COSC 122
Computer Fluency

Internet

Dr. Abdallah Mohamed

Key Points

1) Networks allow computers to communicate information.

2) Communication requires a shared medium, a common language, and a protocol.

3) TCP/IP is the standard protocol for computers on the Internet.

4) The Internet and computers have made a significant impact on our lives, both positive and negative.

What is Communication?

Communication is the act of sending information from one party to another.

A *sender* transmits the information to one or more *receivers*.

For communication to be effective we need:

- a *shared medium* accessible to both senders and receivers
- a *language* or encoding for representing the information sent
- a *protocol* or set of rules explaining how the medium is used by both the sender and the receiver

Example: What are the medium, language, and protocol used in a classroom lecture like this one?

Types of Communication
Synchronous vs. Asynchronous

Communication can be categorized in several ways.

Synchronous communication is when the sender and receiver are active at the same time.

◆ e.g. telephone call, instant messaging

Asynchronous communication is when the sending and receiving occur at different times.

◆ e.g. email

Types of Communication
Broadcast vs. Point-to-Point

In *broadcast communication* (or multicast) there is **a single sender and many receivers**.

◆ e.g. cable and satellite television

In *point-to-point communication* there is a **single sender and a single receiver**.

◆ e.g. telephone calls

Aside: Spam

Spam is **unsolicited electronic communications**.

Marketers can flood a person with unwanted messages. By design, the receiver is forced to accept these messages which are most likely **deleted or filtered out** by their software (**spam filter**).

◆ *White list* – accepted sender ;

◆ *black list* – rejected sender

Spam may have derived from a Monty Python skit in which the word "spam" (spiced ham) was chanted by Vikings in a restaurant and their chants drowned out other conversation.

Medium, Language, and Protocol

Question: Fill in the blanks: The Vikings' repetition of "spam" drowned out other conversations because they were not following the _____ for the _____.

a) language / medium

b) medium / protocol

c) language / protocol

d) protocol / medium

Practice Questions

Determine if the following are *synchronous* or *asynchronous* and *broadcast* or *point-to-point*:

- radio
- classroom lecture
- e-mail
- telephone call
- postcard
- whispering to another person
- Others?

What is the Internet?

The ***Internet*** is a collection of computers and networks that transmit data using the standard **Internet Protocol (IP)**.
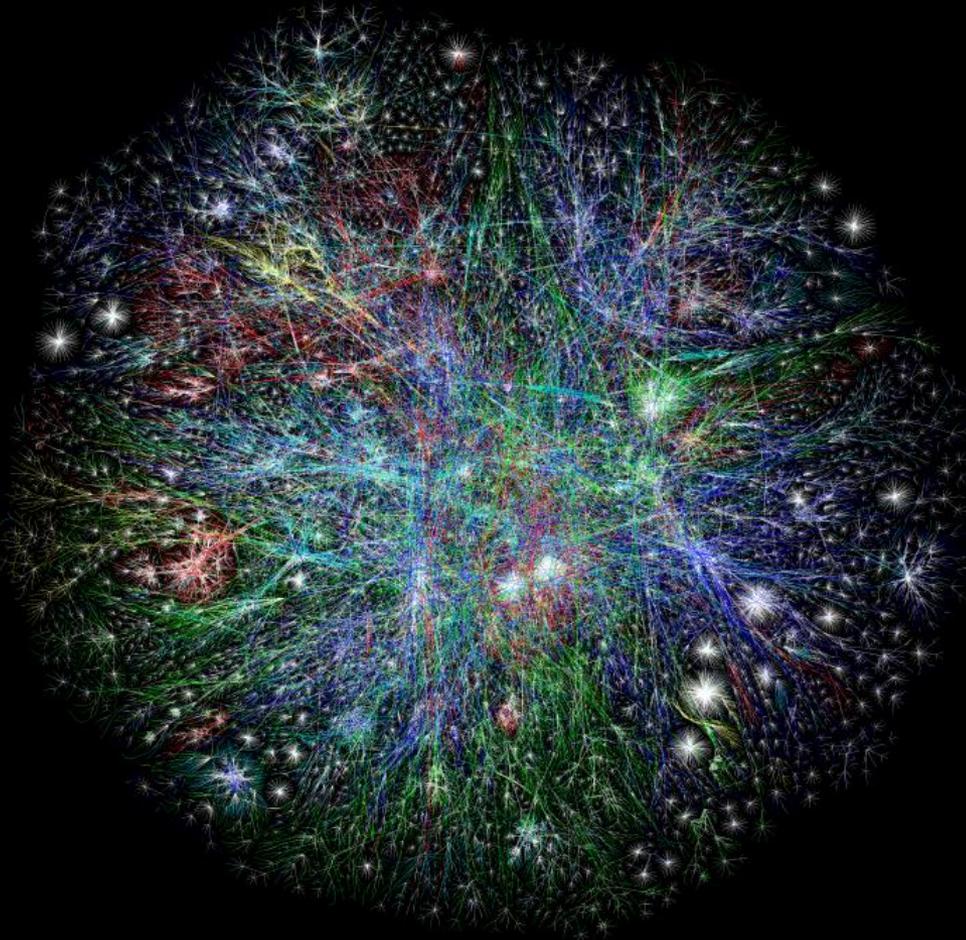
Visualization of Internet (2005)
Colors represent domains
Lengths represent delays

Internet Layers

The Internet is divided into layers to handle its complexity.

Layering is a technique for dealing with complex systems where **each layer**

◆ **provides services** for the layers above and

◆ **uses services** of the layers below.

Example: Postal mail system. Three layers:

◆ person layer (send letter)

◆ mail carrier layer (distribute mail in area)

◆ distribution layer (planes, trucks)

⇨ You as a person sending a letter do not worry about how the letter gets to its destination as long as you address it correctly.

⭐ *Five Internet Layers*

application: supports messages between programs

- ◆ e.g. HTTP between browser and server

transport: process-to-process data transfer

- ◆ e.g. TCP – guaranteed message delivery

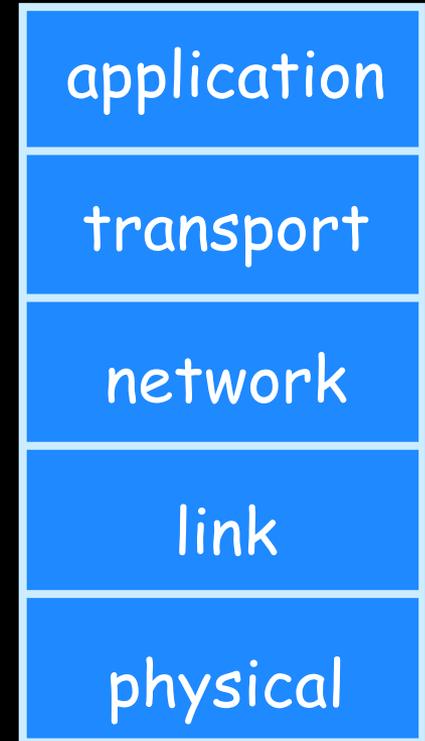*Network (internet)*: send *packets* from *source* to *destination*

- ◆ e.g. IP – send message to any machine
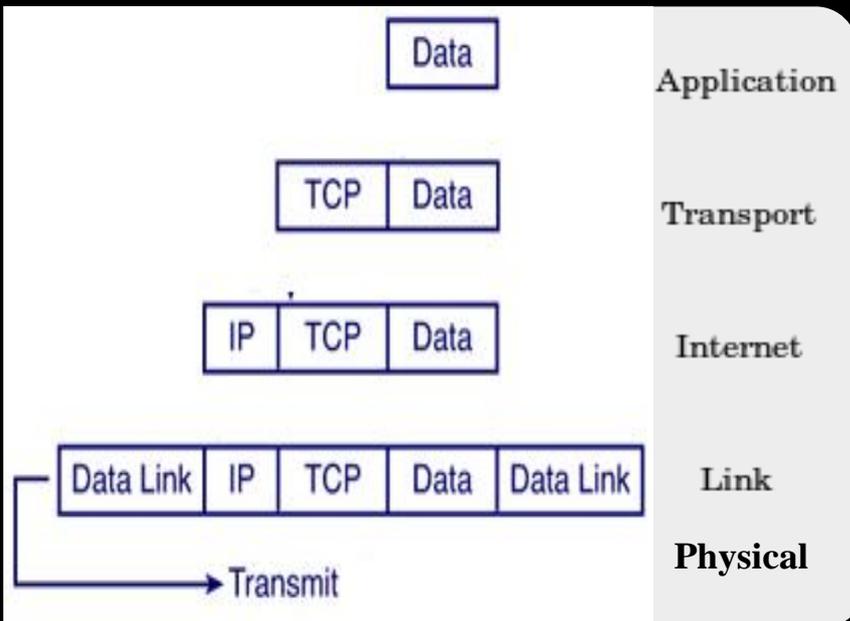
link: data transfer between *neighbors*

- ◆ e.g. Ethernet – communicate within building

physical: encoding of bits on medium
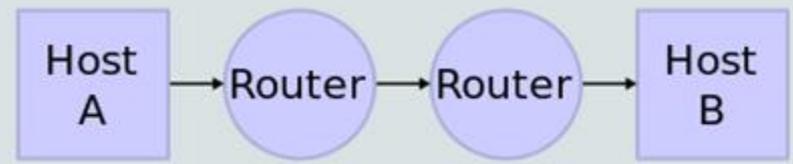
- ◆ e.g. send signals over CAT5 wire

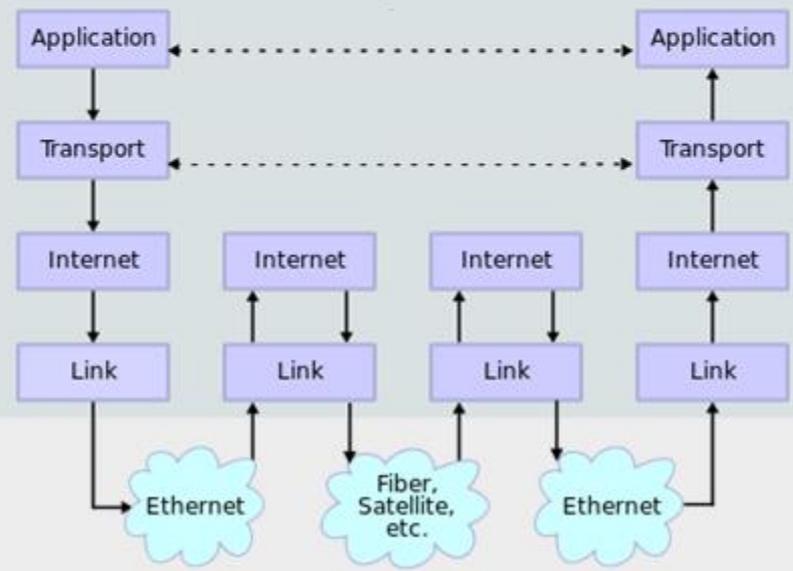| application |
| --- |
| transport |
| network |
| link |
| physical |

Five Internet Layers

Encapsulation of application data descending through the layers

Two Internet hosts connected via two routers and the corresponding layers used at each hop.

The Internet's Communication Properties

The Internet is an *asynchronous*, *point-to-point* communication system.

However, the speed of electronic communications allows for the development of applications on the Internet that appear synchronous and for information to be broadcast to many users.

Examples:
- ◆ Point-to-point, asynchronous – email
- ◆ Point-to-point, synchronous – instant messaging
- ◆ Broadcast, asynchronous – web pages, blogs
- ◆ Broadcast, synchronous – chat rooms

How do you get on the Internet?

For your computer to communicate with others on the Internet, **you need the three basic communication components:**

◆ shared medium, language, and protocol

The **medium** is either a wire or wireless link.

◆ Your ISP* may be a phone, cable, or satellite company.

⇨ The medium is the phone/cable wire entering your house or air waves if satellite/cellular transmission.

⇨ Your computer communicates with an ISP computer which relays information to and from the Internet on your behalf.

The communication format (**language**) and **protocol** used is called **TCP/IP** (Transmission Control Protocol/Internet Protocol).

I am on the Internet… Now what?
IP Addresses

A computer on the Internet is given a unique identifier called an *Internet Protocol* (*IP*) *address*.

- An IP address is **similar to your telephone number.**

An IP version 4 (IPv4) address consists of 4 numbers in the range of 0 to 255. The numbers are separated by dots.

- Example: 142.231.95.1

Since there are an increasing number of computers and devices being added to the Internet, there is an **ongoing transition to IP version 6 (IPv6)** addresses which have 16 numbers from 0-255 represented in hexadecimal.

- Example: 2002:CE57:25A2:0000:0000:0000:CE57:25A2

The Internet

Video (1:47): https://www.youtube.com/watch?v=qv0XCaUkfNk

IP Addresses and Domain Names

Although IP addresses are unique, they are even **harder to remember than phone numbers.**

Domain names are **text names for computers** that are easier to remember.

- A *domain* **is a related group of networked computers**.

Domain names are organized *hierarchically*. The most general part of the hierarchy is at the end of the name.
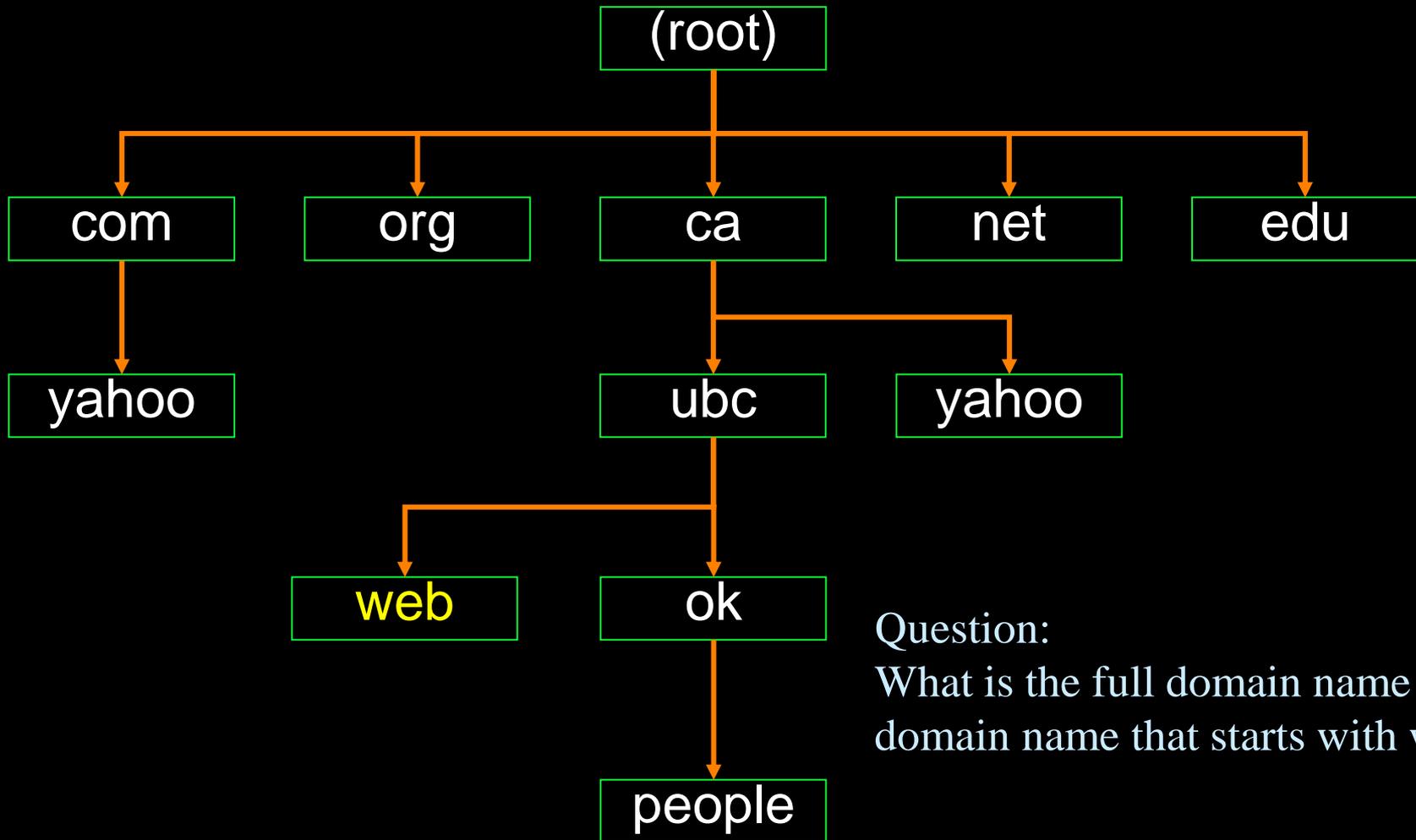
Example: `people.ok.ubc.ca`

- `ca` – Canadian domain
- `ubc` – University of British Columbia
- `ok` – Okanagan campus
- `people` – name of computer/server on campus

Visualizing the Domain Hierarchy

```
                        ┌──────────┐
                        │  (root)  │
                        └────┬─────┘
        ┌──────────┬─────────┼──────────┬──────────┐
    ┌───▼───┐  ┌───▼───┐ ┌───▼───┐ ┌────▼───┐ ┌────▼───┐
    │  com  │  │  org  │ │  ca   │ │  net   │ │  edu   │
    └───┬───┘  └───────┘ └───┬───┘ └────────┘ └────────┘
    ┌───▼───┐          ┌─────┴─────┐
    │ yahoo │      ┌───▼───┐   ┌───▼───┐
    └───────┘      │  ubc  │   │ yahoo │
                   └───┬───┘   └───────┘
               ┌───────┴───┐
           ┌───▼───┐   ┌───▼───┐
           │  web  │   │  ok   │
           └───────┘   └───┬───┘
                       ┌───▼────┐
                       │ people │
                       └────────┘
```

Question:
What is the full domain name of the
domain name that starts with web?

Top-level Domains

Top-level domains appear in the last part of domain name:

- com   commercial organizations
- edu   educational institutions
- org   organizations
- net   networks
- mil   military
- gov   government agencies
- ca   Canadian domain

Top-level domains are controlled by *ICANN*.

- e.g. many new domains have been approved in 2011 allowing for considerable freedom in naming. e.g. .soda or .pizza

DNS Servers

The ***Domain Name System*** (***DNS***) **translates the human-readable names into IP addresses**. There are DNS servers on the Internet which provide this mapping function.

◆A DNS server has a similar function as a **phonebook**.

Each Internet computer knows the IP address of its nearest *DNS server*. When you use a domain name in a request, your computer asks the DNS server to look up the IP address.

If the closest DNS server does not know the IP address, it asks a ***root name server***, which keeps the **master list** of name-to-address relationships.

◆There are **13 root name servers** (with multiple mirrored instances) distributed across the globe.

DNS Root Servers

Single instance (named A to M)

X multiple instances

Review
Synchronous vs. Asynchronous

Question: Select **one** that performs synchronous communication.

A) email

B) letter

C) telephone call

D) television

Review
Broadcast vs. Point-to-Point

Question: Select *one* that performs broadcast communication.

A) radio

B) classroom lecture

C) telephone call

D) email

Review
IP Address

Question: Which one of the following is a valid IP4 address?

A) 0.0.0.0

B) 255.255.255.255.255.255

C) 1.2.3.256

D) 111.222.3456

Review
Domain Names

Question: Which part of the address **people.ok.ubc.ca** is the largest (most general) domain?

A) people

B) ok

C) ubc

D) ca

The World Wide Web (WWW)

The *World Wide Web* (*WWW*) is an application built on top of the Internet that allows for the display and transmission of documents called web pages.

◆ Developed by Tim Berners-Lee at CERN in 1991.

A *web page* is a *document* that contains *mark-up* that allows it to be displayed graphically by a *web browser*. The page may also contain *hyperlinks* to link to related web pages.

A *web server* is a computer on the Internet with the task of storing web pages and responding to clients' requests for them.

The World Wide Web (WWW) is the **web servers** and the **files** they store.

Clients and Servers

A *server* is a computer that stores information such as a web page, e-mail, database, etc.

A *client* is a computer that requests information stored at a server.

When you click a hyperlink in your browser, your computer becomes the client and requests the appropriate web page from the server that stores that page (*web server*).

Once the web page is sent to you, the client-server interaction is complete. The server fulfills many brief requests from clients very rapidly.

Requesting a Web Page

A web page is requested by the user by either:

- typing in a *Universal Resource Locator* (*URL*) into the web browser's address field OR

- clicking on a hyperlink in a document that contains a URL

A request for a URL has three parts:

- Protocol: http:// - Hypertext Transfer Protocol
  - ⇨ Tells the computer how to handle the file

- Server computer's domain name or IP address

- Page's path and file name:
  - ⇨ Tells the server which file (page) is requested and where to find it.

URL Example

Here is an example URL:

`http://people.ok.ubc.ca/abdalmoh/teaching/122/index.html`

http protocol   server domain name   location of file on server

Another example with an IP address:

`http://64.40.106.38/okanagan/psycomp/welcome.html`

http protocol   server IP address   location of file on server

File Structure

A *file structure* is a method for organizing files on a computer.

The common file structure is a *hierarchy* of directories or folders.

Each *folder* has a name and can contain any number of files or subdirectories.

Each *file* has a name.

File Structure Usage

The file structure or file system is used when storing files locally on your computer.

It is also used as part of a URL as it provides the location on the server computer of the file that is requested by the client.

Example:

```
http://people.ok.ubc.ca/abdalmoh/teaching/122/index.html
```

- The directory path is: `abdalmoh/teaching/122/`
- The file with name: `index.html` is in the above directory.
- Note that directory names are separated by slashes "/".

Hypertext

The **HTTP** protocol can transmit any **file type**, not just documents. However, it is most commonly used to transmit documents written in Hypertext Markup Language (**HTML**).

HTML describes the layout of a document including fonts, text style, image placement, and hypertext links.

Hypertext links provides a way to jump from point to point in documents (non-linear). Links may jump within a document, between documents on a server, and to documents on other servers.

We will see how to write HTML documents soon.

```
HTTP: Hypertext Transfer Protocol
HTML: Hypertext Markup Language
```

TCP/IP (Transmission Control Protocol/ Internet Protocol) ★

TCP/IP (Transmission Control Protocol/Internet Protocol) is the structure (language) and protocol used for communication between computers on the Internet.

This is how TCP/IP works:

- **Information is broken** into a sequence of small fixed-size units called IP *packets*.
  - ⇨ Each packet has space for the unit of data, the source and destination IP addresses, and a sequence number.
- The packets are sent over the Internet **one at a time** using whatever route is available.
- Because each packet can take a different route, congestion and service interruptions do not delay transmissions.
- Receiver re-assembles packets using sequence numbers.

Moving Packets

How packets are transmitted between computers on the Internet is dependent on the medium of transmission.

The ***Internet backbone*** has the largest capacity (***bandwidth***) and is ***optical fiber***.

Smaller bandwidth connections may use copper fiber or cable to connect machines to hubs and routers.

End users may use phone lines, cable lines, or even fiber optic connections as their ***first hop*** (connection) to the Internet.

Verizon Backbone

Telus Network

Source: http://about.telus.com/networktechnology/images/NationalNetwork.gif

Trace Route Example

Practice Question
Sending a Message Using TCP/IP

Using two small pieces of paper. Write on each paper:

◆ The *destination row and seat #* of your receiver (IP address).

◆ Your *sender row and seat #*. (IP address)

◆ Row numbers start at 1 from the front of the class.

◆ Seat numbers start at 1 from the left and increase going right. Only count seats where people are sitting.

◆ The number 1 or 2 for the first or second piece of paper.

◆ The first part of a sentence on piece #1 and second part on piece #2.

◆ Example:

| **Src:** Row 4 Seat 5 **Dest:** Row 6 Seat 3 **Seq#:** 1	**Src:** Row 4 Seat 5 **Dest:** Row 6 Seat 3 **Seq#:** 2
This is how	TCP/IP works.

Practice Question
Sending a Message Using TCP/IP (2)

Routing:

- The person at the left end of the row is your gateway to the Internet. They will be called the *gateway router*. All packets are sent to the gateway router even if the destination is in your row.

- When a gateway router receives a packet they will look at the destination address. If it is in their row, they will route it back down the row. If not, they will route it to an internal router.

- Several people are selected to be *internal routers*. They are responsible for some number of rows (e.g. 1 to 5). When they receive a message, and the destination is in their set of rows, they will give it to the gateway for that row. Otherwise, they will give it to the adjacent *core router* (up or down).

- When ready, send your message. (You may break it up).

- When you get a message, reassemble. On the back of the paper, send a reply back to the sender.

R5
R4
R3
R2
R1

C1 C2 C3 C4 C5
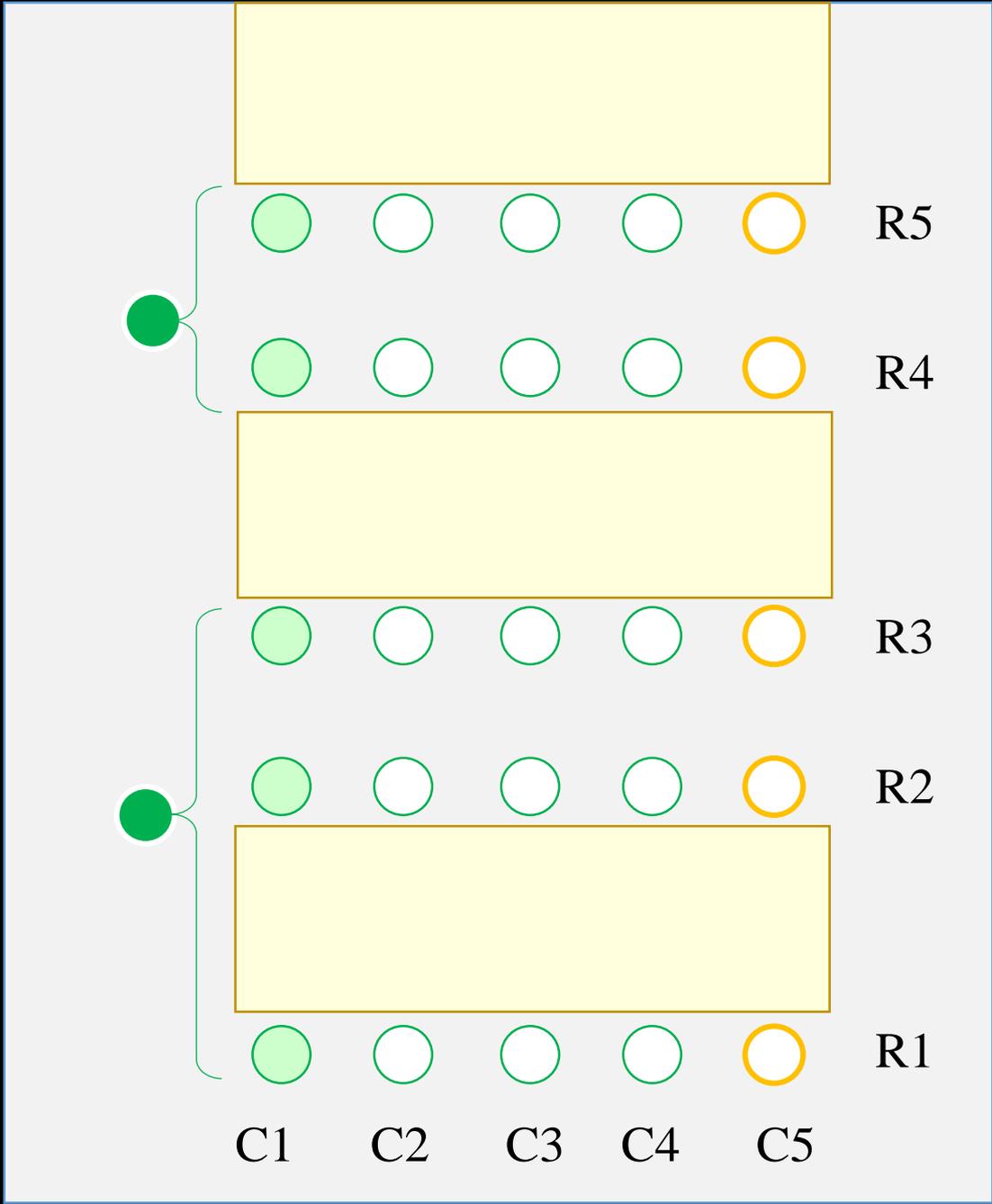
Internal router

Gateway

Computer node

Sender

R# Row number

C# Column number

TCP/IP

Question: Put the following steps in order to describe transmitting information on the Internet using TCP/IP.

1) put sequence #, sender and destination IP addresses on packets

2) route packets through whatever route is available

3) re-assemble packets using sequence numbers

4) break data into fixed-sized packets

a) 1,2,3,4
b) 4,2,3,1
c) 4,1,2,3
d) 4,2,3,1

Aside: What is DHCP?
What are tiny URLs?

Dynamic Host Configuration Protocol or **DHCP** automatically assigns an IP address when a device is attached to the network.

- Your IP address will change on your device whenever you connect to a different access point.
- A static IP address for servers does not change.
- DHCP also assigns local DNS server addresses.

Tiny URLs like http://tinyurl.com/cosc122 or http://bit.ly/pqTvJw used in Twitter and short messages **are not full URLs**. The first part of the URL http://tinyurl.com or http://bit.ly is a valid URL that goes to tinyurl.com or bit.ly. The last part "cosc122" or "pqTvJw" is a hash code that is used to look up the full URL that was originally entered and redirects to it.

Aside: Where is the www?

Does a name have to contain `www` (e.g. <u>`www.yahoo.com`</u>)?

◆ If the server name contains `www`, then `www` would be the name of the computer in that domain.

◆ A computer can be a web server even if its name does not contain `www`. (e.g., <u>`technet.microsoft.com`</u>)

Other Frequently Asked Questions (FAQs):

◆ 1) Do I have to type `www` or can I just type `yahoo.com`?

⇨ You can normally just type `yahoo.com` without the `http` and `www`, and the browser will fill it in for you. Note that the site still is <u>**`www.yahoo.com`**</u>.

◆ 2) What is http?

⇨ Http stands for hypertext transfer protocol. It is the name of the protocol for sending web pages over the Internet. **It is the use of the http protocol that makes a computer a web server.**

◆ 3) Are all computers on the Internet web servers?

⇨ No.

Aside: What is a protocol?

A *protocol* is a set of rules for transmitting information.

The underlying transmission protocol on the Internet is TCP/IP, but there are a variety of protocols for computer networks.

Many applications on the Internet have their own protocol built **on top** of TCP/IP to transfer information specific to them.

◆ We have already seen HTTP. Others include FTP and email.

In addition to the protocol, servers and clients need to know a *port* (or **address of software on server**) when multiple server applications may be running on the same machine.

◆ The port for HTTP is usually 80.

Aside: The anatomy of an e-mail address

An e-mail address has a structure similar to a URL:

`abdallah.moahmed@ubc.ca`

↑ unique name of user on server (local-part)

↑ server domain name

FAQ:

- ◆ 1) Why the @ symbol?
  - ⇨ At "@" symbol was used as a separator **because it is not in the user or server name** and naturally implies that user is associated with the server.
  - ⇨ Developed by Ray Tomlinson in 1971.

- ◆ 2) Are e-mail addresses case-sensitive?
  - ⇨ Domain names are not case-sensitive.
  - ⇨ Local user names are case-sensitive,
    - ◆ but most servers ignore case to avoid errors. Lower case characters are used for readability and by convention.

Aside: What is a firewall?

A *firewall* is a network device that **is installed on the edge of a network** to prevent unauthorized network traffic from entering a local network.

- ◆ A firewall uses information in the packets (IP addresses, ports) to determine good traffic from bad traffic.
- ◆ Administrators can restrict access to certain sites or applications using a firewall.

A *proxy server* or *gateway* is a computer on the network that your computer must **communicate through** to get out onto the Internet.

Other Types of Networks
WAN and LAN

The Internet is a "***network of networks***" as it connects independent networks together using a common protocol.

A ***Wide Area Network*** (***WAN***) is designed to send information between widely separated locations.

A ***Local Area Network*** (***LAN***) connects computers close enough to be linked by a single cable or wire pair.

- ◆ **Ethernet** is the main technology for LAN.

Our campus has a WAN to connect all buildings together, and within each building is one or more LANs.

LAN Overview

A local area network has a **shared channel** (wire, wire pair, or optical fiber) that **connects a set of computers**.

Each computer is connected to the channel, allowing it to send a signal that can be detected by all computers connected to the channel.

When you plug in a network cable to a wall socket, you are connecting your computer to the shared channel. Behind the walls are cables running to a central hub that connects all plug locations.

Ethernet Protocol

The commonly used protocol for communicating between computers on a LAN is called *Ethernet*.

Ethernet is a very simple, decentralized protocol. To send:

- A computer listens to the channel. If it's quiet, it's free.
- The computer starts sending on the channel.
- While sending, the computer listens to make sure it is the only one sending. If not, it stops for a random amount of time then starts again.

Try it: Select several people. Each tries to say "I am transmitting a message" at roughly the same time. Try until everyone has sent the message.

Internet History

- **1960s:** Packet switching developed
- **1972:** ARPAnet had 15 nodes and a host-to-host protocol. First public demo. Ray Tomlinson at BBN wrote e-mail program.
- **1973:** Ethernet at Xerox PARC (Metcalfe)
- **1974:** Cerf and Kahn – TCP/IP (Turing Award)
- **1979:** ARPAnet has 200 nodes
- **1982:** SMTP e-mail protocol defined
- **1983:** DNS defined for name-to-IP-address translation
- **1990:** Internet has 100,000 nodes
- **1991:** **WWW** invented by Tim Berners-Lee.
- **1994:** Mosaic (later Netscape) developed by Marc Andreesen.
- **1995:** ARPANet decommissioned. Replaced backbone by commercial Internet service providers.

Internet Current

- **1995-present:** Tremendous growth in the number of users and applications. Common applications:
    - ⇨ Communications (text, voice, video) – email, Skype
    - ⇨ News/advertising/entertainment
    - ⇨ Massive Data Sets – Google Maps/Earth, scientific data, directories
    - ⇨ Social sites – MySpace, Facebook, LinkedIn
    - ⇨ E-commerce – Amazon, Walmart
    - ⇨ Search – Google, Yahoo, Bing
    - ⇨ Online gaming/social gaming
    - ⇨ Mobile access and smart phones – iPhone, Android, Blackberry
- Improved network **bandwidth** and reduction in **cost** has made these applications possible. Standardization is also important.
- Bandwidth and hardware is now a commodity (*cloud computing*).
- Easier than ever to build an Internet application.

Impact of the Internet on Society

- Nowhere is remote.
  - A person in Kelowna has the **same access** to Internet information as someone in Toronto.
- People are interconnected.
  - Can **interact with people** around the world.
- Social relationships are changing.
  - We are **spending more time online** and doing less in-person activities.
- English is becoming a universal language.
  - The influence of American culture since World War II has led to rapid adoption of English as the default language for global commerce, science and technology.
- Freedom of speech and assembly have expanded.
  - The Internet is *technically* unmediated allowing freedom of expression (both positive and negative). **Anyone can publish at almost no cost.**
  - Countries like China can restrict access to information on the Internet.

What is The Value of Information on the Internet?

Since anyone can publish a web page with information (fact or fiction), this introduces several important issues:

- ◆ *Information overload* – **too much information** which makes it difficult to find relevant information

- ◆ *Information organization* – how is information organized so that it can be **easily found and used**

- ◆ *Information quality* – the lack of independent editing creates an issue of **trustworthiness and completeness**

Measuring Information Quality

The quality of information can be measured in several ways:

◆1) **Investigating the source** – Trusted sources with an online presence should have quality information.

⇨ It is possible to look up the organization that publishes a web site using its domain name and the Whois facility.

⇨ Canada Whois: http://whois.cira.ca/public

◆2) **Realistic site content** – A site is more believable if it contains *physical addresses*, *phone numbers* **and** *credentials*, and if it appears *current* and *professionally* done.

◆3) **Search engine ranking and external links** – Although not fool proof, higher search engine rankings and links from other sites are an indication that others value the information on the site.

Aside: How does a search engine work?

All popular search engines such as Google and Bing have two basic parts:

- ◆ *Crawler*: Visits sites on the Internet, discovering Web pages and building an index to the Web's content.
  - ⇨ A search engine has crawlers running continuous to refresh and update its index database of web pages.
  - ⇨ When a crawler visits a page it identifies the terms on the page and then processes any outgoing links.

- ◆ *Query processor*: Looks up user-submitted keywords in the index and reports back which Web pages the crawler has found containing those words.
  - ⇨ The query processor does not search the Internet – it only returns answers previously found by the crawlers.
  - ⇨ The ranking algorithm to identify important pages is critical to success of the search engine. Google uses the PageRank algorithm.

Aside: How does a search engine work?
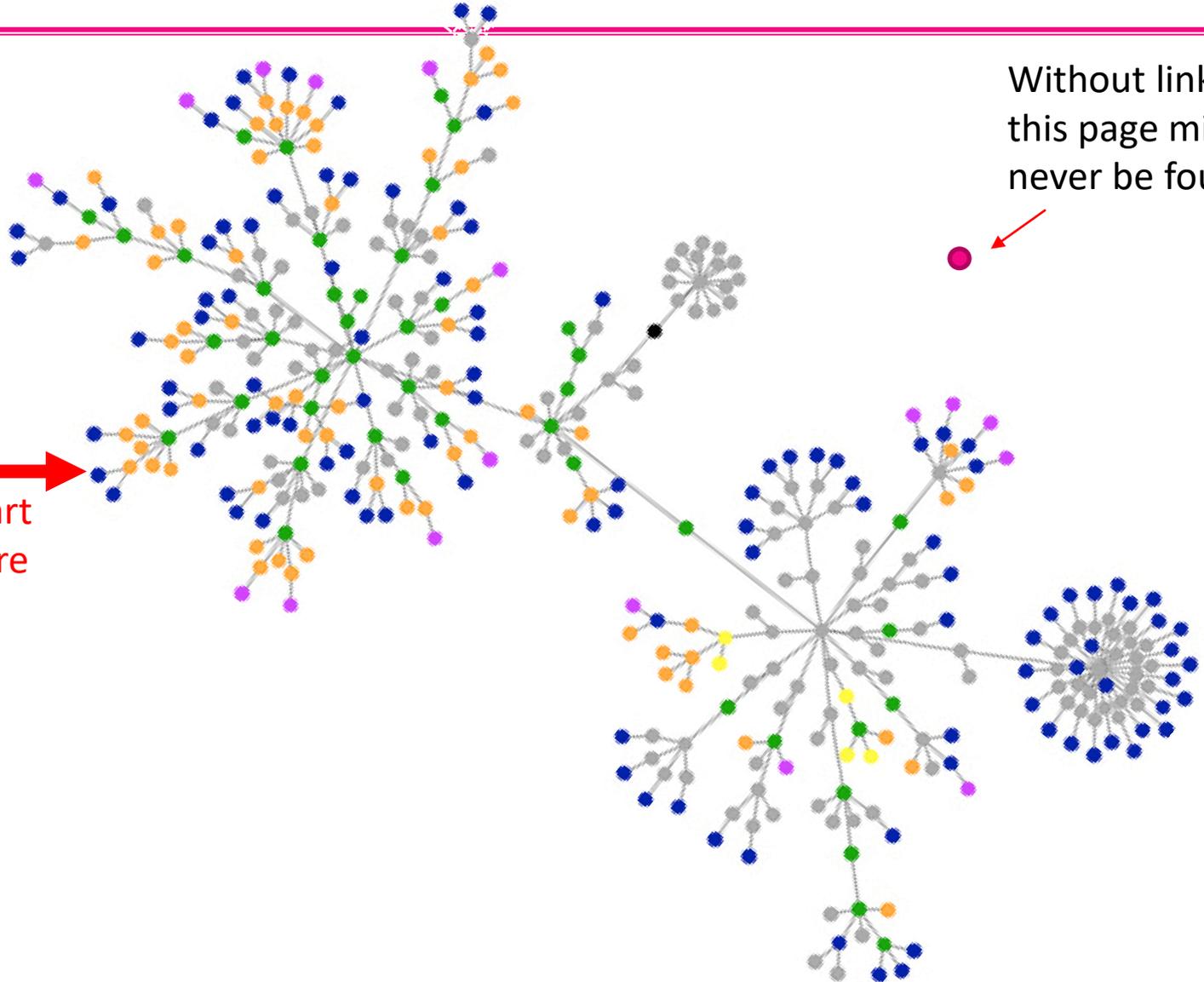
Without links, this page might never be found!

Start
here

Web Crawler

Aside: How to Search Effectively

Search engines allow basic keyword search and more advanced search features. Some things you should know:

- Search for a phrase by putting **double quotes** around it.
  - ⇨ e.g. "Computer Fluency" instead of Computer Fluency
- By default, there is a logical `AND` connecting terms. This means that **all terms must appear** in the document.
  - ⇨ e.g. Computer Fluency means both Computer and Fluency must appear.
- You can also use `OR` to indicate either term is suitable:
  - ⇨ e.g. (Book `OR` Magazine)   - parenthesis are optional
- You can use `NOT` or (-) to indicate work should not appear:
  - ⇨ e.g. `NOT` Fluency, -Fluency
- You can use plurals, but the search engine will normally discard them. (Called *stemming*) E.g. trees becomes tree
- Question: What is a cached page in a search engine?

Survey
Essential Technology

Question: What technology could you **absolutely** not live without?

A) television

B) cell phone

C) social network sites

D) email/text messaging/chat

E) none of the above

Discussion
Effect of Internet and IT on Society

In small groups, discuss what you think are the most important positive and negative effects of the Internet and computers on society.

◆ Are there issues that we have not covered?

Be prepared to give a short summary of your discussions.

Conclusion

The Internet is an *asynchronous*, *point-to-point* communication tool. However, due to its speed, *broadcast* and *synchronous* applications are also supported.

The three components of any communication are a shared medium, a common language, and an agreed upon protocol.

An *IP address* is a unique address that identifies a computer on the Internet. *Domain names* are used as they are easier to remember, and are mapped to IP addresses by a DNS server.

The *World Wide Web* allows for the storage, transmission, and display of information in documents called web pages.

The Internet and IT in general has made a significant impact on society and our daily lives.

Objectives

- Compare and contrast: synchronous and asynchronous
- Compare and contrast: broadcast and point-to-point
- Identify what types of communication common devices use.
- List and define the 3 components of communication.
- Define: Internet
- Explain how you can get on the Internet.
- Explain the format and purpose of an IP address. IPv4 vs. IPv6.
- Describe the hierarchical structure of a domain name.
- Explain the purpose and role of a DNS server.
- Explain the key features of the TCP/IP protocol.
- Define: client, server
- Define: WWW, web page, web server, web browser
- List and explain the components of a URL.

Objectives (2)

- Provide the unique feature of HTML documents compared to other documents.

- Define: file structure, file, directory

- Compare and contrast: WAN and LAN

- Provide an overview of the Ethernet protocol.

- List and discuss some of the impacts of the Internet and IT on society.

- List 3 challenges with the vast amounts of information available on the Internet.

- Discuss how you can evaluate the quality of information found online.

- List the two components of a search engine.