The Internet

Nelson Padua-Perez

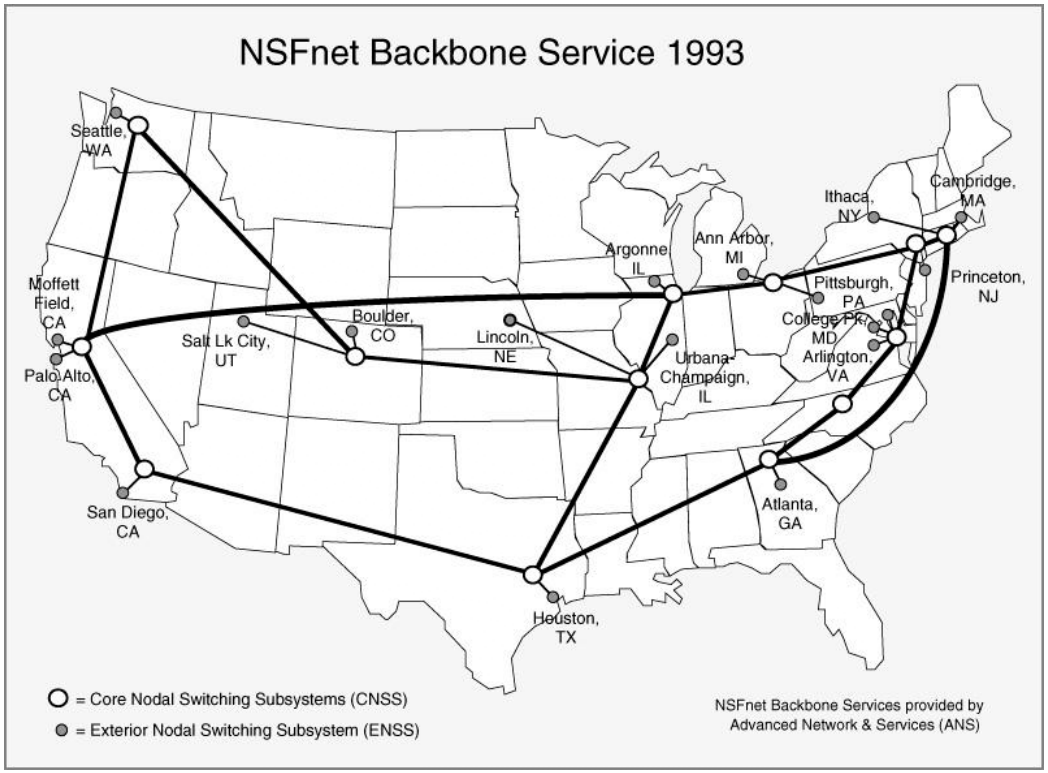
William Pugh

Department of Computer Science

University of Maryland, College Park

Networking Background

- **Definition**
  - **Set of computers using common protocols to communicate over connecting media**
- **History**
  - **1969 ARPANET**
  - **1986 NSFnet**
  - **1995 Internet**



Networking Concepts

- Internet addresses
  - IPv4 vs IPv6
- Ports
- TCP vs. UDP
  - Reliability
  - Connection vs. packet oriented
- Sockets
- URLs
- NAT boxes
- Firewalls

Internet Protocol (IP) Address

- Unique address for machine on internet
  - Get from ISP when connecting to internet
  - Allows network to find your machine
  - 32-bit unsigned integer ⇒ 128.8.128.8
  - Domain name service maps name to ip address
- Name and address for local machine
  - localhost
  - 127.0.0.1

Internet (IP) Address

- Domain Name System (DNS)
  - DNS servers on internet
  - Can look up IP address associated with name
  - DNS server may need to query other DNS servers
    - edu DNS server queries umd.edu server to find cs.umd.edu
- Machine can have multiple IP addresses
  - Virtual machines

Internet (IP) Address

- **Problem**
  - Running out of 32-bit IP addresses
  - Exacerbated by initial address allocation
    - Stanford & MIT given more IP addresses than China
- **Switching to 128-bit IP addresses in IPv6**
  - 1+ million addresses per square meter on Earth

Ports

- Abstraction to identify (refine) destination
  - Provide multiple communications channels/services at single IP address
  - think port # = extension #
- Format
  - Unsigned 16-bit integer (0 to 65,535)
  - Ports 0 to 1023 are privileged ports
- Many ports pre-assigned to important services
  - 21 ftp (file transfer)
  - 23 telnet (remote terminal)
  - 25 SMTP (email)
  - 80 http (web)
  - …

Ways to communicate

- **TCP**
  - two-way streaming connection between two machines
- **UDP**
  - send of one packet of information from one machine to another
- **UDP multicast**
  - send of one packet of information from one machine to all machines on the local area network

UDP is unreliable

- **UDP is build on top of the basic internet protocol**
  - **Send a packet, hope it gets there**
  - **No notification if it gets there**
  - **No notification if it gets dropped**
  - **Packets can arrive late or out of order**

  - **Intermediate routers *will* drop packets if the network is congested.**

TCP is reliable

- TCP is a reliable system built out of unreliable parts
- Two way stream of bytes
- Uses sequence numbers and acknowledgements and retransmissions to ensure that the packets do arrive and arrive in order
  - you might have a failure, but you will be told that there was a communication failure

Wikipedia: Internet Sockets

- Asocket is a software abstraction, designed to provide a standard application programming interface (API) for sending and receiving data across a computer network.
- Sockets are designed to accommodate virtually any networking protocol, though in practice are used mostly for the internet suite of protocols (i.e. TCP/IP.)
- Sockets are implemented in many different computer languages and for most operating systems. In RFC documents relating to TCP or UDP, a socket on a certain host is defined as the combination of an IP address, a protocol, and a port number.
- The BSD operating system introduced network sockets in 1983.

User/Unreliable Datagram Protocol

- Just name the ip address and port you want to ship the data to
- Provide the data
  - typically < 512 bytes, but can be up to 65Kbytes
- UDP packet contains a sending ip address and port
  - can be used to indicate where responses should be sent
  - can be spoofed

Transmission Control Protocol

- Server listens on a particular port
  - only one process can listen for TCP connections on each port
- Client says: Please connect to port 80 on www.cnn.com
  - a whole bunch of packets get exchanged to establish a communication channel
  - temporary port numbers are generated on both ends and used for this communication channel
  - After setup completes, server continues to listen for new requests to establish communication channel
  - when channel no longer needed, closed and ports recycled.

UDP / TCP

- UDP is lower overhead
  - much more efficient for small msgs
- Is late data useful?
  - many VOIP systems use UDP, missing packets are compensated for
    - noise/silence

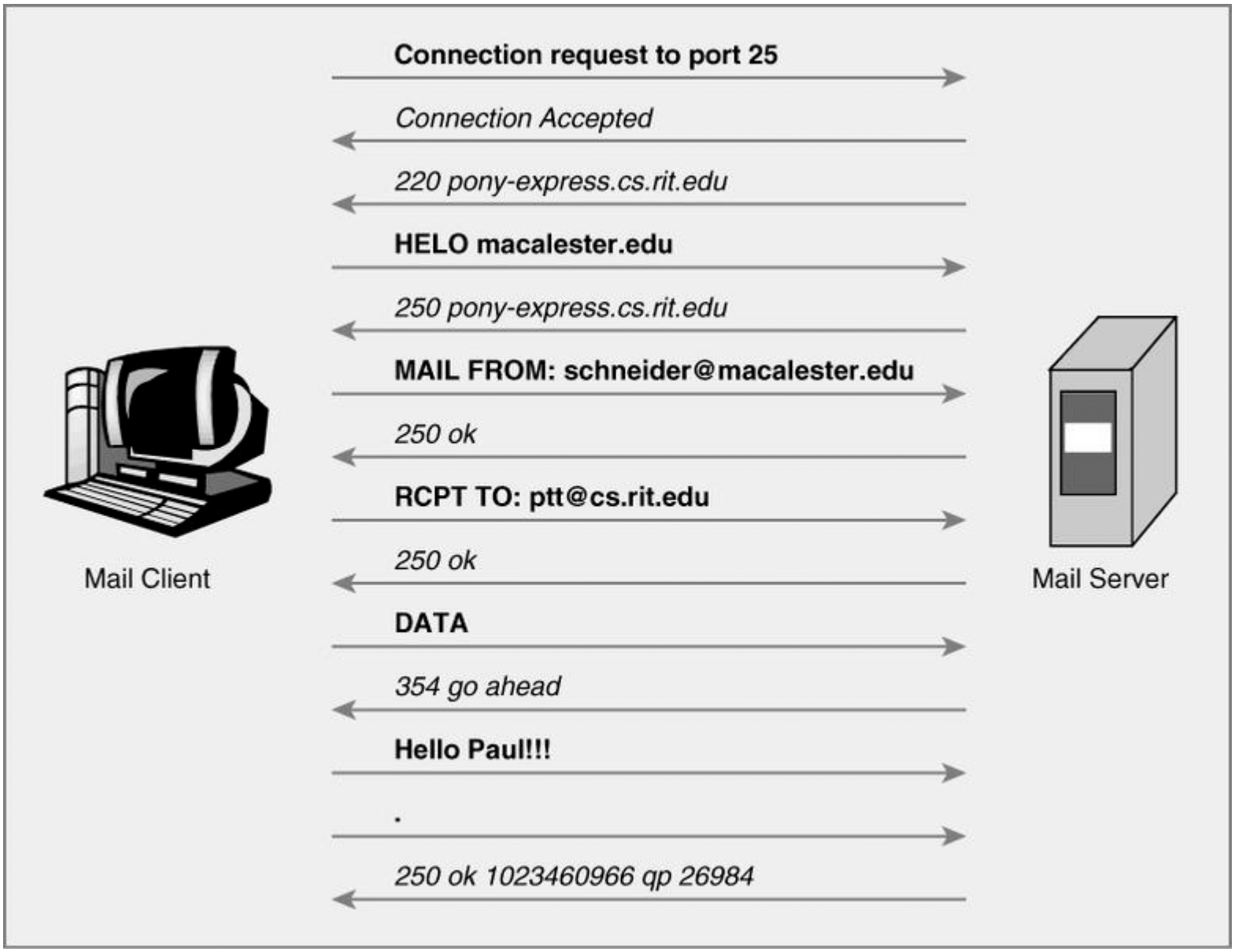
Uniform Resource Locators (URLs)

- **Represent web resources**
  - **Web pages**
  - **Arbitrary files**
  - **…**
- **Examples**
  - **http://www.cs.umd.edu/index.html**
  - **ftp://www.cs.umd.edu/pub/doc/csd_policies.pdf**
  - **https://login.yahoo.com/**
  - **file://dir/my.txt**

Uniform Resource Locators (URLs)

- Consists of
  - Protocol
    - http
    - ftp
    - https (secure http)
    - file
    - …
  - IP address (or domain name)
  - Port (optional)
    - http://www.cs.umd.edu:80/
  - protocol specific information

Protocols – Email Delivery



Protocol - HTTP GET

- **Client connects to server on port 80**

```
GET /~pugh/index.html HTTP/1.0
blank line
```

- **Server responses with HTTP headers**

```
HTTP/1.1 200 OK
Date: Mon, 20 Feb 2006 03:47:44 GMT
Server: Apache
Last-Modified: Wed, 15 Feb 2006 01:17:09 GMT
ETag: "9b2b1c-948-1222af40"
Accept-Ranges: bytes
Content-Length: 2376
Connection: close
Content-Type: text/html; charset=ISO-8859-1
```

- **Followed by blank line, then contents of response**

NAT boxes

- Network address translation
- Used, for example, in your house, allows several different computers to all have their own internal IP address, and the NAT box mergers and manages these so that they appear to be one IP address on the Internet
  - the one assigned to you by your ISP
- Computers on the other side of the NAT box generally can't initiate communication with you
  - for bad and good

More terms and acronyms

- **Firewall**
  - system that allows only certain communications to pass through part of a network
- **DHCP**
  - dynamic assignment of IP address