INFO/CS 4302
Web Information Systems

FT 2012
Week 2: Internet History & Architecture (Part1)

Theresa Velden
Introduction

• Popular opinions on how the Internet was invented:
Introduction

- Popular opinions on how the Internet was invented:
  - O Tim Berner’s Lee
Introduction

• Popular opinions on how the Internet was invented:
  O Tim Berner’s Lee
  O Al Gore
Introduction

• Popular opinions on how the Internet was invented:
  O Tim Berner’s Lee
  O Al Gore
  O Xerox
Introduction

• Popular opinions on how the Internet was invented:
  O Tim Berner’s Lee
  O Al Gore
  O Xerox
  O The military
Introduction

• Popular opinions on how the Internet was invented:
  O Tim Berner’s Lee
  O Al Gore
  O Xerox
  O The military
  O Many people: government funded researchers working with corporations
Introduction

• What are some of the technical key components that make up the Internet?
The History of The Internet
The History of The Internet

Movie by Melih Bilgil

http://www.youtube.com/watch?v=qpcU25OAcj0
Internet Architecture
De-centralized network design

(Paul Baran, RAND early 1960s)

Images: RAND (http://www.rand.org/about/history/baran.list.html)
De-centralized network design

"hot-potato routing"
(Paul Baran, RAND early 1960s)
packet switching (early 1960’s)

Kleinrock (MIT)    | Paul Baran (RAND) | Donald Davies (NPL, UK)

- Communication divided in equal sized packages (up to 1 Kilobyte)
- Each packet with a header with packet, source and destination information
- Each Internet router determines path
- Message (re)assembly at destination
Circuit switching
circuit switching vs. packet switching

- Dedicated line for entire conversation incl. silences (-)
- Less efficient (-)
- More reliable (+)

- Full band with for each packet (+)
- More efficient (+)
  - Users share network
  - Resource use only when needed
- Less reliable (-)
Routing algorithms

TCP & IP
(1978)

Vinton Cerf
Bob Kahn
TCP/IP Protocol Suite:

- User application
- Packet re-assembly
- Packet delivery
- Hardware address resolution

Diagram:

- Application Layer
  - DHCP, DNS, FTP, HTTP, IMAP, POP, Telnet, etc.

- Transport Layer
  - TCP, UDP, SCTP, RSVP, etc.

- Internet Layer
  - IP (IPv4, IPv6), ICMP, IPsec, etc.

- Link Layer
  - ARP/inARP, Tunnels (L2TP), MAC, etc.
IP = Internet Protocol

• Rules for sending and receiving data: **addressing** and **routing**
• Datagram: header + payload

Source: RFC 791
IP = Internet Protocol

- Hides underlying heterogeneity (ether, cable, wireless…)
- Connection-less (packets routed individually)
- ‘best effort at delivery’, unreliable (packets may be dropped)
- Unique hierarchical addressing scheme for computers on network (needed for routing)
IP Addresses:
5 regional internet registries (RIRs)

Global pool of IPv4 addresses was exhausted early in 2011
IANA: Internet Assigned Numbers Authority
## IP Addresses

<table>
<thead>
<tr>
<th></th>
<th>Internet Protocol version 4 (IPv4)</th>
<th>Internet Protocol version 6 (IPv6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deployed</td>
<td>1981</td>
<td>1999</td>
</tr>
<tr>
<td>Address Size</td>
<td>32-bit number</td>
<td>128-bit number</td>
</tr>
<tr>
<td>Address Format</td>
<td>Dotted Decimal Notation: 192.149.252.76</td>
<td>Hexadecimal Notation: 3FFE:F200:0234:AB00:0123:4567:8901:ABCD</td>
</tr>
<tr>
<td>Prefix Notation</td>
<td>192.149.0.0/24</td>
<td>3FFE:F200:0234::/48</td>
</tr>
<tr>
<td>Number of Addresses</td>
<td>$2^{32} = \sim 4,294,967,296$</td>
<td>$2^{128} = \sim 340,282,366,920,938,463,463,374,607,431,768,211,456$</td>
</tr>
</tbody>
</table>

TCP = Transfer Control Protocol

- Flow controlled (to avoid congestion)
- Reliable
  - No data lost or duplicated (re-transmit)
  - Serial numbers for packets
  - Different from UDP: ‘state-less’ (careless) used for streaming
- Connection-oriented (‘virtual circuits’)

UDP

TCP

Connection-oriented

Virtual circuits

Reliable

Flow controlled

Avoids congestion

No data loss or duplication

Sequential packet numbers

UDP

State-less

Careless

Streaming
DNS = Domain Name Resolution

- Invented in 1983 by Paul A Mockapetris at University of Southern California
- Allows us to memorize ‘names’ rather than 32 bit or 128 bit numbers
- Domain names: define realm of autonomy
- Hierarchical:
  - Top level domains: country codes, e.g. .jp, .uk, or generic: .com, .edu, .org etc.
  - Second-level domains: cornell.edu
  - Third-level domains: e.g. cs.cornell.edu
- Distributed and dynamic database (name servers)
  - 13 root name servers http://www.root-servers.org/
DNS – Domain Name Resolution

Realization of The Internet

- 1962 J.C.R. Licklider: series of memos ‘Galactic Network’ (MIT), ARPA IPTO Director
- 1967 ‘plan for ARPANET’ by Lawrence G. Roberts (MIT), Designer and Manager
  - Motivation: time sharing and communication to avoid duplication of efforts
- 1969 first 4 computers networked together (UCLA, Stanford, Utah, Santa Barbara)
- 1973 Ethernet (optical cable)
- 1989 ARPANET renamed ‘Internet’, commercial ISPs emerge
- 1995 NSFNET decommissioned, commercial restrictions lifted
Internet Governance & Standardization
Internet Governance

By independent, non-for-profit membership organizations:
• Open
• Self-regulated

Source:
http://www.internetsociety.org/internet/how-it-works/technical-aspects
Internet Governance

- Internet Corporation for Assigned Names and Numbers (ICANN)
  - DNS administration | IP # assignment | Protocol #’s | port #’s
  - Operates under contract with U.S. Department of Commerce (this is controversial!) [http://www.icann.org/](http://www.icann.org/)

Further reading: ‘A concise guide to the major internet bodies’ by Alex Simonelis (2005) [http://ubiquity.acm.org/article.cfm?id=1071915](http://ubiquity.acm.org/article.cfm?id=1071915)
FILE TRANSFER PROTOCOL

INTRODUCTION

The objectives of FTP are 1) to promote sharing of files (computer programs and/or data), 2) to encourage indirect or implicit (via programs) use of remote computers, 3) to shield a user from variations in file storage systems among Hosts, and 4) to transfer data reliably and efficiently. FTP, though usable directly by a user at a terminal, is designed mainly for use by programs.

The attempt in this specification is to satisfy the diverse needs of users of maxi-Hosts, mini-Hosts, and TIPs, with a simple, and easily implemented protocol design.

This paper assumes knowledge of the following protocols described in the ARPA Internet Protocol Handbook.

The Transmission Control Protocol
The TELNET Protocol

DISCUSSION

In this section, the terminology and the FTP model are discussed. The terms defined in this section are only those that have special significance in FTP. Some of the terminology is very specific to the FTP model; some readers may wish to turn to the section on the FTP model while reviewing the terminology.

TERMINOLOGY

ASCII
The ASCII character set as defined in the ARPA Internet Protocol Handbook. In FTP, ASCII characters are defined to be the lower half of an eight-bit code set (i.e., the most significant bit is zero).

access controls
Access controls define users' access privileges to the use of a system, and to the files in that system. Access controls are necessary to prevent unauthorized or accidental use of files. It is the prerogative of a server-FTP process to invoke access controls.
The Network Working Group seems to consist of Steve Carr of Utah, Jeff Rulifson and Bill Duvall at SRI, and Steve Crocker and Gerard Deloche at UCLA. Membership is not closed.

The Network Working Group (NWG) is concerned with the HOST software, the strategies for using the network, and initial experiments with the network.

Documentation of the NWG’s effort is through notes such as this. Notes may be produced at any site by anybody and included in this series.
The end-to-end principle

The function in question can completely and correctly be implemented only with the knowledge and help of the application standing at the end points of the communication system. Therefore, providing that questioned function as a feature of the communication system itself is not possible. (Sometimes an incomplete version of the function provided by the communication system may be useful as a performance enhancement.) [1].

- Functionality added at end points and transparent at mid-points
- Maximal flexibility for client to customize and innovate at the end points
Recap (take away)
Recap (take away):

- Internet ≠ World Wide Web

- Principles of Internet Design
  - Distributed
  - Open
  - End-to-end
Outlook

• Thursday:
  o A look at the Internet today
  o Hands-on: Internet protocols

• Next Week:
  o The History and the Architecture of the World Wide Web
Reminders

• Homework 1 is due on Sunday 11:59pm (submission via the CMS)
• Subscribe to the course on piazza
Resources used

- Internet Pioneers: [http://www.ibiblio.org/pioneers/index.html](http://www.ibiblio.org/pioneers/index.html)
- RAND corporation website
- Lecture slides INFO/CS4302 by Carl Lagoze