2. Application Layer
Chapter 2: Application Layer

application
transport
network
link
physical
Chapter 2: Application Layer

Our Goal:

• Conceptual, implementation aspects of network application protocols
  • Transport-layer service models
  • Client-server paradigm
  • Peer-to-peer paradigm
  • Content distribution networks

• Learn about protocols by examining popular application-level protocols
  • HTTP
  • SMTP / POP3 / IMAP
  • DNS

• Creating network applications
  • socket API
Chapter 2: Outline

- Principles of network applications
- Socket Programming with UDP and TCP
- Web and HTTP
- Electronic Mail (SMTP, POP3, IMAP)
- DNS
- P2P Applications
- Video Streaming and Content Distribution Networks
Creating a network app

Write programs that:

• run on (different) *end systems*

• communicate over network

• e.g., web server software communicates with browser software

No need to write software for network-core devices

• network-core devices do not run user applications

• applications on end systems allows for rapid app development, propagation
Application architectures

Possible structure of applications:

• Client-server
• Peer-to-peer (P2P)
• Hybrid
Client-server architecture

**server:** always-on host
- permanent IP address
- data centers for scaling

**clients:**
- communicate with server
- may be intermittently connected
- may have dynamic IP addresses
- do not communicate directly with each other
P2P architecture

• *no* always-on server

• arbitrary end systems directly communicate

• peers request service from other peers, provide service in return to other peers
  
  • *self scalability* – new peers bring new service capacity, as well as new service demands
P2P architecture

• *no* always-on server
• arbitrary end systems directly communicate
• peers request service from other peers, provide service in return to other peers
  • *self scalability* – new peers bring new service capacity, as well as new service demands
P2P architecture

• *no* always-on server
• arbitrary end systems directly communicate
• peers request service from other peers, provide service in return to other peers
  • *self scalability* – new peers bring new service capacity, as well as new service demands
• Peers are intermittently connected and change IP addresses
  • *Complex Management*
Hybrid of client-server and P2P

Skype
- Internet telephony app
- Finding address of remote party: centralized server(s)
- Client-client connection is direct (not through server)

Instant messaging
- Chatting between two users is P2P
- Presence detection/location centralized:
  - User registers its IP address with central server when it comes online
  - User contacts central server to find IP addresses of buddies
Case Study: Napster Vs Gnutella

Any problem with this architecture?
Processes communicating

**process:** program running within a host

- within same host, two processes communicate using inter-process communication (defined by OS)
- processes in different hosts communicate by exchanging messages

**clients, servers**

**client process:** process that initiates communication

**server process:** process that waits to be contacted

aside: same process can be both a client and a server for different connections; e.g., in P2P networks
Sockets

- process sends/receives messages to/from its **socket**
- socket analogous to mailbox
  - App A puts message in mailbox/socket
  - App A relies on transport infrastructure to pick up message from A’s socket/mailbox and deliver it to B’s socket
Addressing processes

• to receive messages, process must have *identifier*

• host device has unique 32-bit IP address

• **Q:** does IP address of host on which process runs suffice for identifying the process?
  
  ▪ **A:** no, *many* processes can be running on same host

• *identifier* includes both IP address and port numbers associated with process on host.

• example port numbers:
  
  • HTTP server: 80
  • mail server: 25

• to send HTTP message to gaia.cs.umass.edu web server:
  
  • IP address: 128.119.245.12
  • port number: 80

• more shortly...
App-layer protocol defines

- types of messages exchanged,
  - e.g., request, response
- message syntax:
  - what fields in messages & how fields are delineated
- message semantics
  - meaning of information in fields
- rules for when and how processes send & respond to messages

open protocols:
- defined in RFCs
- allows for interoperability
- e.g., HTTP, SMTP

proprietary protocols:
- e.g., Skype
What transport service does an app need?

**Data loss**
- some apps (e.g., file transfer, web transactions) require 100% reliable data transfer
- other apps (e.g., audio) can tolerate some loss

**Timing**
- some apps (e.g., Internet telephony, interactive games) require low delay to be “effective”

**Throughput**
- some apps (e.g., multimedia) require minimum amount of throughput to be “effective”
- other apps (“elastic apps”) make use of whatever throughput they get

Why is bandwidth different from timing constraints?
## Transport service requirements: common apps

<table>
<thead>
<tr>
<th>application</th>
<th>data loss</th>
<th>throughput</th>
<th>time sensitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>file transfer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e-mail</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web documents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>real-time audio/video</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stored audio/video</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>interactive games</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>text messaging</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Transport service requirements: common apps

<table>
<thead>
<tr>
<th>application</th>
<th>data loss</th>
<th>throughput</th>
<th>time sensitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>file transfer</td>
<td>no loss</td>
<td>elastic</td>
<td>no</td>
</tr>
<tr>
<td>e-mail</td>
<td>no loss</td>
<td>elastic</td>
<td>no</td>
</tr>
<tr>
<td>Web documents</td>
<td>no loss</td>
<td>elastic</td>
<td>no</td>
</tr>
<tr>
<td>real-time audio/video</td>
<td>loss-tolerant</td>
<td>audio: 5kbps-1Mbps, video:10kbps-5Mbps</td>
<td>yes, 100’s msec, yes, few secs</td>
</tr>
<tr>
<td>stored audio/video</td>
<td>loss-tolerant</td>
<td>same as above</td>
<td>yes, 100’s ms</td>
</tr>
<tr>
<td>interactive games</td>
<td>loss-tolerant</td>
<td>few kbps up</td>
<td>yes and no</td>
</tr>
<tr>
<td>text messaging</td>
<td>no loss</td>
<td>elastic</td>
<td>no</td>
</tr>
</tbody>
</table>
Internet transport protocols services

TCP service:

- **reliable transport** between sending and receiving process
- **flow control**: sender won’t overwhelm receiver
- **congestion control**: throttle sender when network overloaded
- **does not provide**: timing, minimum throughput guarantee, security
- **connection-oriented**: setup required between client and server processes

UDP service:

- **unreliable data transfer** between sending and receiving process
- **does not provide**: reliability, flow control, congestion control, timing, throughput guarantee, security, or connection setup,

**Q**: why bother? Why is there a UDP?
**Internet apps: application, transport protocols**

<table>
<thead>
<tr>
<th>application</th>
<th>application layer protocol</th>
<th>underlying transport protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>e-mail</td>
<td>SMTP [RFC 2821]</td>
<td></td>
</tr>
<tr>
<td>remote terminal access</td>
<td>Telnet [RFC 854], SSH</td>
<td></td>
</tr>
<tr>
<td>Web</td>
<td>HTTP [RFC 2616]</td>
<td></td>
</tr>
<tr>
<td>file transfer</td>
<td>FTP [RFC 959]</td>
<td></td>
</tr>
<tr>
<td>streaming multimedia</td>
<td>HTTP (e.g., YouTube), RTP [RFC 1889]</td>
<td></td>
</tr>
<tr>
<td>Internet telephony</td>
<td>SIP, RTP, proprietary (e.g., Skype)</td>
<td></td>
</tr>
<tr>
<td>naming</td>
<td>DNS</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 2: Outline

✓ Principles of network applications
❑ Socket Programming with UDP and TCP
❑ Web and HTTP
❑ Electronic Mail (SMTP, POP3, IMAP)
❑ DNS
❑ P2P Applications
❑ Video Streaming and Content Distribution Networks
Socket programming

**goal:** learn how to build client/server applications that communicate using sockets

**socket:** door between application process and end-to-end transport protocol
Two socket types for two transport services:

- **UDP**: unreliable datagram (User Datagram Protocol)
- **TCP**: reliable, byte stream-oriented (Transmission Control Protocol)

Application Example:
1. client reads a line of characters (data) from its keyboard and sends data to server
2. server receives the data and converts characters to uppercase
3. server sends modified data to client
4. client receives modified data and displays line on its screen
Socket programming *with UDP*

UDP: no “connection” between client & server

- no handshaking before sending data
- sender explicitly attaches IP destination address and port # to each packet
- receiver extracts sender IP address and port# from received packet

UDP: transmitted data may be lost or received out-of-order

Application viewpoint:
- UDP provides *unreliable* transfer of groups of bytes (“datagrams”) between client and server
Client/server socket interaction: UDP

**server (running on serverIP)**

create socket, port= x:
serverSocket =
socket(AF_INET,SOCK_DGRAM)

read datagram from
serverSocket

write reply to
serverSocket
specifying
client address,
port number

**client**

create socket:
clientSocket =
socket(AF_INET,SOCK_DGRAM)

Create datagram with server IP and
port=x; send datagram via
clientSocket

read datagram from
clientSocket

specifying
client address,
port number

close
clientSocket
Example app: UDP client

C UDPClient

```c
#include<sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
#include <arpa/inet.h>
#include <netdb.h>

serverPort = 12000
clientSocket = socket(AF_INET, SOCK_DGRAM, 0)

sendto(clientSocket, message, msg_length, dest_info, dest_info_len);

close (clientSocket)
```

Port to create at the client
create an endpoint for communication (socket file descriptor)
Send message to server

Include C's socket headers

Server's information

Close socket
Example app: UDP server

**C UDP Server**

```c
#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
#include <arpa/inet.h>
#include <netdb.h>

serverPort = 12000
serverSocket = socket(AF_INET, SOCK_DGRAM, 0)
bind(serverSocket, own_addr_info, addr_len)

while (1) {
    numbytes = recvfrom(serverSocket, buf, buf_len, 0, (struct sockaddr *)&their_addr, &addr_len)
    print buf
}
```

local port number 12000 ➔ serverPort = 12000
create UDP socket FD ➔ serverSocket = socket(AF_INET, SOCK_DGRAM, 0)
Bind to the socket FD ➔ bind(serverSocket, own_addr_info, addr_len)
Read from UDP socket into buf, getting client’s address (client IP and port) ➔ numbytes = recvfrom(serverSocket, buf, buf_len, 0, (struct sockaddr *)&their_addr, &addr_len)
print buf ➔ Who did we receive message from?
Socket programming with TCP

client must contact server
• server process must first be running
• server must have created socket (door) that welcomes client’s contact

client contacts server by:
• Creating TCP socket, specifying IP address, port number of server process
• when client creates socket: client TCP establishes connection to server TCP

• when contacted by client, server TCP creates new socket for server process to communicate with that particular client
  • allows server to talk with multiple clients
  • source port numbers used to distinguish clients (more in Chap 3)

application viewpoint:
TCP provides reliable, in-order byte-stream transfer ("pipe") between client and server
Client/server socket interaction: TCP

server (running on hostid)  client

create socket, port=x, for incoming request:
serverSocket = socket()

wait for incoming connection request
connectionSocket = serverSocket.accept()

TCP connection setup

read request from connectionSocket
write reply to connectionSocket
close connectionSocket

create socket, connect to hostid, port=x
clientSocket = socket()

send request using clientSocket
read reply from clientSocket
close clientSocket
Example app: TCP client

C TCP Client

#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
#include <netdb.h>
#include <arpa/inet.h>

serverName = 'servername'
serverPort = 12000

clientSocket = socket(AF_INET, SOCK_STREAM, 0)
connect(clientSocket, server_info, server_info_len))

numbytes = recv(clientSocket, buf, buf_size, 0)
close(clientSocket)
Example app: TCP server

C TCP Server

```c
#include <sys/types.h>
#include <sys/socket.h>
#include <netinet/in.h>
#include <netdb.h>
#include <arpa/inet.h>

serverPort = 12000
serverSocket = socket(AF_INET, SOCK_STREAM, 0)
bind(serverSocket, serverAddr, serverAddr_len)
listen(serverSocket, queueSize)

while 1 {
    new_fd = accept(serverSocket, theirAddr, theirAddr_len)
    send(new_fd, "Hello, world!", 13, 0)
    close(new_fd);
}

close(serverSocket)
```
Some Details

Communication with diverse systems

Utility function for address and service translation

```
int getaddrinfo(const char *node, const char *service, const struct addrinfo *hints, struct addrinfo **res);
```

Utility functions for byte-ordering

```
uint32_t htonl(uint32_t hostlong);
uint16_t htons(uint16_t hostshort);
uint32_t ntohl(uint32_t netlong);
uint16_t ntohs(uint16_t netshort);
```

Networks use big-endian byte ordering

Little endian

Big endian

Utility functions for converting IPv4 and IPv6 addresses from binary to text form

```
const char *inet_ntop(int af, const void *src, char *dst, socklen_t size);
```
Chapter 2: Outline

- Principles of network applications
- Socket Programming with UDP and TCP
- Web and HTTP
- Electronic Mail (SMTP, POP3, IMAP)
- DNS
- P2P Applications
- Video Streaming and Content Distribution Networks
First, a review...

- **web page** consists of *objects*
- object can be HTML file, JPEG image, Java applet, audio file,...
- web page consists of *base HTML-file* which includes *several referenced objects*
- each object is addressable by a *URL*, e.g.,

```
www.someschool.edu/someDept(pic.gif)
```

*host name*  
*path name*
HTTP overview

HTTP: hypertext transfer protocol

• Web’s application layer protocol
• client/server model
  • client: browser that requests, receives, (using HTTP protocol) and “displays” Web objects
  • server: Web server sends (using HTTP protocol) objects in response to requests
HTTP overview

*Uses TCP:*
- client initiates TCP connection (creates socket) to server, port 80
- server accepts TCP connection from client
- HTTP messages (application-layer protocol messages) exchanged between browser (HTTP client) and Web server (HTTP server)
- TCP connection closed

*HTTP is “stateless”*
- server maintains no information about past client requests

Aside
protocols that maintain “state” are complex!
- past history (state) must be maintained
- if server/client crashes, their views of “state” may be inconsistent, must be reconciled
HTTP connections

**non-persistent HTTP**
- at most one object sent over TCP connection
  - connection then closed
- downloading multiple objects required multiple connections

**persistent HTTP**
- multiple objects can be sent over single TCP connection between client, server
Non-persistent HTTP

suppose user enters URL: www.someSchool.edu/someDepartment/home.index

1a. HTTP client initiates TCP connection to HTTP server (process) at www.someSchool.edu on port 80

1b. HTTP server at host www.someSchool.edu waiting for TCP connection at port 80. “accepts” connection, notifying client

2. HTTP client sends HTTP request message (containing URL) into TCP connection socket. Message indicates that client wants object someDepartment/home.index

3. HTTP server receives request message, forms response message containing requested object, and sends message into its socket

4. HTTP server closes TCP connection.

5. HTTP client receives response message containing html file, displays html. Parsing html file, finds 10 referenced jpeg objects

6. Steps 1-5 repeated for each of 10 jpeg objects
Non-persistent HTTP: response time

RTT (Round Trip Time): time for a small packet to travel from client to server and back

HTTP response time:

- one RTT to initiate TCP connection
- one RTT for HTTP request and first few bytes of HTTP response to return
- file transmission time
- non-persistent HTTP response time
  
  = 2RTT + file transmission time
Persistent HTTP

Persistent HTTP:

• server leaves connection open after sending response

• subsequent HTTP messages between same client/server sent over open connection

• client sends requests as soon as it encounters a referenced object

• as little as one RTT for all the referenced objects
Other optimizations

• Pipelining
  • Send several requests at once

![Diagram showing pipelining with time and RTT labels]

- initiate TCP connection
- RTT
- request file
- RTT
- request new files
Other optimizations

• Pipelining
  • Send several requests at once

• HTTP/2
  • Push resources
Other optimizations

• Pipelining
  • Send several requests at once

• HTTP/2
  • Push resources

• QUIC
  • Eliminate first RTT

- initiate QUIC connection
- request file

RTT

send file

push objects

time

time

HTTP request message

- two types of HTTP messages: request, response

- HTTP request message:
  - ASCII (human-readable format)

```
GET /index.html HTTP/1.1\r\nHost: www-net.cs.umass.edu\r\nUser-Agent: Firefox/3.6.10\r\nAccept: text/html,application/xhtml+xml\r
Accept-Language: en-us,en;q=0.5\r\nAccept-Encoding: gzip,deflate\r\nAccept-Charset: ISO-8859-1,utf-8;q=0.7\r\nKeep-Alive: 115\r\nConnection: keep-alive\r\n\r
```

 carriers return, carriage return character
line feed at start line-feed character
de of line indicates
e of header lines
HTTP request message: general format

- **Request line**
  - method
  - URL
  - version

- **Header lines**
  - header field name
  - value

- **Entity body**
Uploading form input

**POST method:**
- web page often includes form input
- input is uploaded to server in entity body

**URL method:**
- uses GET method
- input is uploaded in URL field of request line:

  www.somesite.com/animalsearch?monkeys&banana
Method types

HTTP/1.0:
- GET
- POST
- HEAD
  - asks server to leave requested object out of response

HTTP/1.1:
- GET, POST, HEAD
- PUT
  - uploads file in entity body to path specified in URL field
- DELETE
  - deletes file specified in the URL field
HTTP response message

status line (protocol status code status phrase)

HTTP/1.1 200 OK
Date: Sun, 26 Sep 2010 20:09:20 GMT
Server: Apache/2.0.52 (CentOS)
Last-Modified: Tue, 30 Oct 2007 17:00:02 GMT
ETag: "17dc6-a5c-bf716880"
Accept-Ranges: bytes
Content-Length: 2652
Keep-Alive: timeout=10, max=100
Connection: Keep-Alive
Content-Type: text/html; charset=ISO-8859-1

data data data data data data data ...

header lines

status line

data, e.g., requested HTML file
HTTP response status codes

- status code appears in 1st line in server-to-client response message.
- some sample codes:
  
  **200 OK**
  - request succeeded, requested object later in this msg
  
  **301 Moved Permanently**
  - requested object moved, new location specified later in this msg
    (Location:)
  
  **400 Bad Request**
  - request msg not understood by server
  
  **404 Not Found**
  - requested document not found on this server
  
  **505 HTTP Version Not Supported**
Trying out HTTP (client side) for yourself

1. Telnet to your favorite Web server:

```
telnet gaia.cs.umass.edu 80
```
opens TCP connection to port 80 (default HTTP server port) at gaia.cs.umass.edu. anything typed in will be sent to port 80 at gaia.cs.umass.edu

2. type in a GET HTTP request:

```
GET /kurose_ross/interactive/index.php HTTP/1.1
Host: gaia.cs.umass.edu
```
by typing this in (hit carriage return twice), you send this minimal (but complete) GET request to HTTP server

3. look at response message sent by HTTP server!
(or use Wireshark to look at captured HTTP request/response)
User-server state: cookies

Many Web sites use cookies **four components:**

1) Cookie header line of HTTP *response* message
2) Cookie header line in next HTTP *request* message
3) Cookie file kept on user’s host, managed by user’s browser
4) Back-end database at Website

**example:**

- Susan always access Internet from PC
- visits specific e-commerce site for first time
- when initial HTTP requests arrives at site, site creates:
  - unique ID
  - entry in backend database for ID
Cookies: keeping “state” (cont.)

One week later:

- eBay 8734
- Amazon 1678

**usual http request msg**

**cookie file**

**usual http response msg**

**set-cookie: 1678**

**usual http request msg**

**cookie: 1678**

**usual http response msg**
Cookies (continued)

**what cookies can be used for:**
- authorization
- shopping carts
- recommendations
- user session state (Web e-mail)

**how to keep “state”:**
- protocol endpoints: maintain state at sender/receiver over multiple transactions
- cookies: http messages carry state

**aside — cookies and privacy:**
- cookies permit sites to learn a lot about you
- you may supply name and e-mail to sites
Web caches (proxy server)

goal: satisfy client request without involving origin server

• user sets browser: Web accesses via cache
• browser sends all HTTP requests to cache
  • object in cache: cache returns object
  • else cache requests object from origin server, then returns object to client
More about Web caching

- cache acts as both client and server
  - server for original requesting client
  - client to origin server
- typically cache is installed by ISP (university, company, residential ISP)

**why Web caching?**
- reduce response time for client request
- reduce traffic on an institution’s access link
- Internet dense with caches: enables “poor” content providers to effectively deliver content
Caching example:

**assumptions:**
- avg object size: **1 Mbit**
- avg request rate from browsers to origin servers: **15/sec**
- avg data rate to browsers: **15 Mbps**
- RTT from institutional router to any origin server: **2 sec**
- access link rate: **15.4 Mbps**

**consequences:**
- LAN utilization: **1.5%**  
- access link utilization = **97%**
- total delay = Internet delay + access delay + LAN delay  
  = **2 sec + minutes + usecs**

Problem!
Caching example:

**assumptions:**

- avg object size: **1 Mbit**
- avg request rate from browsers to origin servers: **15/sec**
- avg data rate to browsers: **15 Mbps**
- RTT from institutional router to any origin server: **2 sec**
- access link rate: **15.4 Mbps**

**consequences:**

- LAN utilization: **1.5%**
- access link utilization = **97%**
- total delay = Internet delay + access delay + LAN delay
  = **2 sec + minutes + usecs**

Cost: increased access link speed (not cheap!)
Caching example:

**assumptions:**
- avg object size: **1 Mbit**
- avg request rate from browsers to origin servers: **15/sec**
- avg data rate to browsers: **15 Mbps**
- RTT from institutional router to any origin server: **2 sec**
- access link rate: **15.4 Mbps**

**consequences:**
- LAN utilization: **1.5%**
- access link utilization = ??
- total delay = Internet delay + access delay + LAN delay = ??

*How to compute link utilization, delay?*
Caching example:

Calculating access link utilization, delay with cache:

- suppose cache hit rate is 0.4
  - 40% requests satisfied at cache, 60% requests satisfied at origin

access link utilization:
- 60% of requests use access link
- data rate to browsers = 0.6*15 Mbps = 9 Mbps
- utilization = 9/15.4 = 0.58 → 58%

total delay
- = 0.6 * (delay from origin servers) + 0.4 * (delay when satisfied at cache)
- = 0.6 (2.01) + 0.4 (~msecs) = ~ 1.2 secs
- < less than with 100 Mbps link

Cost: web cache (cheap!)
Conditional GET

- **Goal:** don’t send object if cache has up-to-date cached version
  - no object transmission delay
  - lower link utilization
- **cache:** specify date of cached copy in HTTP request
  - `If-modified-since: <date>`
- **server:** response contains no object if cached copy is up-to-date:
  - HTTP/1.0 304 Not Modified
- `HTTP request msg
 If-modified-since: <date>`
- `HTTP response
 HTTP/1.0
 304 Not Modified`
- `HTTP request msg
 If-modified-since: <date>`
- `HTTP response
 HTTP/1.0
 200 OK
 <data>`
Chapter 2: Outline

✓ Principles of network applications
✓ Socket Programming with UDP and TCP
✓ Web and HTTP
❑ Electronic Mail (SMTP, POP3, IMAP)
❑ DNS
❑ P2P Applications
❑ Video Streaming and Content Distribution Networks
Electronic mail

Three major components:
• user agents
• mail servers
• SMTP: Simple Mail Transfer Protocol

User Agent
• a.k.a. “mail reader”
• composing, editing, reading mail messages
• e.g., Outlook, Thunderbird, iPhone mail client
• outgoing, incoming messages stored on server
Electronic Mail: Mail Servers

Mail Servers:

- **mailbox** contains incoming messages for user
- **message queue** of outgoing (to be sent) mail messages
- **SMTP protocol** between mail servers to send email messages
  - client: sending mail server
  - “server”: receiving mail server

![Diagram of mail servers and their connections with SMTP protocol]
Electronic Mail: SMTP [RFC 2821]

• uses TCP to reliably transfer email message from client to server, port 25
• direct transfer: sending server to receiving server
• three phases of transfer
  • handshaking (greeting)
  • transfer of messages
  • closure
• command/response interaction (like HTTP)
  • commands: ASCII text
  • response: status code and phrase
• messages must be in 7-bit ASCII
Scenario: Alice sends message to Bob

1) Alice uses UA to compose message “to” bob@someschool.edu

2) Alice’s UA sends message to her mail server; message placed in message queue

3) Client side of SMTP opens TCP connection with Bob’s mail server

4) SMTP client sends Alice’s message over the TCP connection

5) Bob’s mail server places the message in Bob’s mailbox

6) Bob invokes his user agent to read message
Sample SMTP interaction

S: 220 hamburger.edu
C: HELO crepes.fr
S: 250 Hello crepes.fr, pleased to meet you
C: MAIL FROM: <alice@crepes.fr>
S: 250 alice@crepes.fr... Sender ok
C: RCPT TO: <bob@hamburger.edu>
S: 250 bob@hamburger.edu ... Recipient ok
C: DATA
S: 354 Enter mail, end with "." on a line by itself
C: Do you like ketchup?
C: How about pickles?
C: .
S: 250 Message accepted for delivery
C: QUIT
S: 221 hamburger.edu closing connection
SMTP: final words

• SMTP uses persistent connections
• SMTP requires message (header & body) to be in 7-bit ASCII
• SMTP server uses CRLF.CRLF to determine end of message

comparison with HTTP:
• HTTP: pull
• SMTP: push
• both have ASCII command/response interaction, status codes
• HTTP: each object encapsulated in its own response message
• SMTP: multiple objects sent in multipart message
Mail message format

SMTP: protocol for exchanging email messages

RFC 822: standard for text message format:

- header lines, e.g.,
  - To:
  - From:
  - Subject: 
    - different from SMTP MAIL FROM, RCPT TO: commands!

- Body: the “message”
  - ASCII characters only
Mail access protocols

- **SMTP**: delivery/storage to receiver’s server
- **mail access protocol**: retrieval from server
  - **POP**: Post Office Protocol [RFC 1939]: authorization, download
  - **IMAP**: Internet Mail Access Protocol [RFC 1730]: more features, including manipulation of stored messages on server
  - **HTTP**: gmail, Hotmail, Yahoo! Mail, etc.
Chapter 2: Outline

✓ Principles of network applications
✓ Socket Programming with UDP and TCP
✓ Web and HTTP
✓ Electronic Mail (SMTP, POP3, IMAP)
❑ DNS
❑ P2P Applications
❑ Video Streaming and Content Distribution Networks
DNS: domain name system

people: many identifiers:
  • SSN, name, passport #

Internet hosts, routers:
  • IP address (32 bit) - used for addressing datagrams
  • “name”, e.g., www.yahoo.com - used by humans

Q: how to map between IP address and name, and vice versa?

Domain Name System:
  • distributed database implemented in hierarchy of many name servers
  • application-layer protocol: hosts, name servers communicate to resolve names (address/name translation)
    • note: core Internet function, implemented as application-layer protocol
    • complexity at network’s “edge”
DNS: services, structure

**DNS services**

- hostname to IP address translation
- host aliasing
  - canonical, alias names
- mail server aliasing
- load distribution
  - replicated Web servers: many IP addresses correspond to one name

**why not centralize DNS?**

- single point of failure
- traffic volume
- distant centralized database
- maintenance

A: *doesn’t scale!*
client wants IP for www.amazon.com; 1st approximation:

- client queries root server to find com DNS server
- client queries .com DNS server to get amazon.com DNS server
- client queries amazon.com DNS server to get IP address for www.amazon.com
DNS: Root Name Servers

• Contacted by local name server that cannot resolve name

• Root name server:
  • Contacts authoritative name server if name mapping not known
  • Gets mapping
  • Returns mapping to local name server

13 logical root name “servers” worldwide
• each “server” replicated many times
TLD, Authoritative DNS Servers

Top-level domain (TLD) servers:

• responsible for com, org, net, edu, aero, jobs, museums, and all top-level country domains, e.g.: uk, fr, ca, jp
• Network Solutions maintains servers for .com TLD
• Educause for .edu TLD

Authoritative DNS servers:

Organization’s own DNS server(s), providing authoritative hostname to IP mappings for organization’s named hosts
  • can be maintained by organization or service provider
Local DNS name server

• does not strictly belong to hierarchy
• each ISP (residential ISP, company, university) has one
  • also called “default name server”
• when host makes DNS query, query is sent to its local DNS server
  • has local cache of recent name-to-address translation pairs (but may be out of date!)
  • acts as proxy, forwards query into hierarchy
DNS name resolution example

- host at cs.stanford.edu wants IP address for csl.illinois.edu

*iterated query:*
- contacted server replies with name of server to contact
- “I don’t know this name, but ask this server”
DNS name resolution example

recursive query:
- puts burden of name resolution on contacted name server
- heavy load at upper levels of hierarchy?
DNS: caching, updating records

• once (any) name server learns mapping, it caches mapping
  • cache entries timeout (disappear) after some time (TTL)
  • TLD servers typically cached in local name servers
    • thus root name servers not often visited

• cached entries may be out-of-date (best effort name-to-address translation!)
  • if name host changes IP address, may not be known Internet-wide until all TTLs expire

• update/notify mechanisms proposed IETF standard
  • RFC 2136
DNS records

**DNS**: distributed database storing resource records (RR)

RR format: (name, value, type, ttl)

- **type=A**
  - name is hostname
  - value is IP address

- **type=NS**
  - name is domain (e.g., foo.com)
  - value is hostname of authoritative name server for this domain

- **type=CNAME**
  - name is alias name for some “canonical” (the real) name
  - value is canonical name
  - **www.ibm.com** is really **servereast.backup2.ibm.com**

- **type=MX**
  - value is name of mail server associated with name
DNS protocol, messages

- **query** and **reply** messages, both with same **message format**

**message header**

- **identification**: 16 bit # for query, reply to query uses same #
- **flags**:
  - query or reply
  - recursion desired
  - recursion available
  - reply is authoritative

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>identification</td>
<td>2 bytes</td>
</tr>
<tr>
<td># questions</td>
<td>2 bytes</td>
</tr>
<tr>
<td># authority RRs</td>
<td>2 bytes</td>
</tr>
<tr>
<td># answer RRs</td>
<td>2 bytes</td>
</tr>
<tr>
<td># additional RRs</td>
<td>2 bytes</td>
</tr>
<tr>
<td>questions</td>
<td>(variable # of questions)</td>
</tr>
<tr>
<td>answers</td>
<td>(variable # of RRs)</td>
</tr>
<tr>
<td>authority</td>
<td>(variable # of RRs)</td>
</tr>
<tr>
<td>additional info</td>
<td>(variable # of RRs)</td>
</tr>
</tbody>
</table>
### DNS protocol, messages

<table>
<thead>
<tr>
<th>Field</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>identification</td>
<td>2 bytes</td>
</tr>
<tr>
<td>flags</td>
<td>2 bytes</td>
</tr>
<tr>
<td># questions</td>
<td></td>
</tr>
<tr>
<td># answer RRs</td>
<td></td>
</tr>
<tr>
<td># authority RRs</td>
<td></td>
</tr>
<tr>
<td># additional RRs</td>
<td></td>
</tr>
<tr>
<td>questions (variable # of questions)</td>
<td></td>
</tr>
<tr>
<td>answers (variable # of RRs)</td>
<td></td>
</tr>
<tr>
<td>authority (variable # of RRs)</td>
<td></td>
</tr>
<tr>
<td>additional info (variable # of RRs)</td>
<td></td>
</tr>
</tbody>
</table>

- **name, type fields for a query**
- **RRs in response to query**
- **records for authoritative servers**
- **additional “helpful” info that may be used**
Inserting records into DNS

• example: new startup “Network Utopia”

• register name networkuptopia.com at DNS registrar (e.g., Network Solutions)
  • provide names, IP addresses of authoritative name server (primary and secondary)
  • registrar inserts two RRs into .com TLD server: (networkutopia.com, dns1.networkutopia.com, NS) (dns1.networkutopia.com, 212.212.212.1, A)

• create authoritative server type A record for www.networkuptopia.com; type NS record for networkutopia.com
Chapter 2: Outline

✓ Principles of network applications
✓ Socket Programming with UDP and TCP
✓ Web and HTTP
✓ Electronic Mail (SMTP, POP3, IMAP)
✓ DNS
❑ P2P Applications
❑ Video Streaming and Content Distribution Networks
Pure P2P architecture

• *no* always-on server
• arbitrary end systems directly communicate
• peers are intermittently connected and change IP addresses

*examples:*
  • file distribution (BitTorrent)
  • Streaming (KanKan)
  • VoIP (Skype)
**File distribution: client-server vs P2P**

**Question:**
how much time to distribute file (size $F$) from one server to $N$ peers?
- peer upload/download capacity is limited resource
File distribution time: client-server

- **Server transmission**: must sequentially send (upload) \( N \) file copies:
  - time to send one copy: \( F/u_s \)
  - time to send \( N \) copies: \( NF/u_s \)

- **Client**: each client must download file copy
  - \( d_{min} = \min \) client download rate
  - min client download time: \( F/d_{min} \)

\[
\text{Time to distribute } F \text{ to } N \text{ clients using client-server approach} = D_{c-s} \geq \max\{NF/u_s, F/d_{min}\}
\]

Increases linearly in \( N \)
File distribution time: P2P

- **Server transmission:** must upload at least one copy
  - time to send one copy: \( F/u_s \)

- **Client:** each client must download file copy
  - min client download time: \( F/d_{\text{min}} \)

- **Clients:** must download total \( NF \) bits \( \rightarrow \) still need to upload \( NF \) bits
  - max upload rate (limiting max download rate) is \( u_s + \sum u_i \)

\[
D_{P2P} \geq \max\{F/u_s, F/d_{\text{min}}, NF/(u_s + \sum u_i)\}
\]

*Time to distribute \( F \) to \( N \) clients using P2P approach*

increases linearly in \( N \) ... 
... but so does this, as each peer brings service capacity
Client-server vs. P2P: example

client upload rate = $u$, $F/u = 1$ hour, $u_s = 10u$, $d_{\min} \geq u_s$
P2P file distribution: BitTorrent

- File divided into 256Kb chunks
- Peers in torrent send/receive file chunks

**tracker:** tracks peers participating in torrent

**torrent:** group of peers exchanging chunks of a file

Alice arrives ...
... obtains list of peers from tracker
... and begins exchanging file chunks with peers in torrent
P2P file distribution: BitTorrent

- peer joining torrent:
  - has no chunks, but will accumulate them over time from other peers
  - registers with tracker to get list of peers, connects to subset of peers (“neighbors”)

- while downloading, peer uploads chunks to other peers
- peer may change peers with whom it exchanges chunks
- *churn*: peers may come and go
- once peer has entire file, it may (selfishly) leave or (altruistically) remain in torrent
BitTorrent: requesting, sending file chunks

**requesting chunks:**
- at any given time, different peers have different subsets of file chunks
- periodically, Alice asks each peer for list of chunks that they have
- Alice requests missing chunks from peers, rarest first

**sending chunks: tit-for-tat**
- Alice sends chunks to those four peers currently sending her chunks *at highest rate*
  - other peers are choked by Alice (do not receive chunks from her)
  - re-evaluate top 4 every 10 secs
- every 30 secs: randomly select another peer, starts sending chunks
  - “optimistically unchoke” this peer
  - newly chosen peer may join top 4
BitTorrent: tit-for-tat

(1) Alice “optimistically unchokes” Bob
(2) Alice becomes one of Bob’s top-four providers; Bob reciprocates
(3) Bob becomes one of Alice’s top-four providers

higher upload rate: find better trading partners, get file faster!
Distributed Hash Table (DHT)

- Hash table
- DHT paradigm
- Circular DHT and overlay networks
- Peer churn
Simple Database

Simple database with (key, value) pairs:

- key: human name; value: social security #

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Washington</td>
<td>132-54-3570</td>
</tr>
<tr>
<td>Diana Louise Jones</td>
<td>761-55-3791</td>
</tr>
<tr>
<td>Xiaoming Liu</td>
<td>385-41-0902</td>
</tr>
<tr>
<td>Rakesh Gopal</td>
<td>441-89-1956</td>
</tr>
<tr>
<td>Linda Cohen</td>
<td>217-66-5609</td>
</tr>
<tr>
<td>.......</td>
<td>........</td>
</tr>
<tr>
<td>Lisa Kobayashi</td>
<td>177-23-0199</td>
</tr>
</tbody>
</table>

- key: movie title; value: IP address
**Hash Table**

- More convenient to store and search on numerical representation of key
- key = hash(original key)

<table>
<thead>
<tr>
<th>Original Key</th>
<th>Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Washington</td>
<td>8962458</td>
<td>132-54-3570</td>
</tr>
<tr>
<td>Diana Louise Jones</td>
<td>7800356</td>
<td>761-55-3791</td>
</tr>
<tr>
<td>Xiaoming Liu</td>
<td>1567109</td>
<td>385-41-0902</td>
</tr>
<tr>
<td>Rakesh Gopal</td>
<td>2360012</td>
<td>441-89-1956</td>
</tr>
<tr>
<td>Linda Cohen</td>
<td>5430938</td>
<td>217-66-5609</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lisa Kobayashi</td>
<td>9290124</td>
<td>177-23-0199</td>
</tr>
</tbody>
</table>

• More convenient to store and search on numerical representation of key
• key = hash(original key)
Distributed Hash Table (DHT)

- Distribute (key, value) pairs over millions of peers
  - pairs are evenly distributed over peers
- Any peer can **query** database with a key
  - database returns value for the key
  - To resolve query, small number of messages exchanged among peers
- Each peer only knows about a small number of other peers
- Robust to peers coming and going (churn)
Assign key-value pairs to peers

• rule: assign key-value pair to the peer that has the closest ID.
• convention: closest is the immediate successor of the key.
• e.g., ID space {0,1,2,3,...,63}
• suppose 8 peers: 1, 12, 13, 25, 32, 40, 48, 60
  • If key = 51, then assigned to peer 60
  • If key = 60, then assigned to peer 60
  • If key = 61, then assigned to peer 1
Silly Strawman Circular DHT

- each peer *only* aware of immediate successor and predecessor.
What is the value associated with key 53?

$O(N)$ messages on average to resolve query, when there are $N$ peers.
Circular DHT with shortcuts (Chord)

- each peer keeps track of IP addresses of predecessor, successor, short cuts.
- reduced from 6 to 3 messages.
- possible to design shortcuts with $O(\log N)$ neighbors, $O(\log N)$ messages in query

What is the value for key 53?
Peer churn

display: peer churn:
❖ peers may come and go (churn)
❖ each peer knows address of its two successors
❖ each peer periodically pings its two successors to check aliveness
❖ if immediate successor leaves, choose next successor as new immediate successor

example: peer 5 abruptly leaves
Peer churn

handling peer churn:

❖ peers may come and go (churn)
❖ each peer knows address of its two successors
❖ each peer periodically pings its two successors to check aliveness
❖ if immediate successor leaves, choose next successor as new immediate successor

example: peer 5 abruptly leaves

• peer 4 detects peer 5’s departure; makes 8 its immediate successor

• 4 asks 8 who its immediate successor is; makes 8’s immediate successor its second successor.
P2P Motivation – Revisited

• Client-server dominates the mainstream. Why?
  • Performance
    • Economies of scale
    • Round trip times
    • (Mass file distribution is a rare exception)

• So, why peer-to-peer?
  • Avoid single points of failure
    • Technological  ➔ Robustness, Survivability
    • ...and social  ➔ Power to the people
Chapter 2: Outline

✓ Principles of network applications
✓ Socket Programming with UDP and TCP
✓ Web and HTTP
✓ Electronic Mail (SMTP, POP3, IMAP)
✓ DNS
✓ P2P Applications
❑ Video Streaming and Content Distribution Networks
Video Streaming and CDNs: context

- video traffic: major consumer of Internet bandwidth
  - Netflix, YouTube: 37%, 16% of downstream residential ISP traffic
  - ~1B YouTube users, ~75M Netflix users
- challenge: scale - how to reach ~1B users?
  - single mega-video server won’t work (why?)
- challenge: heterogeneity
  - different users have different capabilities (e.g., wired versus mobile; bandwidth rich versus bandwidth poor)
- solution: distributed, application-level infrastructure
Multimedia: video

- video: sequence of images displayed at constant rate
  - e.g., 24 images/sec
- digital image: array of pixels
  - each pixel represented by bits
- coding: use redundancy *within* and *between* images to decrease # bits used to encode image
  - spatial (within image)
  - temporal (from one image to next)

**Spatial coding example:** instead of sending $N$ values of same color (all purple), send only two values: color value (*purple*) and number of repeated values ($N$)

**Temporal coding example:** instead of sending complete frame at $i+1$, send only differences from frame $i$
### Multimedia: video

- **CBR: (constant bit rate):** video encoding rate fixed
- **VBR: (variable bit rate):** video encoding rate changes as amount of spatial, temporal coding changes
- **examples:**
  - MPEG 1 (CD-ROM) 1.5 Mbps
  - MPEG2 (DVD) 3-6 Mbps
  - MPEG4 (often used in Internet, < 1 Mbps)

**spatial coding example:** instead of sending $N$ values of same color (all purple), send only two values: color value (purple) and number of repeated values ($N$)

**temporal coding example:** instead of sending complete frame at $i+1$, send only differences from frame $i$
Video Compression

- Color Encoding:
  - RGB
  - YUV
Video Compression

- Color Encoding: YUV 4:2:0 Sampling
Video Compression

- Spatial Encoding: Frequency domain compression using DCT (Discrete Cosine Transform)

<table>
<thead>
<tr>
<th>Original pixel data</th>
</tr>
</thead>
<tbody>
<tr>
<td>114 108 100 99 109 129 152 166</td>
</tr>
<tr>
<td>109 102 95 94 104 124 146 161</td>
</tr>
<tr>
<td>99 93 85 84 94 114 137 151</td>
</tr>
<tr>
<td>86 80 72 71 82 102 124 138</td>
</tr>
<tr>
<td>73 66 58 57 68 88 110 125</td>
</tr>
<tr>
<td>60 53 46 45 55 75 97 112</td>
</tr>
<tr>
<td>50 43 36 35 45 65 88 102</td>
</tr>
<tr>
<td>45 38 31 30 40 60 82 97</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DCT coefficient data</th>
</tr>
</thead>
<tbody>
<tr>
<td>700 200 0 0 0 0 0 0</td>
</tr>
<tr>
<td>-150 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>110 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>
Video Compression

- **Spatial Encoding**: Frequency domain compression using DCT (Discrete Cosine Transform)
  - Divide into 8x8 Blocks
  - Take DCT of block
  - Quantize and Compress block

![Diagram](image-url)
Video Compression

- Spatial Encoding: Frequency domain compression using DCT (Discrete Cosine Transform)
  - Divide into 8x8 Blocks
  - Take DCT of block
  - Quantize and Compress block
Video Compression

- Temporal Encoding: Motion Compensation
Video Compression

- Temporal Encoding: Motion Compensation

![Diagram showing temporal encoding with motion compensation](image)
Video Compression

- Temporal Encoding: Motion Compensation
Streaming stored video:

simple scenario:

video server
(Stored video)

client

Internet
Streaming multimedia: DASH

- **DASH: Dynamic, Adaptive Streaming over HTTP**
- **server:**
  - divides video file into multiple chunks
  - each chunk stored, encoded at different rates
  - *manifest file:* provides URLs for different chunks
- **client:**
  - periodically measures server-to-client bandwidth
  - consulting manifest, requests one chunk at a time
    - chooses maximum coding rate sustainable given current bandwidth
    - can choose different coding rates at different points in time (depending on available bandwidth at time)
Streaming multimedia: DASH

- **DASH:** Dynamic, Adaptive Streaming over HTTP
- **“intelligence” at client:** client determines
  - *when* to request chunk (so that buffer starvation, or overflow does not occur)
  - *what encoding rate* to request (higher quality when more bandwidth available)
  - *where* to request chunk (can request from URL server that is “close” to client or has high available bandwidth)
Content distribution networks

• **challenge:** how to stream content (selected from millions of videos) to hundreds of thousands of simultaneous users?

• **option 1:** single, large “mega-server”
  • single point of failure
  • point of network congestion
  • long path to distant clients
  • multiple copies of video sent over outgoing link

....quite simply: this solution **doesn’t scale**
Content distribution networks

- **challenge:** how to stream content (selected from millions of videos) to hundreds of thousands of simultaneous users?

- **option 2:** store/serve multiple copies of videos at multiple geographically distributed sites (*CDN*)
  - **enter deep:** push CDN servers deep into many access networks
    - close to users
    - used by Akamai, 1700 locations
  - **bring home:** smaller number (10’s) of larger clusters near (but not within) access networks
    - used by Limelight
Content Distribution Networks (CDNs)

- CDN: stores copies of content at CDN nodes
  - e.g. Netflix stores copies of “House of Card”

- subscriber requests content from CDN
  - directed to nearby copy, retrieves content
  - may choose different copy if network path congested
Chapter 2: summary

our study of network apps now complete!

• application architectures
  • client-server
  • P2P

• application service requirements:
  • reliability, bandwidth, delay

• Internet transport service model
  • connection-oriented, reliable: TCP
  • unreliable, datagrams: UDP

• specific protocols:
  • HTTP
  • SMTP, POP, IMAP
  • DNS
  • P2P: BitTorrent

• video streaming, CDNs
• socket programming:
  TCP, UDP sockets
Chapter 2: summary

Most importantly: learned about protocols!

- Typical request/reply message exchange:
  - Client requests info or service
  - Server responds with data, status code
- Message formats:
  - Headers: fields giving info about data
  - Data: info(payload) being communicated

Important themes:
- Control vs. messages
- Centralized vs. decentralized
- Stateless vs. stateful
- Reliable vs. unreliable message transfer
- “Complexity at network edge”