The Internet

History & Current Applications

Connecting computers to other computers

- Share data
- Join computing forces
- Ensure resiliency

The idea of connecting a computer to other computers is as old as the computer itself.

Connecting computers allows transferring data without having to physically move storage devices, it aggregates the compute power of individual computers, and it allows recovering from the failure of an individual computer.

Types of Communication

- *Synchronous*:
  - sender and receiver active at the same time (e.g. phone call)
- *Asynchronous*:
  - receiver active sometime after sender has sent a message (e.g., postcard)
- *Broadcast*:
  - one sender, many receivers (e.g. TV)
- *Multi-cast*:
  - one sender, many receivers, but not everyone is a receiver (e.g. Magazine subscription)
- *Point-to-point*:
  - one sender communicating with one specific receiver (Uni-cast)

Computers communicate using many types of communication.

Internet history

- Computers connected in (local) network
- Connecting networks, networks of networks, interconnected networks, *internet*
- Examples of early networks of networks
  - ARPANET, Mark I, CYCLADES, Telenet

4

First, computers were connected into local networks.

Then, networks were connected to each other

Packet switching

- Divide messages into packets
- Per-packet routing
  - Packets store destination address, message id, and index within message
- Store and forward networking
  - Servers receive packets
  - Store them temporarily
  - Forward them to other servers

5

One breakthrough was the idea of splitting up message into packets, and routing packets individually. Routing is done based on destination but also on dynamic load, i.e. which route to destination is least congested. Communication path is also split up into hops—there is no direct line from sender to receiver.

Packet switching

- Divide messages into packets
- Per-packet routing
- Store and forward networking
- Better than circuit switching (telephony)
  - Circuit switching implies rigid routing
  - Packet switching: better bandwidth utilization and response times on resource-limited networks
- Better than message switching
  - Message switching: routing entire messages
  - Failure requires resending entire message

6

Packet-switching is a better strategy than circuit switching, traditionally used in telephony, where one establishes a communication path between sender a receiver.

Packet switching is also a better solution than message switching. A problem when sending a large message requires resending the entire message. With packets, only the packet that didn't make it needs to be resent.

Early years

- Transmission Control Protocol / Internet Protocol
  - 1982
  - Internet : world-wide web of fully interconnected TCP/IP networks
- ARPANET (military) and NSFNET (National Science Foundation
  - 1980's
  - Access to supercomputers

7

Communication on network of networks was standardized through TCP/IP. All communication would use this protocol. The term internet is coined. TCP/IP was another breakthrough—standardization of communication will internet growth.

In the early to mid 80's only national agencies had the internet. They were providing access to supercomputers for education and research.

"We set up a telephone connection between us and the guys at SRI ...", Kleinrock ... said in an interview: "We typed the L and we asked on the phone,

"Do you see the L?"
"Yes, we see the L," came the response.
We typed the O, and we asked, "Do you see the O."
"Yes, we see the O."
Then we typed the G, and the system crashed ...

Yet a revolution had begun" ....[10]

"Roads and Crossroads of Internet History" by Gregory Gromov. 1995

Commercial traffic

- Internet Service Providers (ISPs) emerge in late 80's and 90's
- ARPANET and NSFNET decommissioned

Commercial viability of internet quickly becomes obvious. No need for separate government networks.

Explosive growth of the internet

- In 1993, Internet carried 1% of information flowing through 2-way telecommunication
- By 2000, 51%
- By 2007, 97%

10

Web Addressing (IPv4)

- Everything connected to the internet gets a web address.

- In IPv4 (Internet Protocol version 4), address is a 4-byte number, such as 128.10.10.250, or 128.10.10.41.

- Example:
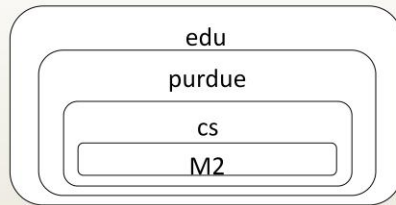
| 25.13.109.72 | 25.13.109.73 | 25.13.109.74 | 25.13.109.75 |

Web Addressing (IPv4)

```
┌──────────────────────────────────────┐
│                 edu                  │
│  ┌────────────────────────────────┐  │
│  │             purdue             │  │
│  │  ┌──────────────────────────┐  │  │
│  │  │            cs            │  │  │
│  │  │  ┌────────────────────┐  │  │  │
│  │  │  │         M2         │  │  │  │
│  │  │  └────────────────────┘  │  │  │
│  │  └──────────────────────────┘  │  │
│  └────────────────────────────────┘  │
└──────────────────────────────────────┘
```

Example: m2.cs.purdue.edu
- **M2**
  - the computer name
- **cs**
  - a member of the **cs** domain, which has many computers
- **purdue**
  - a member of the **purdue** domain, which has many departments
- **edu**
  - A member of the **edu** domain, which has many schools

But it would be unpleasant if everyone had to use and remember those numbers. Instead, a structured name is used, such as M2.cs.purdue.edu, the desktop of an office PC at Purdue CS (128.10.10.41).

Web Addressing (IPv4), cont.

- DNS servers are computers that translate textual addresses to IP addresses.

www.google.com → [server] → 74.125.225.81
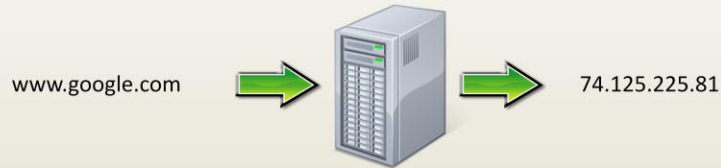
- However, a 4-byte address structure is no longer sufficient. The world has run out of IP addresses. IPv6, gradually deployed in parallel, uses 16-byte addresses, such as 2531:0db8:85a3:0000:0000:8a2e:0370:7334

DNS servers are computers that translate textual addresses (e.g. www.google.com) to IP addresses (74.125.225.81).

If the DNS server does not know, it asks a root server (13 across the globe). If they don't know they ask the authoritative name server of the domain.

iClicker question

- How many internet addresses can be defined with 4 bytes?

A. 4x8

B. 8x8x8x8

C. 256x256x256x256

D. 2^4

E. None of the above

14

Internet applications

- Information repository
- Telecommunication
- Health care
- Education
- Commerce
- Stock exchange trading
- Gaming
- Social networking

The internet is used virtually in all aspects of life. Here are some of the most popular applications.

Internet—information repository

- Server side
  - Web pages stored on web servers
  - Web pages are defined using specialized languages
    - e.g. HTML (Hyper Text Markup Language)
    - Content (e.g. text, images, audio sequences, videos) + hyperlinks (links to other webpages)
  - Connected web pages form a graph
    - 1 trillion (1,000,000,000,000, or one million millions) web pages by 2008

A web server is a computer connected to the internet and who has the mission of answering data requests. A web page is a file encoded in a special language that allows inserting multi media content as well as links to other pages.

Internet—information repository

- Client side
  - Web browser—a software application
  - Allows displaying content, including multimedia
  - Extendable through plugins
  - Allows following hyperlinks on webpages to navigate or surf the web

17

Users can surf the web as long as they are connected to the internet, and as long as they have a web browser.

A web browser is a software tool with an open architecture, which allows adding capability a posteriori, after it was developed. For example we could invent our own video file format, e.g. with extension .177, and then implement a library that teaches Internet Explorer how to play back .177 files. Such a library is called a plugin.

Search engines

- With a web browser you can only go to known web addresses
- Search engines allow finding addresses of web pages that might have the content needed
- How to find web pages in a graph with 1 trillion nodes?

Page rank algorithm

- Named after Larry Page, used by Google
- A link analysis algorithm
- Assigns weights to each web page to measure its relative importance
- Page receives higher weight when
  - Many pages link to it
  - A page with a high weight points to it

Page Rank Algorithm example

A
3.3%

B
38.4%

C
34.3%

D
3.9%

E
8.1%

F
3.9%

1.6%

1.6%

1.6%

1.6%

1.6%
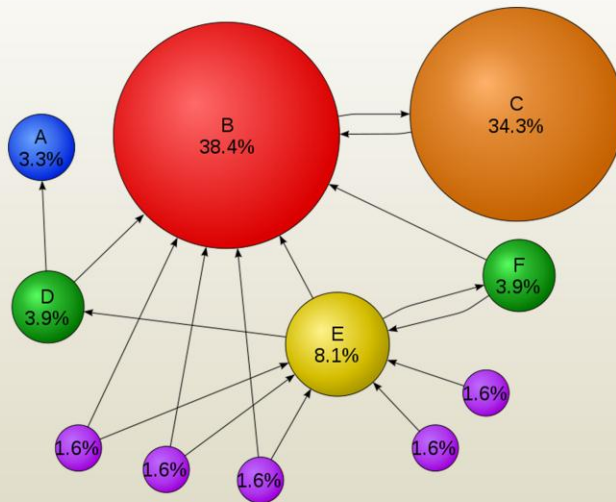
20

Page C has a higher rank than page E because it has a link from B, which is an important page, since it has many pages linking to it.

Page Rank is vulnerable to manipulation—creating fake web pages to point at a page whose rank one wants to boost.

Internet—information repository

- So far we have the following ingredients
  - Web page—*encodes content*
  - Web server—*stores content and sends it when requested*
  - Web browsers—*allow users to go to web addresses*
  - Search engines—*allow users to find information on the web*
- Who provides the content?

21

Wikipedia

- Multilingual, web-based, free-content encyclopedia
- Openly editable model
  - Written collaboratively by largely anonymous internet volunteers
  - Results in information of remarkably good quality
- 4 million articles in English

Internet for telecommunication

- Email, chat, blog
  - Text, images, and videos, asynchronously
- Voice and video synchronously
  - Voice over IP, video over IP
  - New protocols developed for live streaming
  - Requires (residential) broadband

Voice over IP and video over IP are communication protocols and technologies involved in transmitting audio and video over a packet-switching network like the internet.

Healthcare

- Provides medical knowledge
  - used by physicians *and* patients
- Remote surgery
  - 2001 Operation Lindbergh, cholecystectomy surgeon in NY and patient in Strasbourg, France

24

Education

- Knowledge repository
  - Is Wikipedia an acceptable reference?
- Online learning
  - Synchronous: remote student attend lecture in real time
  - Asynchronous: students access prerecorded lectures
  - This course has an online section
  - Benefits: scheduling flexibility, reaching remote students, guest lectures by remote instructors

25

Because of the fact that anyone can contribute to Wikipedia, and because articles are not necessarily reviewed and approved by an expert in the field, Wikipedia is rejected by some as a scholarly reference. However, a good Wikipedia article does have traditional references, which could be used instead.

iClicker Question

In the distance learning system described, remote students are displayed in a virtual extension of the physical classroom (see photo). Should a remote student always be assigned to the same seat? Should nearby students be able to talk to teach other privately?

A. Yes, they'll become friends and work together, improving learning outcomes.
B. No, they'll talk to each other and not pay attention to class.
C. Yes, but remote students should choose where to sit.
D. A and C.
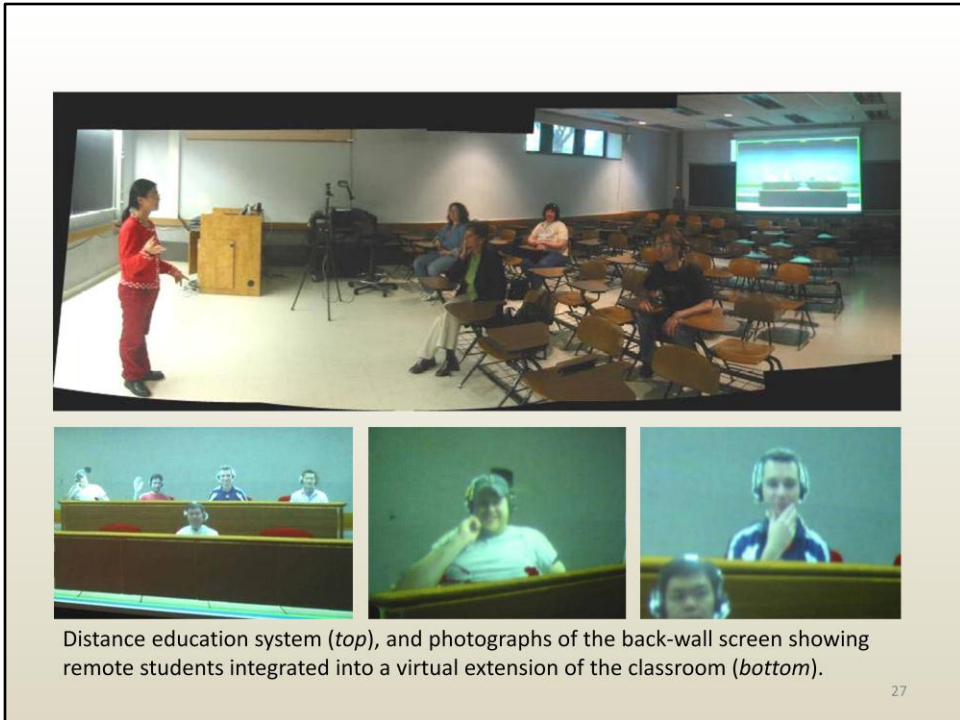E. No, that's why they take the remote course to begin with.

Reference

R. Dondera, C. Jia, V. Popescu, C. Nita-Rotaru, M. Dark, C. York, "Virtual Classroom Extension for Effective Distance Education", IEEE Computer Graphics and Applications, 2008.

Abstract

We present the design, implementation, and initial results of a system for remote lecture attendance based on extending on-campus classrooms to accommodate remotely located students. A remote student is modeled with a real-time video sprite. The sprites are integrated into a geometric model that provides a virtual extension of the classroom. The virtual extension is rendered and projected onto the back wall of the classroom. The remote students are displayed at a natural location within the field of view of the instructor, who can conveniently get a sense of their body language and of their facial expression. The system has been deployed in a first classroom and a pilot study indicates that the system promises to deliver quality education remotely. The system relies exclusively on commodity components, therefore it can be deployed in any classroom to allow any course to offer distance education seats.

Distance education system (*top*), and photographs of the back-wall screen showing remote students integrated into a virtual extension of the classroom (*bottom*).

Reference

R. Dondera, C. Jia, V. Popescu, C. Nita-Rotaru, M. Dark, C. York, "Virtual Classroom Extension for Effective Distance Education", IEEE Computer Graphics and Applications, 2008.

Abstract

We present the design, implementation, and initial results of a system for remote lecture attendance based on extending on-campus classrooms to accommodate remotely located students. A remote student is modeled with a real-time video sprite. The sprites are integrated into a geometric model that provides a virtual extension of the classroom. The virtual extension is rendered and projected onto the back wall of the classroom. The remote students are displayed at a natural location within the field of view of the instructor, who can conveniently get a sense of their body language and of their facial expression. The system has been deployed in a first classroom and a pilot study indicates that the system promises to deliver quality education remotely. The system relies exclusively on commodity components, therefore it can be deployed in any classroom to allow any course to offer distance education seats.

Commerce

- Online stores
  - Lower costs: no physical stores, no state sales taxes*
  - Early on only some products were bought online (e.g. books, music), not anymore
  - Cyber Monday—Monday following Thanksgiving
- Online auctions
  - System of seller rating by costumers
  - Online bidding over days
  - Anyone can own a "store"
  - eBay

28

*
"Amazon.com will begin collecting Indiana's 7 percent sales tax from customers in the state in 2014, under an agreement announced Monday."

"Gov. Mitch Daniels' office said Indiana will become the fourth state with such a tax collection agreement with Seattle-based Amazon. It follows a lawsuit by Indianapolis-based shopping mall owner Simon Property Group against the state over the issue and a lobbying push on state legislators by traditional retailers to end what they call an unfair price advantage for online retailers."

Source: huffingtonpost.com

Stock Exchange Trading

- Anyone can trade securities from anywhere
- Computer simply executes trade
- Computer decides which trades to make according to sophisticated algorithms
- High frequency trading
  - Investment position held only for very brief periods of time, from seconds to hours
  - Might have contributed to the 2010 Flash Crash

29

Dow Jones Flash Crash of May 6 2010
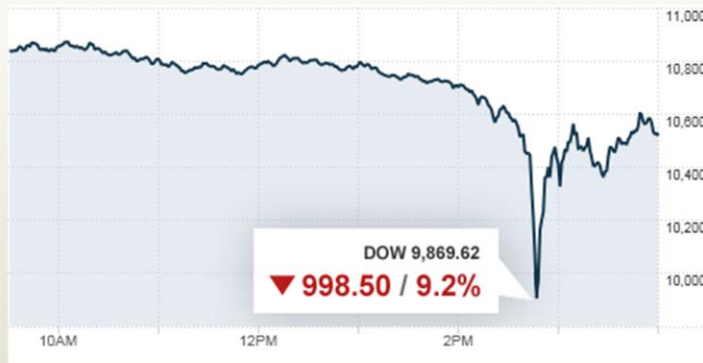
- At 2:42pm the Dow Jones starting falling rapidly
- Dropped more than 600 points in 5 minutes
- Biggest intra-day decline, second largest point swing
- Losses recovered within minutes

30

Gaming

- MMORPG (Massively Multiplayer Online Role-Playing Game)
- Second Life
  - Players interact with each other through avatars
  - Avatar (here): An electronic image that represents and is manipulated by a player of a computer game (M-Webster)
  - 80,000 players online at the same time
- Half-life, World of Warcraft

31

Social networking

- Facebook
  - 1 trillion page views in June 2011
  - 1b users soon
  - Filed for Initial Public Offering on Feb 1, 2012

Total active users[N 1]

Date	Users (in millions)	Days later	Monthly growth[N 2]
August 26, 2008	100[30]	1,665	178.38%
April 8, 2009	200[31]	225	13.33%
September 15, 2009	300[32]	160	9.38%
February 5, 2010	400[33]	143	6.99%
July 21, 2010	500[34]	166	4.52%
January 5, 2011	600[35][N 3]	168	3.57%
May 30, 2011	700[36]	145	3.45%
September 22, 2011	800[37]	115	3.73%

32

Social networking

- Facebook
- Twitter
- What is the innovation?
  - Everyone can create their own webpage trivially (no HTML, no Java, no Flash programming required)
  - Exponential growth through friends of friends
  - Deployed on all devices, including smartphones

33