3. The Internet
Reference

What is the Internet?

- A global network that allows one computer to connect with other computers in the world

What can be done on the Internet?
- exchange text, data files, programs, multimedia information, etc.
- remote control a computer (if allow) on the side of the world

Internet ≠ Web
- Strictly speaking, Web is just a kind of service in the Internet that makes use of the Hypertext Transfer Protocol (HTTP)
The Beginning

- The seeds of Internet were planted in 1969
  - Advanced research Projects Agency (ARPA) of the US Department of Defense wanted to connect computers at different universities and defense contractors
  - Result in a network called ARPANET

- Goal:
  - Create a computer network of multiple paths (by means of telephone lines)
  - Hence survive any attack (nuclear!) or disaster
  - Allow remote users to share scarce computing resource
3. The Internet

- At the beginning, only 4 primary hosts
  - **Host**: machine that provides computing service
- Grew quickly and spread widely
  - Jump across Atlantic to Norway and England in 1973
  - In 1973, **National Science Foundation (NSF)** continued the task to support the development
  - Established the **NSFnet**
- Link between NSFnet, ARPANET and other networks was called **the Internet**
- Both NSFnet and APRANET are for academic purposes, not allow private business functions
ARPANET was shut down in 1990
Government funding for NSFnet was discontinued in 1995
However, the Internet has attracted much attention to the business world
Commercial Internet backbones were established to replace ARPANET and NSFnet
They become the current Internet
According to eMarketer (May 2002),
- number of Internet users in US in 2002 is 153 millions
Internet: networks of networks

- Many connected computing devices: PCs, workstations, servers, PDA, phones, …
- Communication links: fiber, copper, radio, satellite, …
Features of the Internet

- Connected by many **small networks** in the world using TCP/IP protocol stack
- No specific path between any two hosts that are communicating
- Path is determined dynamically by **Router**, that relays the data from the source host to the destination host
- Data may need to go through many routers before they reach the destination
- When data from different sources need to go through a particular router, it will become busy and may **delay** or even **lost** the data
- Hence **Internet is NOT** a reliable network
“On the Internet”

- What does it mean to be “on the Internet”?  
  - Your computer has an “address”  
    - so that other computers can find you  
  - Your computer “speaks” one particular type of language of the Internet world  
    - so that you can communicate with the others  
  - Your computer knows how to work with other hosts on the Internet, e.g., exchange files, data, programs  
    - a mechanism needs to be devised to ensure cooperation between hosts
A. Address

- Way to identify people / computers
- On the Internet, the term “address” is used loosely
  - Can mean many different things from an email address to a URL (Uniform resource locator)
- More specifically, 2 types:
  - Internet protocol (IP) address (or network address): 4-part numeric address
    e.g. 158.132.148.28
  - Domain name system (DNS) address
    e.g. www.polyu.edu.hk
A.1 IP Address

- Each IP address is 32-bit long (four bytes)
- The four-byte address is written out as a.b.c.d

```
<table>
<thead>
<tr>
<th>Byte 1</th>
<th>Byte 2</th>
<th>Byte 3</th>
<th>Byte 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>158</td>
<td>132</td>
<td>161</td>
<td>99</td>
</tr>
</tbody>
</table>
```

- IP addresses are hierarchical
  - network I.D. and host I.D
  - Each Network I.D. on the Internet needs to be registered to the Internet Assigned Number Authority (IANA)
Class A – for very large networks

1 bit 7 bits 24 bits

<table>
<thead>
<tr>
<th></th>
<th>Net I.D.</th>
<th>Host I.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• Only $2^7$ (127) networks can belong to this class
• Each network, there are $2^{24}$ hosts or computers
• Very few class A networks in the world
  e.g. Arpanet – the earliest packet switched WAN (started 40 years ago)
Class B – for medium size networks

- 2^14 (16384) networks can belong to this class
- Each network, there are 2^16 (65536) hosts or computers
- Polyu’s address belongs to this group
  e.g. 158.132.14.1

<table>
<thead>
<tr>
<th>1001 1110</th>
<th>1000 0100</th>
<th>0000 1110</th>
<th>0000 0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network I.D.</td>
<td>Host I.D.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Class C – for small networks

<table>
<thead>
<tr>
<th>3 bits</th>
<th>21 bits</th>
<th>8 bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Net I.D.</td>
<td>Host I.D.</td>
<td></td>
</tr>
</tbody>
</table>

- $2^{21}$ networks can belong to this class
- Each network, there are only $2^8$ (256) hosts or computers
Class D – for multicast network

<table>
<thead>
<tr>
<th>4 bits</th>
<th>28 bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 1 0</td>
<td>Group no.</td>
</tr>
</tbody>
</table>

- Packets are addressed to a multicast group
- Not often supported on Internet
### Special Addresses

- **Network ID** | **Host ID** | Explanation
- All 1s | All 1s | Broadcast to local network
- Not all 1s | All 1s | Broadcast to the specified network
- All 1s | Not all 1s | Broadcast the specified host on all networks
- All 0s | All 0s | This host
- All 0s | Not all 0s | The specified host on this network
- Not all 0s | All 0s | The specified network

- **127.x.y.z**: x, y, z could be anything. Used for loopback, i.e., the packet will be received back by the host that transmits it. For debugging purpose.
A.2 Domain Name

• Every computer has a network address
e.g. 158.132.161.99
• To access a computer, we need to specify its network address
• Human beings are weak in memorizing numbers
• We prefer computer name or domain name
e.g. hkpu10.polyu.edu.hk
• Need a machine on the Internet to convert name to number
Domain name hierarchy

Example:

hkpu10.polyu.edu.hk

- Computer name
- Root domain name
  - other examples:
    - com – commercial company
    - org – general organization
    - net – major network centre
    - gov – government org.
    - mil – military group
    - edu – education org.

- The domain within edu.hk
- The domain within hk
- Note: edu.hk is not the same as edu

- One of the educational institutions in H.K.
• An organization needs to register its domain name
e.g. PolyU has registered its name to the domain of edu.hk
• Once a domain name is assigned, the organization is free to assign other names belong to its domain
e.g. we can have
  hkpu10.polyu.edu.hk
  smtp.polyu.edu.hk
  mail.polyu.edu.hk
3. The Internet

Where is www.yahoo.com?

Address of www.yahoo.com

Domain Name Server (DNS) of polyu.edu.hk

Usually UDP

Where is www.yahoo.com?

Address of yahoo.com?

Where is www.yahoo.com?

Address of the DNS of Yahoo.com

DNS of Yahoo.com

DNS of com
• Nevertheless, such a complicated procedure needs not perform in most cases
• Client computers usually remember the answers that it got before
• It reduces the loading to the root DNS
• To further reduce loading, there can be many root DNS on the Internet
e.g. there are a few “com” root DNS
B. Language of Internet

- How computer exchange data?
  - Needs a set of rules and procedures to control the way data is transmitted between computers
- Technical term: protocol
- The major protocol used on the Internet: TCP/IP: transmission control protocol / internet protocol
  - Unix machine can exchange data with a Mac machine
C. Client / Server Operations

- The Internet is only a general communication infrastructure
- Mechanism needs to be devised to ensure cooperation between hosts
- Traditionally, the client / server model is adopted on the Internet
- An application on one computer attempts to communicate with an application on another (client)
- An application on the other computer answers the incoming request (server)
- Server – waits passively for contact
- Client – initiates communication actively
Client
- invoked directly by users
- actively initiates contact’s with a server
- usually does not require a special hardware nor a sophisticated OS

Server
- Special purpose, privileged computer dedicated to providing one service, but can handle multiple remote clients at the same time
- Waits passively for contact from remote clients
- Requires powerful hardware and a sophisticated OS
High End Server Example: Sun Fire™ 15K

**Architecture Highlight**

- **CPU:** SUN UltraSparc III 900MHz x 72
- **Memory:** max 0.5 TB (Tera Bytes means Mega-Mega Bytes)
- **I/O:** 72 Hot-swappable PCI slots
- **Storage:** max 144 Ultra SCSI Disk Drive (about 2.6 TB)
- **OS:** Solaris 8
- **Other features:** Dynamic reconfiguration and hot-swappable component
- **18 virtual domains**
- **Cost:** about HK$5 million
C.1 Major applications on the Internet

- Web wide web (WWW)
- Electronic mail (Email)
- News – for message posting
- Telnet – for remote access of host
- FTP – for file transfer
- Internet Relay Chat (IRC) – for real-time group discussion

- Server for each of the above applications can be found on the Internet
- They can be implemented within a computer or in different computers
C.2 World Wide Web (WWW)

- World wide web (or the Web): Allow people to locate and view multimedia-based documents, e.g., text, graphics, animations, audios, videos, etc. through the Internet using HTTP.
- Created in 1989.
- At European Particle Physics Lab in Geneva, Switzerland: as a method for incorporating footnotes, figures and cross-references into online hypertext documents using HTTP (hypertext transfer protocol).

- Hypertext document: encoded file that uses the HTML (hypertext markup language).
- Hypertext document = web page.
TCP/IP and HTTP

**TCP/IP**
- Define the rules of how a byte of data transmitted from one computer can be correctly received by another computer
- Does not define anything about the meaning this byte of data

**HTTP**
- Define the meaning of those received data
- Define the data that a client should send to a Web server to request services
- Define also the data that a Web server should send to the client such that the client can understand
Client software such as Web browser, sends out a "request". Find server, which is connected to the Internet via a Bridge/router and TCP/IP. Web Server responses are sent to the Client. Browser displays the page.
Features of HTTP

- HTTP is text based
  - All communications between server and client are done with ASCII character string

- Use request / respond mechanism
  - Request is always made by client
  - Server gives a response for every request

- Stateless
  - Server tries to push the data to client as fast as possible
  - Will not keep track with the state of the client once the material is sent (hence stateless)
  - Important to Internet system since it is impractical to maintain a long connection between server and client due to the low reliability of Internet
3. The Internet

```
HTTP/1.0 200 OK
Date: Fri, 31 Dec 1999 23:59:59 GMT
Content-Type: text/html
Content-Length: 1354

<html>
<body>
<h1>Happy New Millennium!</h1>
(more file contents)
...
</body>
</html>
```

Example
HTTP and HTML

- **HTTP**
  - Define the meaning of the received data
  - Define the data that a client should send to a Web server to request services
  - Define also the data that a Web server should send to the client such that the client can understand

- **HTML**
  - Define the content inside a Web page
  - Define how the text, graphic, image, etc. should be placed on the Web browser
Web browser

- The success of WWW largely due to the popularity of Web browser
- Different kind of OS (Mac, Linux, Windows, etc) has different way to display multimedia information
- Two solutions:
  - Ask all OS to come up with a standard approach: Very difficult
  - Develop an application program that every OS must install: Sound also difficult but surprisingly achieved by Netscape
- Web browser: A type of software (client) for navigating the WWW
- 1993: develop Mosaic
- More popular Web browsers: Netscape and Microsoft Internet Explorer
- Become the *de facto* standard of every client computer
URL Details

- A Web browser seeks for a web page based on its URL (Uniform Resource Locator)
- 2 basic parts
  - A portion specifying the method of file access
  - A portion specifying the Internet location of the file to be accessed
- First portion: ending with ://
  - Tells the type of file access
  - http:// says that the server that is going to give the file to you following the http protocol
  - Besides http, can be ftp://, rtsp://, or simply File://
- Second portion: specify the Internet address
  - IP or DNS address
3. The Internet

http://www.eie.polyu.edu.hk
3. The Internet

- A ftp server is also implemented at www.eie.polyu.edu.hk. Hence it can provide ftp service
- Most Web browsers can also be used to handle ftp functions
- The files of the user are shown on the browser
- Files can be uploaded to or retrieved from the server by simply click-and-pick

ftp://enpklun@www.eie.polyu.edu.hk

- Service required
- user name
- DNS address