CM0133 Internet Computing

Computer Networks and the Internet

Lecturer and Course Info

Dr. Florian Twaroch

Rm S 3.07

Email: f.a.twaroch@cs.cf.ac.uk

Office hours: Monday 9:15 – 11:15
or individual agreement

Online Course Materials

http://users.cs.cf.ac.uk/F.A.Twaroch/teach.html

Course Overview

- 24 lectures (each 1 hour)
  - Thursday, Rm S/1.25, 15:10 – 16:00
  - Friday, Rm T/2.09, 11.10 – 12:00
- 10 labs
  - Monday C / 2.10 11:10 – 12:-00 (CS) 15:10 – 16:00 (IS), 16:10 – 17:00 (IS)
  - Tuesday C/2.10 12:10 – 13:00 (CS), 13:10 – 14:00 (CS), 14.10 – 15:00 (CS)
  - Friday 9:10 – 10:00 (IS), 10:10 – 11:00 (IS)
- 6 Tutorials
  - 2 Tutorials
  - Remaining 4 are used to switch lecture slots on demand

Examination/Assessment Information

- Written Exam (2 hours) – 60%
- Two Courseworks – 40%
  - FTP Assignment

    Handout Date: 12.02.2010
    HandIn Date: 01.03.2010, 8.30am – 9.30am
  - Web Site Assignment

    Hand out Date: 11.03.2010
    Hand In Date: 26.04.2010, 8.30am - 9.30am

Labs

- **Multimedia Lab C2.10:** Macintosh PowerPC to be used as the primary machines for practical work
- The School's Unix network will be required for storage and running of PHP Scripts. The setting up and programming required for advanced WWW pages can only be facilitated in this manner
- If you want to carry out this work on your on machine it is your own responsibility to create the necessary infrastructure. I will give you information in a tutorial on how to setup your own webserver.
- MySQL access will be granted via a Unix Server Students will therefore require access to file space etc on the Unix network
- Students can access this network remotely via several means

Question

- What is the Internet ?
- Pair up in two – discuss two minutes
- Note keywords!

Military Motivation of an Information Network

- A system capable of supporting (text-based) communication between dissimilar computers and networks
- Minimal central administration to make the system robust
- Flexible in allowing the addition or removal of nodes to or from the network
- The network should be decentralised and therefore not vulnerable to the loss of any individual computer site

Annoyed Scientist's Motivation of an Information Network

"For each of these three terminals, I had three different sets of user commands. So if I was talking online with someone at S.D.C. and I wanted to talk to someone I knew at Berkeley or M.I.T. about this, I had to get up from the S.D.C. terminal, go over and log into the other terminal and get in touch with them [...]".

I said, it's obvious what to do (But I don't want to do it): If you have these three terminals, there ought to be one terminal that goes anywhere you want to go where you have interactive computing. That idea is the ARPAnet."

Robert W. Taylor, co-writer with Licklider of "The Computer as a Communications Device", in an interview with the New York Times (http://partners.nytimes.com/library/tech/99/12/biztech/articles/122099outlook-bobb.html)

What Is The Internet?

- It is complex and therefore there is no single one sentence description. You might say today I learned about:
  - A network of networks
  - The information super highway
  - The semantic web
  - The world wide web
  - Email
  - Protocols and Networking
  - Services

What is the Internet ?

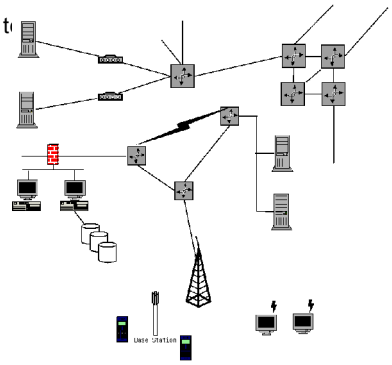
- A worldwide computer network
- connecting millions of devices (e.g. phones,laptops, PCs), two types server and clients
- serves to transmit information via Webpages and Email
- Non traditional end systems: PDAs, TVs, automobiles, digital cameras and toasters are being connected to the Internet:

  CHECK: http://news.bbc.co.uk/1/hi/sci/tech/1264205.stm

  and http://www.savetz.com/yic/YIC11FI_6.html

Example for an Internet Network

- Click t



Communication in the Network

- **End systems** are connected by **communication links,** which can be made from many different physical media (coaxial cable, copper wire, fiber optics, radio spectrum).
- Different links transmit at different transmission rates, called **bandwidth**. The bandwidth is currently measured in bits/second.
  - Assignment: Find out the bandwidth of your home broadband.
- Routers connect different end systems by forwarding **packets**, i.e. chunks of information from their incoming communication links to their outgoing communication links.

'The Internet'

- **'The Internet'** is the public network (global network of networks)
- **Intranets** are private networks as run by the government or corporate networks, whose hosts can not exchange messages with hosts outside the private networks. Except those messages passing through firewalls, which control the access to outside resources.

Routes

- A **route** is the **path** a packet takes through the network from the sending end system to the receiving end system.
- Today's Internet is based on the idea of **packet switching.** Path or parts of paths can be shared by multiple communicating end systems.
- **Packet switching** was one of the big innovations that lead to the Internet and succeeded the traditional circuit networks used for telephones that have dedicated path.

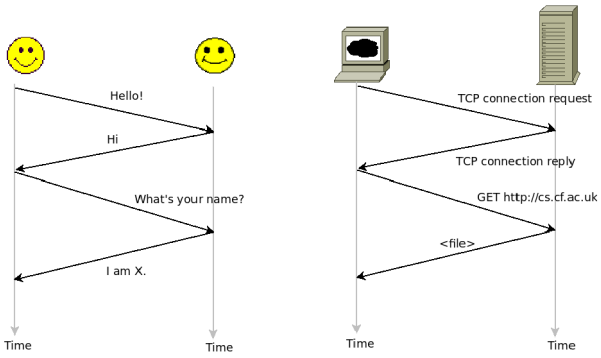
Internet Service Providers (ISP)

- An ISP is a network of routers and communication links that allows the communication between internet users.
- Examples: AOL, MSN, BT, Cardiff University, …
- Different bandwidth are available, 56Kbps dial-up modem, cable modem, DSL, high speed LAN, WLAN
- Lower tier ISP are connected to upper tier ISP. Upper tier ISP are often national or even internationally organized and consist of high speed routers.
- ISP networks are managed independently, run the **IP protocol** and conform to certain naming and address conventions.

What is a Protocol?

- Say “Hello” to your neighbour and then ask for the name.
- Humans use protocols all the time, e.g. introducing, ordering in a restaurant, etc.
- Two communicating entities have to run the same protocol in order to achieve the task (same language, shared vocabulary, same structure, etc.)
- A question answer protocol controls the transmission and receipt of messages, and a set of conventional actions taken when these messages are sent and received.
- Network protocols are similar to human protocols, the entities are hardware or software components.

A Network Protocol



http://www.feeblenmds-gifs.com/smiley-faces.html, figure adapted from Kurose and Ross (2003, p 6)

Protocols

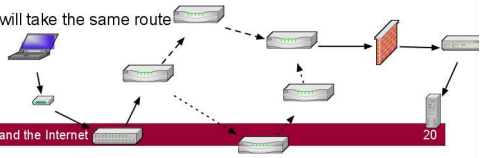
- **The Internet makes extensive use of protocols, we will look just a three :**
  - **IP – Internet Protocol**
  - **TCP – Transmission Control Protocol**
  - **HTTP – Hypertext Transfer Protocol**
- **There are many others, if you are interested in more details consult the literature links on the last slide and the Web.**

Standards for the Internet - IETF

- Protocols such TCP, IP, HTTP and SMTP have been defined through requests for comments (RFCs). There are more than 3000 RFCs (2003).
- The Internet Standards are developed by the **Internet Engineering Task Force** (IETF). These standards have made the technical development of the Internet possible.
- The procedure: A general request for comments serves to resolve architecture problems which can lead to a RFC.

IP Protocol

- Best-effort protocol
- Data is transmitted in packets
  - Header: From address, to address, packet size, checksum, …
  - Data
- IPv4 - 32bit addresses - 4,294,967,296 addresses
- IPv6 - 128bit addresses - Lots of addresses (2^{128})
- At each hop the device decides how to send on the packet
- Packets are transported in a series of hops
- Not all packets will take the same route



Converting Domain Names

- Domain names are for humans to read.
- The Internet actually uses numbers called **IP addresses** to describe network addresses.
- The Domain Name System (DNS) – resolves IP addresses into easily recognizable names
- For example:

  131.251.169.1 = cs.cf.ac.uk
- A domain name and its IP address refer to the same Web server.

Domain Name Syntax

- Domain names are read right to left, from general to more specific locations
- For example, *www.xyz.com* can be interpreted as follows:
  - com — commercial site top-level domain
  - xyz — registered company domain name
  - www — host name (it is a convention to name web server hosts "www" which stands for "world wide web")

Slide adapted from Marti Hearst's course: http://courses.ischool.berkeley.edu/i141/f07/lectures/se-course-intro.ppt

Domain Name Example

- Click to add an outline

http://www.yourplacenames.com

Server (host) name	Registered company domain name	Domain category (top-level domain)

Domain names are part of URLs (Uniform Resource Locator), used in web pages.

Slide adapted from Marti Hearst's course: http://courses.ischool.berkeley.edu/i141/f07/lectures/se-course-intro.ppt

Top Level Domains

- com, biz, cc — commercial or company sites
- edu — educational institutions, typically universities
- org — organizations; originally meant for clubs, associations and nonprofit groups
- net — network sites, including ISPs
- int — international organizations (rarely used)
- .uk – country domain
  - ac.uk - academic
  - nhs.uk
  - parliament.uk
  - and many more

Many other top level domains are available

Slide adapted from Marti Hearst's course: http://courses.ischool.berkeley.edu/i141/f07/lectures/se-course-intro.ppt

TCP – Transmission Control Protocol

- RFC 793
- TCP provides to an application reliable transport, flow control, and congestion control.
  - Reliable: A application can rely on the connection to deliver all of its data without error and in proper order.
  - Flow Control: Neither side of the connection overwhelms the other side by sending to many packets to fast
  - Congestion Control: If routers get congested, end systems are forced to decrease their send out rates for packets. End systems do not receive anymore acknowledgments for packets they have sent out and are so alerted about a network congestion.
- An application is oblivious about how TCP guarantees reliability, flow and congestion control.
- TCP is a connection-oriented service and used by many well known services such as Telnet, SMTP, FTP and HTTP.

TCP based Internet Services

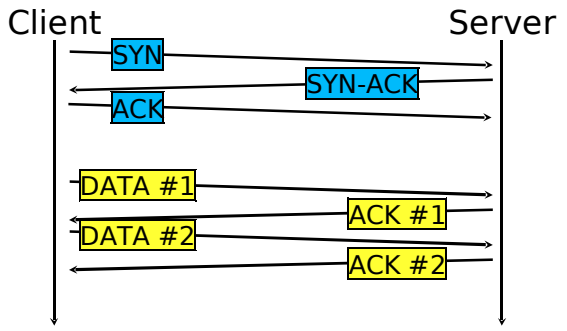
- Email & MIME (Multipurpose Internet Mail Extensions)
  text (text/html), image, video, etc.
- Telnet & ssh
- FTP – File Transfer Protocol
- Gopher
- IRC – Internet Relay Chat
- Newsgroups
- WWW – World Wide Web

  HTTP (Hypertext transfer protocol) uses a Question-Answer-Scheme, i.e. a browser sends a request und gets a response from a server. Note the serverdoes not send out anything without a request.

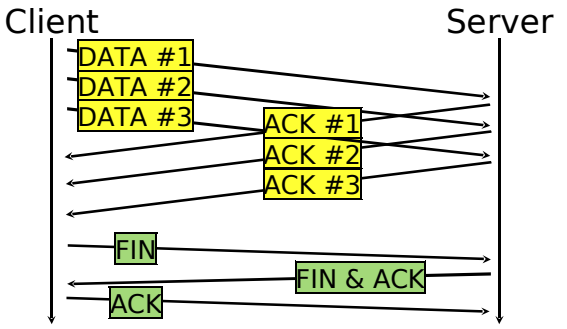
Transmission Control Protocol (TCP)

- Closely tied to the Internet Protocol => TCP/IP
- Guaranteed transfer protocol
- Uses sequence numbers to identify packets
- Receiver sends an acknowledgement for each packet
- Lost packets can be identified and re-sent
- Each packet is routed individually
  - Client must re-order packets before re-assembling the data

Transmission Control Protocol (TCP)



Transmission Control Protocol (TCP)



HyperText Transport Protocol (HTTP)

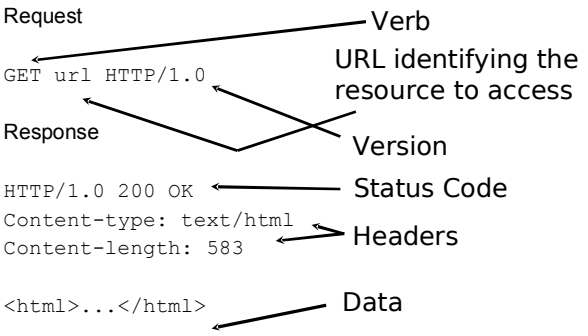
- Request - Response based
- Pure-text protocol
- Stateless protocol
- Request: Request line (, Headers), Empty line (, data)
- Response: Response code, Headers, Empty line, data

Response Codes (RFC 2616)

Some examples:

- 301 Moved Permanently
- 400 Bad Request: The request could not be understood by the server due to malformed syntax. The client SHOULD NOT repeat the request without modifications.
- 401 Unauthorized: The request requires user authentication. The response MUST include a WWW-Authenticate header field (section 14.47) containing a challenge applicable to the requested resource.

HyperText Transport Protocol (HTTP)



HTTP Verbs

- Safe Verbs
  - GET
  - HEAD
- Unsafe Verbs
  - POST
  - PUT
  - DELETE

The default verb for accessing any webpage

The default verb for sending larger amounts of data to a webpage

Connectionless Services

- Contrary a connectionless service does not involve any handshaking between applications.
- It is faster than a connection oriented service because it lacks communication overheads
- It is unreliable compared to a connection oriented service.
- UDP – User Data Gram Protocol (RFC 768) is used by applocations such as Internet phone and video conferencing

Connection oriented vs. connectionless Service

- *A **Connection Oriented Reliable Service** guarantees that data transmitted from a sender to a receiver will eventually be delivered to the receiver in order and in its entirety. ([1, page 5]).*
- As opposed to a **connectionless unreliable service**
- The Internet supports both
- But it is not known how long it takes data to be delivered from a receiver to a sender.

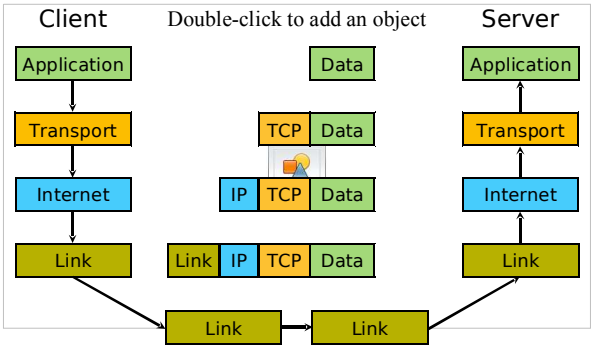
The Internet
A Service Oriented Description

- The Internet allows to develop **distributed applications** running on different end systems exchanging data.
  - Remote login
  - Email
  - Web Surfing
  - Instant Messaging
  - Audio and Video Streaming
  - Internet Telephony
  - Distributed games
  - Peer-to-peer file sharing (P2P)
  - … and many more

Clients and Servers

- End systems have been generally termed hosts as they host application software. such as web server applications or email readers applications.
- A client program (e.g. browser, email client) receives and sends requests to a server program (web server, email server). Usually they are located on different hosts.
- By definition client/server Internet applications are distributed applications.
- In peer-to-peer applications (Napster, Gnutella) the user end system acts as both a client and a server. You can also configure you PC or laptop to run a web server and a browser on the same machine. Network protocols are similar to human protocols.
- Assignment: Check the SETI@Home Project!

Data Transport



Telnet

- TCP/IP client
- Uses
  - Remote login - Historical, do not use as username/password transmitted as clear text
  - Connection testing
    - Any protocol that builds on TCP/IP and is text-based can be tested
    - Telnet client acts as the client for the protocol

Secure SHell

- A secure alternative to Telnet
  - ssh -l username hostname
  - ssh -l scxxxxx lapis.cs.cf.ac.uk
  - Windows: Putty
- Can also be used to transfer files securely
  - scp filename username@host:remotename
  - Windows: WinSCP

Browser

- The client we will focus on is the browser.
- But software will reside on either side the client (browser) and the server (web server).
- We will learn how client and server can interact to create interesting and dynamic web applications.

History of the Internet

- **1945 –** Vannevar Bush first proposed the basics of hypertext. Motivation for the work of Tim Berners Lee.
- **1962 -** J.C.R. Licklider, appointed head of ARPA. Licklider was a visionary who pioneered the idea of networked computers for collaborated working. Funded many projects related to his idea.
- **1969 -** forerunner to the internet, the ARPAnet was developed. United states military wanted a robust command and control center that could withstand nuclear attack. Leads to TCP/IP.
- **1972 –** Advanced Research Projects Agency (ARPA) was renamed to Defense Advanced Research Project Agency (DARPA)
- **1974 -** Publication of TCP/IP by Vinton Cerf and Robert Kahn at Stanford University
- **1978** - First international packet-switched network (Canada, UK)

Continued History of the Internet

- **1983** - ARPANET switches to using TCP/IP

  JANET, based on the SERCnet X. 25 research network, went live. JANET is a network to support education and research in the UK.
- **1985** - NSFNET a 56 kBit/sec university network
- **1990 -** Tim Berners-Lee and CERN in Geneva implements a hypertext system, invents HTML, HTTP and URLs, for the international high-energy physics community
- **1992 –** World-Wide Web released by CERN.

History

- **1988** - Opening of the NSFNET to commercial partners
- **1991** - Tim Berners Lee publishes the HTML standard, which underpins the World Wide Web
- **1993 -** Marc Andreessen and NCSA and the University of Illinois develops a graphical user interface to the WWW, called "Mosaic for X". First graphical Web browser (rendering engine behind Tim Berners-Lee is currently the Director of the World Wide Web Consortium (W3C).
- **1996 to 2001** - Massive growth period (.com Bubble)
- **2005** - Google maps sparks the development of rich Web applications
- **2003-2007** Web 2.0, blogging, social networking
- **Present – Web** 3.0 Tim Berners-Lee inspired Semantic Web. The meaningful web, human and MACHINE consumption of web content.

Trends & Future of the Web

- Commercialisation of the Web and Advertising will grow
- Social media and Networking will further develop
- Webinars allow Virtual meetings and are all purpose online conference tools. Further developments in this area are expected.
- Cloud computing: Is that trend that applications and data from your desktop move to web servers (“The Cloud”). Users are not aware of were the data is actually stored and processed, e.g. Dropbox, Google Docs
- Web 3.0

http://www.mastermewmedia.org/top-internet-trends-2010-a-guide-to-the-best-predictions-from-the-web-part-1/

Trends & Statistics

- Plethora of clients and platforms to support
- Use of Browsers (e.g. Dec 2009 statistics – w3schools):

2009	IE 8	IE 7	IE 6	Firefox	Chrome	Safari	Opera
Dec	13.5%	12.8 %	10.9 %	46.4 %	9.8%	3.6 %	2.3 %

http://www.w3schools.com/browsers/browsers_stats.asp

- Other web sites and statistics give a different picture. The Internet Explorer is predominant with 80% of use on various pages! Therefore investigate which clients your users are going to utilize but also be careful where statistical information comes from.

Trends & Statistics

- In 2005: Over 200 Million Hosts (avg) and 150 Million WWW sites
- In 2007: Around 1.09 Billion use the internet
- In 2005: Over 1 Billion home pages
- 15% World Population Online, 46% English speaking world
- In 2005: 80% of the Web users are between the ages of 16 and 44
- In 2005: 57% of Web users have a college degree, 28% have graduate degrees
- In 2005: 80% business users surf the web
- In 2007: Browse the internet from your TV set and Mobile phone
- In 2007: Broadband (fast) internet for all
- In 2007: More than half of all email is spam (junk)

INTERNET USAGE STATISTICS
The Internet Big Picture
World Internet Users and Population Stats

WORLD INTERNET USAGE AND POPULATION STATISTICS

World Regions	Population (2009 Est.)	Internet Users Dec. 31, 2000	Internet Users Latest Data	Penetration (% Population)	Growth 2000-2009	Users % of Table
Africa	991,002,342	4,514,400	67,371,700	6.8 %	1,392.4 %	3.9 %
Asia	3,808,070,503	114,304,000	738,257,230	19.4 %	545.9 %	42.6 %
Europe	803,850,858	105,096,093	418,029,796	52.0 %	297.8 %	24.1 %
Middle East	202,687,005	3,284,800	57,425,046	28.3 %	1,648.2 %	3.3 %
North America	340,831,831	108,096,800	252,908,000	74.2 %	134.0 %	14.6 %
Latin America/Caribbean	586,662,468	18,068,919	179,031,479	30.5 %	890.8 %	10.3 %
Oceania / Australia	34,700,201	7,620,480	20,970,490	60.4 %	175.2 %	1.2 %
WORLD TOTAL	6,767,805,208	360,985,492	1,733,993,741	25.6 %	380.3 %	100.0 %

NOTES: (1) Internet Usage and World Population Statistics are for September 30, 2009. (2) CLICK on each world region name for detailed regional usage information. (3) Demographic (Population) numbers are based on data from the US Census Bureau . (4) Internet usage information comes from data published by Nielsen Online, by the International Telecommunications Union, by GfK, local Regulators and other reliable sources. (5) For definitions, disclaimer, and navigation help, please refer to the Site Surfing Guide. (6) Information in this site may be cited, giving the due credit to www.internetworldstats.com. Copyright © 2001 - 2009, Miniwatts Marketing Group. All rights reserved worldwide.

Links

- Internet Engineering Task Force (IETF): http://www.ietf.org
- World Wide Web Consortium: http://www.w3.org
- The Association for Computing Machinery (ACM): http://www.acm.org
- Institute of Electrical and Electronics Engineers (IEEE): http://www.ieee.org

Literature

- Kurose, James F. (2003) Computer networking : a top-down approach featuring the Internet / James F. Kurose, Keith W. Ross. Location: Trevithick: Short Loan (1 week loan) TK5105.875.I57.K8

  http://wps.aw.com/aw_kurose_network_5/111/28536/7305312.cw/index.html

  http://www.awl.com/kurose-ross
- Stallings, William (2001). Computer networking with internet protocols and technology. Location: Trevithick TK5105.875.I57.S8

  http://williamstallings.com/CNIP/CNIP1e.html
- Bates, Chris (2006) Web programming : building internet applications / Chris Bates. Location: Trevithick TK5105.888.B2

Summary - Outlook

- Definition of the Internet
- Mini Introduction to Networking
- Protocols
- Services / Client and Servers
- History of the Internet
- Next Time: Your First Website