%	7%
G.I. Generation	Born -1936, Age 73+	9%	4%

Source: Pew Internet & American Life Project December 2008 survey. N=2,253 total adults, and margin of error is ±2%. N=1,650 total internet users, and margin of error is ±3%.

http://www.pewinternet.org/pdfs/PIP_Generations_2009.pdf

Teens and Gen Y dominant activities

Generational Differences in Online Activities

	Online Teens[a] (12-17)	Gen Y (18-32)	Gen X (33-44)	Younger Boomers (45-54)	Older Boomers (55-63)	Silent Generation (64-72)	G.I. Generation (73+)	All Online Adults[b]
Go online	93%	87%	82%	79%	70%	56%	31%	74%
Teens and Gen Y are more likely to engage in the following activities compared with older users:								
Play games online	78	50	38	26	28	25	18	35
Watch videos online	57	72	57	49	30	24	14	52
Get info about a job	30~	64	55	43	36	11	10	47
Send instant messages	68	59	38	28	23	25	18	38
Use social networking sites	65	67	36	20	9	11	4	35
Download music	59	58	46	22	21	16	5	37
Create an SNS profile	55	60	29	16	9	5	4	29
Read blogs	49	43	34	27	25	23	15	32
Create a blog	28	20	10	6	7	6	6	11
Visit a virtual world	10	2	3	1	1	1	0	2

Dominant activities for Gen X and older

	Online Teens[a] (12-17)	Gen Y (18-32)	Gen X (33-44)	Younger Boomers (45-54)	Older Boomers (55-63)	Silent Generation (64-72)	G.I. Generation (73+)	All Online Adults[b]
Activities where Gen X users or older generations dominate:								
Get health info	28	68	82	74	81	70	67	75
Buy something online	38	71	80	68	72	56	47	71
Bank online	*	57	65	53	49	45	24	55
Visit gov't sites	*	55	64	62	63	60	31	59
Get religious info	26~	31	38	42	30	30	26	35
And for some activities, the youngest and oldest cohorts may differ, but there is less variation overall:								
Use email	73	94	93	90	90	91	79	91
Use search engines	*	90	93	90	89	85	70	89
Research products	*	84	84	82	79	73	60	81
Get news	63	74	76	70	69	56	37	70
Make travel reservations	*	65	70	69	66	69	65	68
Research for job	*	51	59	57	48	33	9	51
Rate a person or product	*	37	35	29	30	25	16	32
Download videos	31~	38	31	21	16	13	13	27
Participate in an online auction	*	26	31	27	26	16	6	26
Download podcasts	19	25	21	19	12	10	10	19

Online shopping activities

Online shopping activities		
	% of internet users who have <u>ever</u> done this	% of internet users who did this <u>yesterday</u> (the day before they responded to our survey)
Research online a product they are thinking of buying	81%	20%
Purchase a product online such as books, music, or clothing	66	6
Bought or made travel reservations online	64	4
Participated in an online auction	26	3
Paid to access or download digital content	17	4
Bought or sold stocks online	11	1
Source: Pew Internet & American Life Project Survey, September 2007, n for internet users=1,684. Margin of error is +/- 3 points.		

Percent of all American adults who have ever bought a product online

Jun 00	Aug 01	Jan 02	Feb 03	Feb 04	May 05	Aug 06	Sep 07
22%	32%	35%	39%	41%	46%	50%	49%
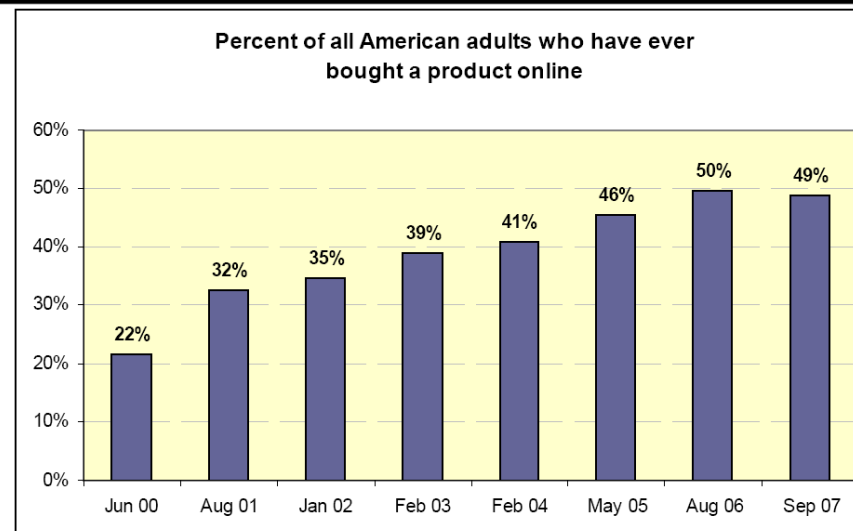
http://www.pewinternet.org/pdfs/PIP_Online%20Shopping.pdf

People Online in Tourism Market

People online in tourism markets (%'s)
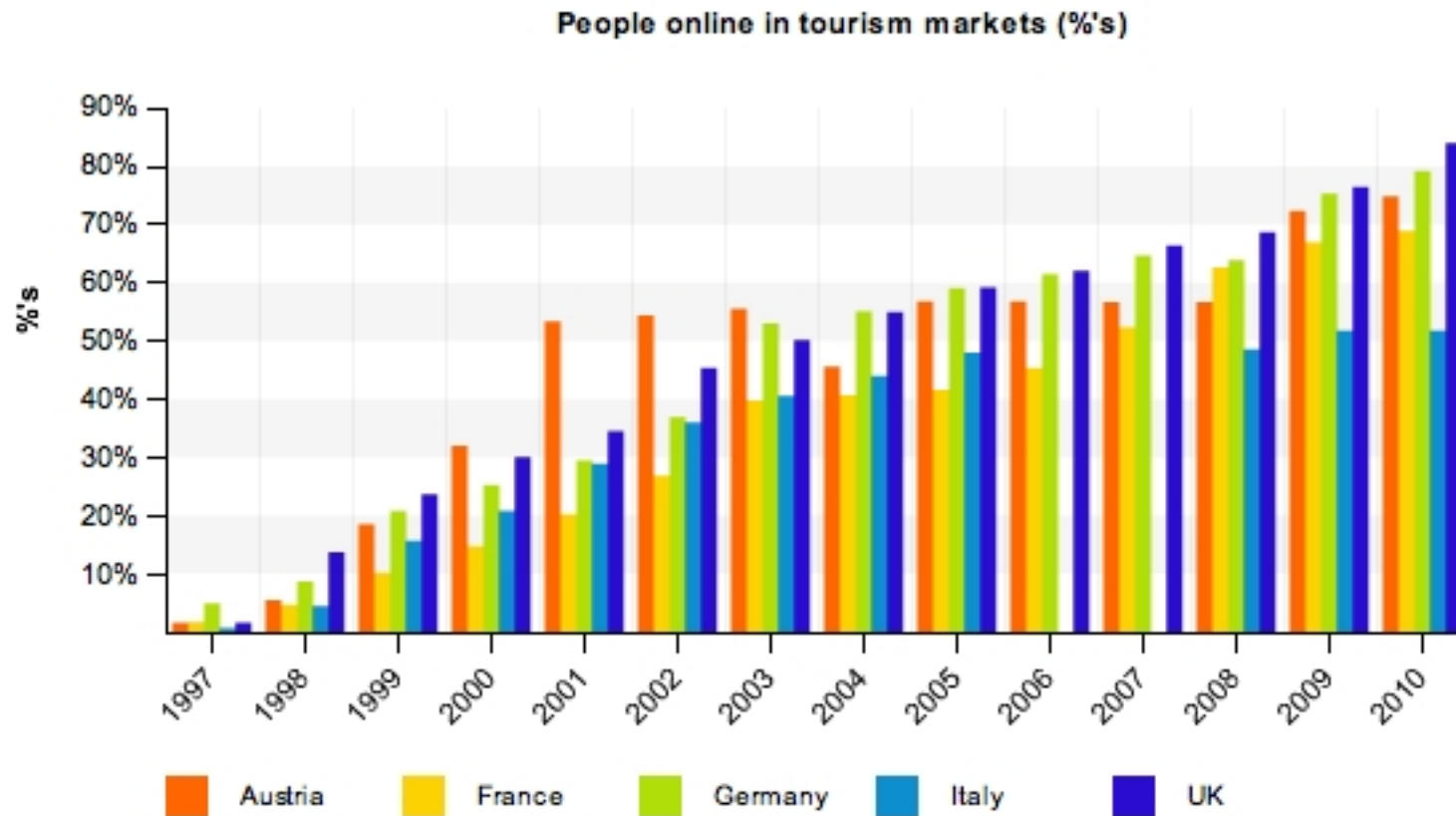
Total Sites Across All Domains August 1995 - November 2007
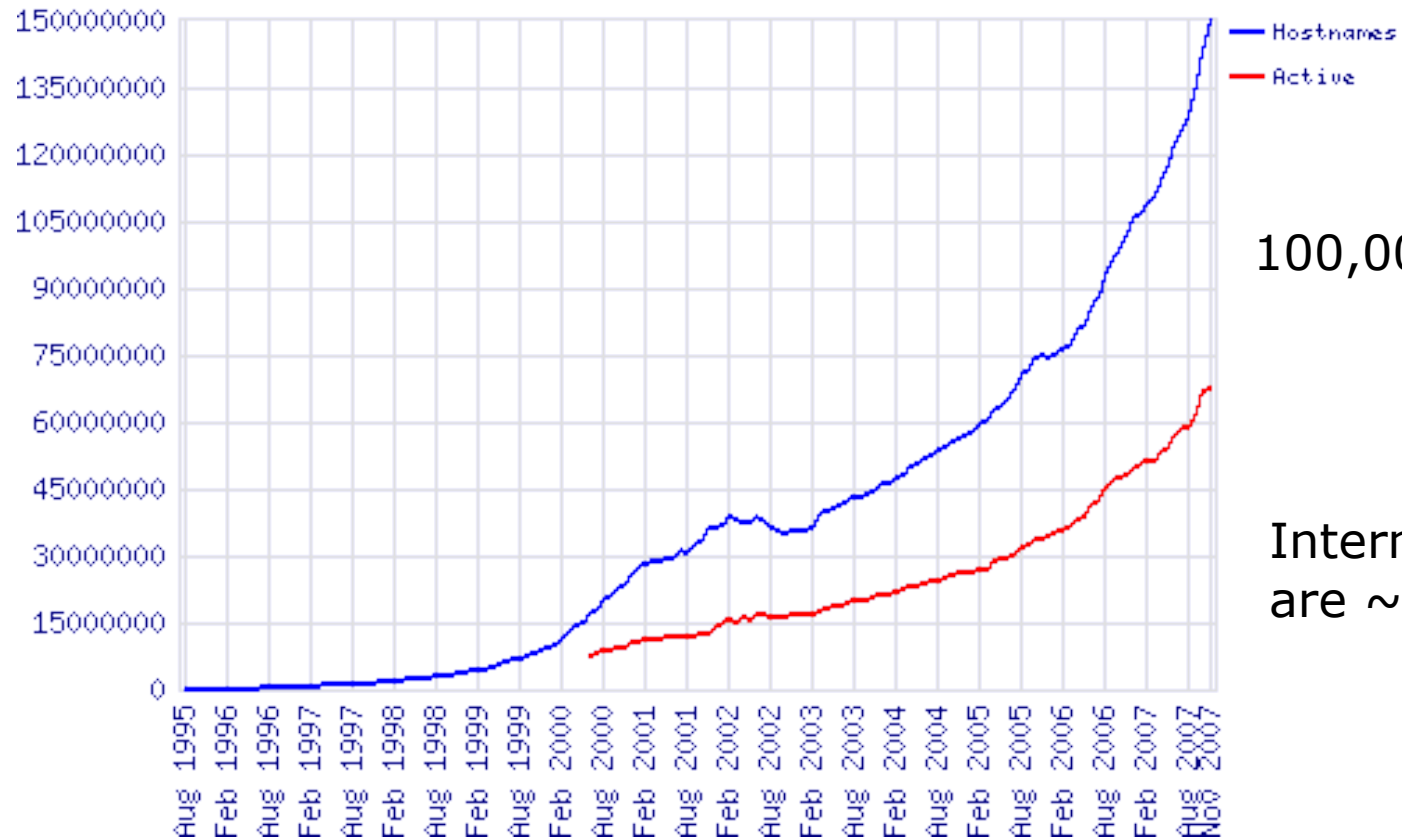
100,000,000

Internet users are ~ 2Billions

Much of the growth in sites this year has come from the increasing number of blogging sites, in particular at Live Spaces, Blogger and MySpace.

An active web site every 18 users!

http://news.netcraft.com

Top Global Web Properties

comSCORE.

Top Global Web Properties Ranked by Total Unique Visitors (000)* May 2008 Total Worldwide, Age 15+ - Home and Work Locations Source: comScore World Metrix		
Property	Total Unique Visitors (000)	% Reach
Google Sites	643,809	75.5
Microsoft Sites	572,016	67.1
Yahoo! Sites	514,831	60.3
Wikipedia Sites	263,120	30.8
AOL LLC	252,394	29.6
eBay	247,791	29.0
Fox Interactive Media	169,301	19.8
Amazon Sites	159,281	18.7
Apple Inc.	140,380	16.5
CNET Networks	133,480	15.6
Ask Network	127,769	15.0
FACEBOOK.COM	123,851	14.5
Adobe Sites	107,361	12.6
Time Warner - Excluding AOL	98,000	11.5
WordPress	96,394	11.3
Viacom Digital	86,546	10.1
Baidu.com Inc.	80,201	9.4
TENCENT Inc.	77,885	9.1
Glam Media	77,391	9.1
New York Times Digital	77,172	9.0

* Excludes traffic from public computers such as Internet cafes
and access from mobile phones or PDAs

What are the top three Web properties?

with respect to the number of visitors

Top Web Sites in Italy

Capture - Recapture

- **SE1=** reported size of search engine 1
- **Q** – set of queries
- **QSE1** and **QSE2** = pages returned for **Q** from two engines
- **OVR** – overlap of **QSE1** and **QSE2**
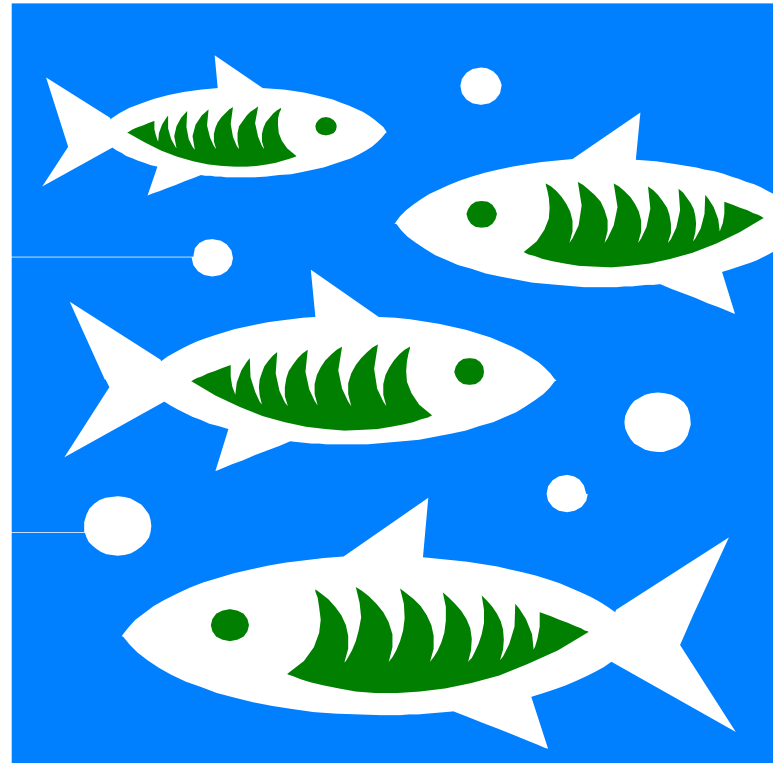
- Estimate of **Web** size:
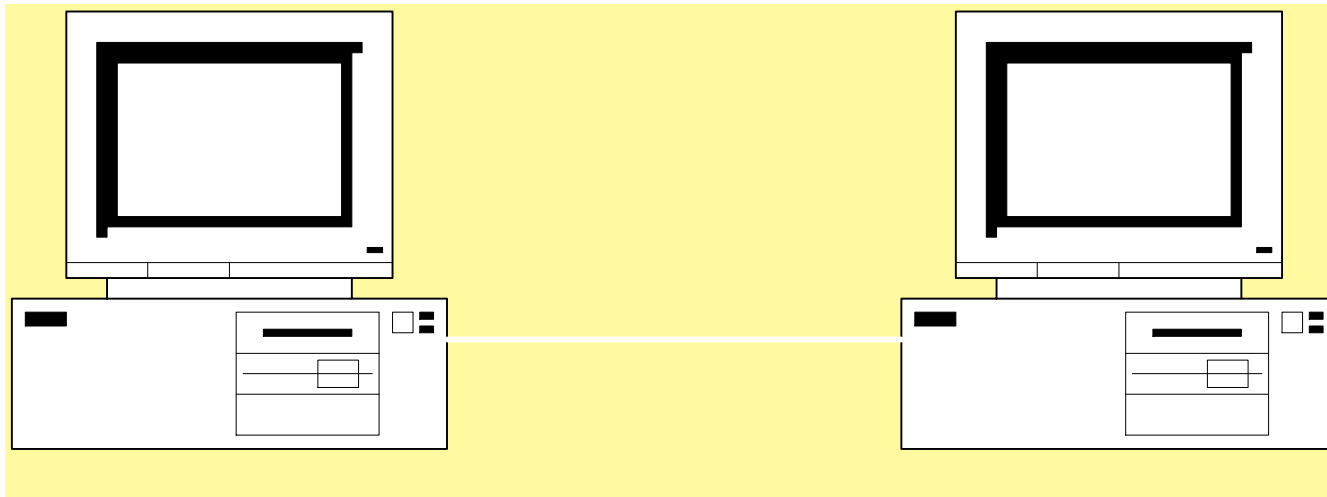- **SE1/Web = OVR/QSE2**
- **Web = (QSE2 * SE1) / OVR**

Concentration
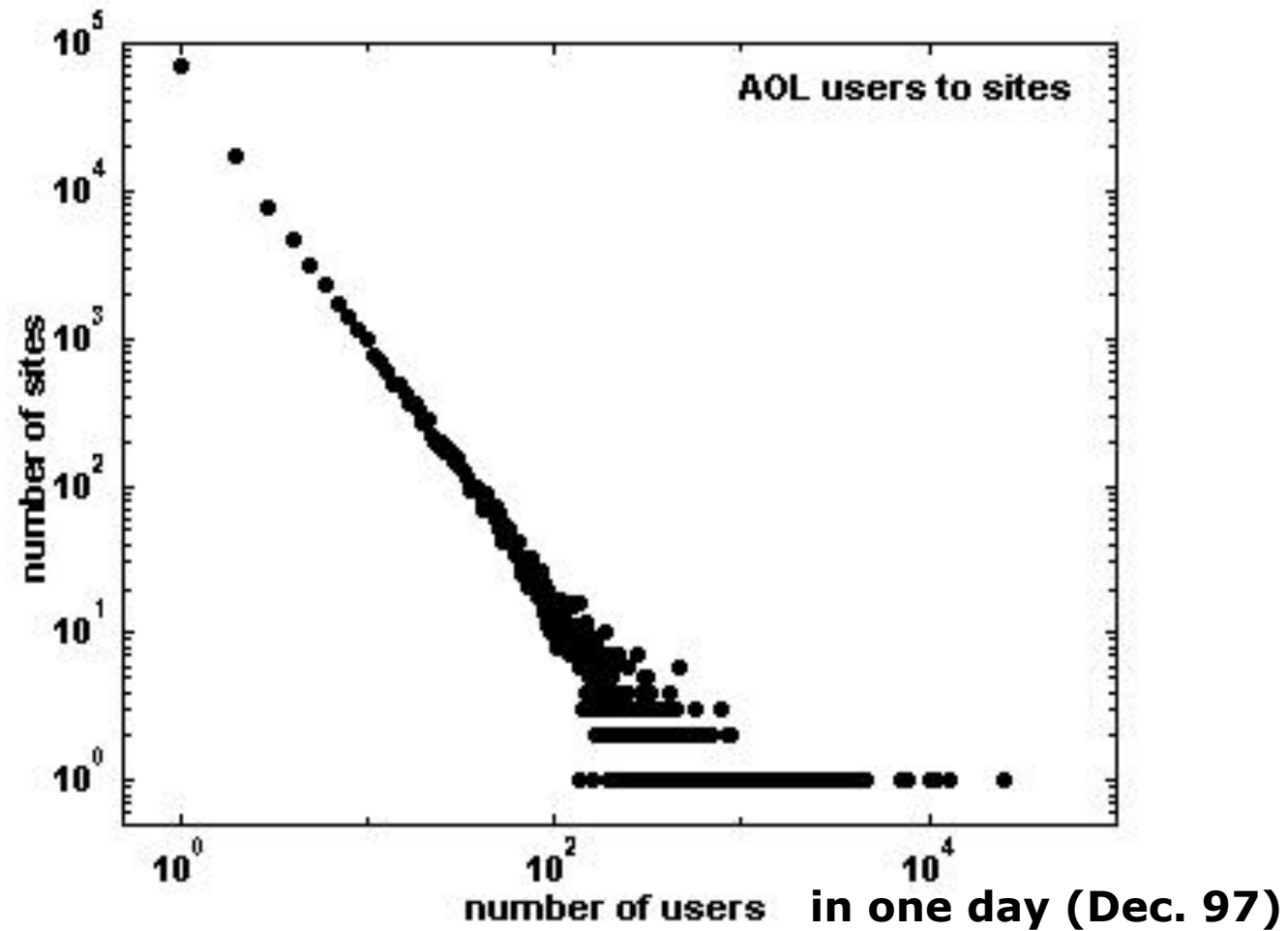
Power-law: $y = Cx^{-a}$

$\log(y) = \log(C) - a \log(x)$

The simplest network?

The computers have their NIC (Network Interface Card) with a socket (RJ-45 jack) and a wire (crossover cable) that goes from one computer to another

Computer Networks

- A **computer network** is two or more computers connected together using a telecommunication system for the purpose of communicating and sharing resources
- Why they are interesting?
  - Overcome geographic limits
  - Access remote data
  - Separate clients and server
- Goal: Universal Communication (any to any)
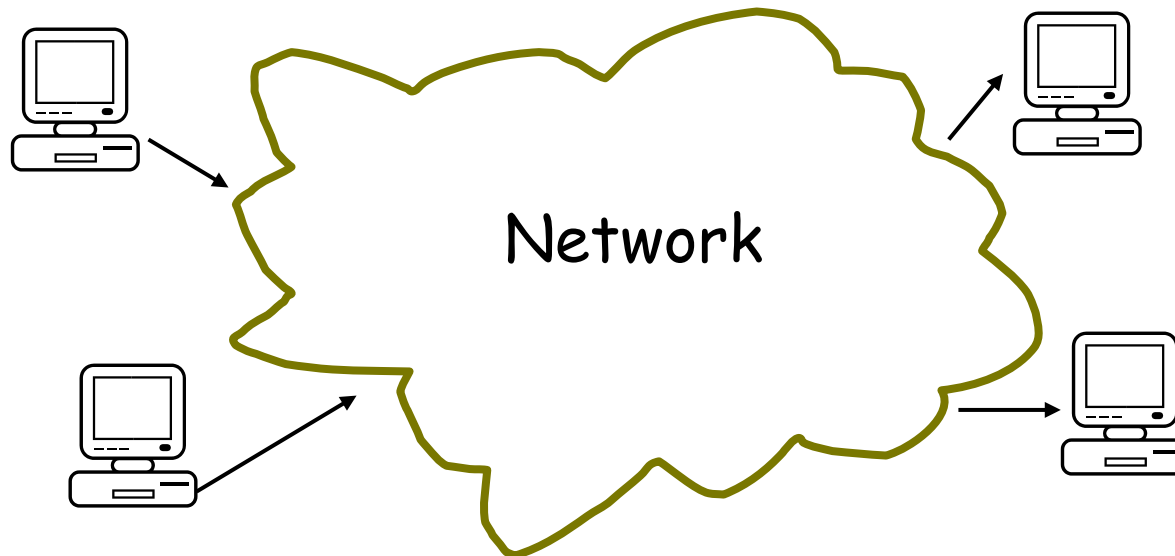
Network

Distributed Systems

- Internet is not a "computer network" – it is a **network of networks**
- **The World Wide Web is a distributed system that runs on top of the Internet**
- *A **distributed system** is a collection of independent computers that appears to its users as a single coherent system*
  - Example: in the WWW everything looks like a document (Web page)
- The distinction between CN and DS lies on the **software** not on the **hardware**.

Client-Server Model

- A network with two clients and one server
- **Server:** store data on some powerful computer
- **Client:** access data on server and process locally on a simpler machine
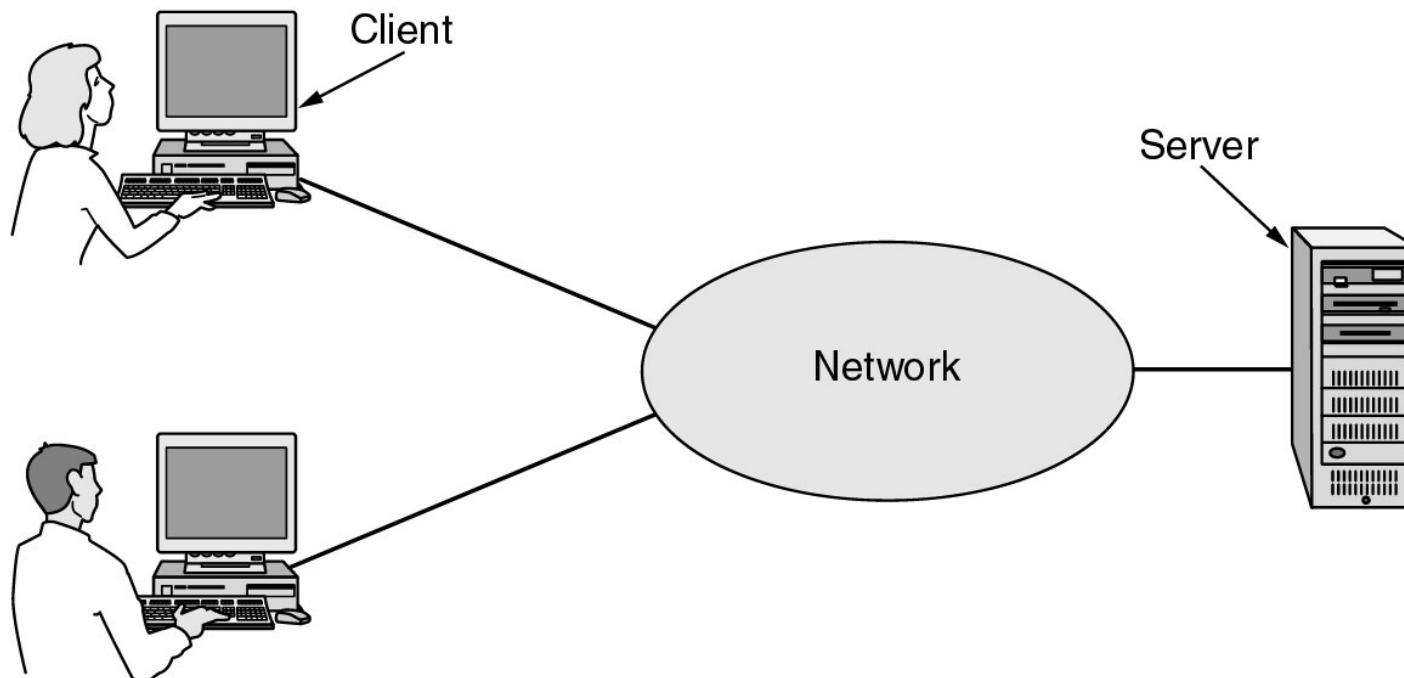
Client-Server Model (2)

- The client-server model involves requests and replies
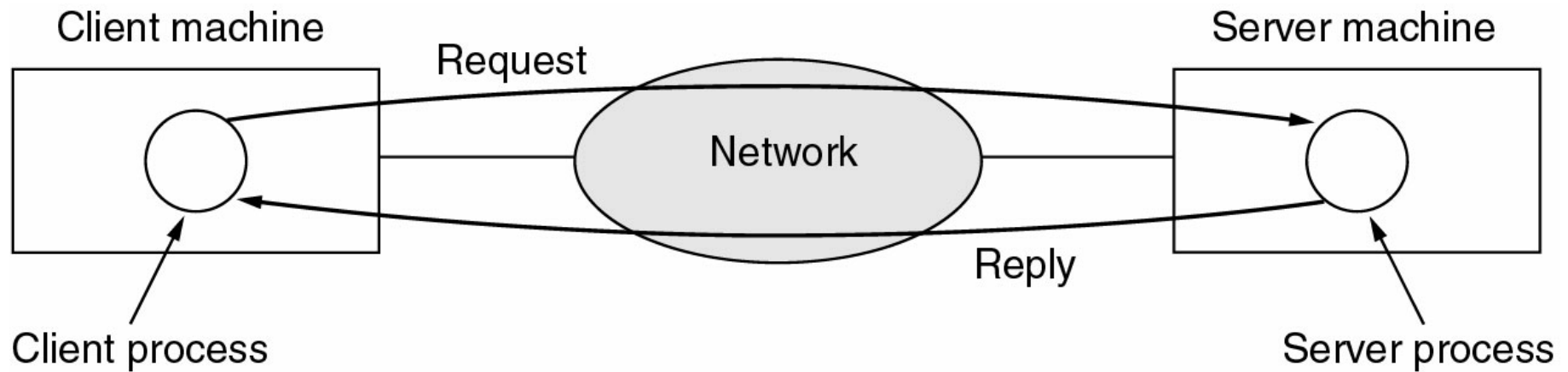
- Examples
  - e-mail
  - Video conferencing
  - File downloading
  - Instant messaging
  - Chatting

Network Applications

- Some forms of e-commerce

Tag	Full name	Example
B2C	Business-to-consumer	Ordering books on-line
B2B	Business-to-business	Car manufacturer ordering tires from supplier
G2C	Government-to-consumer	Government distributing tax forms electronically
C2C	Consumer-to-consumer	Auctioning second-hand products on-line
P2P	Peer-to-peer	File sharing

amazon.com.

Città di Bolzano
Stadt Bozen

G2C

ebaY

C2C

P2P

skype

B2B and B2C

Peer-to-Peer Systems

- In peer-to-peer system there are no fixed clients and servers
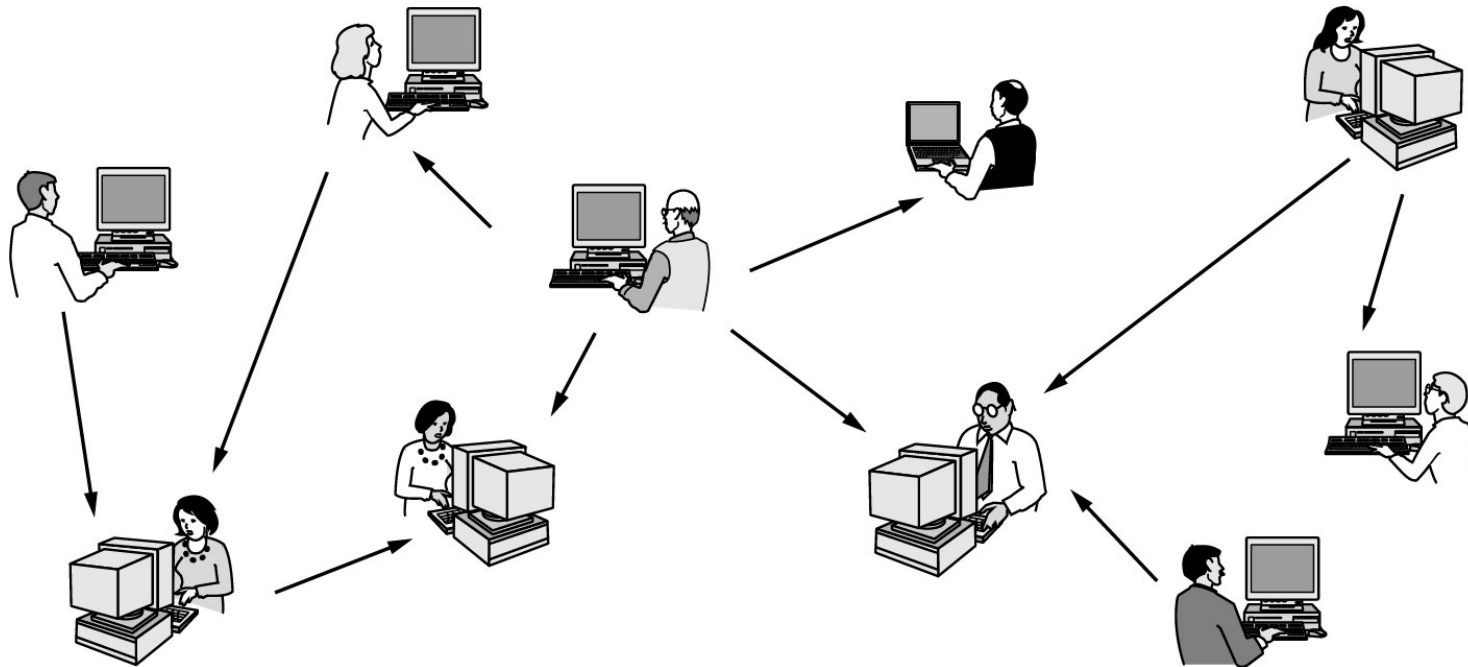
- Examples?
- Skype, Kazaa, eMule, exchanging business cards with bluetooth, …

Transmission Technologies

- **Broadcast links**
  - A **single** communication channel is shared by all the machines on the network
  - Packets sent by a machine to another brings the address of the recipient
  - Avoid collision in sharing the medium (channel)
- **Point-to-point links**
  - Many connections between **individual pairs** of machines
  - A packet is **routed** from one machine to another

Networks scale

- Classification of interconnected processors by scale.

Interprocessor distance	Processors located in same	Example
1 m	Square meter	Personal area network
10 m	Room	Local area network
100 m	Building	Local area network
1 km	Campus	Local area network
10 km	City	Metropolitan area network
100 km	Country	Wide area network
1000 km	Continent	Wide area network
10,000 km	Planet	The Internet

Type of Networks

- **PAN:** A **personal area network** is a computer network (CN) used for communication among computer devices (including telephones and personal digital assistants) close to one person
  - **Technologies:** USB and Firewire (wired), IrDA and Bluetooth (wireless)
- **LAN:** A **local area network** is a CN covering a small geographic area, like a home, office, or group of buildings
  - **Technologies:** Ethernet (wired) or Wi-Fi (wireless)
- **MAN: Metropolitan Area Networks** are large CNs usually spanning a city
  - **Technologies:** Ethernet (wired) or WiMAX (wireless)
- **WAN: Wide Area Network** is a CN that covers a broad area, e.g., cross metropolitan, regional, or national boundaries
  - **Examples:** Internet
  - **Wireless Technologies:** HSDPA, EDGE, GPRS, GSM.

Local Area Networks
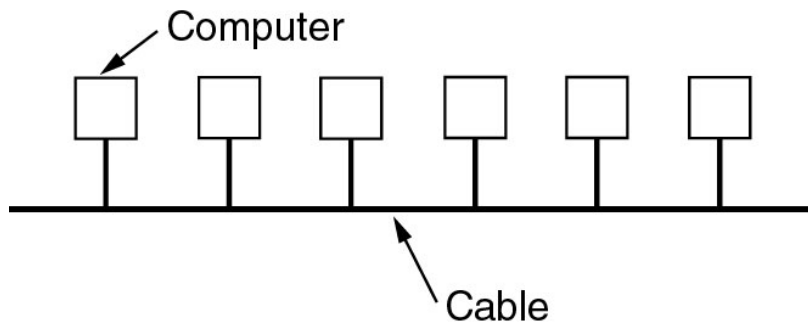
(a)
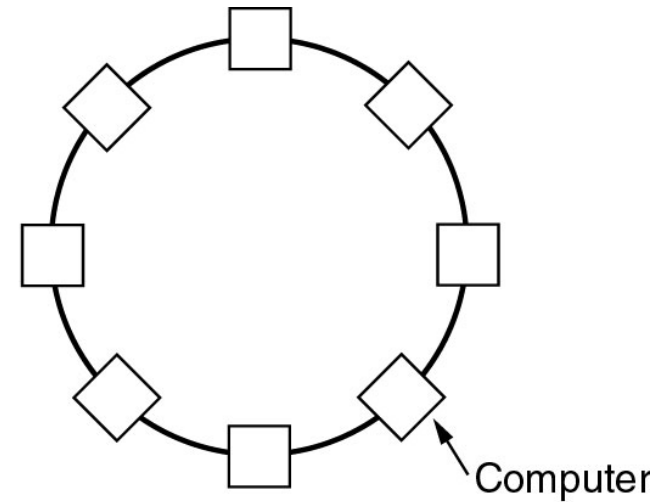
(b)

- Two broadcast networks: (a) Bus, (b) Ring
- LAN can use cables (Ethernet protocol) or electromagnetic waves (Wi-Fi) to transmit information

Wireless Networks
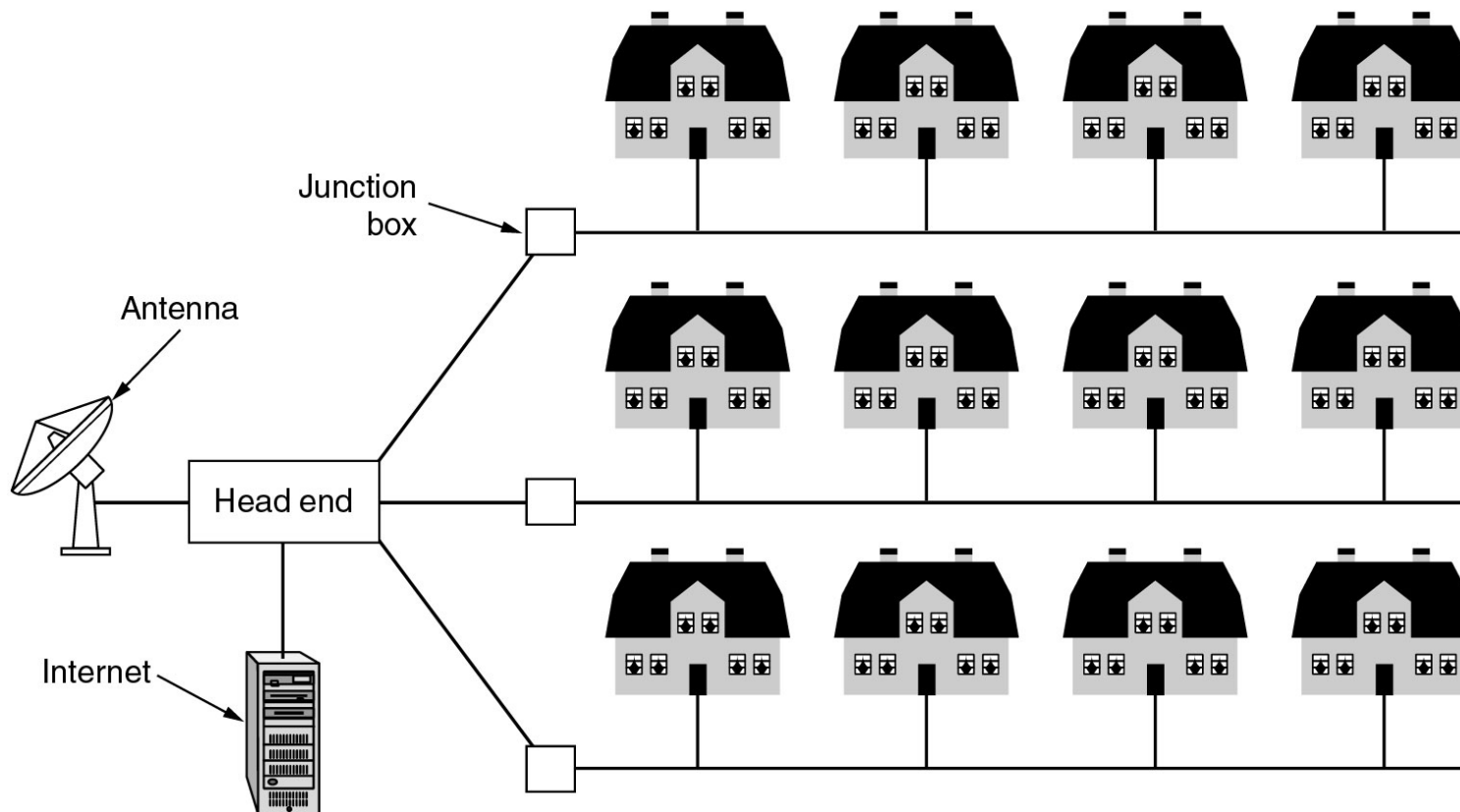
- (a) Bluetooth configuration
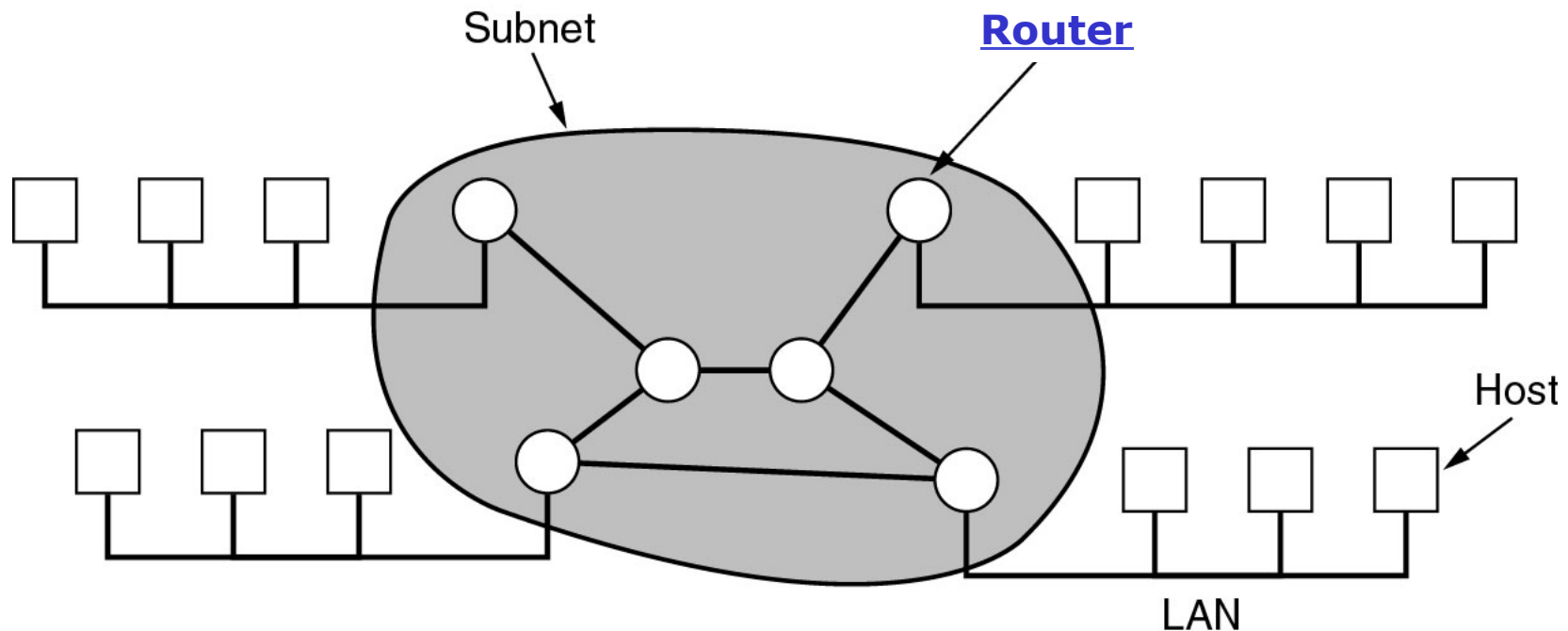- (b) Wireless LAN

(a)

Base station | To wired network
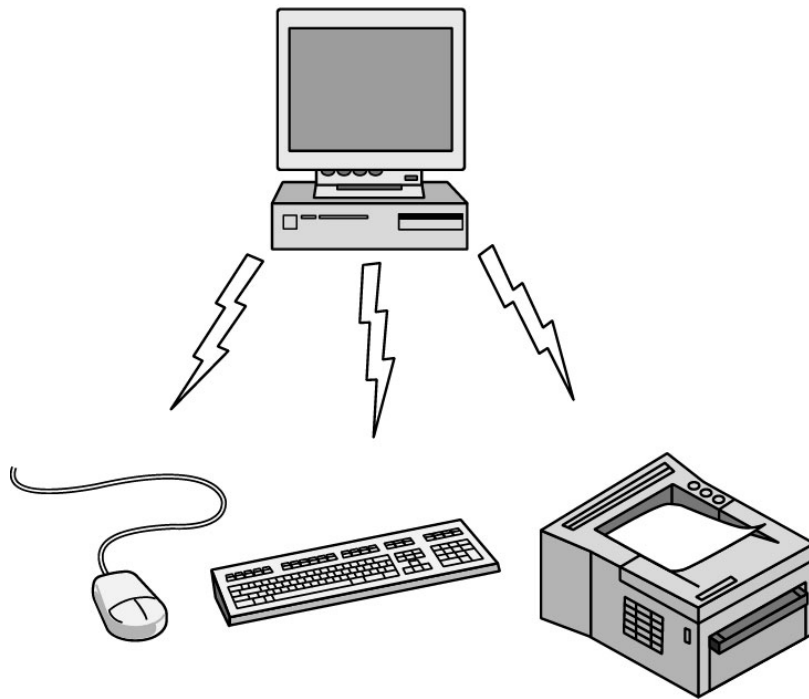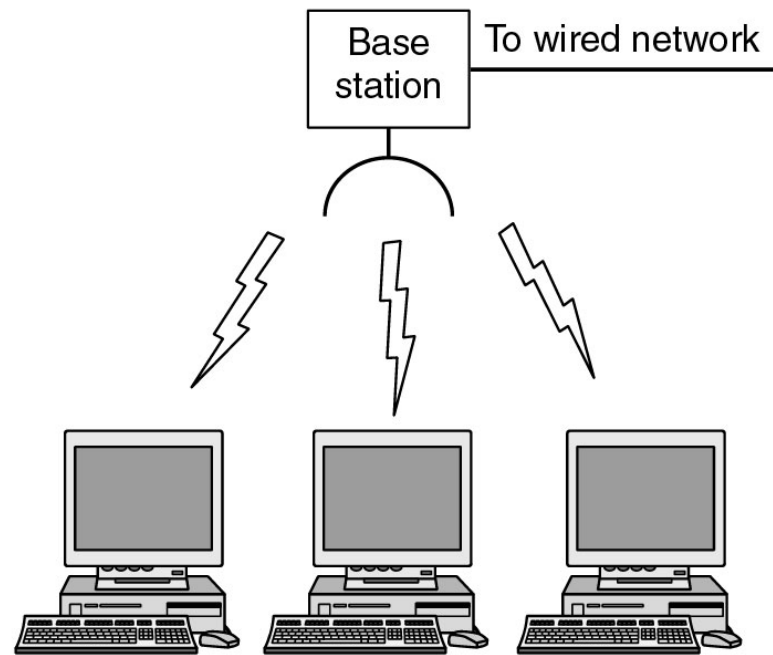
(b)

Metropolitan Area Networks

- A metropolitan area network based on cable TV

Wide Area Networks

- **Host** are owned by users
- **Subnet** is owned by the telephone company or an Internet service provider
- A subnet is composed by **transmission lines** connecting two **switching** elements (**router**) – not the hosts.

Wide Area Networks (2)

- A stream of packets from sender to receiver
- A routers **store-and-forward** each packet
- The decision of where to send a packet is taken according to a **routing algorithm**
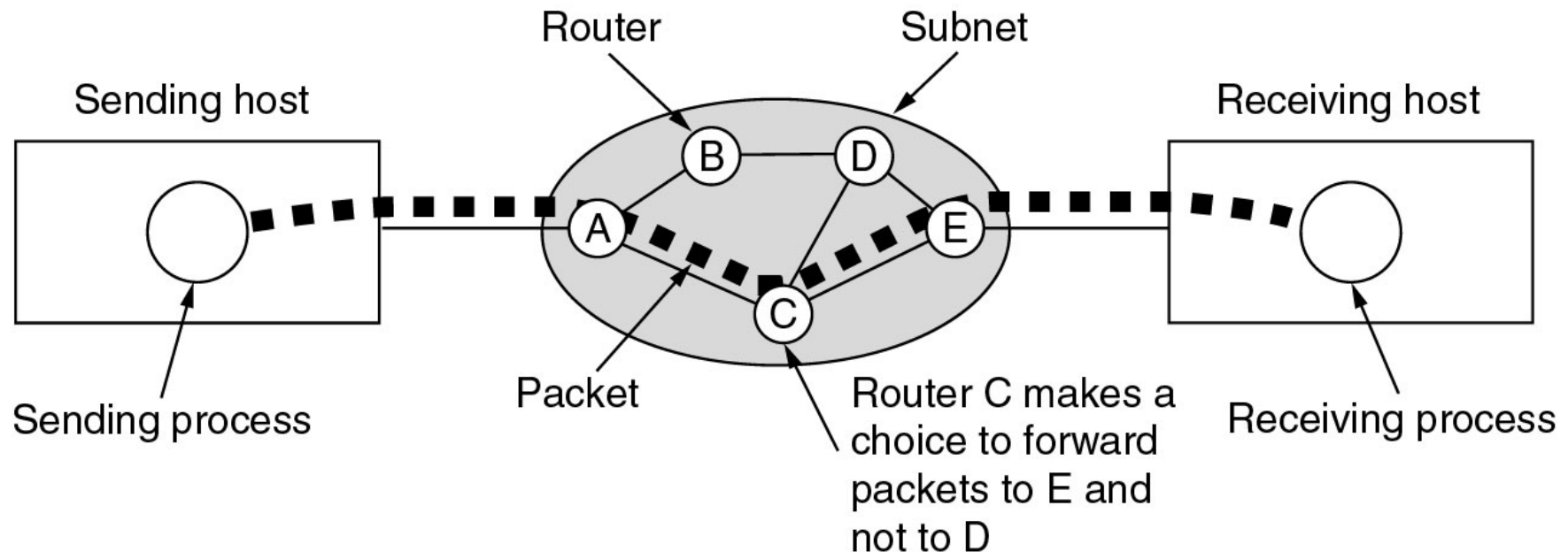
Internetworks and Internet

- Many networks exists in the world
- In order to establish a communication between to "different" networks (hardware e software) there are **gateways**
- A **collection of interconnected networks** is called **internetworks** or **internet**
- The **Internet**, with a capital "I", is the network of networks which either use the TCP/IP protocol or can interact with TCP/IP networks via gateways (the interpreters)
- The Internet presents these networks as one, seamless network for its users
- **Internet is a particular internetwork**

ARPANET

- 1950 Department of Defence wanted a command-and-control network that could survive nuclear war
- At that time, there was only the telephone network
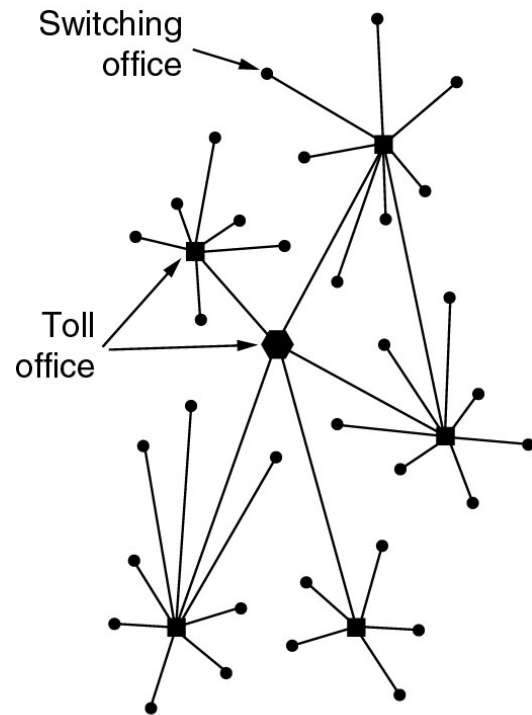

(a)
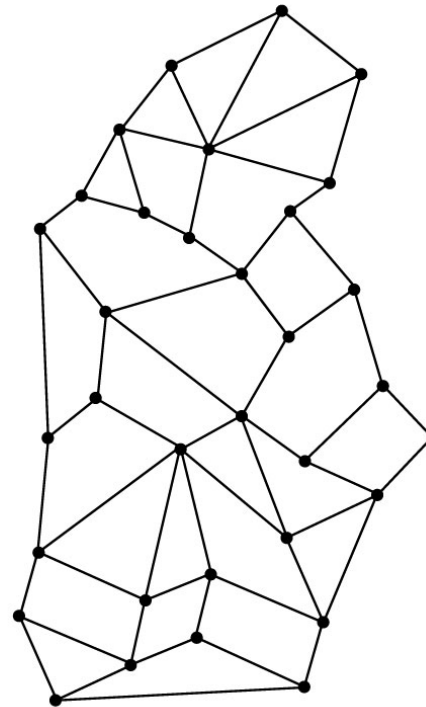
(b)

(a) Structure of the telephone system – vulnerable!

(b) Baran's proposed distributed switching system.

The original ARPANET

- IMP (Interface Message Processors) are minicomputers connected by 56-Kbps transmission lines (*the grandfathers of the routers*)

- Each IMP is **connected** with (at least) **2** IMPs (*why? is this enough?*)

- A host is connected to a IMP – it sends to it a message that is split into packets (1008 bits) forwarded independently to destination.
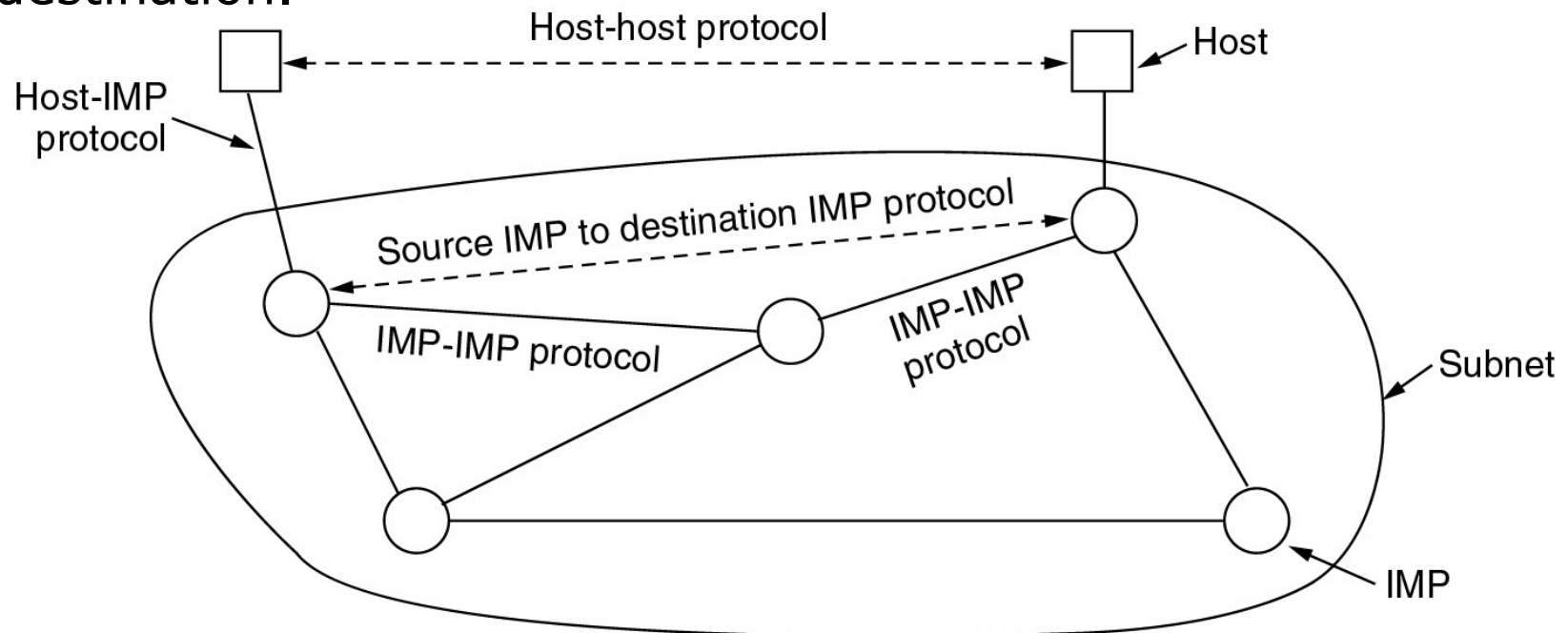
The Growth of ARPANET

(a) (b) (c)

(d) (e)
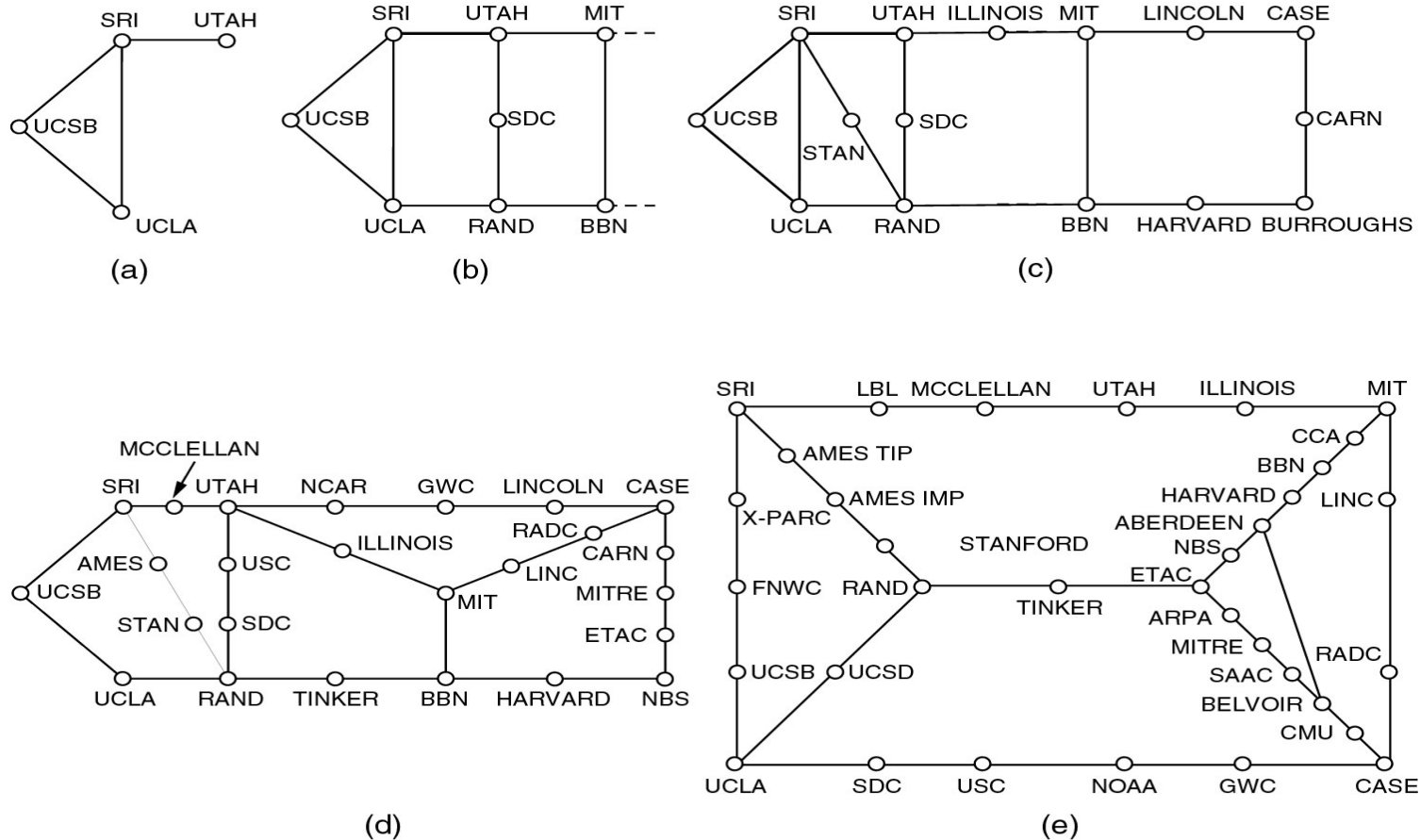
- (a) December 1969, (b) July 1970, (c) March 1971, (d) April 1972, (e) September 1972
- TCP/IP invented by Cerf and Khan in 1974.

NSFNET

- The NSFNET backbone in 1988
- The computers (fuzzball) where connected with TCP/IP (56 Kbps lines)
- Then to 448Kbps, then 1.5-Mbps, then 45-Mbps (ANSNET – then sold to America Online)
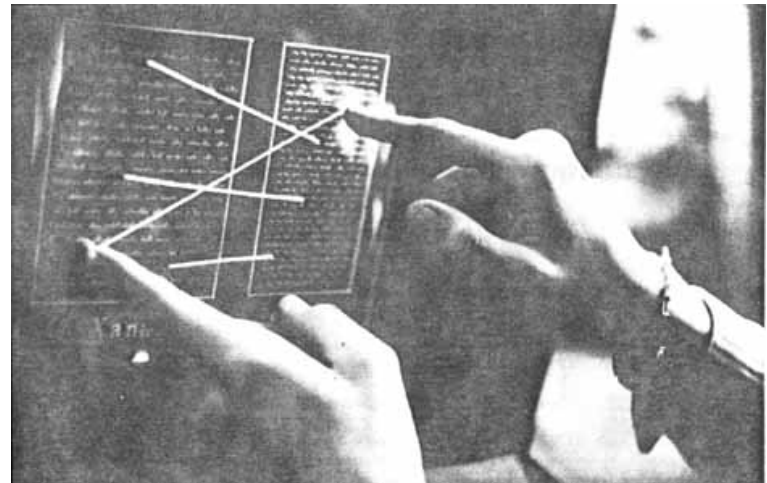- Connected to ARPANET trough a link between an IMP and a fuzzball in the Carnegie Mellon computer room.

○ NSF Supercomputer center
⊗ NSF Midlevel network
● Both
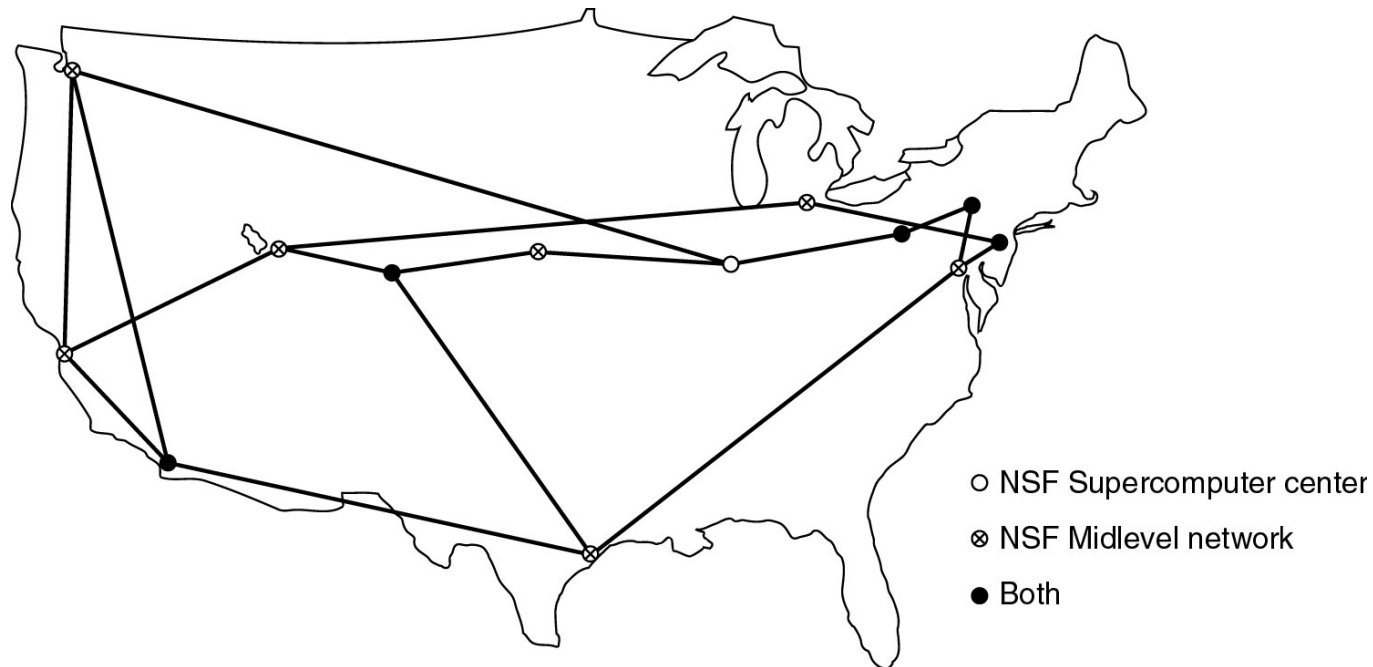
History of the Internet

- 1969 - **RFCs** begun by S. Crocker (http://rfc.sunsite.dk/)
- 1972 – First **email** by Ray Tomlinson & Larry Roberts
- 1970's - TCP by Vint Cerf & Bob Kahn
  - Evolved into TCP/IP, and UDP
- 1980s – Hardware Explosion (LANs, PCs, and workstations)
  - 1983 – Ethernet by Metcalfe
- DNS – Distributed and scalable mechanism for resolving host names into IP addresses
- UC Berkeley implements TCP/IP into Unix BSD
- 1985 – Internet used by researchers and developers.
- 1993 – the first Web Browser (NCSA Mosaic)

History of the Internet

- Tim Berners-Lee at CERN in 1989
  - Proposal for WWW in 1990
  - First web page on November 13, 1990
- Hypertext - Text that contains links to other text.
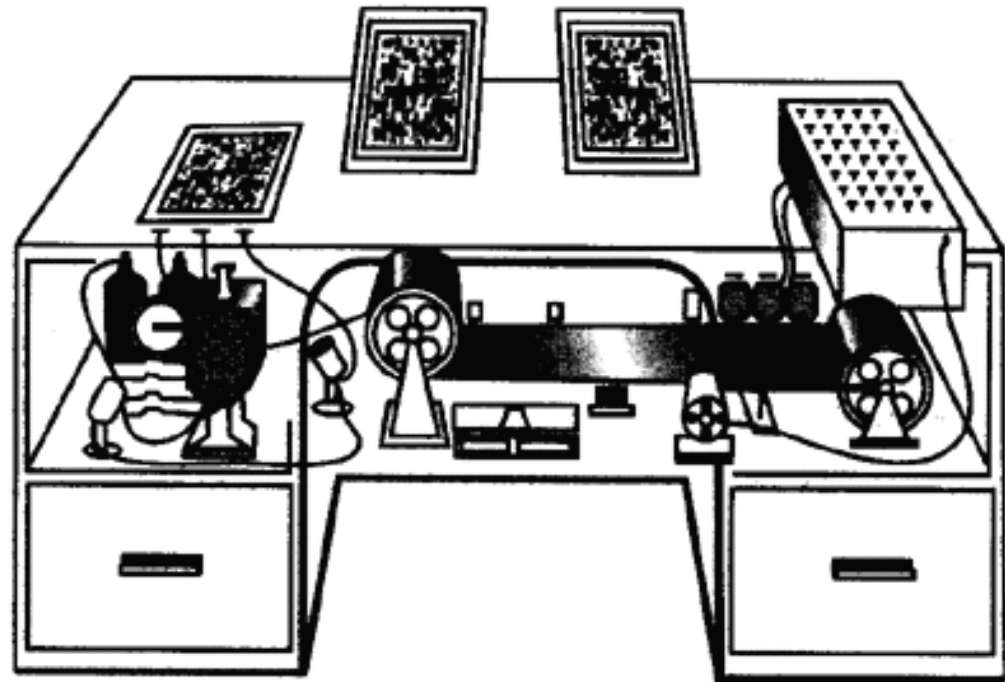 - Ted Nelson's Xanadu

  - Vannevar Bush's Memex

(http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm)

- W3C
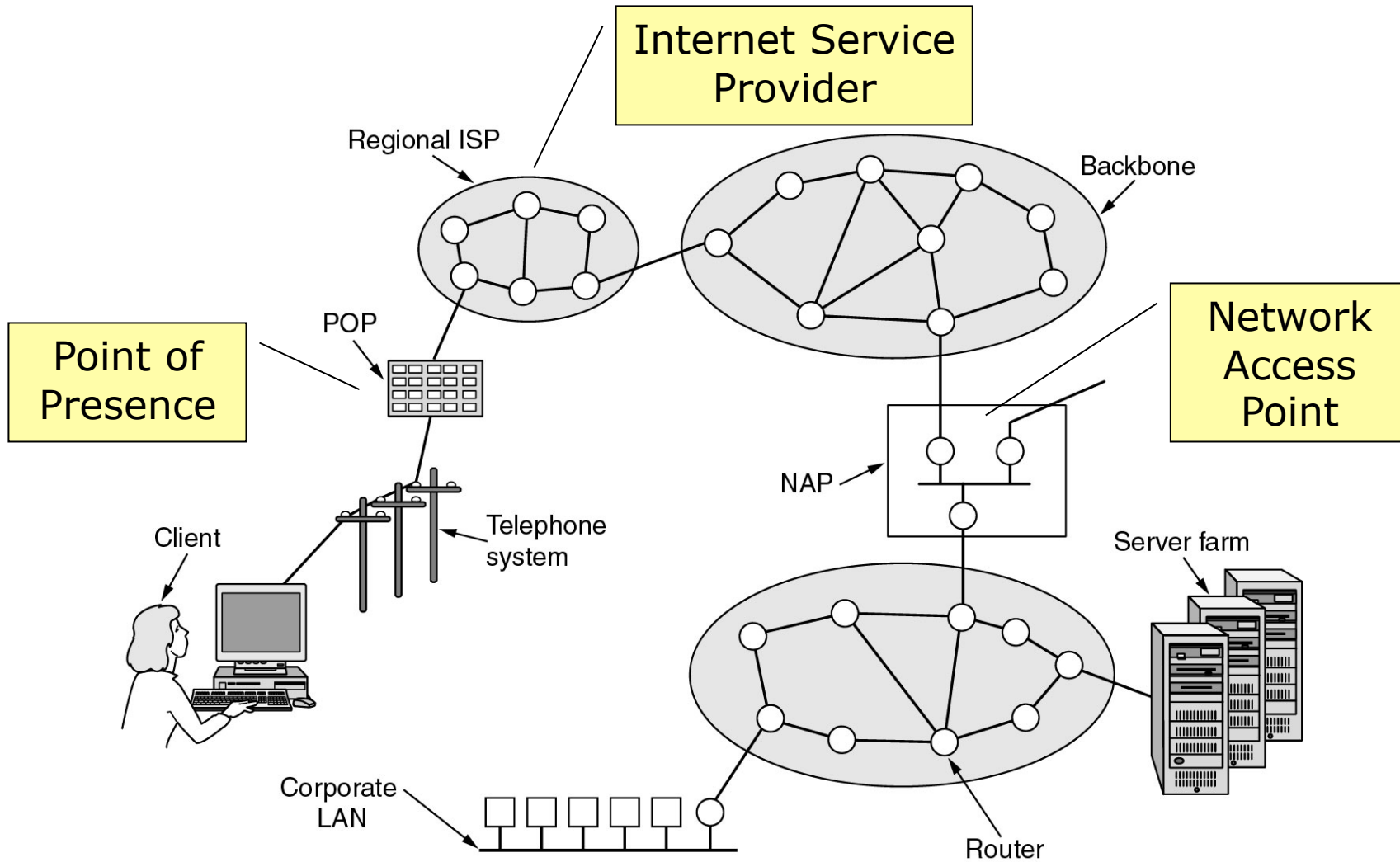
Get more info at:
http://www.isoc.org/internet/history/

Bush 1945 – As We May Think

The memex is a desktop machine, consisting of:

1) A user interface
2) A repository of documents
3) A search engine
4) A linking mechanism
5) Memex II can learn from its experience.

# Architecture of the Internet

Internet Service Provider

Regional ISP

Backbone

Point of Presence

POP

Network Access Point

NAP

Client

Telephone system

Server farm

Corporate LAN

Router

Important Terms

- **POP (Point of Presence):** an access point to the Internet. It is a physical location that houses servers, routers, ATM switches and digital/analog call aggregators

- **ISP (Internet Service Provider):** business or organization that provides consumers or businesses access to the Internet and related services

- **Backbone:** a large collection of interconnected commercial, government, academic and other high-capacity data routes and core routers that carry data across the countries, continents and oceans of the world

- **NAP (Network Access Point):** 4 Network access points where a packet switches from one backbone to another.

IP Addresses Network
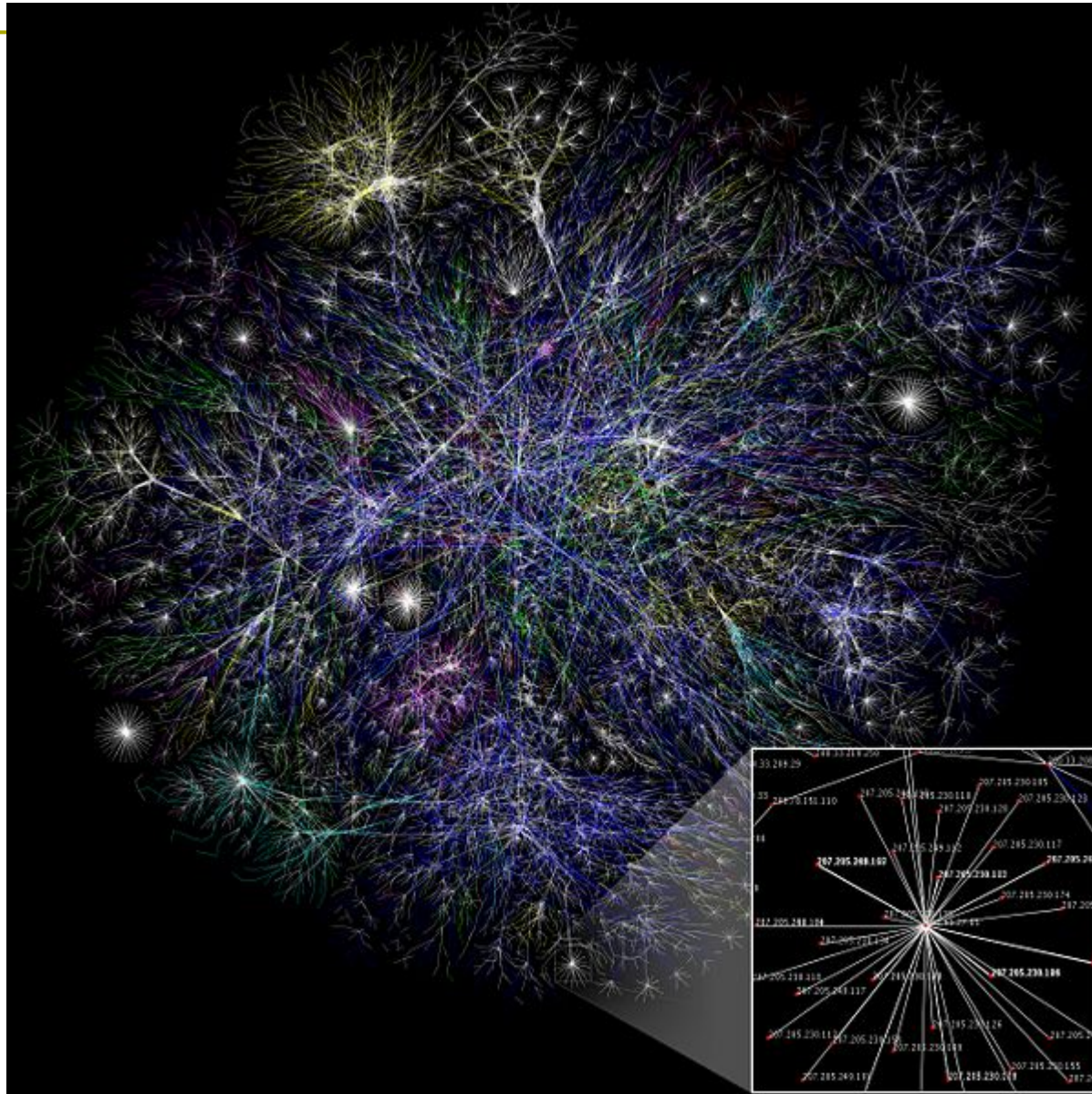
A line represent a connection between 2 IP addresses. The length represent the time delay between the 2 nodes.

http://en.wikipedia.org/wiki/Image:Internet_map_1024.jpg

How to Make the Internet a Lot Faster

- Feb. 2010 - Google announced its plans to build an experimental fiber network that would offer **gigabit-per-second** broadband speeds to U.S. homes

- This will make possible: transfer of very large files, streaming high-definition (and possibly 3-D) video, video conferencing, and gaming

- **BUT** the transmission control protocol (TCP), the 20-year-old algorithm that governs most of the traffic flow over the Internet, doesn't work well at gigabit-per-second speeds

- How it make sure it isn't losing data cause it to use too little of the bandwidth available)(Steven Low, Caltech).

http://www.technologyreview.com/web/24605/