

# Data and Computer Communications

Tenth Edition  
by William Stallings

# **CHAPTER 24**

**Electronic Mail, DNS, and HTTP**

*“Consider a future device for individual use, which is a sort of mechanized private file and library. It needs a name, and, to coin one at random, “memex” will do. A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory. All this is conventional, except for the projection forward of present-day mechanisms and gadgetry. It affords an immediate step, however, to associative indexing, the basic idea of which is a provision whereby any item may be caused at will to select immediately and automatically another. This is the essential feature of the memex. The process of tying two items together is the important thing.”*

**—As We May Think, Vannevar Bush, *The Atlantic*, July 1945**

# Electronic Mail

- A facility that allows users at workstations and terminals to compose and exchange messages
- Messages need never exist on paper unless the user desires a paper copy of the message
- Some e-mail systems only serve users on a single computer, while others provide service across a network of computers

# Internet Mail Architecture

- Currently defined in RFC 5598
- Fundamental level consists of:
  - A user world, in the form of Message User Agents (MUA)
  - The transfer world, in the form of the Message Handling Service (MHS), which is composed of Message Transfer Agents (MTA)



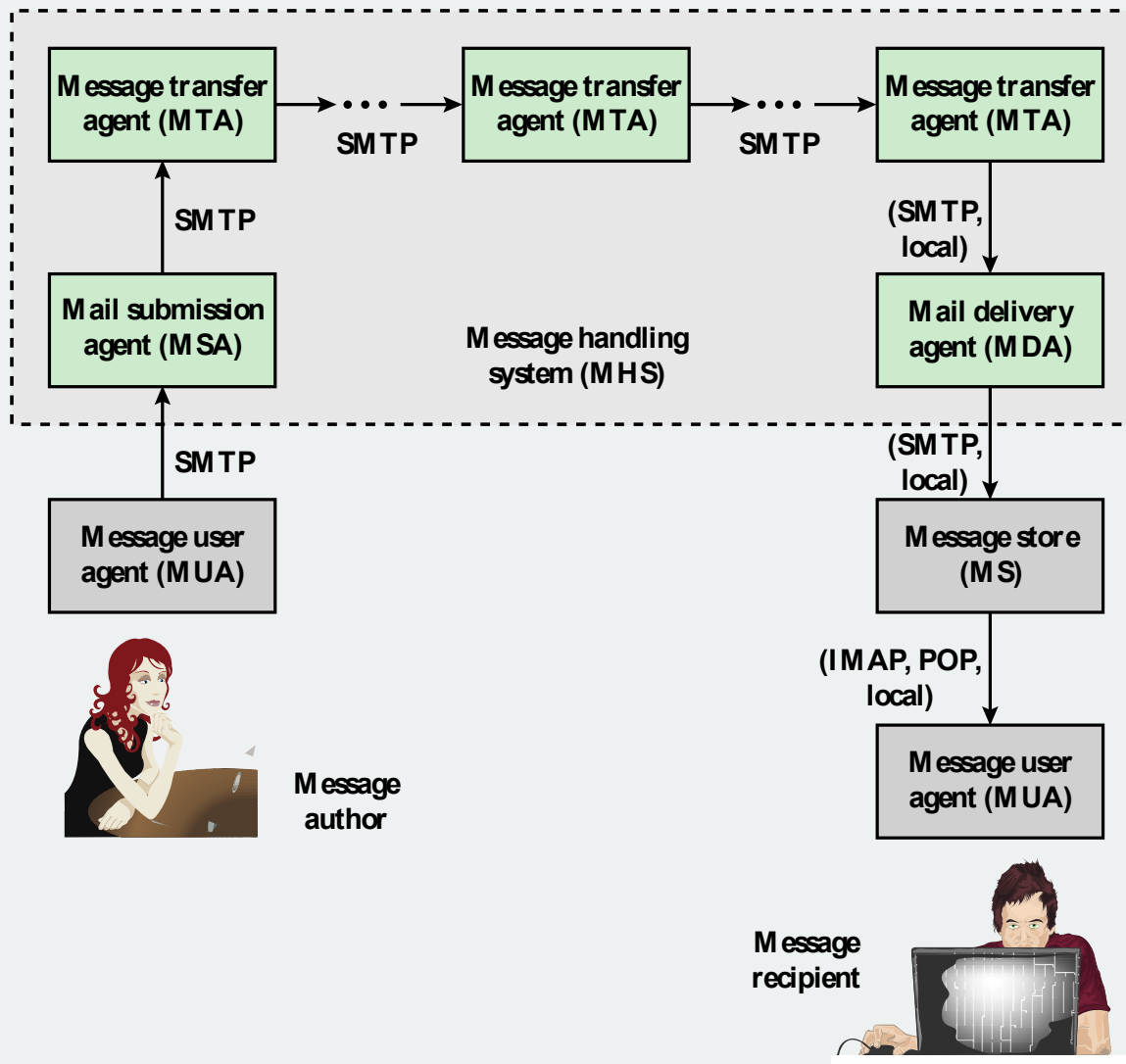


Figure 24.1 Function Modules and Standardized Protocols Used Between Them in the Internet Mail Architecture

# Administrative Management Domain (ADMD)

- An Internet e-mail provider



- Each ADMD can have different operating policies and trust-based decision making

# Domain Name System (DNS)

- A directory lookup service that provides a mapping between the name of a host on the Internet and its numerical address





# Internet Mail Standards

## Post Office Protocol (POP3)

- Allows an e-mail client to download an e-mail from an e-mail server
- Connects via TCP/IP to the server

## Internet Mail Access Protocol (IMAP)

- Provides stronger authentication than POP3 and provides other functions not supported by POP3

## Simple Mail Transfer Protocol (SMTP)

- Protocol used for transfer of mail from a user agent to an MTA and from one MTA to another

## Multipurpose Internet Mail Extensions (MIME)

- Supplements SMTP and allows the encapsulation of multimedia messages inside of a standard SMTP message

# SMTP



- Defined in RFC 821
- The standard protocol for transferring mail between hosts in the TCP/IP suite
- Is not concerned with the format or content of messages themselves
  - Exceptions:
    - SMTP standardizes the message character set as 7-bit ASCII
    - SMTP adds log information to the start of the delivered message that indicates the path the message took

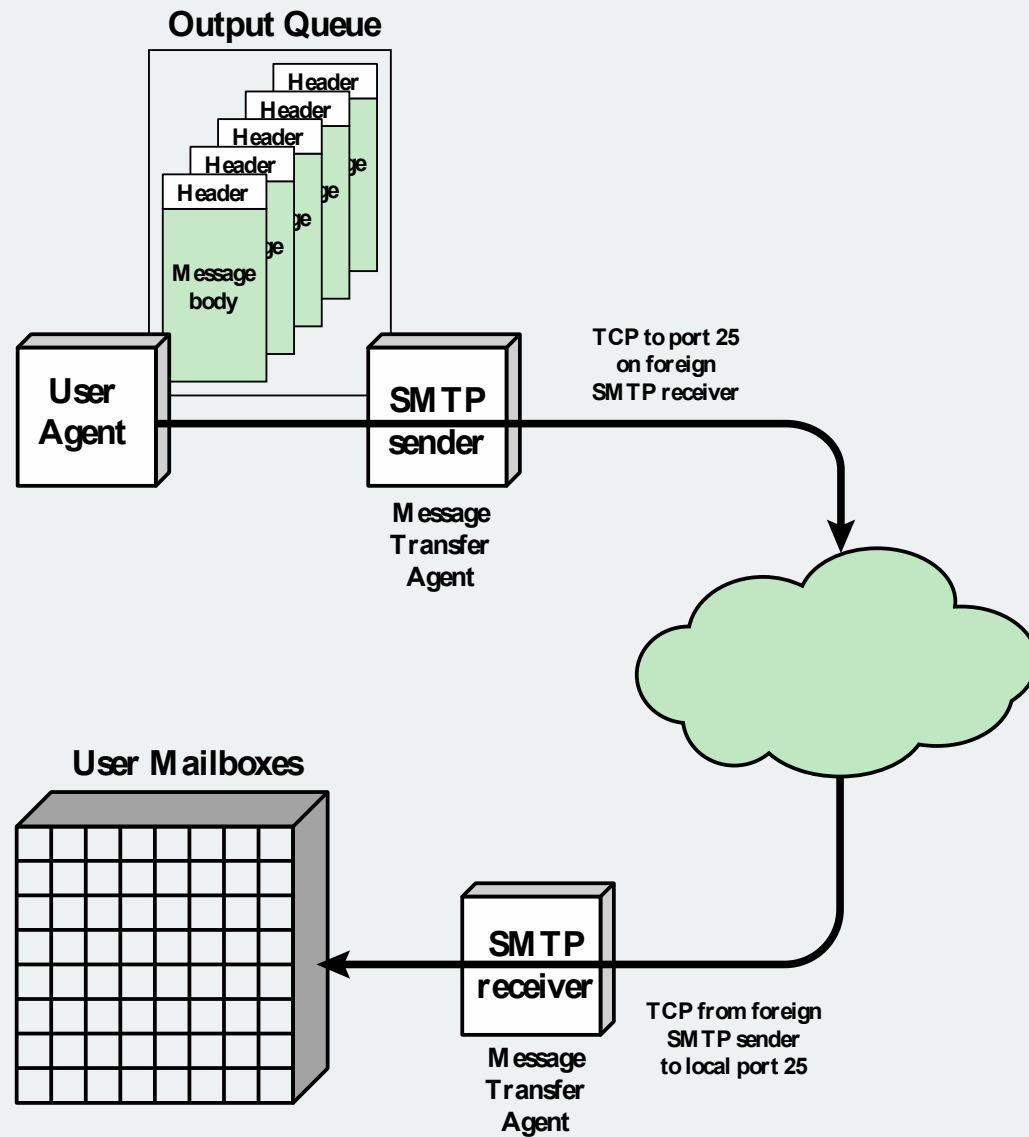
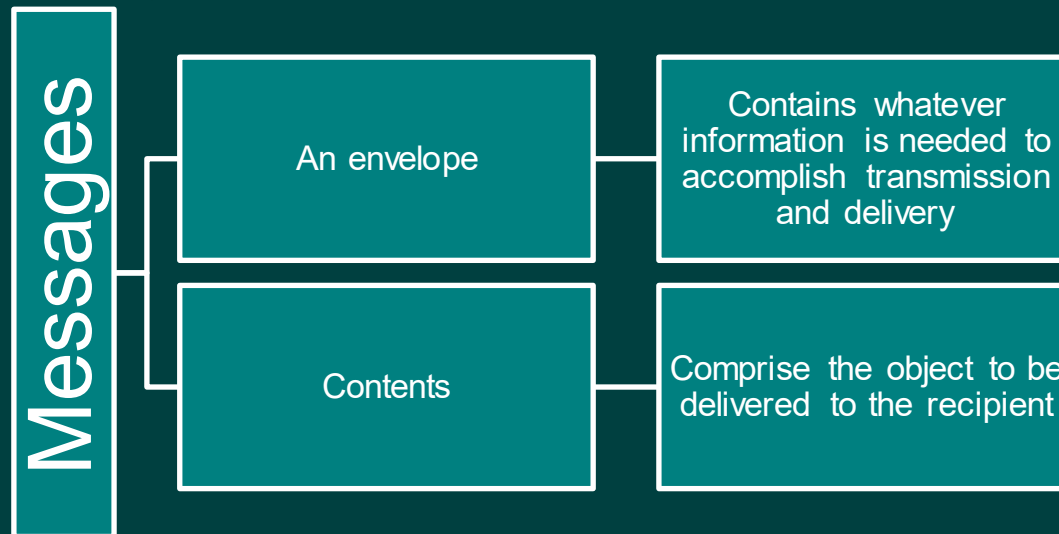


Figure 24.2 SMTP Mail Flow

# RFC 822

- Defines a format for text messages that are sent using electronic mail



- Adopted by the SMTP standard as the format for use in constructing messages for transmission via SMTP
- Message consists of a sequence of lines of text and uses a general “memo” framework
  - A number of header lines, which follow a rigid format, followed by a body portion consisting of arbitrary text
  - Most frequently used header keywords are From, To, Subject, and Date

# Limitations of SMTP/822

SMTP cannot transmit  
executable files or other binary  
objects

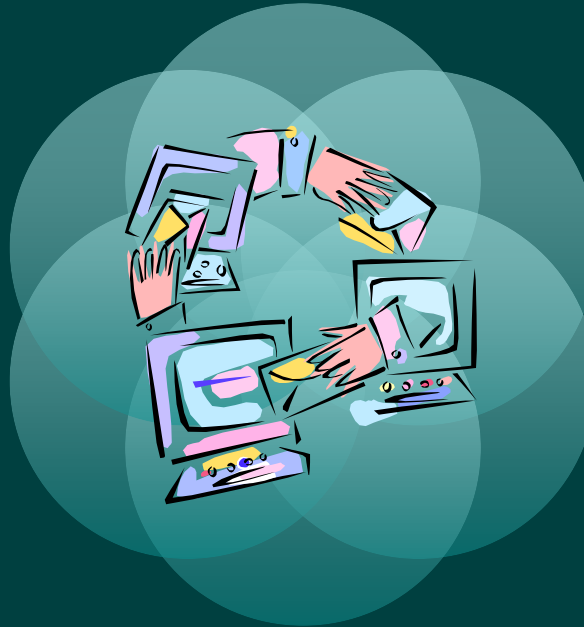
Some SMTP  
implementations do not  
adhere completely to the  
SMTP standards defined in  
RFC 821

SMTP cannot transmit text  
data that includes national  
language characters

SMTP gateways to X.400  
electronic mail networks  
cannot handle nontextual  
data included in X.400  
messages

SMTP servers may reject  
mail messages over a certain  
size

SMTP gateways that translate  
between the character codes  
ASCII and EBCDIC do not use  
a consistent set of mappings,  
resulting in translation  
problems



# MIME

- Extension to the RFC 822 framework that is intended to address some of the problems and limitations of the use of SMTP and RFC 822 for electronic mail
- Includes the following elements:
  - Five new message header fields are defined, which may be included in an RFC 822 header
  - A number of content formats are defined, thus standardizing representations that support multimedia electronic mail
  - Transfer encodings are defined that enable the conversion of any content format into a form that is protected from alteration by the mail system

# MIME Header Fields

- MIME-version
- Content-type
- Content-transfer-encoding
- Content-ID
- Content-description



# Table 24.1 MIME Content Types

(Table can be found on page 812 in textbook)

Type	Subtype	Description
Text	Plain	Unformatted text; may be ASCII or ISO 8859.
Multipart	Mixed	The different parts are independent but are to be transmitted together. They should be presented to the receiver in the order that they appear in the mail message.
	Parallel	Differs from Mixed only in that no order is defined for delivering the parts to the receiver.
	Alternative	The different parts are alternative versions of the same information. They are ordered in increasing faithfulness to the original and the recipient's mail system should display the "best" version to the user.
	Digest	Similar to Mixed, but the default type/subtype of each part is message/rfc822
Message	rfc822	The body is itself an encapsulated message that conforms to RFC 822.
	Partial	Used to allow fragmentation of large mail items, in a way that is transparent to the recipient.
	External-body	Contains a pointer to an object that exists elsewhere.
Image	jpeg	The image is in JPEG format, JFIF encoding.
	gif	The image is in GIF format.
Video	mpeg	MPEG format.
Audio	Basic	Single-channel 8-bit ISDN mu-law encoding at a sample rate of 8 kHz.
Application	PostScript	Adobe Postscript
	octet-stream	General binary data consisting of 8-bit bytes.



# Table 24.2

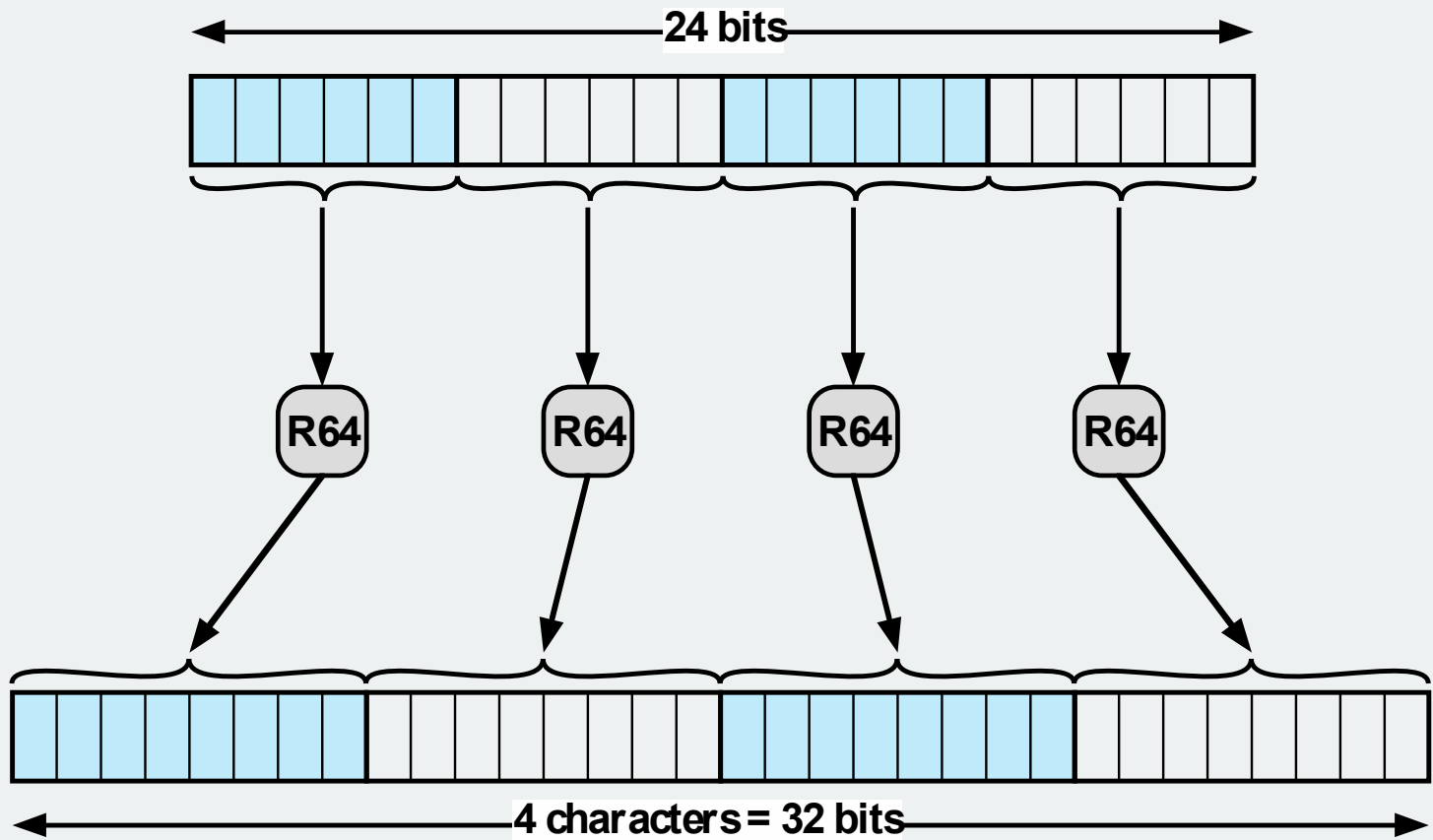
## MIME Transfer Encodings

<b>7bit</b>	The data are all represented by short lines of ASCII characters.
<b>8bit</b>	The lines are short, but there may be non-ASCII characters (octets with the high-order bit set).
<b>binary</b>	Not only may non-ASCII characters be present but the lines are not necessarily short enough for SMTP transport.
<b>quoted-printable</b>	Encodes the data in such a way that if the data being encoded are mostly ASCII text, the encoded form of the data remains largely recognizable by humans.
<b>base64</b>	Encodes data by mapping 6-bit blocks of input to 8-bit blocks of output, all of which are printable ASCII characters.
<b>x-token</b>	A named nonstandard encoding.

6-Bit Value	Character Encoding	6-Bit Value	Character Encoding	6-Bit Value	Character Encoding	6-Bit Value	Character Encoding
0	A	16	Q	32	G	48	w
1	B	17	R	33	H	49	x
2	C	18	S	34	I	50	y
3	D	19	T	35	J	51	z
4	E	20	U	36	K	52	0
5	F	21	V	37	L	53	1
6	G	22	W	38	M	54	2
7	H	23	X	39	N	55	3
8	I	24	Y	40	O	56	4
9	J	25	Z	41	P	57	5
10	K	26	a	42	Q	58	6
11	L	27	b	43	R	59	7
12	M	28	c	44	S	60	8
13	N	29	d	45	T	61	9
14	O	30	e	46	U	62	+
15	P	31	f	47	V	63	/
						(pad)	=

**Table**  
**24.3**  
**Radix-64**  
**Encoding**

(Table can be found on  
 page 815 in textbook)



**Figure 24.3 Printable Encoding of Binary Data into Radix-64 Format**

# POP

- Internet standard defined in RFC 1939
- Supports the basic functions of download and delete for e-mail retrieval
- States:

## Authentication state

Client must authenticate itself to the user

Often done with a user ID/password combination

## Transaction state

Once the server successfully authenticates the client, the client can access the mailbox to retrieve and delete messages

## Update state

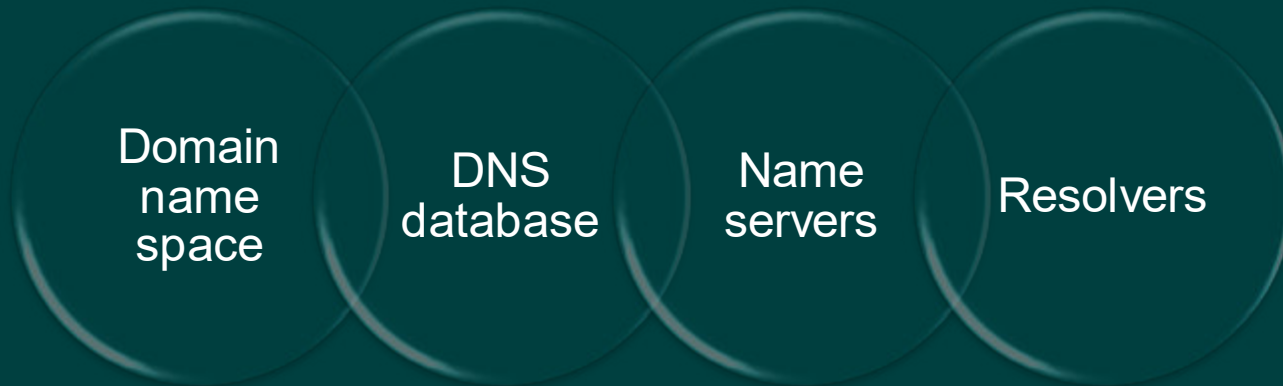
During this state, the server enacts all of the changes requested by the client's commands and then closes the connection

# IMAP

- Defined by RFC 3501
- Servers store messages for multiple users to be retrieved upon client requests
- Features:
  - Clients can have multiple remote mailboxes from which messages can be retrieved
  - Clients can also specify criteria for downloading messages, such as not transferring large messages over slow links
  - IMAP always keeps messages on the server and replicates copies to the clients
  - IMAP4 allows clients to make changes both when connected and when disconnected

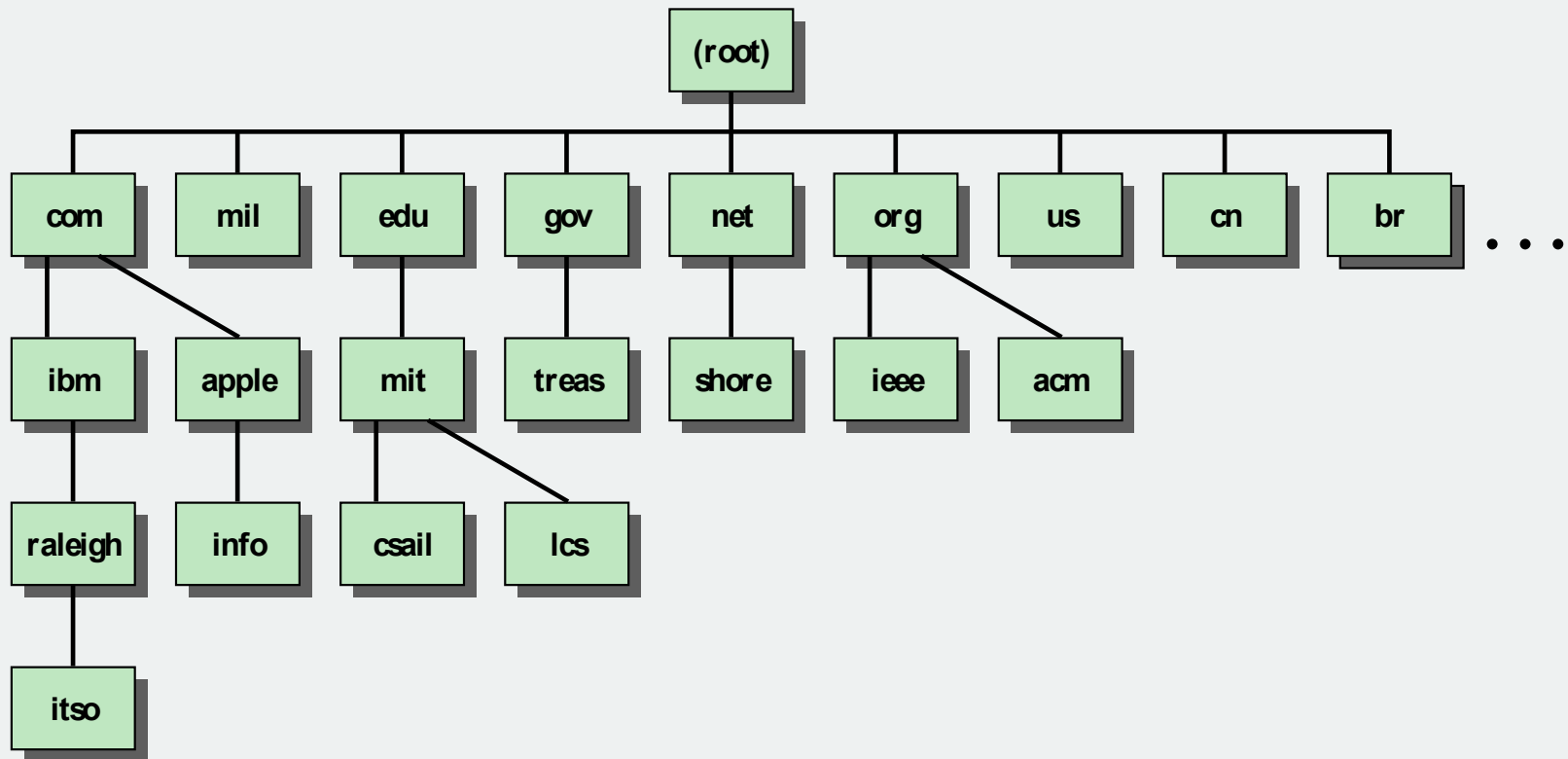
# Internet Directory Service (DNS)

- A directory lookup service that provides a mapping between the name of a host on the Internet and its numerical address
- Essential to the functioning of the Internet
- Defined in RFCs 1034 and 1035
- Four elements comprise the DNS:



# Domain

- Refers to a group of hosts that are under the administrative control of a single entity
- Organized hierarchically, so that a given domain may consist of a number of subordinate domains
- Names are assigned and reflect the hierarchical organization



**Figure 24.4 Portion of Internet Domain Tree**



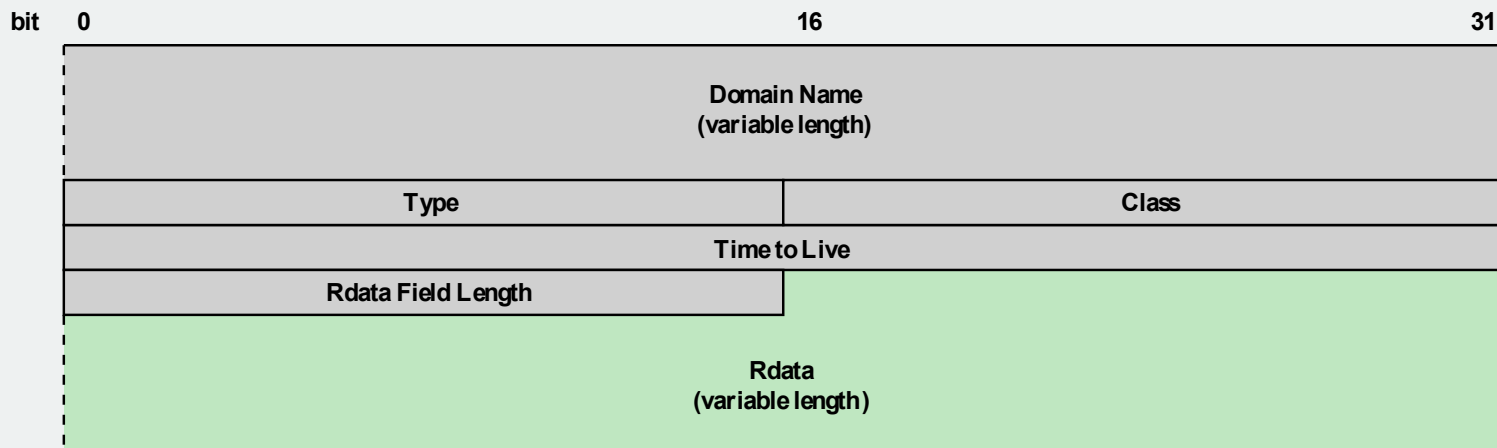
Domain	Contents
com	Commercial organizations
edu	Educational institutions
gov	U.S. federal, state, and local government agencies
mil	U.S. military
net	Network support centers, Internet service providers, and other network-related organizations
org	Nonprofit organizations
us	U.S. state and local government agencies, schools, libraries, and museums
country code	ISO standard 2-letter identifier for country-specific domains (e.g., au, ca, uk)
biz	Dedicated exclusively for private businesses
info	Unrestricted use
name	Individuals, for email addresses and personalized domain names.
museum	restricted to museums, museum organizations, and individual members of the museum profession
coop	Member-owned cooperative organizations, such as credit unions
aero	Aviation community
pro	Medical, legal, and accounting professions
arpa	Address and routing parameter area; used for technical infrastructure purposes, such as reverse domain name resolution
int	International organizations

**Table  
24.4**

**Top-  
Level  
Internet  
Domains**

# DNS Database

- Based on a hierarchical database containing resource records (RRs) that include the name, IP address, and other information about hosts
- Key features:
  - Variable-depth hierarchy for names
  - Distributed database
  - Distribution controlled by the database

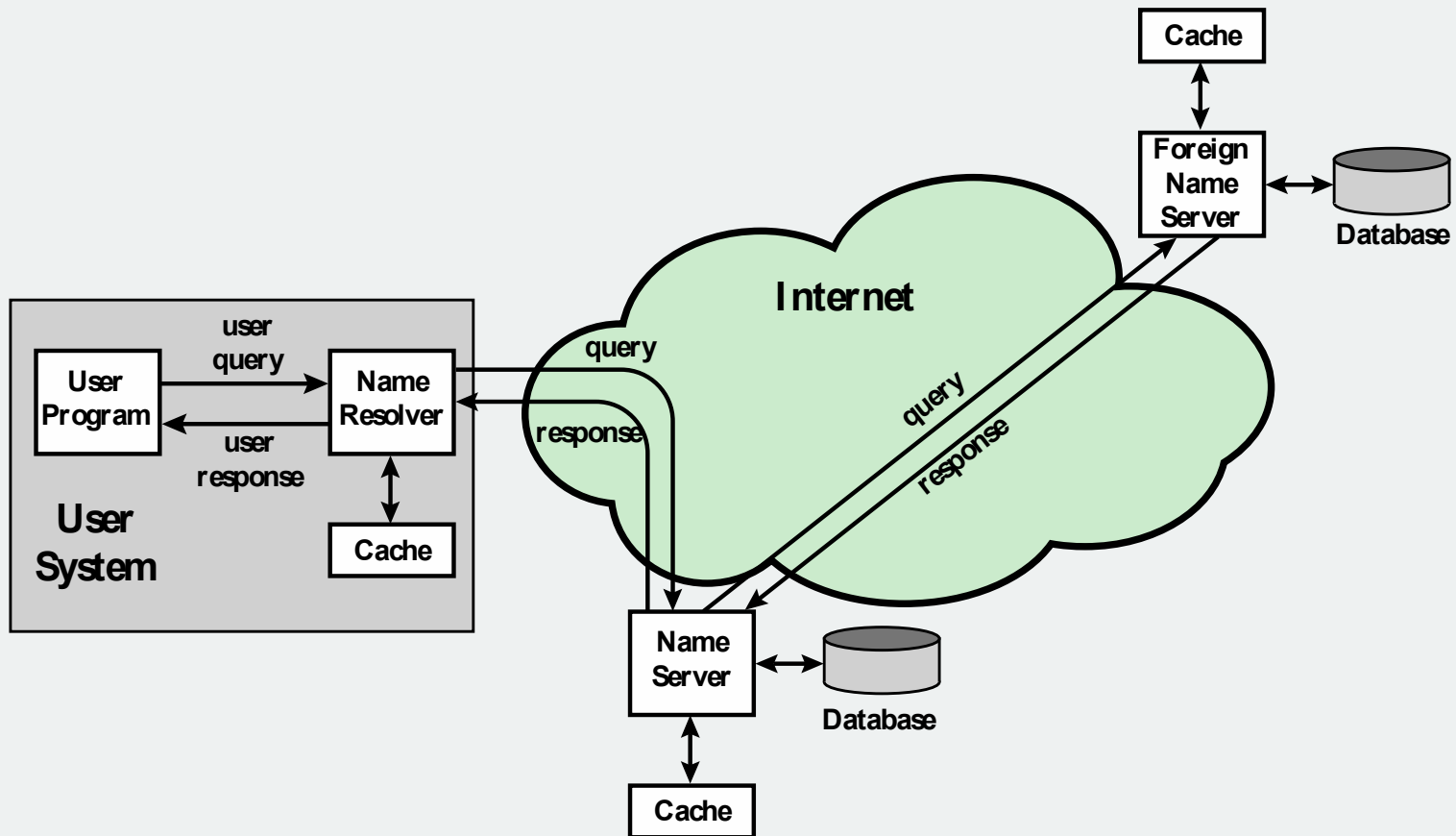


**Figure 24.5 DNS Resource Record Format**

# Table 24.5

## Resource Record Types

Type	Description
A	A host address. This RR type maps the name of a system to its IPv4 address. Some systems (e.g., routers) have multiple addresses, and there is a separate RR for each.
AAAA	Similar to A type, but for IPv6 addresses.
CNAME	Canonical name. Specifies an alias name for a host and maps this to the canonical (true) name.
HINFO	Host information. Designates the processor and operating system used by the host.
MINFO	Mailbox or mail list information. Maps a mailbox or mail list name to a host name.
MX	Mail exchange. Identifies the system(s) via which mail to the queried domain name should be relayed.
NS	Authoritative name server for this domain.
PTR	Domain name pointer. Points to another part of the domain name space.
SOA	Start of a zone of authority (which part of naming hierarchy is implemented). Includes parameters related to this zone.
SRV	For a given service provides name of server or servers in domain that provide that service.
TXT	Arbitrary text. Provides a way to add text comments to the database.
WKS	Well-known services. May list the application services available at this host.



**Figure 24.6 DNS Name Resolution**

```

telnet locis.loc.gov
Trying 140.147.254.3...
Connected to locis.loc.gov.
Escape character is '^]'.
      L O C I S:  LIBRARY OF CONGRESS INFORMATION SYSTEM

      To make a choice: type a number, then press ENTER

      1  Copyright Information      -- files available and up-to-date
      2  Braille and Audio         -- files frozen mid-August 1999
      3  Federal Legislation       -- files frozen December 1998

*  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

      The LC Catalog Files are available at:
      http://lcweb.loc.gov/catalog/

*  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

      8  Searching Hours and Basic Search Commands
      9  Library of Congress General Information
     10  Library of Congress Fast Facts

     12  Comments and Logoff
      Choice:
    9

      LIBRARY OF CONGRESS GENERAL INFORMATION

      LC is a research library serving Congress, the federal government, the
      library community world-wide, the US creative community, and any researchers
      beyond high school level or age.  On-site researchers request materials by
      filling out request slips in LC's reading rooms; requesters must present a
      photo i.d.  Staff are available for assistance in all public reading rooms.

      -----
      The following phone numbers offer information about hours and other services:

      General Research Info:      202-707-6500      Reading Room Hours:      202-707-6400
      Exhibits/Tours/Gift Shop:  202-707-8000      Location/Parking:       202-707-4700
      Copyright Information:      202-707-3000      Cataloging Products:    202-707-6100
      Copyright Forms:            202-707-9100      " " fax:                202-707-1334

      -----
      For information on interlibrary loan, see:  http://lcweb.loc.gov/rr/loan/

      12  Return to LOCIS MENU screen

      Choice:

```

**Figure 24.7 A Telnet Session**

Server	Operator	Cities	IP Addr
A	VeriSign Global Registry Services	6 sites in the United States, Germany, Hong Kong	IPv4: 198.41.0.4 IPv6: 2001:503:BA3E::2:30
B	Information Sciences Institute	Marina Del Rey, CA, U.S.	IPv4: 192.228.79.201 IPv6: 2001:478:65::53
C	Cogent Communications	6 sites in the United States, Germany, Spain	192.33.4.12
D	University of Maryland	College Park, MD, U.S.	128.8.10.90
E	NASA Ames Research Center	Mountain View, CA, U.S.	192.203.230.10
F	Internet Software Consortium	49 sites in the United States and other countries	IPv4: 192.5.5.241 IPv6: 2001:500::1035
G	U.S. DOD Network Information Center	6 sites in United States, Japan, Germany, Italy	192.112.36.4
H	U.S. Army Research Lab	Aberdeen, MD, U.S. San Diego, CA, USA	IPv4: 128.63.2.53 IPv6: 2001:500:1::803f:235
I	Netnod	38 sites in the United States and other countries	IPv4: 192.36.148.17 IPv6: 2001:7fe::53
J	VeriSign Global Registry Services	70 sites in the United States and other countries	IPv4: 192.58.128.30 IPv6: 2001:503:C27::2:30
K	Reseaux IP Europeens - Network Coordination Centre	18 sites in the United States and other countries	IPv4: 193.0.14.129 IPv6: 2001:7fd::1
L	Internet Corporation for Assigned Names and Numbers	55 sites in the United States and other countries	IPv4: 199.7.83.42 IPv6: 2001:500:3::42
M	WIDE Project	6 sites in the United States, Japan, Korea, France	IPv4: 202.12.27.33 IPv6: 2001:dc3::35

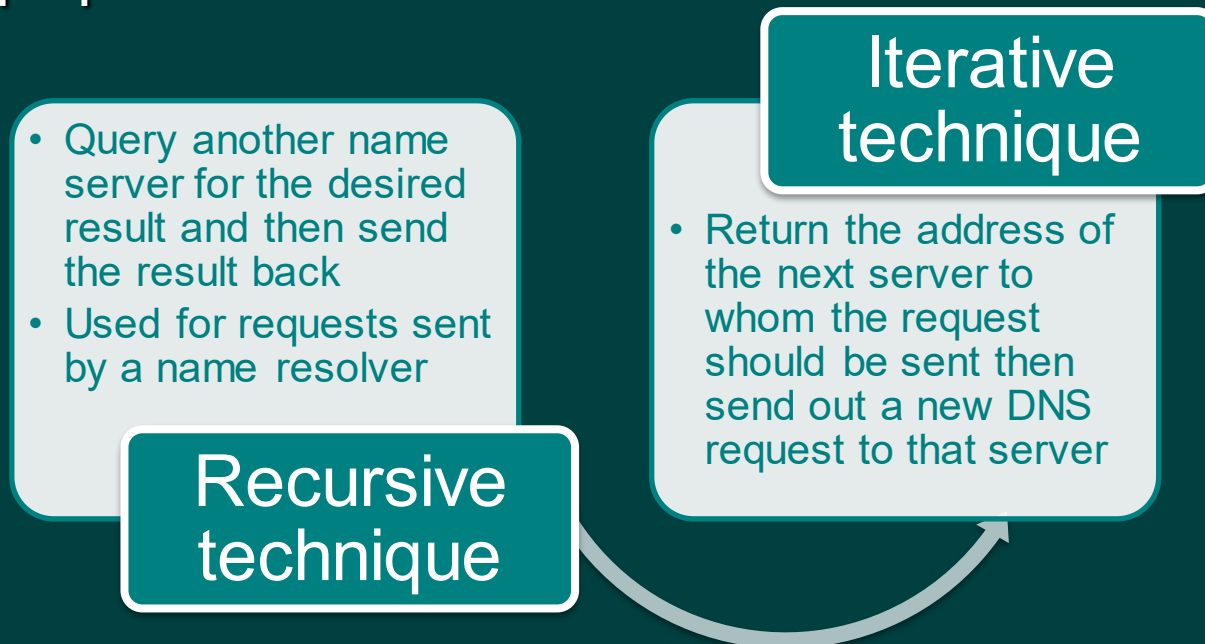
# Table 24.6

## Internet Root Servers

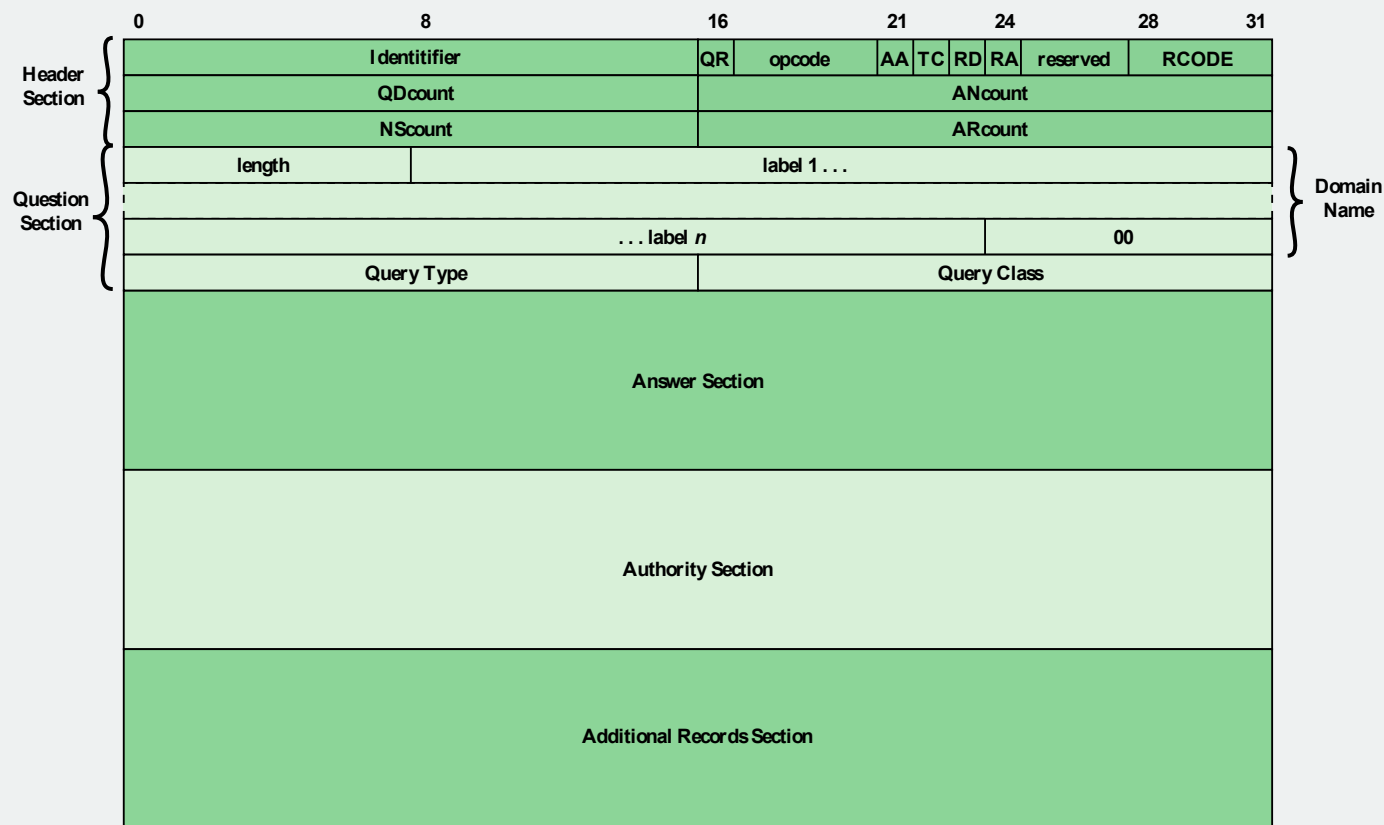
(Table can be found on page 824 in textbook)

# Name Resolution

- Each query begins at a name resolver located in the user host system
- If the resolver does not have the requested name in its cache, it sends a DNS query to the local DNS server
- Resolvers use UDP for single queries and TCP for group queries







QR = query/response bit  
 AA = authoritative answer  
 TC = truncated  
 RD = recursion desired  
 RA = recursion available

RCODE = response code  
 QDcount = number of entries in question section  
 ANcount = number of resource records in answer section  
 NScount = number of name server resource records in authority section  
 ARcount = number of resource records in additional records section

**Figure 24.8 DNS Message Format**

<p><b>Cache</b></p> <p>A program's local store of response messages and the subsystem that controls its message storage, retrieval, and deletion. A cache stores cacheable responses in order to reduce the response time and network bandwidth consumption on future, equivalent requests. Any client or server may include a cache, though a cache cannot be used by a server while it is acting as a tunnel.</p>	<p><b>Origin Server</b></p> <p>The server on which a given resource resides or is to be created.</p>
<p><b>Client</b></p> <p>An application program that establishes connections for the purpose of sending requests.</p>	<p><b>Proxy</b></p> <p>An intermediary program that acts as both a server and a client for the purpose of making requests on behalf of other clients. Requests are serviced internally or by passing them, with possible translation, on to other servers. A proxy must interpret and, if necessary, rewrite a request message before forwarding it. Proxies are often used as client-side portals through network firewalls and as helper applications for handling requests via protocols not implemented by the user agent.</p>
<p><b>Connection</b></p> <p>A transport layer virtual circuit established between two application programs for the purposes of communication.</p>	<p><b>Resource</b></p> <p>A network data object or service which can be identified by a URI.</p>
<p><b>Entity</b></p> <p>A particular representation or rendition of a data resource, or reply from a service resource, that may be enclosed within a request or response message. An entity consists of entity headers and an entity body.</p>	<p><b>Server</b></p> <p>An application program that accepts connections in order to service requests by sending back responses.</p>
<p><b>Gateway</b></p> <p>A server that acts as an intermediary for some other server. Unlike a proxy, a gateway receives requests as if it were the original server for the requested resource; the requesting client may not be aware that it is communicating with a gateway. Gateways are often used as server-side portals through network firewalls and as protocol translators for access to resources stored on non-HTTP systems.</p>	<p><b>Tunnel</b></p> <p>An intermediary program that is acting as a blind relay between two connections. Once active, a tunnel is not considered a party to the HTTP communication, though the tunnel may have been initiated by an HTTP request. A tunnel ceases to exist when both ends of the relayed connections are closed. Tunnels are used when a portal is necessary and the intermediary cannot, or should not, interpret the relayed communication.</p>
<p><b>Message</b></p> <p>The basic unit of HTTP communication, consisting of a structured sequence of octets transmitted via the connection.</p>	<p><b>User Agent</b></p> <p>The client that initiates a request. These are often browsers, editors, spiders, or other end-user tools.</p>

# Table 24.7

## Key Terms Related to HTTP

(Table can be found on page 827 in textbook)

# HTTP

- Transaction-oriented client/server protocol
- Most typical use is between a Web browser and a Web server
- Makes use of TCP to provide reliability
- Is a stateless protocol
- Each transaction is treated independently
- Flexible in the formats that it can handle

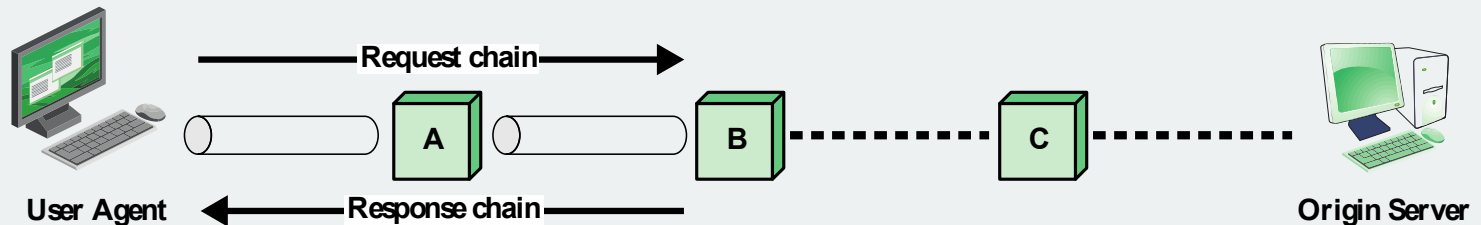
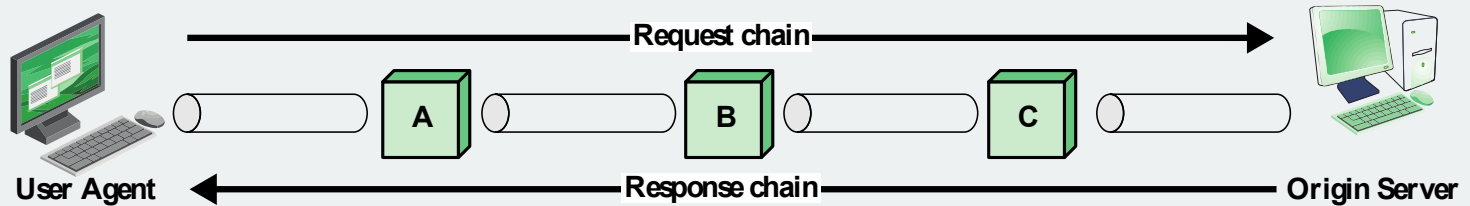
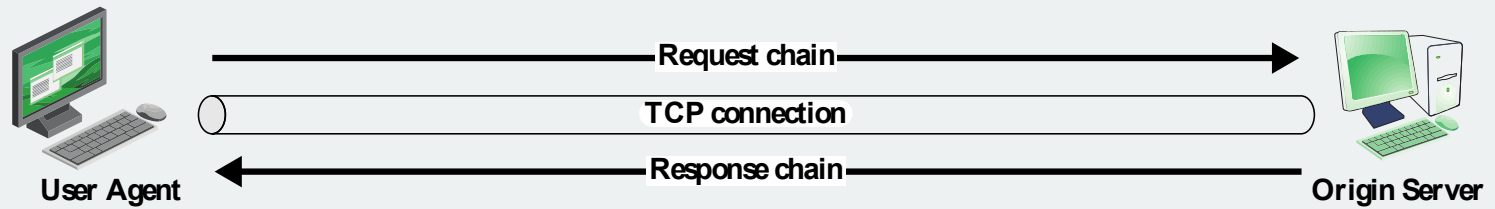
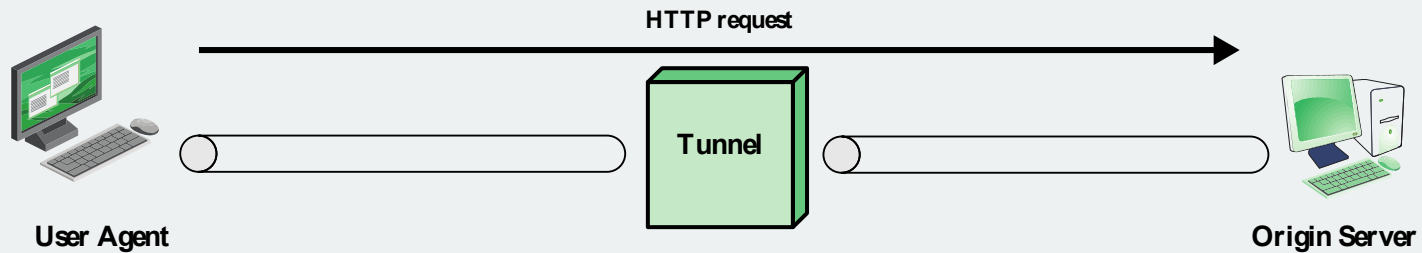
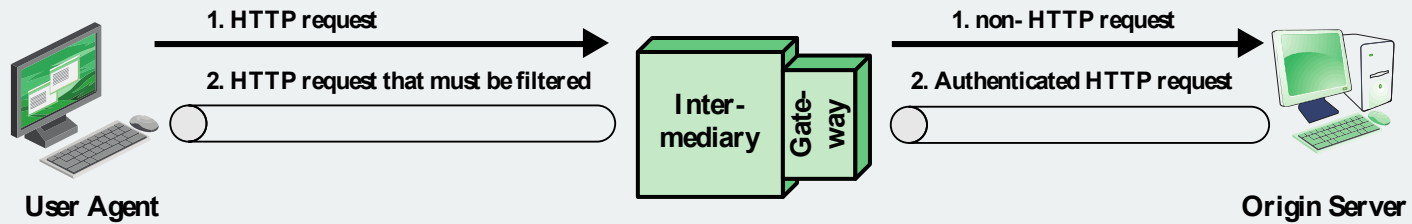
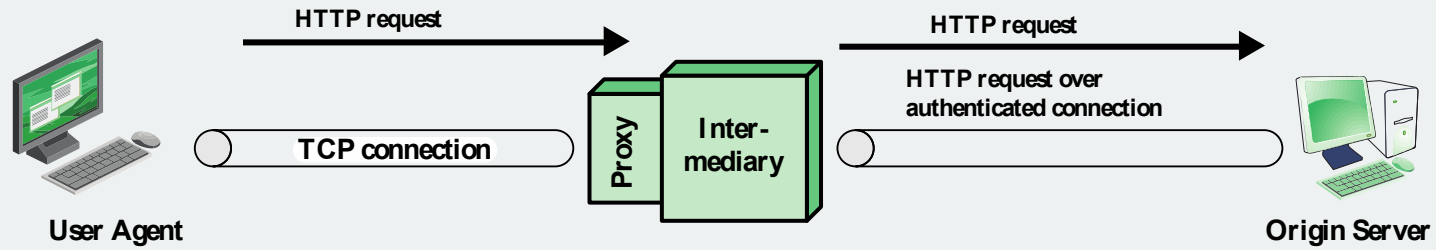


Figure 24.9 Examples of HTTP Operation



**Figure 24.10 Intermediate HTTP Systems**

# Proxy

- A forwarding agent, receiving a request for a URL object, modifying the request, and forwarding the request toward the server identified in the URL
  - Acts as a server in interacting with a client and as a client in interacting with a server
- Scenarios that call for the use of a proxy:
    - Security intermediary
      - Client and server may be separated by a security intermediary such as a firewall, with the proxy on the client side of the firewall
    - Different versions of HTTP
      - If the client and server are running different versions of HTTP, the proxy can implement both versions and perform the required mapping

# Gateway

- A server that appears to the client as if it were an origin server
- Acts on behalf of other servers that may not be able to communicate directly with a client
- Scenarios in which gateways can be used:
  - Security intermediary
  - Non-HTTP server



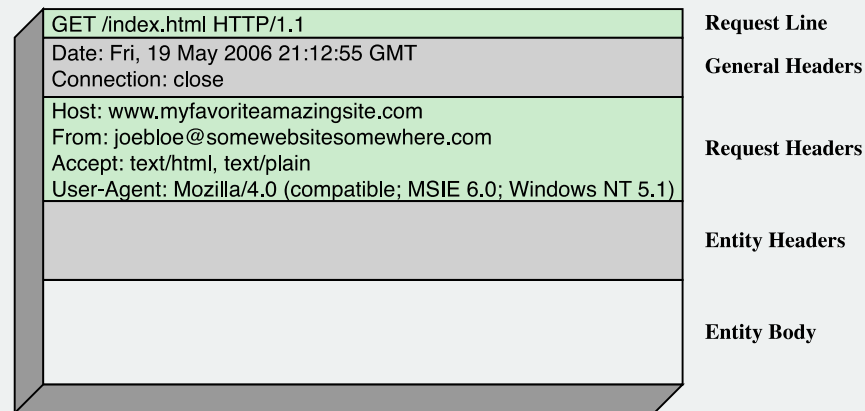
# Tunnel

- Performs no operations on HTTP requests and responses
- Is simply a relay point between two TCP connections
- Used when there must be an intermediary system between client and server but it is not necessary for that system to understand the contents of messages

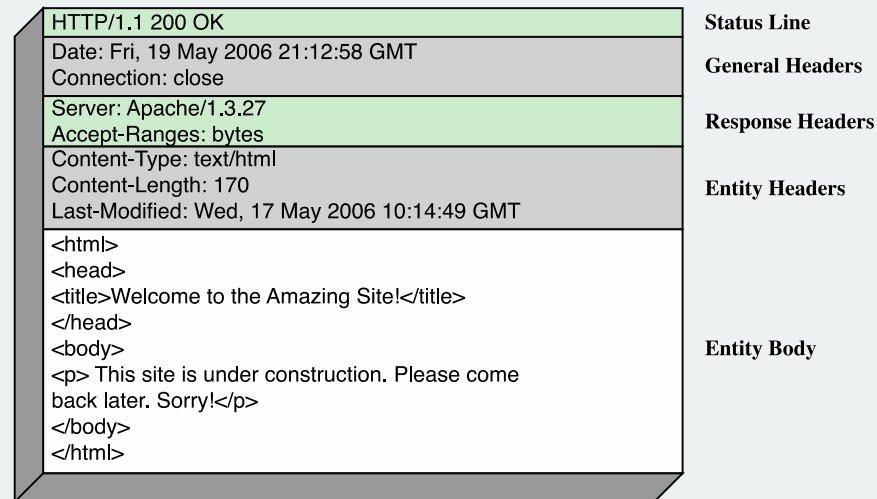


# Cache

- A facility that may store previous requests and responses for handling new requests
  - If a new request arrives that is the same as a stored request then the cache can supply the stored response rather than accessing the resource indicated in the URL
- Can operate on a client or server or on an intermediate system other than a tunnel
- Not all transactions can be cached



**(a) HTTP Request**



**(b) HTTP Response**

**Figure 24.11 Examples of HTTP Message Format**

- Words in lowercase represent variables or names of rules.
- A rule has the form

name = definition

- DIGIT is any decimal digit; CRLF is carriage return, line feed; SP is one or more spaces.
- Quotation marks enclose literal text.
- Angle brackets, "<" ">", may be used within a definition to enclose a rule name when their presence will facilitate clarity.
- Elements separated by bar ("|") are alternatives.
- Ordinary parentheses are used simply for grouping.
- The character "\*" preceding an element indicates repetition. The full form is:

<I>\*<J>element

indicating at least I and at most J occurrences of element. \*element allows any number, including 0; 1\*element requires at least one element; and 1\*2element allows 1 or 2 elements; <N>element means exactly N elements.

- Square brackets, "[" "]", enclose optional elements.
- The construct "#" is used to define, with the following form:

<I>#<J>element

indicating at least I and at most J elements, each separated by a comma and optional linear white space.

- A semicolon at the right of a rule starts a comment that continues to the end of the line.

# Table 24.8

## Augmented BNF Notation Used in URL and HTTP Specifications

# General Header Fields

- Can be used in both request and response messages
- Are applicable in both types of messages and contain information that does not directly apply to the entity being transferred
- Fields are:
  - Cache-control
  - Connection
  - Date
  - Forwarded
  - Keep-alive MIME-version
  - Pragma
  - Upgrade



# Summary

## ➤ Electronic Mail – SMTP and MIME

- Internet mail architecture
- Simple Mail Transfer Protocol (SMTP)
- Multipurpose Internet Mail Extensions (MIME)
- POP and IMAP

## ➤ Internet Directory Service: DNS

- Domain names
- The DNS database
- DNS operation

## ➤ Web access and HTTP

- HTTP overview
- Messages
- Request messages
- Response messages
- Entities