RAMCO INSTITUTE OF TECHNOLOGY
Approved by AICTE, New Delhi & Affiliated to Anna University
Accredited by NAAC & An ISO 9001: 2015 Certified Institution
NBA Accredited UG Programs: CSE, EEE, ECE and MECH

Department of Artificial Intelligence and Data Science

Academic Year: 2022 - 2023 (Even Semester)

Degree, & Semester: B. TECH /IV/

Regulation: Anna University R 2021

Course Code & Title: CS3591- Computer Networks

Category of the Course: Professional Core Course

Name of the Faculty member(s) with Designation & Department:

Dr.M. Kaliappan

Professor and Head

Department of Artificial Intelligence and Data Science

Acknowledgement:
James F. Kurose, Keith W. Ross, Computer Networking, A Top-Down Approach Featuring the Internet, Eighth Edition, Pearson Education, 2021

Review Questions & Answers
Problems & Answer

Chapter2: Application layer

1. List five non-proprietary Internet applications and the application-layer protocols that they

use.

Non-Proprietary Internet Applications:

Non-Proprietary Internet Applications are those internet applications which are not registered or protected as a trademark or brand name and they use certain application layer protocols.

Some of the Non-proprietary Internet applications along with the application layer protocols are mentioned below:

- **The web:**
  - The web application is a client-server program in which the client runs in a web browser.
  - The web application uses Hyper Text Transfer Protocol (HTTP) as application layer protocol.
  - HTTP is an application protocol for distributive, collaborative, and hypermedia information systems.
- **File Transfer:**
- The File Transfer web application is an application that allows services to the users to share files over the web.
- The File Transfer web application uses the File Transfer Protocol (FTP) as application layer protocol.
- The FTP is a standard application layer tool in a computer network which is used to transfer files between client and server.

- **Remote Login:**
  - The Remote login related applications are the applications that are used to control one computer from another computer by the use of a remote.
  - The Remote login related application uses the Telecommunications Network (Telnet) as application layer protocol.
  - Telnet is a bidirectional interactive text oriented communication facility which is used as a protocol on the internet by using a virtual terminal connection.
- **E-mail:**
- E-mail applications are used to send emails over the internet.
- The Simple Mail Transfer Protocol (SMTP) is used in E-mail related applications.
- The SMTP is only an application layer delivery protocol which is an internet standard for electronic mail transformation.
- **Bit Torrent File Sharing:**
- Bit Torrent file sharing application is a communication protocol that is used for peer to peer file sharing to share and distribute files and data over the internet.
- The Bit Torrent File sharing applications use Bit Torrent Protocol as application layer protocol.
- The Bit Torrent Protocol is a protocol that is used for distributing electronic files.

2. What information is used by a process running on one host to identify a process running on another host?

The IP address of the destination host and the port number of the destination socket.

3. What is the difference between network architecture and application architecture?

Network architecture refers to the organization of the communication process into layers (e.g., the five-layer Internet architecture). Application architecture, on the other hand, is designed by an application developer and dictates the broad structure of the application (e.g., client-server or P2P).

4. For a communication session between a pair of processes, which process is the client and which is the server?

In the context of a communication session between a pair of processes, the process that initiates the communication (that is, initially contacts the other process at the beginning of the session) is labeled as the client. The process that waits to be contacted to begin the session is the server.

5. For a P2P file-sharing application, do you agree with the statement, "There is no notion of client and server sides of a communication session"? Why or why not?
No. In a P2P file-sharing application, the peer that is receiving a file is typically the client and the peer that is sending the file is typically the server.

6. Suppose you wanted to do a transaction from a remote client to a server as fast as possible. Would you use UDP or TCP? Why?

You would use UDP. With UDP, the transaction can be completed in one roundtrip time (RTT) - the client sends the transaction request into a UDP socket, and the server sends the reply back to the client's UDP socket.

7. Suppose Alice, with a web-based email account (such as hotmail or gmail), sends a message to Bob, who accesses his mail from his mail server using POP3. Discuss how the message gets from Alice's host to Bob's host. Be sure to list the series of application-layer protocols that are used to move the message between the two hosts.

Message is sent from Alice's host to her mail server over HTTP. Alice's mail server then sends the message to Bob's mail server over SMTP. Bob then transfers the message from his mail server to his host over POP3.

8 Give four reasons (arguments) against having one DNS server.

1. Single point of failure: if the DNS server crashes, so does the entire Internet
2. Traffic concentration: the single server would have to handle all DNS queries for all HTTP requests and email messages for hundreds of millions of hosts.

Chapter2: Application layer

3. Delayed responses: since the single server can only be close to a very few hosts, most of the hosts will have to travel large distances (and experience propagation delay), and traverse many links (some of which maybe congested) to reach the server.
4. Book-keeping and updates (maintenance): the DNS server would have to keep track of every new host or every removed host in the Internet. This doesn't scale.

9. What is the current architecture of DNS? (mention the various types of servers and their function)

The current architecture of DNS is a distributed, hierarchical database, with 3 levels (and server types) of hierarchy: 1. the Root DNS servers: there are 13 root servers around the world, each consists of a cluster of replicated servers for security and reliability purposes. 2. top-level domain servers (TLD) responsible for top-level domains (e.g., co, org, net, edu, gov) and country top-level domains (e.g., uk, fr, ca, jp). 3. authoritative DNS servers: keep the mapping for publicly accessible resources at organizations (e.g., web and mail servers).
4. Local name server: does not belong strictly to the hierarchy and is queried first when a host requests to resolve an address.

10. What are the two types of query/search propagation in DNS? What is the main difference between them? <5 points>

the two types of queries are: iterative queries and recursive queries. Iterative (or iterated) queries propagate from the host to its local DNS server and from then on to a root server (which replies to the local DNS server), then from the local server to the TLD server (which replies to the local DNS server), then from the local server to the authoritative server (which replies to the local DNS server). The recursive query, by contrast, puts the burden on the contacted server and may increase the burden on the high level servers (e.g., the root server has to contact the TLD which in turn contacts the authoritative server. The latter query method may incur less delay.

11- Discuss a mechanism we studied to improve DNS performance and elaborate on how the performance can improve.

Using DNS caching is one way to improve DNS performance, first by reducing the delay required to get the address resolution (since the cache servers are now closer to the requesting hosts), and by reducing the overall load of DNS going to the higher level DNS servers.

12. Discuss three different architectures of the peer-to-peer applications. Give examples of real applications for each architecture and discuss the advantages and disadvantages of each architecture.

1. Centralized directory of resources/files, as in Napster. Advantage is that search for resources is simple with min overhead (just ask the centralized server). The disadvantages are: single point of failure, performance bottleneck and target of lawsuit.

2. Fully distributed, non-centralized architecture, as in Gnutella, where all peers and edges form a 'flat' overlay (without hierarchy). Advantages: robustness to failure, no performance bottleneck and no target for lawsuit. Disadvantages is that search is more involved and incurs high overhead with query flooding.

3. Hierarchical overlay, with some nodes acting as super nodes (or cluster heads), or nodes forming loose neighborhoods (sometimes referred to as loose hierarchy, as in BitTorrent). Advantages, robust (no single point of failure), avoids flooding to search for resources during queries. Disadvantages, needs to keep track of at least some nodes using the 'Tracker' server. In general, this architecture attempts to combine the best of the 2 other architectures.

13. In BitTorrent, suppose Alice provides chunks to Bob throughout a 30-second interval. Will Bob necessarily return the favor and provide chunks to Alice in this same interval? Why or why not?

It is not necessary that Bob will also provide chunks to Alice. Alice has to be in the top 4 neighbors of Bob for Bob to send out chunks to her (or through random selection); this might not occur even if Alice provides chunks to Bob throughout a 30-second interval.

13. Why do HTTP, FTP, SMTP, and POP3 run on top of TCP rather than on UDP?

TCP is more reliable than udp, udp may have failures or data loss, so we can't afford to have losses inhttp,smtp,pop3 and so on.

14. Consider an e-commerce site that wants to keep a purchase record for each of its customers. Describe how this can be done with cookies

When a user visits the e-commerce site for the first time, the website will return a cookie number, which is stored on the user's host, managed by the browser. The cookie number is present in cookie header and is generated by the server of the e-commerce website, and the number is unique for every customer. The client receives the response along with the header and the number with a line appended to a special cookie file. The file contains the server name and the user's associated ID number.

In the subsequent request to the same server the client includes a cookie header which consists of a header line which specifies the id number for that server. During each visit or purchase from the e-commerce website, the browser sends the cookie number back to the website server. This cookie number is used to identify the user (or browser) who is visiting the site.

15. Describe how Web caching can reduce the delay in receiving a requested object. Will Web caching reduce the delay for all objects requested by a user or for only some of the objects? Why?

Web caching can bring the desired content "closer" to the user, possibly to the same LAN to which the user's host is connected. Web caching can reduce the delay for all objects, even objects that are not cached, since caching reduces the traffic on links.

16. Print out the header of an e-mail message you have recently received. How many Received: header lines are there? Analyze each of the header lines in the message.

Received:
from 65.54.246.203 (EHLO bay0-omc3-s3.bay0.hotmail.com) (65.54.246.203) by mta419.mail.mud.yahoo.com with SMTP; Sat, 19 May 2007 16:53:51 -0700
Received:
from hotmail.com ([65.55.135.106]) by bay0-omc3-s3.bay0.hotmail.com with Microsoft SMTPSVC(6.0.3790.2668); Sat, 19 May 2007 16:52:42 -0700
Received:
from mail pickup service by hotmail.com with Microsoft SMTPSVC; Sat, 19 May 2007 16:52:41 -0700 Message-ID: <BAY130-F26D9E35BF59E0D18A819AFB9310@phx.gbl>
Received: from 65.55.135.123 by by130fd.bay130.hotmail.msn.com with HTTP; Sat, 19 May 2007 23:52:36 GMT From: "prithula dhungel" <prithuladhungel@hotmail.com>
To: prithula@yahoo.com
Bcc:
Subject: Test mail
Date: Sat, 19 May 2007 23:52:36 +0000 Mime-Version: 1.0
Content-Type: Text/html; format=flowed
Return-Path: prithuladhungel@hotmail.com
Figure: A sample mail message header

Received: This header field indicates the sequence in which the SMTP servers send and receive the mail message including the respective timestamps. In this example there are 4 "Received:" header lines. This means the mail message passed through 5 different SMTP servers before being delivered to the receiver's mail box. The last (forth) "Received:" header indicates the mail message flow from the SMTP server of the sender to the second SMTP server in the chain of servers. The sender's SMTP server is at address 65.55.135.123 and the second SMTP server in the chain is by130fd.bay130.hotmail.msn.com. The third "Received:" header indicates the mail message flow from the second SMTP server in the chain to the third server, and so on. Finally, the first "Received:" header indicates the flow of the mail messages from the forth SMTP server to the last SMTP server (i.e. the receiver's mail server) in the chain.
Message-id: The message has been given this number BAY130-F26D9E35BF59E0D18A819AFB9310@phx.gbl (by bay0-omc3-s3.bay0.hotmail.com. Message-id is a unique string assigned by the mail system when the message is first created.
From: This indicates the email address of the sender of the mail. In the given example, the sender is "prithuladhungel@hotmail.com"

To: This field indicates the email address of the receiver of the mail. In the example, the receiver is "prithula@yahoo.com"
Subject: This gives the subject of the mail (if any specified by the sender). In the example, the subject specified by the sender is "Test mail"
Date: The date and time when the mail was sent by the sender. In the example, the sender sent the mail on 19th May 2007, at time 23:52:36 GMT.
Mime-version: MIME version used for the mail. In the example, it is 1.0.
Content-type: The type of content in the body of the mail message. In the example, it is "text/html".
Return-Path: This specifies the email address to which the mail will be sent if the receiver of this mail wants to reply to the sender. This is also used by the sender's mail server for bouncing back undeliverable mail messages of mailer-daemon error messages. In the example, the return path is "shreejaymall97@gmail.com".

17. Is it possible for an organization's Web server and mail server to have exactly the same alias for a hostname (for example, *foo.com*)? What would be the type for the RR that contains the hostname of the mail server?

An organization's Web server and mail server to have exactly the same alias for a hostname is possible. The MX record is used to map the mail server's host name to its IP address.

18. Look over your received emails, and examine the header of a message sent from a user with an .edu email address. Is it possible to determine from the header the IP address of the host from which the message was sent? Do the same for a message sent from a gmail account.

You should be able to see the sender's IP address for a user with an .edu email address. But you will not be able to see the sender's IP address if the user uses a gmail account.

19. In BitTorrent, suppose Alice provides chunks to Bob throughout a 30-second interval. Will Bob necessarily return the favor and provide chunks to Alice in this same interval? Why or why not?

In BitTorrent, It is not necessary that Bob will also provide chunks to Alice.

- Alice has to be in the top four neighbours of Bob and send message to her, this might not occur even if Alice is provides chunks to Bob throughout a 30-second interval.

20. Consider a new peer Alice that joins BitTorrent without possessing any chunks. Without any chunks, she cannot become a top-four uploader for any of the other peers, since she has nothing to upload. How then will Alice get her first chunk?

Chapter2: Application layer

Recall that in BitTorrent, a peer picks a random peer and optimistically unchokes the peer for a short period of time. Therefore, Alice will eventually be optimistically unchoked by one of her neighbors, during which time she will receive chunks from that neighbor.

21. True or false?

a. A user requests a Web page that consists of some text and three images. For this page, the client will send one request message and receive four response messages.

b. Two distinct Web pages (for example, www.mit.edu/research.html and www.mit.edu/students.html) can be sent over the same persistent connection.

c. With non-persistent connections between browser and origin server, it is possible for a single TCP segment to carry two distinct HTTP request messages.

d. The Date: header in the HTTP response message indicates when the object in the response was last modified.

e. HTTP response messages never have an empty message body.

a) False

b) True

c) False

d) False

e) False

22. Read RFC 959 for FTP. List all of the client commands that are supported by the RFC.

The client commands that are supported by the RFC as follows:

- **Access control commands:** USER, PASS, ACT, CWD, CDUP, SMNT, REIN, QUIT.
- **Service commands:** RETR, STOR, STOU, APPE, ALLO, REST, RNFR, RNTO, ABOR, DELE, RMD, MRD, PWD, LIST, NLST, SITE, SYST, STAT, HELP, NOOP.

- **Transfer parameter commands:** PORT, PASV, TYPE STRU, MODE.

23. Consider an HTTP client that wants to retrieve a Web document at a given URL. The IP address of the HTTP server is initially unknown. What transport and application-layer protocols besides HTTP are needed in this scenario?

Transport layer protocols:

- TCP for HTTP
- UDP for DNS;

Application layer protocols:

- DNS
- HTTP

24. Consider the following string of ASCII characters that were captured by Wireshark when the browser sent an HTTP GET message (i.e., this is the actual content of an HTTP GET message). The characters <cr><lf> are carriage return and line-feed characters (that is, the italized character string <cr> in the text below represents the single carriage-return character that was con- tained at that point in the HTTP header). Answer the following questions, indicating where in the HTTP GET message below you find the answer.

GET /cs453/index.html HTTP/1.1<cr><lf>Host: gai a.cs.umass.edu<cr><lf>User-Agent: Mozilla/5.0 (Windows;U; Windows NT 5.1; en-US; rv:1.7.2) Gec ko/20040804 Netscape/7.2 (ax) <cr><lf>Accept:ex t/xml, application/xml, application/xhtml+xml, text /html;q=0.9, text/plain;q=0.8,image/png,*/*;q=0.5 <cr><lf>Accept-Language: en-s,en;q=0.5<cr><lf>Accept- Encoding: zip,deflate<cr><lf>Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.7<cr><lf>Keep-Alive: 300<cr> <lf>Connection:keep-alive<cr><lf><cr><lf>
a. What is the URL of the document requested by the browser?
b. What version of HTTP is the browser running?
c. Does the browser request a non-persistent or a persistent connection?
d. What is the IP address of the host on which the browser is running?
e. What type of browser initiates this message? Why is the browser type needed in an HTTP request message?

a.The document request was http://gaia.cs.umass.edu/cs453/index.html. The Host : field indicates the server's name and /cs453/index.html indicates the file name.

b) The browser is running HTTP version 1.1, as indicated just before the first pair.

c) The browser is requesting a persistent connection, as indicated by the Connection: keep-alive.

d) This is a trick question. This information is not contained in an HTTP message anywhere. So there is no way to tell this from looking at the exchange of HTTP messages alone. One would need information from the IP datagrams (that carried the TCP segment that carried the HTTP GET request) to answer this question .

e) Mozilla/5.0. The browser type information is needed by the server to send different versions of the same object to different types of browsers.

25. The text below shows the reply sent from the server in response to the HTTP GET message in the question above. Answer the following questions, indicat- ing where in the message below you find the answer.

HTTP/1.1 200 OK<cr><lf>Date: Tue, 07 Mar 2008 12:39:45GMT<cr><lf>Server: Apache/2.0.52 (Fedora) <cr><lf>Last- odified: Sat, 10 Dec2005 18:27:46 GMT<cr><lf>ETag: "526c3-f22- 88a4c80"<cr><lf>Accept- Ranges: bytes<cr><lf>Content-Length: 3874<cr><lf> Keep-Alive: timeout=max=100<cr><lf>Connection: Keep- live<cr><lf>Content-Type: text/html; charset= ISO-8859- <cr><lf><cr><lf><!doctype html public "- //w3c//dtd html 4.0 transitional//en"><lf><html><lf> <head><lf> <meta http-equiv="Content- ype" content="text/html; charset=iso-8859-1"><lf> <meta name="GENERATOR" content="Mozilla/4.79 [en] (Windows NT 5.0; U) Netscape]"><lf> <title>CMPSCI 453 / 591 / NTU-ST550A Spring 2005 homepage</title><lf></head><lf> <much more document text following here (not shown)>

a. Was the server able to successfully find the document or not? What time was the document reply provided?

b. When was the document last modified?

c. How many bytes are there in the document being returned?

d. What are the first 5 bytes of the document being returned? Did the server agree to a persistent connection?

a) The server was able to locate the document successfully. Reason is that The status code of 200 and the phrase OK.

- The reply was provided on Tuesday, 07 Mar 2006 12:39:45 Greenwich Mean Time.

b) The document index.html was last modified on Saturday 10 Dec 2005 18:27:46 GMT.

c) There are 3874 bytes in the document being returned.

d) The first five bytes of the returned document are : <!doc. The server agreed to a persistent connection, as indicated by the Connection: Keep-Alive field

26. Obtain the HTTP/1.1 specification (RFC 2616). Answer the following questions: a. Explain the mechanism used for signalling between the client and server to indicate that a persistent connection is being closed. Can the client, the server, or both signal the close of a connection?

b. What encryption services are provided by HTTP?

c. Can a client open three or more simultaneous connections with a given server?

d. Either a server or a client may close a transport connection between them if either one detects the connection has been idle for some time. Is it pos- sible that one side starts closing a connection while the other side is transmitting data via this connection? Explain.

a)

Persistent connections are discussed in section 8 of RFC 2616 (the real goal of this question was to get you to retrieve and read an RFC). Sections 8.1.2 and 8.1.2.1 of the RFC indicate that either the client or the server can indicate to the other that it is going to close the persistent connection. It does so by including the connection-token "close" in the Connection-header field of the http request/reply.

b)

HTTP does not provide any encryption services.

c)

(From RFC 2616) “Clients that use persistent connections should limit the number of simultaneous connections that they maintain to a given server. A single-user client SHOULD NOT maintain more than 2 connections with any server or proxy.”

d)

Yes. (From RFC 2616) “A client might have started to send a new request at the same time that the server has decided to close the "idle" connection. From the server's point of view, the connection is being closed while it was idle, but from the client's point of view, a request is in progress.”

Chapter2: Application layer

27. Suppose within your Web browser you click on a link to obtain a Web page. The IP address for the associated URL is not cached in your local host, so a DNS lookup is necessary to obtain the IP address. Suppose that n DNS servers are visited before your host receives the IP address from DNS; the successive visits incur an RTT of RTT 1 , . . ., RTT n . Further suppose that the Web page associated with the link contains exactly one object, consisting of a small amount of HTML text. Let RTT 0 denote the RTT between the local host and the server containing the object. Assuming zero transmission time of the object, how much time elapses from when the client clicks on the link until the client receives the object?

Consider the IP address of total amount of time:

$RTT_1+RTT_2+\cdots+RTT_n$.

Time elapses from when the client clicks on the link until the client receives the object :

$2RTT_o+RTT_1+RTT_2+\cdots+RTT_n$

28. Referring to Problem 27, suppose the HTML file references eight very small objects on the same server. Neglecting transmission times, how much time elapses with

a. Non-persistent HTTP with no parallel TCP connections?

b. Non-persistent HTTP with the browser configured for 5 parallel connections?

c. Persistent HTTP?

Consider the IP address of total amount of time (Refer problem P7):

$RTT_1+RTT_2+\cdots+RTT_n$.

a. Non-persistent HTTP with no parallel TCP connections as follows:

$RTT_1+\cdots+RTT_n+2RTT_o+3\cdot 2RTT_o = 8RTT_o+RTT_1+\cdots+RTT_n$.

b. Non-persistent HTTP with the browser configured for 5 parallel connections as follows:

$RTT_1+\cdots+RTT_n+2RTT_o+2RTT_o = 4RTT_o+RTT_1+\cdots+RTT_n$.

c. Persistent HTTP as follows: $RTT_1+\cdots+RTT_n+2RTT_o+RTT_o$

29. Consider a short, 10-meter link, over which a sender can transmit at a rate of 150 bits/sec in both directions. Suppose that packets containing data are 100,000 bits long, and packets containing only control (e.g., ACK or hand- shaking) are 200 bits long. Assume that N parallel connections each get 1/N of the link bandwidth. Now consider the HTTP protocol, and suppose that each downloaded object is 100 Kbits long, and that the initial downloaded object contains 10 referenced objects from the same sender. Would parallel downloads via parallel instances of non-persistent HTTP make sense in this case? Now consider persistent HTTP. Do you expect significant gains over the non-persistent case? Justify and explain your answer.

Transmission rate(R)= 150 bits/sec
Packet length(L)=100,000 bits long
Control data=200 bits
Object data= 100 Kbits
Distance(d) =10 meter
N =10

$d = d_p$(propagation delay) $+ d_t(transmission\ delay)$

d_t=L/R seconds

d_p = d/s $= T_p$
Bandwidth = 150 bits/sec
Number of connections (N) = 10 [As 10 referenced objects]

$$\text{Bandwidth} = \frac{150}{10}\ \text{bits/sec}$$
$$= 15\,\text{bits/sec}$$

Total time for all received objects:

$$\left(\frac{200}{150}+T_p+\frac{200}{150}+T_p+\frac{200}{150}+T_p+\frac{100,000}{150}+T_p\right)+$$
$$\left(\frac{200}{15}+T_p+\frac{200}{15}+T_p+\frac{200}{15}+T_p+\frac{100,000}{15}+T_p\right)$$

$$=\left(\frac{200+200+200+100,000}{150}+4T_p\right)+\left(\frac{200+200+200+100,000}{15}+4T_p\right)$$
$$=\left(\frac{100,600}{150}+4T_p\right)+\left(\frac{100,600}{15}+4T_p\right)$$
$$=\left(670+4T_p\right)+\left(6706+4T_p\right)$$
$$=7377+8\times T_p\ Seconds$$

Total time for persistent HTTP connection:

$$\left(\frac{200}{150}+T_p+\frac{200}{150}+T_p+\frac{200}{150}+T_p+\frac{100,000}{150}+T_p\right)$$
$$+10\times\left(\frac{200}{150}+T_p+\frac{100,000}{150}+T_p\right)$$

$$=\left(\frac{200+200+200+100,000}{150}+4T_p\right)+10\times\left(\frac{200+100,000}{150}+2T_p\right)$$
$$=\left(\frac{100,600}{150}+4T_p\right)+10\times\left(\frac{100,200}{150}+2T_p\right)$$
$$=\left(670+4T_p\right)+\left(6680+20T_p\right)$$
$$=7350+24\times T_p$$

Let us that the propagation speed of the medium is 300×10^6 m/sec,

Then $T_p = \frac{10}{\left(300\times10^6\right)}$

$$=0.03\,micro\ seconds$$

29. Consider the scenario introduced in the previous problem. Now suppose that the link is shared by Bob with four other users. Bob uses parallel instances of non-persistent HTTP, and the other four users use non-persistent HTTP with- out parallel downloads.
a. Do Bob's parallel connections help him get Web pages more quickly? Why or why not?
b. If all five users open five parallel instances of non-persistent HTTP, then would Bob's parallel connections still be beneficial? Why or why not?

Given data in the question:

Suppose that the link is shared by Bob with four other users. Bob uses parallel instances of non-persistent HTTP, and the other four users use non-persistent HTTP with- out parallel downloads.

a) **Yes,** Bob's parallel connections help him get Web pages more quickly.

- The reason is that, the Bob shared total band with to four other users.
- So, Bob uses parallel instances of non-continual HTTP and other four customers are used non-persistent HTTP without parallel downloads.

b) **Yes,** If all five users open five parallel instances of non-persistent HTTP, then would Bob's parallel connections still be beneficial.

- The reason is that, every customer gets equal size of band width due to Bob does not parallel connection.

30. Write a simple TCP program for a server that accepts lines of input from a client and prints the lines onto the server's standard output. (You can do this by modifying the TCPServer.py program in the text) Compile and execute your program. On any other machine that contains a Web browser, set the proxy server in the browser to the host that is running your server program; also configure the port number appropriately. Your browser should now send its GET request messages to your server, and your server should display the messages on its standard output. Use this platform to determine whether your browser generates conditional GET messages for objects that are locally cached.

TCP Program code:

```
import java.io.*;

import java.net.*;



class TCPServerDemo

{
```

```
public static void main(String argv[]) throws Exception
{
String clientMessage;
ServerSocket messageSocket = new ServerSocket(6789);
while(true) {
Socket connectionSocket = messageSocket.accept();

                    BufferedReader bd = new BufferedReader(new InputStreamReader(
                                connectionSocket.getInputStream( ) ) );
                    clientMessage = bd.readLine();
                    System.out.println("Client message recieved: " + clientMessage + "\n");
}
}
}
```

31. What is the difference between MAIL FROM: in SMTP and From: in the mail message itself?

The difference between MAIL FROM: in SMTP and From: in the mail message as follows:

- The MAIL FROM: in SMTP is a message from the SMTP client that identifies the sender of the mail message to the SMTP server.
- The From: on the mail message itself is NOT an SMTP message, but rather is just a line in the body of the mail message.

32. How does SMTP mark the end of a message body? How about HTTP? Can HTTP use the same method as SMTP to mark the end of a message body? Explain.

Simple Mail Transport Protocol (SMTP) uses a line containing only a period to mark the end of a message body.

Hyper Text transport Protocol (HTTP) uses "Content-Length header field" to indicate the length of a message body.

No, Hyper Text transport Protocol (HTTP) cannot use the method used by Simple Mail Transport Protocol (SMTP), because Hyper Text transport Protocol (HTTP) message could be binary data, where as in Simple Mail Transport Protocol (SMTP), the message body must be in 7-bit ASCII format.

33. Read RFC 5321 for SMTP. What does MTA stand for? Consider the follow- ing received spam email (modified from a real spam email). Assuming only the originator of this spam email is malacious and all other hosts are honest, identify the malacious host that has generated this spam email.

From - Fri Nov 07 13:41:30 2008

Return-Path: <tennis5@pp33head.com>

Received: from barmail.cs.umass.edu

(barmail.cs.umass.edu [128.119.240.3]) by cs.umass.edu (8.13.1/8.12.6) for <hg@cs.umass.edu>; Fri, 7 Nov 2008 13:27:10 -0500

Received: from asusus-4b96 (localhost [127.0.0.1]) by

barmail.cs.umass.edu (Spam Firewall) for <hg@cs.umass.edu>; Fri, 7 Nov 2008 13:27:07 -0500 (EST)

Received: from asusus-4b96 ([58.88.21.177]) by

barmail.cs.umass.edu for <hg@cs.umass.edu>; Fri,

07 Nov 2008 13:27:07 -0500 (EST)

Received: from [58.88.21.177] by

inbnd55.exchangeddd.com; Sat, 8 Nov 2008 01:27:07 +0700

From: "Jonny" <tennis5@pp33head.com>

To: <hg@cs.umass.edu>

Subject: How to secure your savings

MTA stand for ***M***ail ***T***ransfer ***A***gent.

- MTA is used for exchange mails between two systems.
- After obeservation of above data, "*asusus-4b96 [58.88.21.177]*" does not found sender mail ID. If it is spam mail, then originator can be dishonest.

So, "*asusus-4b96 [58.88.21.177]*" is the spam mail.

34. Read the POP3 RFC, RFC 1939. What is the purpose of the UIDL POP3 command?

Read the POP3 RFC, RFC 1939. What is the purpose of the UIDL POP3 command?

- UIDL means "unique-ID listing".
- When a POP3 client issues the UIDL command, the server responds with the unique message ID in mailbox.
- It is useful for "download and keep".

35. a. What is a whois database?

b. Use various whois databases on the Internet to obtain the names of two DNS servers. Indicate which whois databases you used.

c. Use nslookup on your local host to send DNS queries to three DNS servers: your local DNS server and the two DNS servers you found in part (b). Try querying for Type A, NS, and MX reports. Summarize your findings.

d. Use nslookup to find a Web server that has multiple IP addresses. Does the Web server of your institution (school or company) have multiple IP addresses?

e. Use the ARIN whois database to determine the IP address range used by your university.

f. Describe how an attacker can use whois databases and the nslookup tool to perform reconnaissance on an institution before launching an attack.

g. Discuss why whois databases should be publicly available.

a)

The database "whois" is a type of database and it is used to store data of registered users of the internet like domain name (Example: www.sr2jr.com), mapped IP address (Example: 156.52.18.237), alex rank (Example: rank 10000 for www.sr2jr.com), etc.

b)

I searched the keyword "whois" database in the google. I found many DNS(Domain Naming System) servers in the google.

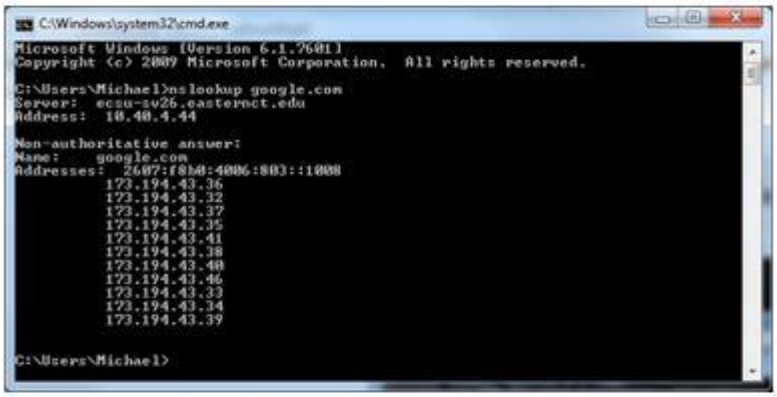
I prefered the following two DNS servers:

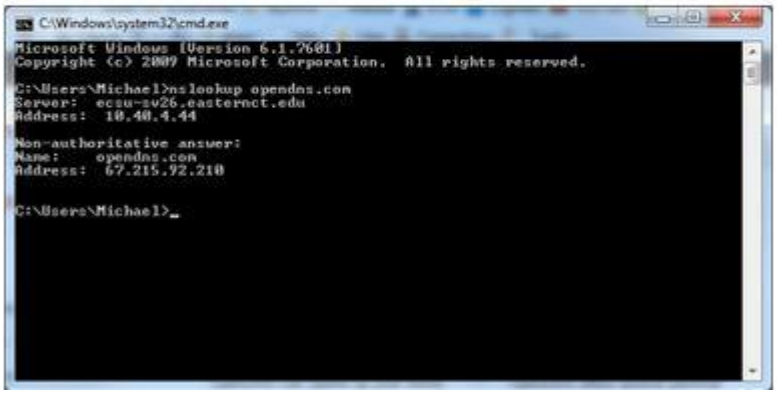
1. http://whois.icann.org
2. http://whois.domaintools.com/

c)

Open command prompt window and type "nslookup".

The following screen shots have Type A, NS, and MX reports:





Chapter2: Application layer

d)
Yes, the Web server of school/company have multiple IP addresses after my observations.

e)
USe the ARIN(American Registry for Internet Numbers) to found the IP address range.

The range of IP address of univerysity (example: sr2jr) is 10.20.3.27 – 12.10.24.24 and 139.111.12.50 – 139.152.16.75.

f)
The attackers collect every IP address the institution using and target those IP Addresses to attack by using whois databases and the nslookup tools. So, every institution require secure connections and strong end-user authentication.

g)
The "whois" databases should be publicly available. It is used to find out registration and IP information about registered domains.

- The main reason is identity of everyone in the world for the communication of growth.
- If people wants to verify data about a particular institutue/company domain, they can use a whois database simply.
- If it is not publicly avilable, then it is difficulty to find the domain data in other ways.

36. Suppose you can access the caches in the local DNS servers of your department. Can you propose a way to roughly determine the Web servers (outside your department) that are most popular among the users in your department? Explain.

We can periodically take a snapshot of the DNS caches in the local DNS servers. The Web server that appears most frequently in the DNS caches is the most popular server. This is because if more users are interested in a Web server, then DNS requests for that server are more frequently sent by users. Thus, that Web server will appear in the DNS caches more frequently.

For a complete measurement study, see:
Craig E. Wills, Mikhail Mikhailov, Hao Shang
“Inferring Relative Popularity of Internet Applications by Actively Querying DNS Caches”, in IMC'03, October 2729, 2003, Miami Beach, Florida, USA

37. Suppose that your department has a local DNS server for all computers in the department. You are an ordinary user (i.e., not a network/system administra- tor). Can you determine if an external Web site was likely accessed from a computer in your department a couple of seconds ago? Explain.

Yes, we can use dig to query that Web site in the local DNS server.

For example, "dig cnn.com" will return the query time for finding cnn.com. If cnn.com was just accessed a couple of seconds ago, an entry for cnn.com is cached in the local DNS cache, so the query time is 0 msec. Otherwise, the query time is large.

38. Consider distributing a file of F = 15 Gbits to N peers. The server has an upload rate of u s = 30 Mbps, and each peer has a download rate of d i = 2 Mbps and an upload rate of u. For N = 10, 100, and 1,000 and u = 300 Kbps, 700 Kbps, and 2 Mbps, prepare a chart giving the minimum distribution time for each of the combinations of N and u for both client-server distribution and P2P distribution.

Consider the data:

F = 15 Gbits =15360 Mbits (Convert 1 Gbits=1024 Mbits)

s = 30 Mbps

d_i = 2 Mbps

minimum distribution time :

$$D_{cs} = \max\left\{\frac{NF}{u_s}, \frac{F}{d_{\min}}\right\}$$
$$= \max\left\{\frac{10\times 15360}{30}, \frac{15360}{2}\right\}$$
$$= \max\{5120, 7680\}$$
$$= 7680\,\text{sec}$$

So, the minimum distribution time for each of the combinations of N and u for both client-server distribution and P2P distribution=7680 seconds

39. Consider an overlay network with N active peers, with each pair of peers having an active TCP connection. Additionally, suppose that the TCP connections pass through a total of M routers. How many nodes and edges are there in the corresponding overlay network?

There are N nodes in the overlay network.

There are N(N-1)/2 edges.

40. Suppose Bob joins a BitTorrent torrent, but he does not want to upload any data to any other peers (so called free-riding).

a. Bob claims that he can receive a complete copy of the file that is shared by the swarm. Is Bob's claim possible? Why or why not?

b. Bob further claims that he can further make his "free-riding" more efficient by using a collection of multiple computers (with distinct IP addresses) in the computer lab in his department. How can he do that?

a. Yes. His first claim is possible, as long as there are enough peers staying in the swarm for a long enough time. Bob can always receive data through optimistic un-choking by other peers
b. His second claim is also true. He can run a client on each host, let each client "free-ride," and combine the collected chunks from the different hosts into a single file. He can even write a small scheduling program to make the different hosts ask for different chunks of the file. This is actually a kind of Sybil attack in P2P networks.

41. Install and compile the Python programs TCPClient and UDPClient on one host and TCPServer and UDPServer on another host.

a. Suppose you run TCPClient before you run TCPServer. What happens? Why?

b. Suppose you run UDPClient before you run UDPServer. What happens? Why?

c. What happens if you use different port numbers for the client and server sides?

a) If you run TCPClient first, then the client will attempt to make a TCP connection with a non-existent server process. A TCP connection will not be made.
b) UDPClient doesn't establish a TCP connection with the server. Thus, everything should work fine if you first run UDPClient, then run UDPServer, and then type some input into the keyboard.
c) If you use different port numbers, then the client will attempt to establish a TCP connection with the wrong process or a non-existent process. Errors will occur.

42. Suppose that in UDPClient.py, after we create the socket, we add the line: clientSocket.bind(('', 5432))

Will it become necessary to change UDPServer.py? What are the port num- bers for the sockets in UDPClient and UDPServer? What were they before making this change?

```
from socket import *

        .

        #write the remaining code here

        .

clientSocket = socket(socket.AF_INET, socket.SOCK_DGRAM)

clientSocket.bind((‘ ’, 5432))

message = raw_input(‘Input lowercase sentence:’)

        .

        #write the remaining code here

        .

clientSocket.close()
```

After modifications then the Port numbers of client and server:

DP Client port number - 5432

UDP Server port number - 12000

Before modifications the Port numbers of client and server :

UDP Client port number - xxxx (number assigned by OS)

UDP Server port number - 12000

43. Can you configure your browser to open multiple simultaneous connections to a Web site? What are the advantages and disadvantages of having a large number of simultaneous TCP connections?

Yes, you can configure many browsers to open multiple simultaneous connections to a Web site. The advantage is that you will you potentially download the file faster. The disadvantage is that you may be hogging the bandwidth, thereby significantly slowing down the downloads of other users who are sharing the same physical links.

44. We have seen that Internet TCP sockets treat the data being sent as a byte stream but UDP sockets recognize message boundaries. What are oneadvantage and one disadvantage of byte-oriented API versus having the API explicitly recognize and preserve application-defined message boundaries?

For an application such as remote login (telnet and ssh), a byte-stream oriented protocol is very natural since there is no notion of message boundaries in the application. When a user types a character, we simply drop the character into the TCP connection.
In other applications, we may be sending a series of messages that have inherent boundaries between them. For example, when one SMTP mail server sends another SMTP mail server several email messages back to back. Since TCP does not have a mechanism to indicate the boundaries, the application must add the indications itself, so that receiving side of the application can distinguish one message from the next. If each message were instead put into a distinct UDP segment, the receiving end would be able to distinguish the various messages without any indications added by the sending side of the application.

45. Many BitTorrent clients use DHTs to create a distributed tracker. For these DHTs, what is the "key" and what is the "value"?

The many BitTorrent clients use DHTs (Distributed Hash Tables) to create a distributed trakcer. The key and value are used in stored peers.

They are as follows:

***key* :Torrent identifier**

Value: IP (Internet Protocol) address of a peer that is participating in the current torrent.

References:

http://www.sr2jr.com/textbook-solutions/computer-science/10201027/computer-networking-a-top-down-approach-application-layer

http://www.sr2jr.com/textbook-solutions/computer-science/10102001/computer-networking-a-top-down-approach-computer-networks-and-the-internet