

CPS221 Lecture: Introduction to Net-Centric Computing

last revised 9/9/14

Objectives

1. To overview the history of computer networks in general and the Internet in particular
2. To overview the major network standards bodies
3. To introduce the notion of a protocol
4. To overview the Internet as it exists today

Materials:

1. Projectable of Figure 1.2 from Kurose Ross
2. Projectable of RFC 2026
3. Projectable of Figure 1.2 from Forouzan
4. Projectable of Figure 1.1 from Forouzan
5. Be sure this lecture is online so it can be accessed via URL below

I. Introduction

- A. The earliest computer systems functioned in isolation from one another; but today it is almost impossible to conceive of a computer system that is not networked in some way to other systems. This includes not only desktop and laptop computers, but also devices such as cell phones, entertainment devices, etc.
- B. In common parlance, we regard “network” and “the Internet” as virtual synonyms. But actually the term “the Internet” originates from the term “internet” (lower-case I), which is simply a network that connects other networks. There were (lowercase i) internets long before the (uppercase I) Internet as we know it today became dominant in the 1990’s.

C. Today, we want to look briefly at how this reality came about.

1. The history leading up to the Internet as we know it.
2. How the Internet copes with diversity.

If you think about it for a moment, the diversity in the Internet is almost incredible.

a) There are a wide variety of types of devices connected to the Internet

(1) Different brands of servers, desktop and laptop computers

(2) Many other kinds of devices.

(a) Cell phones

(b) Entertainment devices such as game consoles, digital VCRs, etc

(c) As a rather extreme example, there is even an Internet toaster that obtains a weather forecast from a meteorological server and then burns an image of the forecast on your morning toast. (Cited by Kurose/Ross p. 10)

b) These devices communicate with each other over a wide variety of types of transmission media (cable systems, phone lines, wireless ...)

c) These devices cooperate to accomplish a wide variety of tasks web access, email, database access ...

D. We will also give a quick overview of the “Internet world” as it exists today. In the process, we will look at the differing service requirements of various kinds of Internet applications.

II. The History of the Internet

A. As I indicated earlier, the idea of a computer network and the idea of an internet actually predate what we know today as the Internet.

B. The earliest computer networks

1. Often set up to connect computers within a single organization (business)

a) Could be classified as LANs (local area networks) or WANs (wide area networks) - e.g. a corporation with multiple sites might use a WAN to support communication between employees at different locations.

b) Supported services like email, information sharing.

Example: Gordon's earliest network provided email between users at Gordon, using usernames that didn't require an "@" symbol - e.g. my email address was simply "BJORK" at one time.

c) Generally used proprietary technologies -

(1) e.g. Gordon's earliest network used a product known as "DECNET", which was a product sold by Digital Equipment Corporation - the manufacturer of our main computer systems.

(2) IBM had a proprietary technology known as SNA (Systems Network Architecture)

(3) Apple had a proprietary technology known as Appletalk

...

2. There were also various dialup networks that supported services like email and newsgroups (e.g. compuserve, genie, prodigy, aol ...)

C. In 1967, the Advanced Research Projects Agency of the Department of Defense (ARPA) proposed the creation of a network to connect researchers working under defense contracts. By 1969, its first implementation connected four universities: UCLA, UCSB, SRI, and the University of Utah.

1. The ARPANET was the first network to use a technique known as “packet switching”. Because this approach is pervasive in networks today, it is useful to compare it to another approach known as circuit switching.

a) The telephone network is an example of a circuit switched network. In such a network

(1) A connection must be set up before communication can take place (e.g. by dialing a phone number at one end and answering at the other end)

(2) Once the connection is established information sent at one end of the connection will be received at the other end without any further identification of the intended recipient.

(3) When communication is done, it is necessary to explicitly terminate the connection (e.g. by hanging up).

b) Packet switching resembles the approach used with postal mail: information is sent over the network in discrete packets.

(1) Each packet carries with it the address of its destination.

(2) Different packets between two communicating systems may actually travel over different routes, and two packets sent at about the same time may arrive in a different order from the order in which they were sent.

(3) Short messages can be sent as a single packet; but longer messages may need to be broken up into several smaller packets.

c) A key difference between the two approaches is that with circuit switching system resources remain dedicated to a particular connection during the entire time it is in existence, whereas with packet switching system resources are used on a per packet basis.

This is why voice telephone service is charged on a “per minute” basis, while postal mail charges per letter and text messaging (which uses packet switching) charges per message.

2. The ARPANET (as it came to be called) grew to over 100 nodes, but was always limited to defense contractor. However, it is considered the major precursor of today’s Internet because many of the key technologies used by the Internet (e.g. TCP/IP) were first developed there.

3. A number of other similar networks arose to support non-defense research, such as the NSFNET and CSNET.

4. What eventually emerged in the early 1990’s was the concept of the Internet as a network of networks. What we know today as the Internet is actually the successor to these earlier networks.

D. Probably no single development contributed as much to the popularization of the Internet as the development of the World-Wide Web.

1. This was originally developed by Tim Berners-Lee in 1989 at CERN (the European Organization for Nuclear Research located in Switzerland.)
2. A key turning point was the development of the Mosaic web browser at the National Center for Supercomputers (NCSA) in 1993. (Many of the original developers of Mosaic worked on a browser known as Netscape which became the precursor of the Mozilla project and Firefox)

III.How Diverse Systems can Communicate: Protocols

- A. So how do diverse systems communicate with one another?

They do so using protocols.

- B. A protocol is simply a set of rules that facilitate communication.

Example: PROJECT Figure 1.2 from Kurose/Ross

- C. In the world of computer networks, there are protocols at multiple levels

1. There are physical protocols, that specify how systems represent binary information. For two devices to connect to each other, they must both represent binary information in the same way.
2. There are protocols that govern how messages are routed over the network..

Example from postal mail: the destination address goes near the center of the envelope, and the return address goes in the upper-left-hand corner. Imagine what would happen if you switched the two on a letter you sent!

3. There are protocols that govern how packets are interpreted by various applications.
 - a) Web requests and pages follow a protocol known as http (hypertext transfer protocol). (That's what the http in a URL stands for. Note that there is a different protocol (https) used for encrypted web operations)
 - b) Email messages follow a protocol known as smtp (simple mail transfer protocol)
 - ...

IV. How Protocols are Agreed Upon: Standards

A. Of course, protocols are only useful if both parties agree to use the same protocol.

B. Standards are of two types:

1. De facto standards are widely used, though they may originate from a single source.

Example: Microsoft Word format is a defacto standard for word-processing, widely accepted because Word is so commonly used.

2. De jure standards are formally promulgated by a standards body, typically after a process that involves opportunity for comment from the parties most affected by the standard.

C. In the Internet world, commonly-used standards come from a variety of standards bodies - many of which deal with many areas, not just the Internet.

1. Standards bodies that deal with a broad range of standards (not just electronics)

a) ISO - International Standards Organization. It is an international body whose membership is drawn mainly from various national standards bodies.

(1) In the world of networks, ISO standardized a reference model for network protocol hierarchies which we will talk about later (the ISO Open Systems Interconnect Model).

(2) In the world of computers, ISO is also responsible for a standard format for CDs and has created standards for some programming languages.

b) ANSI (American National Standards Institute) is the US affiliate of ISO.

(1) In the world of computers, it is the body that has standardized the ASCII code that is commonly used for the English alphabet.

(2) ANSI has also standardized a number of programming languages.

2. Standards bodies that deal primarily with electronics broadly (not limited to telecommunications).

a) IEEE - the Institute of Electrical and Electronics Engineers.

b) EIA - Electronics Industry Association

3. Standard bodies that deal primarily with telecommunications and Internet-related standards

a) ITU-T - The International Telecommunications Union, and its committee CCITT - the Consultive Committee for International Telegraphy and Telephony.

b) W3C - The World-Wide-Web Consortium.

D. In the Internet world, standards go through a process that is managed by the Internet Society through its Internet Activities Board (IAB) and the Internet Engineering Task Force (IETF).

1. The standardization process is built on documents known as “Requests for Comment” (RFC). (The name reflects the history of this process).

2. The process as it exists today is itself defined by an RFC:

PROJECT RFC 2026. (Note references to updates that pertain to specific portions)

3. Not all RFC’s pertain to standards. (In fact, there is a tradition of a joke RFC being published each year on April 1; an obituary of John Postel, the original editor of the RFC series, is RFC 2468!)

4. Any RFC can be accessed by a URL of the form `http://tools.ietf.org/html/rfcXXXX`

(Demo accessing RFC2468 this way)

5. When an RFC pertains to a standard, it goes through a process (known as a sequence of maturity levels) that may - or may not - result in it becoming a standard

PROJECT Forouzan figure 1.2.

V. The World of the Internet

A. The Internet - as it exists today - is a network of networks that can be depicted as follows:

PROJECT Forouzan figure 1.1

B. Nodes on the Internet are identified in one of two ways.

1. Each node has a unique address known as an IP number.

a) Most IP addresses are based on a standard known as ipv4, which uses 4 bytes. (However, v4 refers to this being version 4 of the protocol - not the fact that it uses 4 bytes!)

(1) IP addresses are “owned” by ISPs and are assigned by higher level ISPs to lower level ISPs and local users. IP addresses are used to actually route traffic over the network.

(2) The most common way to write an IP address is as dotted octets. (The term octet comes from the fact that each number is the decimal equivalent of 8 bits)

SHOW Network pane in System Preferences - my IP

(3) While this potentially provides 4 billion addresses, not all addresses are actually used.

(a) Some are reserved for special purposes.

(b) An IP number is actually regarded as being divided into two parts - a network number and the IP number of a specific host on that network. (Recall that the Internet is a network of networks).

SHOW how this is specified by the subnet mask for my computer.

(4) Moreover, even 4 billion addresses are actually not enough for all of the hosts connected to the internet (not just computers, but smart phones, appliances ...)

(5) The usefulness of ipv4 has been extended by using a technique known as Network Address Translation (NAT). I will use the specific way this works out at Gordon as an example:

(a) Gordon actually owns only 70 IP addresses
(216.236.251.129 .. 198)

(b) Machines on campus have IP numbers of the form
172.XXX.YYY.ZZZ or 192.168.YYY.ZZZ or
10.10.XXX.YYY - all in ranges of IP numbers reserved
by ipv4 for private networks.

(c) When a machine on campus wants to contact an off-campus site, Gordon's gateway replaces its "on campus" IP number with one of the external IPs Gordon owns. When a response comes back, Gordon's gateway replaces the "external" IP with the "internal" IP number and sends it to the originating system.

(d) Gordon also permanently assigns several of its external IPs to systems on campus (e.g. an incoming message on 216.236.251.139 is sent to the CS web server, known by the internal IP address 100.10.150.17 and by the URL www.cps.gordon.edu.)

b) One will sometimes see a range of IP addresses denoted by a notation such as 172.16.0.0/12. This specifies a range in which the upper 12 bits correspond to 172.16.0.0 but the remaining 20 bits can take on any value - so the range goes from

10101010 00010000 00000000 00000000 (172.16.0.0) to
10101010 00011111 11111111 11111111 (172.31.255.255)

- c) There is a newer standard - called ipv6 - that uses 128 bit IP numbers. While this standard was defined in 1998, it is only slowly being implemented. (Actually, when I first wrote this lecture in 2011, I used the phrase "slowly being implemented" then. Alas, that remains the case today!)
2. Though the IP address of a node is what is actually used in Internet routing, hosts are more often identified by a URL (e.g. www.gordon.edu or go.gordon.edu).
- a) A full URL typically consists of three or four parts.

Example: When this lecture is posted online, its URL will be <http://www.cs.gordon.edu:80/courses/cps221/lectures-2014/Introduction%20to%20Net-Centric%20Computing.pdf>

`http:` specifies the protocol

`www.cs.gordon.edu` specifies the host machine

`80` specifies the port number (generally omitted since often implied by the protocol - e.g 80 is the port normally used for http)

`courses/cps221/lectures-2014/Introduction%20to%20Net-Centric%20Computing.pdf` specifies a path on the host machine

- b) To actually route traffic to a system identified by a URL, the host name portion of the URL must be translated into an IP number. This is handled by a system designated as a name server, which maintains a table of URLs and the corresponding IP numbers.

(1) To prevent the name tables from becoming very cluttered, a hierarchical scheme based on domains is actually used. Each domain has one or more domain name servers (DNS)

(a) There is a DNS for the `edu` domain, which provides the IP number for a DNS for the `gordon.edu` domain.

(b) There is a DNS for the `gordon.edu` domain, which provides the IP number for a DNS for the `cs.gordon.edu` domain.

(c) There is a DNS for the `cs.gordon.edu` domain, which actually provides the IP number for the `www.cs.gordon.edu` host.

(2) While Gordon is responsible for all URLs within the `gordon.edu` domain (including the `cs` subdomain), top-level domains are managed by the Internet Corporation for Assigned Names and Numbers (ICANN) through a large number of domain name registrars that are allowed to register names in various domains.

c) Note that, while the host name portion of a URL must be mapped to a specific IP number, several different host name portions of a URL might all refer to the same IP - e.g.

`files.cs.gordon.edu` and `joshua.cs.gordon.edu` both refer to the same machine (`10.100.150.16`)

C. Of course, there are a huge variety of applications that utilize the Internet. One way of categorizing them is in terms of their service requirements.

1. Bandwidth - the rate of information transfer needed - often specified in units such as MB / second.
2. Tolerance for end-to-end delay (the interval between the time the information is sent and the time it is received.)
3. Tolerance for loss of information in transit (dropped packets).
4. Examples:
 - a) Electronic mail requires relatively small bandwidth; has high tolerance for delay, but low tolerance for information loss.

(In fact, in contrast with many other applications where larger-size units of information are transmitted, the loss of a packet could result in the lost of an entire email with no indication to either the sender or the receiver that this has occurred.)
 - b) The bandwidth requirements for web page access varies depending on the page contents; there is moderate tolerance for delay (up to a few seconds), and some information loss can be tolerated (the requester can request the page again; if a portion of a page is lost the recipient's browser will typically detect this and request a re-send of the lost portion.)
 - c) For streaming multimedia the bandwidth requirements tend to be much higher and the tolerance for delay tends to be smaller, though some loss of information can be tolerated.
 - d) For interactive applications (e.g. Skype), high bandwidth is again needed, tolerance for delay is very small, though again some loss of information can be tolerated.