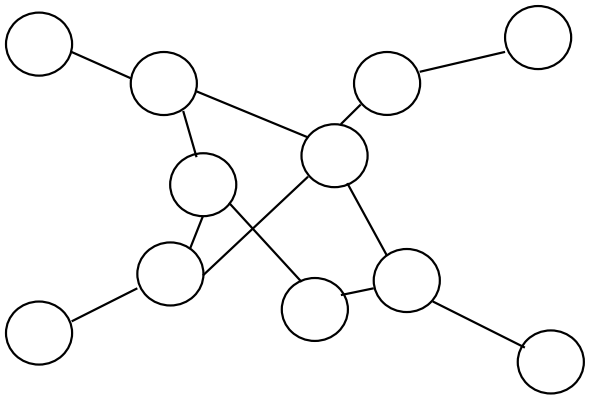
Making the Connection

or: How does the internet work?

(continued from last class)

computer network



designing a simple routing scheme

- How to choose a good route between two points in the network?
- How can nodes “know” where to send a message?

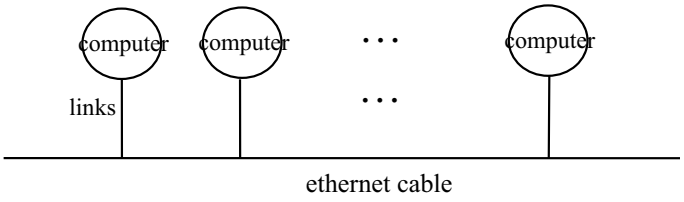
problems with simple scheme?

solutions?

broadcast systems

- In a *broadcast system*, each message is delivered to all recipients within range, e.g., voice, radio.
- Norm Abramson at U. Hawaii linked together a computer and terminals on the Hawaiian islands using a radio channel in 1970.
- Today, *ethernet channels* (wires or optical fibres) link together computers in a small area.

the ethernet



- the ethernet provides broadcast communication
- the *party protocol* implements point-to-point communication on the ethernet

a simple party protocol

1. **repeat**
2. send packet on channel
3. listen for packet on channel
4. **if** packet is detected on channel **then**
5. you are done!
6. **else** (noise is detected on channel)
7. wait a random number of time steps
8. **until** you are done

observations on party protocol

- Randomness helps! (Why?)
- It works well when there is little traffic on the channel.
- There has been much research on improved protocols that can do better in the face of high traffic – ideas?

the world wide web: web pages and links

- a web page is a file
- each web pages has an address; these addresses are called *URLs (uniform / universal resource locators)*
- some web pages are written in a language called *HTML (hypertext markup language)*
- HTML pages have *titles*
- HTML pages may have *links* to other web pages

url examples

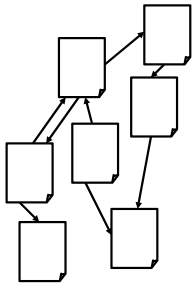
- http://www.google.ca/index.html
  - *http* refers to a protocol for transferring files
  - *www.google.ca* is a domain name
  - *index.html* is a file name (at the google domain)
- http://www.ugrad.cs.ubc.ca/~cs101/current-term/Labs/Getting-Started/index.html
  - ~cs101/current-term/Labs/Getting-Started/index.html specifies the main web page for the lab

web page links

- a link has:
  - an *anchor*: the underlined text you click on
  - a *hyperlink reference*: the URL of the web page you see when you click on the link
- *example:* http://www.ugrad.cs.ubc.ca/~cs101/

organization of the web

- the web is a network, with *directed* links
  - nodes are web pages or documents
  - nodes and links are constantly changing
- *search engines* find information on the web, e.g., Google



search engine

- a search engine is a collection of computer programs for finding information on the WWW
- a typical search engine has three components:
  - a *crawler* (spider, or robot)
  - a *query processor*
  - an *interface* (typically a web page)

crawler, query processor, interface

- The *crawler* creates a *(keyword, URL)* table; keywords are taken from the title of the document and from the *anchors* of links to the document.
- The *query processor* uses this table to find URLs that match the keywords entered by the user in the search engine *interface.*

crawler example

www.attractions.ubc.ca

Attractions and Recreation at UBC
...
Chan Centre
...
....

table

attractions,	www.attractions.ubc.ca
recreation,	www.attractions.ubc.ca
ubc,	www.attractions.ubc.ca
chan,	www.chancentre.com
centre,	www.chancentre.com
performing,	www.chancentre.com
arts,	www.chancentre.com

crawler example

www.attractions.ubc.ca

Attractions and Recreation at UBC
...
Chan Centre
...
....

www.chancentre.com

Chan Centre for the Performing Arts
…
About the Chan
....

table

attractions,	www.attractions.ubc.ca
recreation,	www.attractions.ubc.ca
ubc,	www.attractions.ubc.ca
chan,	www.chancentre.com
centre,	www.chancentre.com
performing,	www.chancentre.com
arts,	www.chancentre.com
about,	www.chancentre.com/about_f.html
chan,	www.chancentre.com/about_f.html

work of crawler never ends…

... since pages are constantly being modified, deleted, or added to the web!

- Next week, you'll create a page!
- I'll link to your page from the class home page.
- Try out Google occasionally to see if and when your page is returned when you do a web search.

research notes

A research team at Google, until recently headed by Monika Henzinger, pioneers new ways to find information on the web:

see http://www.google.com/jobs/inside.html