UNIVERSITY OF DUBLIN

TRINITY COLLEGE

Faculty of Engineering and Systems Sciences

Department of Computer Science

B.A. (Computer Science)
SENIOR FRESHMAN EXAMINATION

Trinity Term 2007

2BA5
Telecommunications and Information Management

Thursday 24th May 2007   GMB   09:30 – 12:30

S. Weber, L. Hederman, D. O'Sullivan, S. Lawless

Answer **five** questions, at least **two** from each section.
Use separate answer books for each **question**.

Section A

1. The data link layer is responsible for the communication between two stations that are connected by a physical medium. Data compression, error detection and error correction are part of the data link layer.

(a) The data transfer between two stations is prone to errors through signal corruption and interference. Assume that you have two stations that are connected by a cable. The cable offers protection against interference e.g. an error may occur every 1000 bytes. Half of the errors are single-bit errors and the other half are burst errors of 4 bit length.

Assess the suitability of the following approaches for error detection of a transfer of 100,000 bytes in the scenario described above: 8-bit parity check, 32-bit CRC for 128 bytes and 32-bit checksum for 128 bytes. Your assessment should outline your reasoning for or against the suitability of each approach and advantages and disadvantages of each approach arising through characteristics such as processing requirements, overhead, etc.

(10 marks)

(b) Flow control is used to prevent a sender from overflowing a receiver by giving the receiver control over the amount of data the sender is allowed to transmit. Contrast an approach based on a Sliding-Window with 4 bits for packet identifiers against an approach using a Stop-And-Wait protocol. Your analysis should examine factors such as throughput, packet delays, and reaction to transmission failures.

(5 marks)

(c) Huffman encoding uses statistical analysis of texts to determine the optimal codes for individual characters. Table 1 lists the distribution of characters from a text. Give the binary tree for Huffman encoding as a diagram and use this binary tree to determine the code for the following sequence of characters: "Internet Protocols". Remember: The tree is built by combining the sub-trees with the smallest weights.

e	10	i	8
c	2	s	7
o	6	n	4
l	5	p	2
t	3	r	2

Table 1: Distribution of characters
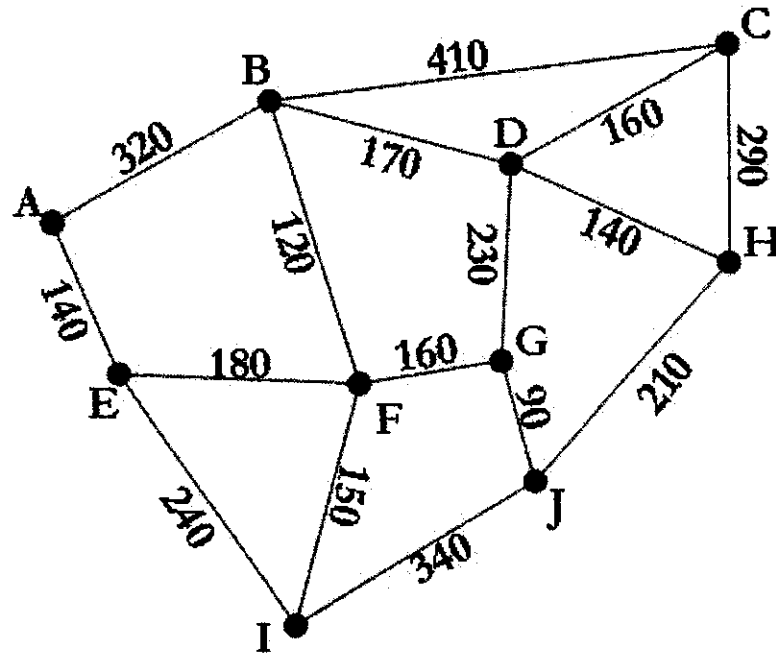
(5 marks)
(Total 20 marks)

2. The communication between two nodes is governed by network layer protocols such as the Internet Protocol (IP) and by routing protocols such as Open Shortest Path First (OSPF).

(a) The Internet Protocol is used to provide connectivity between nodes that are connected by a variety of physical networks. Assume that node A wants to transfer a packet of 5000 bytes to node B. The network that node A is connected to has a maximum transfer unit (MTU) of 4000 bytes. Router C connects this network to another network that has a MTU of 1500 bytes. Router D connects this network to yet another network that has a MTU of 3000. Router E in turn connects this network to the network on which node B is located. The network that router B is connected has a MTU of 1500 bytes.

Analyse the behaviour of IP in terms of the transport of packets over networks with varying MTUs in the given scenario, highlight the advantages and disadvantages of the approach taken by IP and propose an alternative approach. Your analysis should be accompanied by a diagram that visualises the infrastructure and the transmission of the data between node A and node B.

(10 marks)

(b) Link State Routing (LSR) represents one of the major routing concepts. Describe the LSR concept in your own words. The description should be accompanied by diagrams where appropriate. Dijkstra's Shortest Path algorithm is used in LSR to determine the routing table of individual nodes. Describe this algorithm and explain it with the help of the following diagram.

(10 marks)
(Total 20 marks)

3. The Transport Control Protocol (TCP) is widely used to transfer data between computers. In order to fulfil this task efficiently a number of mechanisms have been developed that adapt TCP to the limitations of networks and computers.

(a) A TCP/IP connection is to be established between node A and node B to transfer data for an HTML interaction. Assume that both TCP and IP headers are 20 bytes long and the smallest maximum transfer unit (MTU) between the two nodes is 1540 bytes i.e. the largest packet can include 1500 bytes payload, a TCP header and an IP header. Every individual data packet is replied to with an acknowledgement.

The first TCP segment send by the node A consists of the HTML request of 500 bytes. Node B replies to this with a web page of 1500 bytes. The web page contains a picture and node A requests this picture in a third segment with a payload of 200 bytes. Node B replies with the picture which is 2800 bytes.

Evaluate the overhead introduced by the packets necessary to maintain the TCP lifecycle against the amount of data that is exchanged. Your evaluation should include a diagram that visualises the packets that are exchanged between the two nodes.

(12 marks)

(b) Evaluate the congestion control mechanisms "Additive Increase Multiplicative Decrease (AIMD)" and basic "Slow Start" in the following two scenarios:

I)

Time /ms	Event	Cw
0	-	
10	ACK	
20	ACK	
30	ACK	
40	ACK	
50	ACK	
60	ACK	
70	ACK	
80	TO	
90	ACK	
100	ACK	

Time /ms	Event	Cw
110	ACK	
120	ACK	
130	ACK	
140	TO	
150	ACK	
160	ACK	
170	ACK	
180	ACK	
190	TO	
200	ACK	

II)

Time /ms	Event	Cw
0	-	
10	ACK	
20	ACK	
30	ACK	
40	TO	
50	ACK	
60	ACK	
70	TO	
80	ACK	
90	ACK	
100	ACK	

Time /ms	Event	Cw
110	TO	
120	ACK	
130	ACK	
140	TO	
150	ACK	
160	ACK	
170	ACK	
180	TO	
190	ACK	
200	ACK	

ACK = Acknowledgement

TO = Time Out

Your evaluation should contrast the two mechanisms with each other for both scenarios and show the behaviour of each mechanism in every scenario graphically.

(8 marks)

(Total 20 marks)

4. Medium access control is used to coordinate the communication over a shared medium. A number of protocols have been developed that provide various degrees of flexibility in the access to the shared medium.

(a) Assume that four stations use a wireless medium to communicate. All stations are within coverage of each other and station may intend to transmit data at the same time as all other stations. The access to the wireless medium is controlled by a Carrier Sense Multiple Access (CSMA) scheme with Collision Avoidance (CA) or a Time Division Multiple Access (TDMA) scheme. Every reservation slot is 10ms long and every contention free slot is 100ms. Contrast the two access control schemes with each other considering 2 scenarios:

I. High network load i.e. every station would like to send data at any given time.

II. Low network load i.e. stations only have data to be transmitted every few seconds.

Your analysis should be accompanied by a diagram that visualises the behaviour of the two schemes in each of the scenarios.

(12 marks)

(b) Four stations in a network are connected by a bus and each station intends to transmit data at the same time as all other stations. The access to the bus controlled by a Carrier Sense Multiple Access (CSMA) scheme with Collision Detection (CD). Describe the process that the stations follow in their attempts to acquire the bus and to send their data. The description should include a diagram and an explanation of the diagram.

(4 marks)

(c) Describe Code Division Multiple Access (CDMA) in your own words. The description should explain the process that is applied by each station using CDMA and the background that the mechanism is based on. The process should also be visualized using a diagram.

(4 marks)

(Total 20 marks)

Section B

5. *Keysort* is a technique for sorting a large number of records with only enough main memory to store the keys of those records.

(a) Outline the keysort algorithm.

(6 marks)

(b) Why can keysort not be applied to files stored on magnetic tape ?

(3 marks)

(c) If a disk read or write to a random location costs R milliseconds, and a sequential read or write operation averages S milliseconds what is the (approximate) cost of keysort for a file of N records, assuming only one disk is available ?

(3 marks)

(d) If a second disk drive is available how can you improve on the answer to (c)?

(3 marks)

(e) Outline very briefly how files are sorted if they are too large for keysort.

(5 marks)

(Total 20 marks)

6. A large file of about 100,000 records is to be used for statistical analysis of demographic data. The records consist of (RSI#, birth-year, town/city, county, business sector, job title, net-annual-pay). RSI# is the primary key for the file but is not relevant for this application. The file is stored as a serial (not necessarily sequential) file on magnetic tape. There is no primary key index.

(a) Access is required to the file by county, age, and net-annual-pay.
Discuss the design and structure of inverted files for these attributes.
Assume that all birth-years are in the range 1931 to 1980 inclusive (i.e. ages range from 16 to 65 inclusive).

(8 marks)

(b) Given the inverted files of part (a), outline the procedure you would use to find the business sector in which most of those under 25 years of age taking home more than €20,000 work.

(6 marks)

(c) Consider replacing the inverted files of part (a) by bit map representations.

  (i)   How many bits per record are required ?

  (ii) What advantage other than space might bit maps have for this kind of application ?

(6 marks)

(Total 20 marks)

7. (a) Explain the difference between a Web Search Engine and a traditional Information Retrieval system.

(5 marks)

(b) Describe the architecture of a typical Web Search Engine.

(5 marks)

(c) Define the evaluation measures recall and precision. Is recall or precision more important in evaluating web retrieval systems and why?

(5 marks)

(d) A collection of web pages "c" contains 10 individual web pages. Table 2 below shows the search result rankings obtained from running query "q" on collection "c" using two different web retrieval systems.

  (i)   Assuming web pages WP4, WP5 and WP6 are the only relevant pages in collection "c" for query "q", calculate the recall and precision values at ranking positions 3, 7, and 8 for both retrieval systems.

  (ii) At rank position 7 which system gives the best performance?

(5 marks)

(Total 20 marks)

Ranking Position	Retrieval System A	Retrieval System B
1	WP1	WP2
2	WP3	WP4
3	WP4	WP6
4	WP6	WP8
5	WP7	WP9
6	WP2	WP3
7	WP8	WP5
8	WP5	WP10
9	WP10	WP1
10	WP9	WP7

Table 2: Search Rankings for Query "q"

8 (a) What is XML used for? Provide three reasons why XML has grown rapidly in popularity.

(3 marks)

(b) Describe the difference between DOM processing and SAX processing of XML. In addition, from the evidence given in the XML documents shown in Figure X, propose a DTD for the document. Explain your design decisions.

(7 marks)

(c) An Xpath expression is made up of a series of "location steps", where each individual location step is separated by either a "/" or "//". In general each individual location step has three parts. Describe the three parts referred to.

In addition, provide for each of the following queries upon the XML document shown in Figure 1, an XPath expression that represents the query. Explain the rationale for each XPath expression used.

(i) Retrieve all streets from ShipTo elements;

(ii) Retrieve all descriptions of items that have a manufacturer;

(iii) Retrieve all items where quantity ordered is greater than 1;

(iv) Retrieve the second comment from the first item.

(10 marks)

(Total 20 marks)

```xml
<?xml version="1.0" encoding="ISO-8859-1"?>

<purchaseOrder>
<shipTo exportCode="1" type="UKAddress">
        <name>Helen Zoe</name>
        <street>47 Eden Street</street>
        <city>Cambridge</city>
        <postcode>CB1 1JR</postcode>
</shipTo>

<billTo type="Address">
        <name>Robert Smith</name>
        <street>8 Oak Avenue</street>
        <city>Old Town</city>
        <county>cork</county>
        <zip>95819</zip>
</billTo>

<items>
        <item partNum="833-AA" >
                <productName>Lapis necklace</productName>
                <quantity>1</quantity>
                <EuroPrice>99.95</EuroPrice>
                <comment>Want this for the holidays!</comment>
                <comment>Must send by express</comment>
                <shipDate>2006-01-10</shipDate>
        </item>

        <item partNum="123-BA" manufacturer="Duracell">
                <productName>Batteries</productName>
                <quantity>10</quantity>
                <shipDate>2006-01-10</shipDate>
                <description>
                    <colour>black</colour>
                    <material>metal</material>
                </description>
        </item>
</items>
</purchaseOrder>
```

Figure 1