The Internet
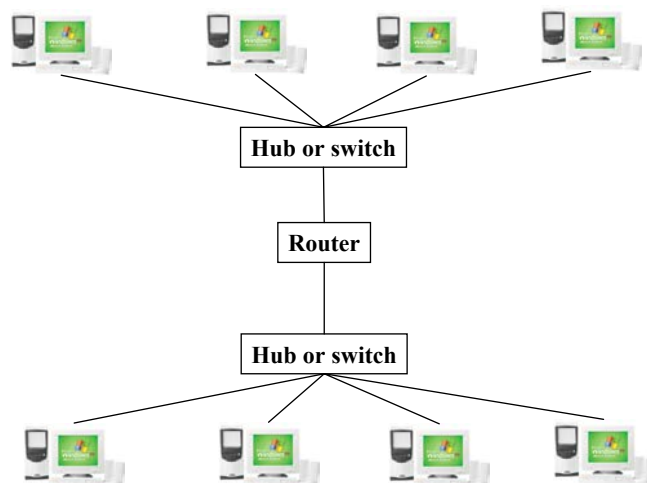
You will learn how the Internet works and some ways of searching for information on it.
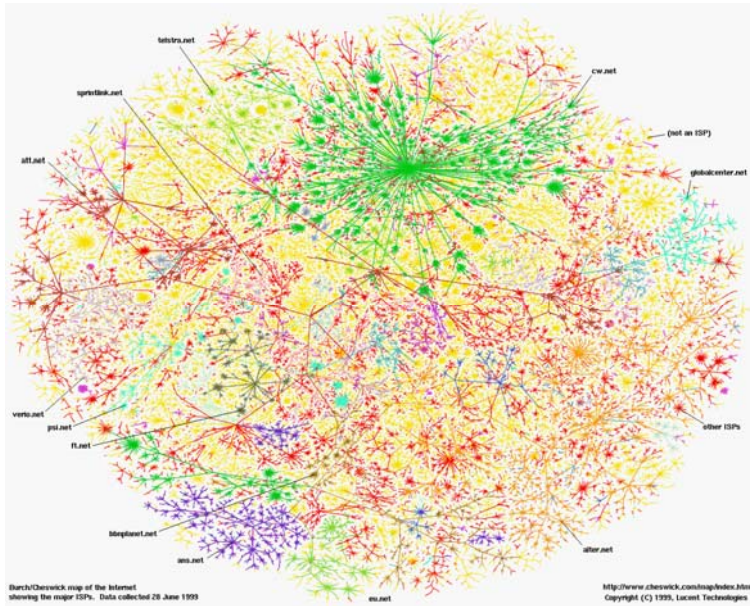

A Network: Computers Which Are Connected

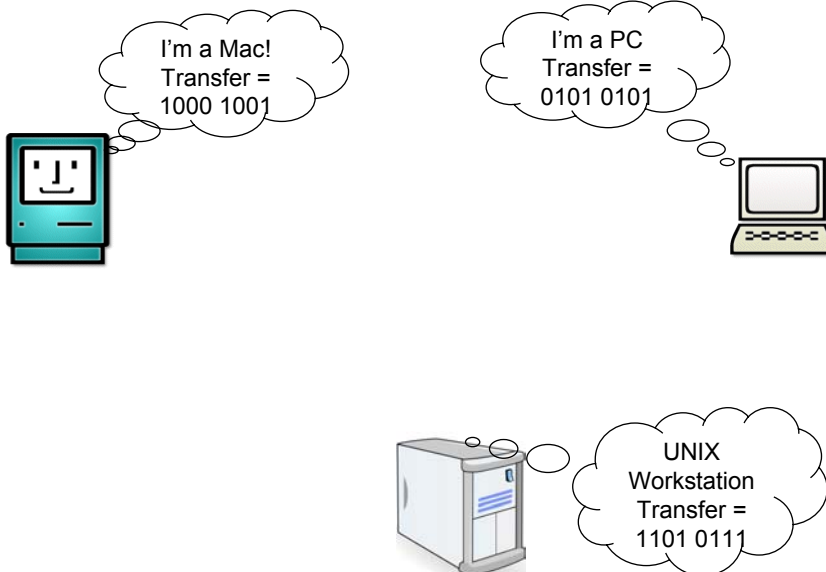Hub or switch

Router

Hub or switch

The Internet: A Gigantic Network of Networks

Computers: Each Architecture Is Different

Protocols: The Common Language Of The Internet

I'm a Mac!
Transfer =
1000 1001

I'm a PC
Transfer =
0101 0101

Transfer = 1111 1111

Transfer = 1111 1111

Transfer = 1111 1111

UNIX
Workstation
Transfer =
1101 0111

James Tam
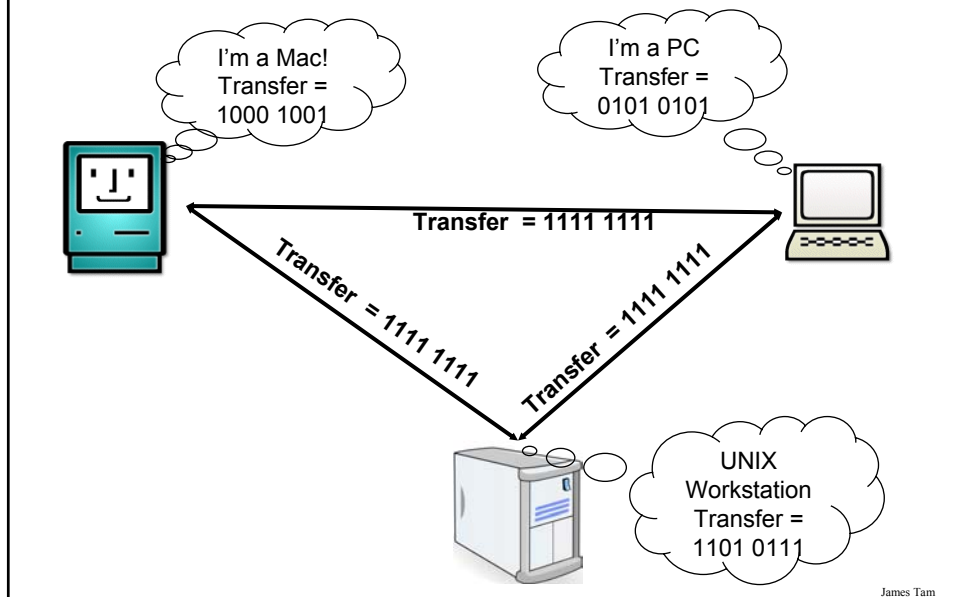
Protocols: The Common Language Of The Internet (2)

- TCP/IP (Transmission control protocol / Internet protocol) is used for the transfer of information on the Internet.
  - TCP: is involved with disassembling/assembling the information being sent.
  - IP: is involved with ensuring the that information reaches it's correct destination.

- Other high-level protocols (easier to use) were developed for specialized tasks
- These protocols still employ TCP/IP but to complete the various tasks the higher-level protocols are the ones employed:
  - Http (Hypertext transfer protocol)
  - SMTP (Simple mail transfer protocol)
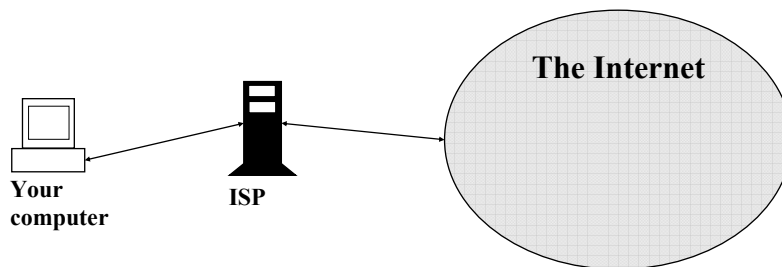 - Telnet
  - FTP (File transfer protocol)

James Tam

Speed Of Your Internet Connection
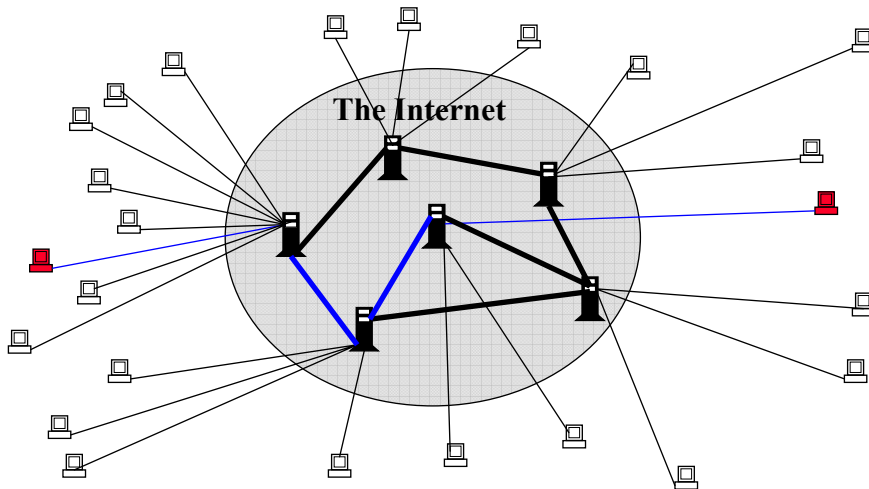
James Tam

Connecting Your Computer To The Internet

- Requires you to sign up with an Internet Service Provider (ISP)

- Types of Internet connections
    1. Old phone dialup connections (narrowband)
    2. Faster phone line connections (broadband)
    3. Cable connections (broadband)
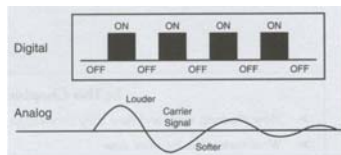   4. Satellite connections (broadband)

James Tam

1. <u>Phone Dialup Connections</u>

- Having your computer call another computer to connect to the Internet:
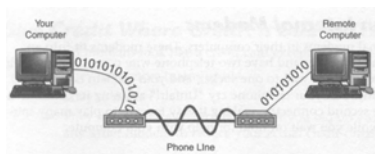

- Problem: Phone lines and computers don't easily mix

From the The Complete Idiot's Guide to
High Speed Internet Connections by Soper

---

1. <u>Phone Dialup Connections (2)</u>

- Having your computer call another computer to connect to the Internet
- Requires a modem (modulator/demodulator)

From the The Complete Idiot's Guide to High Speed Internet Connections by Soper

- Important characteristics:
  - Least expensive method for getting an Internet connection (free for you, for now): http://www.ucalgary.ca/it/node/426)
  - Very slow (300 – 56,000 bps/56Kbps)
  - Ties up a phone line
  - Not usable for many situations / not a continuous connection
  - Recommended usage: low bandwidth applications (text-only)

2. <u>Faster Phone Line Connections (DSL/ADSL)</u>

- DSL (Digital Subscriber Line) / ADSL (Asymmetrical DSL)
- Uses a single wire for voice and data (computer)
- The phone lines are still used to transmit data but a different signal is used from standard dialup (faster)
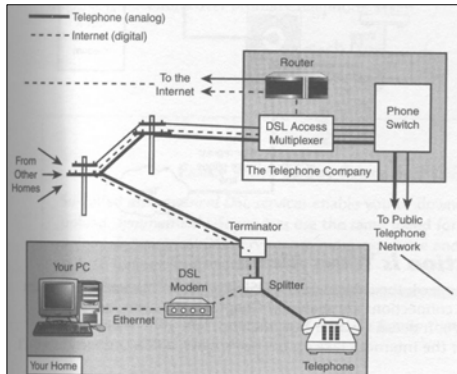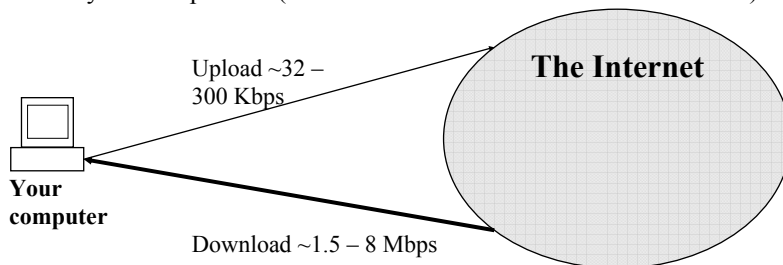

From the The Complete Idiot's Guide to High Speed Internet Connections by Soper

James Tam

## 2. <u>Faster Phone Line Connections (DSL/ADSL)</u>

- Important characteristics:
  - Fast connection: ~ 100's Kbps – several Mbps
  - Doesn't use up a phone line
  - Allows for an always on (continuous) connection

- ADSL (Asymmetrical DSL)
  - You can download information (get information from the internet) faster than you can upload it (send information somewhere to the internet)

Upload ~32 – 300 Kbps

The Internet

Your computer

Download ~1.5 – 8 Mbps

James Tam

3. Cable Connections

- Uses a coaxial (cable TV) connection.
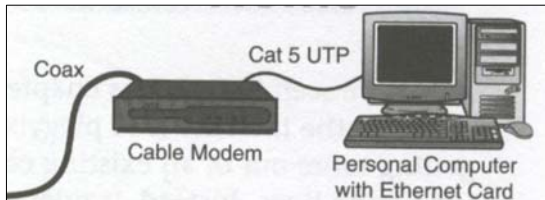
From the The Complete Idiot's Guide to High Speed Internet Connections by Soper

- The Internet connection occurs on the same coaxial cable used for television.
- Important characteristics
  - Download: ~several Megabits (~1 – 6+ Mbps) per second.
  - Upload: approximately several hundred Kilobits (200 – 300 Kbps).
  - Allows for an always on (continuous) connection

# ADSL/DSL Vs. Cable

- **ADSL/DSL**
  - Typical transfer rates are slower
  - Internet connection is not shared
  - ADSL/DSL Internet service is highly restricted
  - (Calgary): The service is targeted more for 'experienced' computer users

- **Cable**
  - Typical transfer rates are faster
  - Internet connection is among shared subscribers
  - Cable Internet service is widely available
  - (Calgary): The service is targeted to the general population.

4. <u>Satellite connections</u>

- Typically if a broadband connection is desired then a standard wired ADSL/DSL or cable connection should be chosen.
- Some remote locations don't allow for either service (no cable wires or DSL/ADSL connections only standard dial-up).
  - In this case the only broadband connection available is via satellite:

Internet provider: Satellite receiver and server

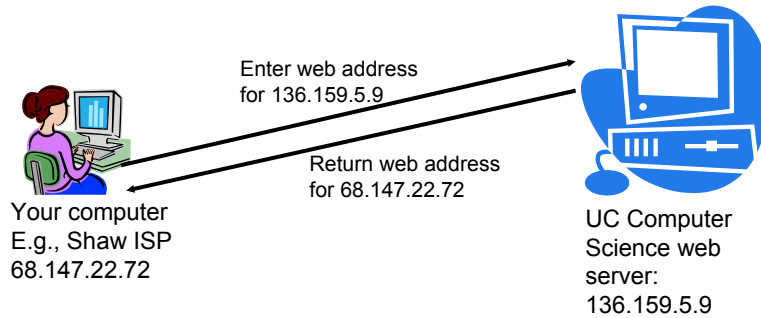Home satellite dish

You
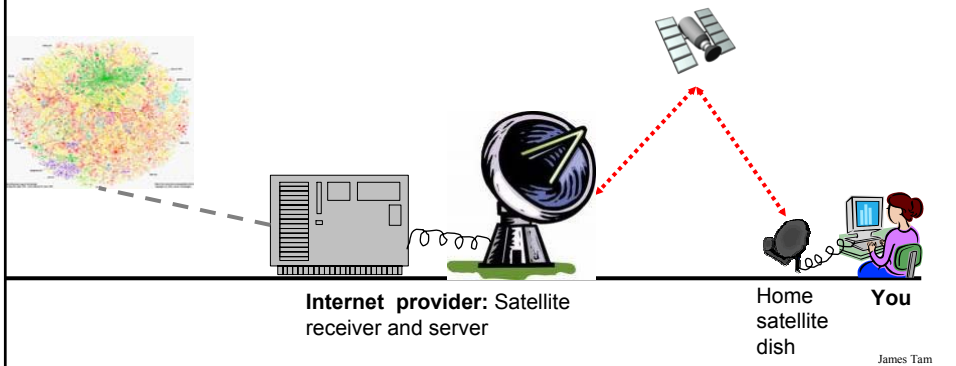
4. <u>Satellite connections (2)</u>

- Satellite connections are faster than dialup but slower than other broadband connections:
  - Upload ~ 100 Kbps
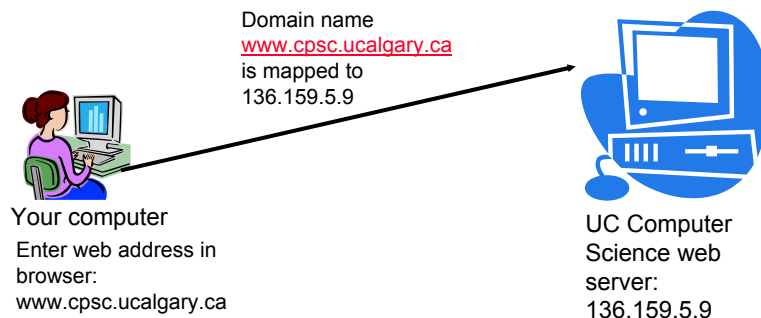 - Download ~ 500 Kbps

IP (Internet Protocol) Addresses

- Needed to properly route information on the Internet.
- Every computer connected to the Internet has an IP Address (IP for short):

Enter web address
for 136.159.5.9

Return web address
for 68.147.22.72

Your computer
E.g., Shaw ISP
68.147.22.72

UC Computer
Science web
server:
136.159.5.9

IP Addresses Are Mapped To Domain Names
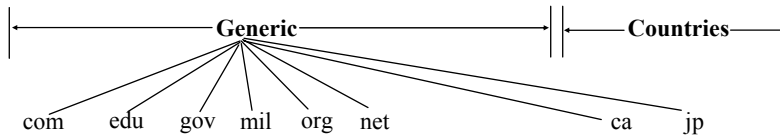
- Because numeric addresses are hard to remember (and may change) domain names are typically used instead

Domain name
www.cpsc.ucalgary.ca
is mapped to
136.159.5.9

Your computer
Enter web address in
browser:
www.cpsc.ucalgary.ca

UC Computer
Science web
server:
136.159.5.9

Types Of Domain Suffixes

• At the end of the web address

Generic domain suffix
•com…commercial
•edu…educational institutes
•gov..the U.S. federal government
•mil…U.S. armed forces
•org…non-profit organizations
•net…Network providers

Country of origin
•ca…Canadian web site
•jp…Japanese web site

Deciphering Email Addresses

• Addresses are similar to web addresses
• Three main parts:
  - **Format:**
  <person's name> @ <host>.<domain>
  - **Example:**
  tamj@cpsc.ucalgary.ca

> **.ca…top level domain (Canada)**
>
> > **ucalgary…domain name for the University of Calgary**
> >
> > > **cpsc…the name of the Computer Science network at the U of C**
> > >
> > > > **tamj..my login name for the CPSC network**

Deciphering IP Addresses

- Consist of 4 numbers

Entire IP address

136.159.5.9

Number
(0 – 255)

- This is the approach originally developed under Internet Protocol version 4 (IPv4).
- With the explosion in the size of the Internet in the 1990's this number of combinations is insufficient.

The Future Addressing Protocol For The Internet

- Internet Protocol version 6 (IPv6).
- Eight 16 bit numbers will be used for Internet addresses.
- Not currently in widespread use yet but it's backward compatible with IPv4.

Assigning IP Addresses

- Different organizations (Registrars) provide Internet Addresses for Internet providers or websites:
  - VeriSign: dot-com (".com"), dot-net (".net")
  - RegistryPro, NeuLevel and Public Interest Registry: dot-biz (".biz"), dot-pro (".pro"), dot-org (".org")

- When your computer gets an Internet connection it's assigned an IP by your service provider which can be: dynamic or static.
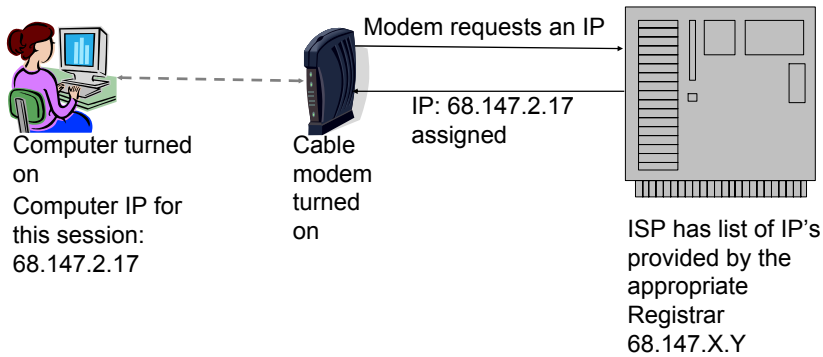
Static Addressing

- When a computer is first connected to a network or the Internet it receives an IP address:

Connection to:
- Network server or ISP's server

A numerical IP address is assigned to the computer

Thereafter this computer will be identified by this address
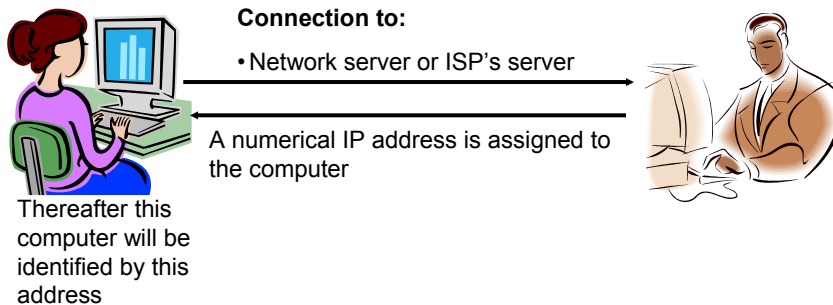
Dynamic Addressing

- Whenever a computer is connected to a network or the Internet it's assigned a IP address from the existing pool of free addresses:

Modem requests an IP

IP: 68.147.2.17 assigned

Computer turned on
Computer IP for this session: 68.147.2.17

Cable modem turned on

ISP has list of IP's provided by the appropriate Registrar 68.147.X.Y

- Consequently the address may or may not be the same as ones that have been previously assigned.

Dynamic Addressing (2)

- **Advantages:**
  - Moving machines around the network to a different (sub)net is trivial. When the machine is connected at it's location it will learn of it's new IP.
  - Resource efficient (not every computer needs it's own IP so fewer IP's are needed).

- **Disadvantages:**
  - Some Internet services (e.g., web, ftp) require a fixed IP address which won't work with dynamic addressing.

Static Addressing

- **Advantages:**
  - Other computers may reliably make connections to this computer using certain protocols (e.g., FTP).

- **Disadvantages:**
  - Machines cannot be simply be physically moved around the network to another (sub)net but must be electronically reconfigured.
  - It's less resource-efficient (every computer must have it's own IP).
  - It's less secure (the computer is "an unmoving target" because once a malicious program has found it, returning to that computer is trivial).

URL's (Uniform Resource Locator)

- Provide a standardize way of describing information.
- Parts of a url (each part is enclosed by angled brackets):

\<protocol> \<machine where document is located> \<path to the document>

e.g., http://www.gamespot.com
- http   indicates a hypertext document (a web page)
- www.gamespot.com   domain on the world wide web (web server)

e.g., ftp://ftp.ucalgary.ca
- ftp   indicates that the file transfer protocol is to be used
- ftp.ucalgary.ca   domain of the server (ftp server)

URL's (Uniform Resource Locator): 2

• Can describe path to documents:

http://pages.cpsc.ucalgary.ca/~tamj/203/assignments/assignment3/index.html

Tamj CPSC account

URL: http://pages.cpsc.ucalgary.ca/~tamj

www

203   http://pages.cpsc.ucalgary.ca/~tamj/203   231

assignments   http://pages.cpsc.ucalgary.ca/~tamj/203/assignments

assignment3

http://pages.cpsc.ucalgary.ca/~tamj/203/assignments/assignment3

Index.html
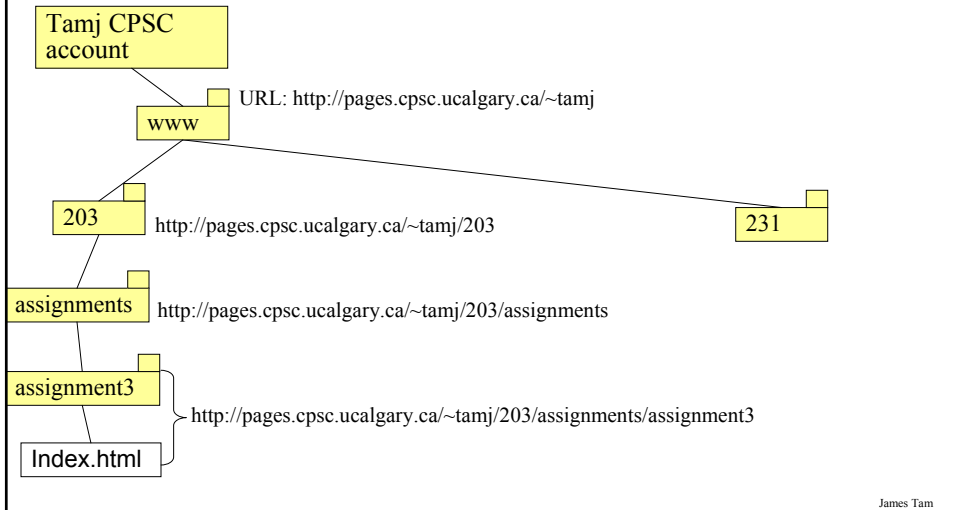
Transfer Of Information Via Packets

• The Internet is not an anonymous network!
• Information is transmitted via packets.
• In order to travel from their source to their destination each packet contains the following information:
• **Header**
  - Source IP
  - Destination IP
  - Length
  - Number
  - Sequence
• **Payload**
• **Footer / Trailer**

Transfer Of Information Via Packets

- The route taken can vary from packet-to-packet

Destination: person viewing the page

Source: web server
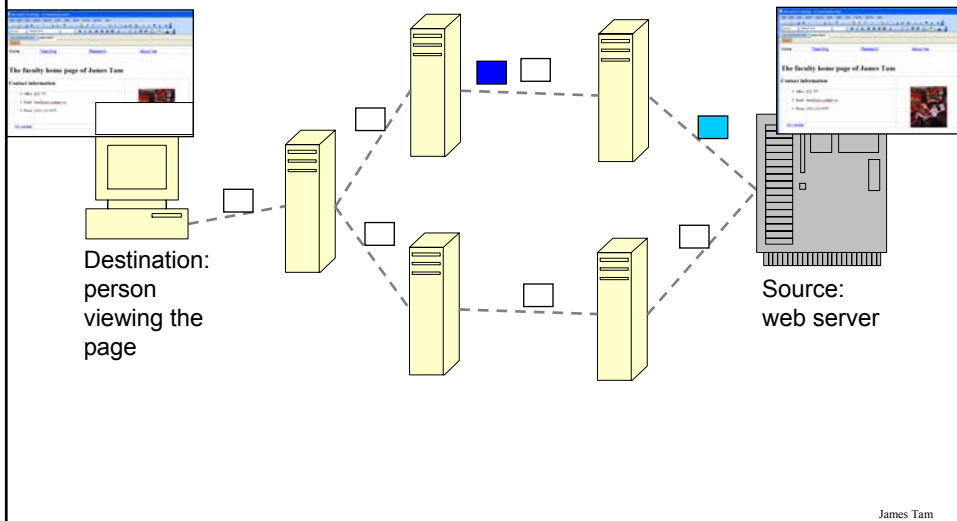
Intranets

- A network that is a mini version of the Internet.
- Provides the features of the Internet (e.g., web pages) but is self enclosed:
  - Cannot be accessed from the outside.
  - Uses the protocols of the Internet.
- Extranet: the part of a company's or organization's intranet that is extended to non-employees or non-members.

How Do Search Engines Work

- Using a search engine involves searching the library of pages built up by that engine.
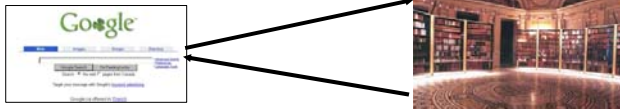
- The library is built up by the search engine's spiders (web)

Popular web site
- Link1
- Link2
- Etc.

Linked page (1st level)
- Link1
- Link2
- Etc.

Linked page (2nd level)
- Link1
- Link2
- Etc.

Search Engine Database

Popular web site	Linked page (1st level)	Linked page (2nd level)
	Linked page (1st level)	
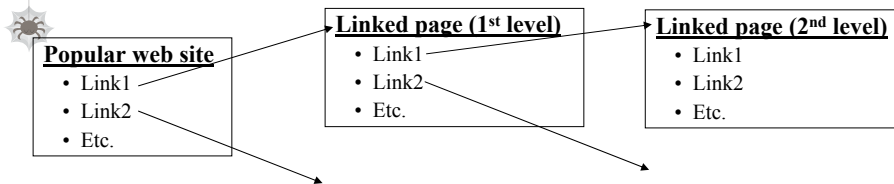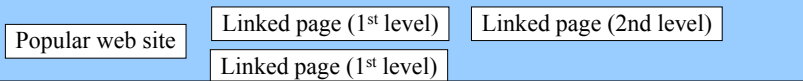
James Tam

Some Search Engines Use People Instead Of Technology

- Rather than using search programs to build the database some search engines organize results based on human scrutiny.

James Tam

Search Results Are Ranked

Making A Site More Noticeable

- Search database built via search spiders:
  - Add relevant keywords to your page.
  - The frequency of keywords may play a role.

- Search databases built via human researchers:
  - Make sure that your site is examined by the people who build the database.

Plain English Searches (Natural Language Searches)

- Easy to formulate the query but may result many hits.
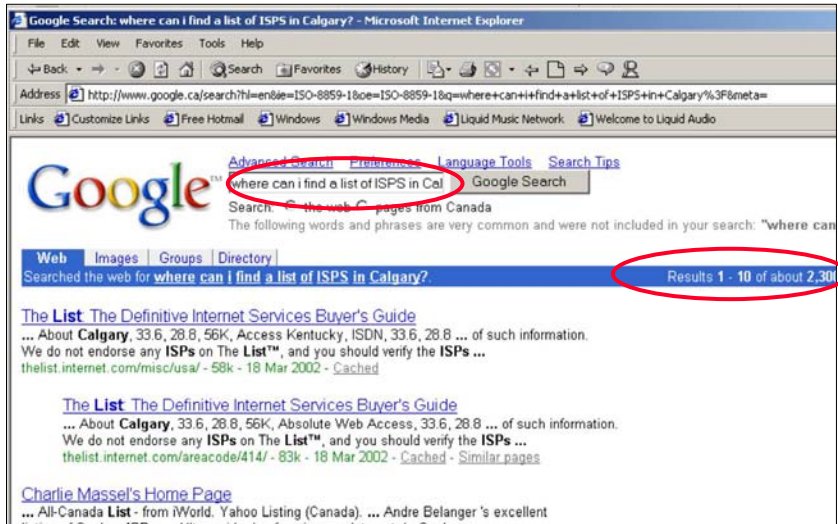
James Tam

Plain English Searches (Natural Language Searches): 2

- Supported by most of the commonly used search engines.
- Some provide better support than others:
  - AskJeeves (www.ask.com)
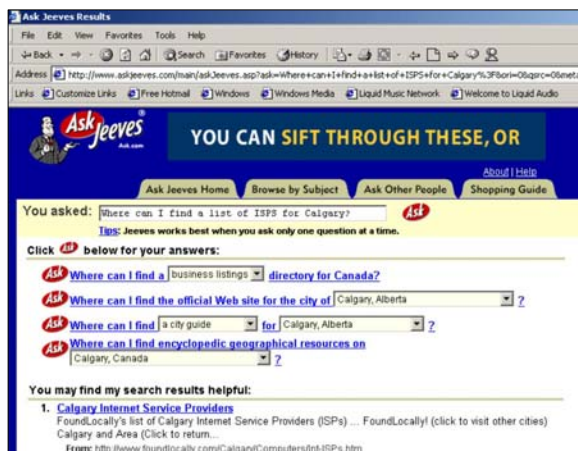
James Tam

Advanced Search Techniques (Google)

- Stopwords/Stop words and quotes
- Searching for synonyms
- Wild card searches
- Searching Ranges
- Site searches

Stop Words

- Ignored by search engines:
  - Common words
  - Reserved words

- The search engine can be forced to include the stop words:
  - E.g., Use quotes (search results must contain whatever is between the quotes).

Searching For Synonyms

- Some concepts can be represented using different words.
- The '~' operator includes synonyms in the search.

Using The Wildcard In Searches

Searching Ranges

- Can be used when searching numerical values within a certain min – max range
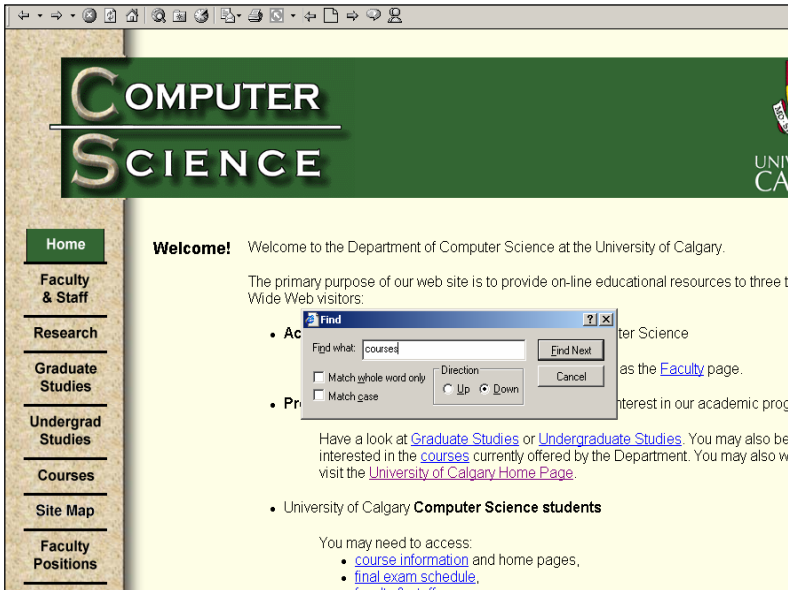- Range operator .. (multiple dots)

Site Searches

- Useful when a webpage is large and/or not well organized:
  - Searching the currently loaded page
  - Searching the entire site

Searching The Currently Loaded Page

Searching An Entire Site

- Involves searching an entire site (and not just the one page from the site that is currently loaded into the web browser).
- One of the 'advanced' search options can be employed.
- Alternatively a shortcut can be used in the main input field.

Meta-Search Engines

- Searches multiple search engines automatically.
- Examples:
  - www.metacrawler.com
  - www.dogpile.com
  - www.profusion.com
  - www.search.com
  - www.mamma.com
- Drawbacks:
  - Searches occur in the simplest form
  - Timeouts
  - Number of results returned

James Tam

You Should Now Know

- What is the Internet
- How protocols allows the Internet to function, when are the different protocols are used
- What are common ways of making an Internet connection, how do they work and what are their strengths and weaknesses
- How IP addresses are used to direct information on the Internet
- What are some of the domains on the Internet
- How to decipher Internet addresses
- What is static vs. dynamic addressing and the strengths and weaknesses of each approach
- What is a url and what information is contained in a url
- How information is transmitted on the Internet via packets

James Tam

You Should Now Know (2)

- What is an Intranet, what is an extranet
- How do search engines gather information
- Some ways of making a website more prominent to a search engine
- What is a natural language search, what is a good search engine to use when employing a natural language search
- Some advanced search techniques using Google:
  - What is a stop word and how to include them in searches
  - How to efficiently search for synonyms
  - How to employ a wildcard search
  - Searching ranges
  - Searching a specific site
- What is a meta-search and what are their weaknesses