UNIVERSITY OF DUBLIN

TRINITY COLLEGE

Faculty of Engineering, Mathematics & Science

School of Computer Science and Statistics

B.A. (Computer Science)

SENIOR FRESHMAN EXAMINATION

Annual Examination 2010

Trinity Term 2010

CS2031 and CS2041

Telecommunications and Information Management

Friday 30th April 2010   RDS – MAIN   14.00 – 17.00

Stefan Weber, John Keeney and Seamus Lawless

Answer **five** questions, at least **two** from each section.

All questions carry equal marks (20 marks).

Use separate answer books for each QUESTION.

Materials permitted for this examination:
Calculator (non-programmable)

Materials omitted from the front page of an examination paper will not be permitted during an examination.

Section A

Question 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.

(1a) The following bit sequence 1001011001111001 was received and is to be checked using CRC with the following polynomial x^5+x^2+x+1.

Show the data bits and CRC bits of the bit sequence that was received and demonstrate the first 4 steps of the polynomial long division. Discuss the interpretation of the possible outcomes of the polynomial long division.

(8 marks)

(1b) Demonstrate the encoding and decoding of the following 7 bits using Hamming code: 1011011 and explain how the result of the decoding may be used by a receiver.

(6 marks)

(1c) Dictionary-based approaches represent one common class of compression algorithms. Explain the central idea of this type of compression algorithms and give and example of the use of a dictionary-based algorithm on a string of characters of your choosing. The description of the example should include an explanation of the individual steps of the algorithm.

(6 marks)

(Total 20 marks)

Question 2

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

(2a) An Internet Service Provider (ISP) has bought the right to use the IP addresses in the range from 213.49.0.0 to 214.57.255.255. It uses Classless Inter-Domain Routing (CIDR) to route traffic to these addresses. It receives a number of requests from companies. First company A buys a block of 15,000 addresses, then company B requests 5,000, followed by company C with 900 addresses and company D requests 350 addresses. The ISP processes these requests in the order it receives them. What is the address range allocated for each client? Give the first and last address of the range, the number of significant bits and the subnet mask.

If CIDR wasn't used, what classes of network addresses would be allocated to each client? How many addresses would be allocated in total? What would be the fraction of addresses actually used by each client? Compare this to the use of CIDR.

(10 marks)

Question 2 is continued on the next page

Question 2 continued from the previous page

(2b) 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. Contrast the approach of LSR to Distance Vector Routing.
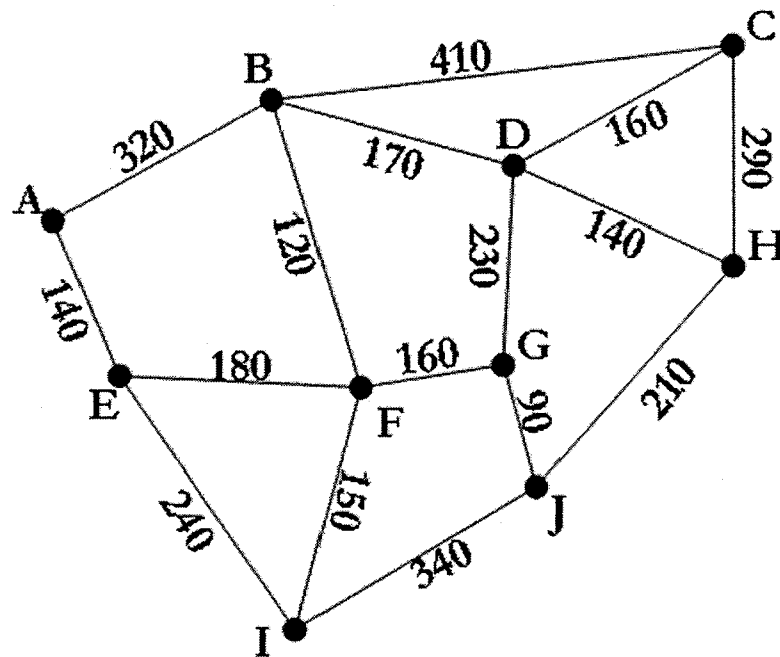
(10 marks)

(Total 20 marks)

Question 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 limitations of networks and computers.

(3a) 4000 bytes are transferred between two hosts, A and B. In one case the Transport Control Protocol (TCP) is used for this transfer and in a second case the User Datagram Protocol (UDP) is used. Describe the packets that are involved in the transfer of the data in each case and compare the overhead that is involved in the transfer. The description of the communication should include a description of the reaction to transmission errors. Discuss the advantage and disadvantage of each protocol using the transfer of the 4000 bytes as an example. Use diagrams to visualise the communication for each case.

(12 marks)

(3b) The terms "Stop and Wait", "Go-back-n" and "Selective Repeat" specify well-known Flow and Error Control mechanisms. Explain each of the mechanisms in turn including the state that needs to be kept at the sender and receiver side of a transfer, the reaction of each mechanism to transmission errors in the transfer of the various packets and the advantages and disadvantages of each mechanism. The explanation of each mechanism should be accompanied by diagrams that visualise the transfer of the packets.

(8 marks)

(Total 20 marks)

Question 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.

(4a) An access point uses the Point Coordination Function (PCF) of 802.11 to communicate with 4 laptops that are associated with it. After the contention free period has been completed, 4 laptops attempted to communicate with one another directly using the Distributed Coordination Function (DCF) of 802.11. Assume that at least a number of transmission attempts lead to collisions.

Describe the frames that are exchanged by the stations and the inter frame spaces that are involved in this exchange. Use diagrams to visualise the chronological exchange of the frames and the inter frame spaces.

(12 marks)

(4b) Assume that four stations use a wired medium to communicate. All stations may intend to transmit data at the same time as all other stations. The access to the medium is controlled by a Carrier Sense Multiple Access (CSMA) scheme with Collision Detection (CD) or a Time Division Multiple Access (TDMA) scheme with a reservation protocol. For the CSMA/CD approach, assume times as you see fit. For the TDMA approach, assume that 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 diagrams that visualise the behaviour of the two schemes in each of the scenarios.
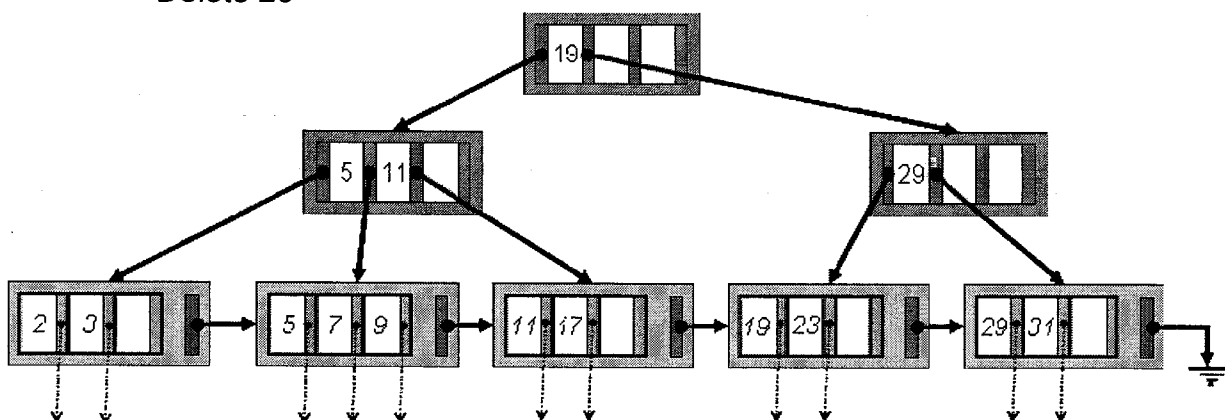
(8 marks)

(Total 20 marks)

Section B

Question 5

5a When is it appropriate to use a multi-level index instead of a single level index?

(2 Marks)

5b Give 4 reasons why B-Trees (and variants) are appropriate for efficiently arranging disk-based data?

(4 Marks)

5c What are the differences between B-Trees and B+Trees? Briefly compare how each difference impacts the structure and efficiency of the two tree types.

(6 Marks)

5d From the picture below, explain if this is a B-Tree or a B+Tree?
Draw and describe what happens to the tree shown for each of the following operations:
  • Insert 10
  • Insert 8
  • Delete 23

(8 Marks)
(Total 20 marks)

Question 6

6a What are the advantages and disadvantages of using a Hashing approach to store data on a disk?

(4 Marks)

6b Briefly compare how to search a large sorted list and a large unsorted list of data stored on a disk.

(4 Marks)

6c Using an example, describe in detail a method to sort a large list of data on a disk, where the total size of the list on disk is much greater than the amount of RAM available.

(12 Marks)
(Total 20 marks)

Question 7

7a Given the following DTD:

```
<!DOCTYPE db [
<!ELEMENT db (person*)>
<!ELEMENT person (name,age,(parent|guardian)?)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT age (#PCDATA)>
<!ELEMENT parent (person)>
<!ELEMENT guardian (person)>
]>
```

Write a valid document for this DTD that contains the following information:
John Doe is 30 years old. His parent is Robert Doe, who is 55 years old.

(6 Marks)

7b   Consider the following XML document:

```
<Vehicles>
  <Car manf="Hyundai">
    <Model>Azera</Model>
    <HorsePower>240</HorsePower>
  </Car>
  <Car manf="Toyota">
    <Model>Camry</Model>
    <HorsePower>240</HorsePower>
  </Car>
  <Truck manf="Toyota">
    <Model>Tundra</Model>
    <HorsePower>240</HorsePower>
  </Truck>
  <Car manf="Hyundai">
    <Model>Elantra</Model>
    <HorsePower>120</HorsePower>
  </Car>
  <Car manf="Toyota">
    <Model>Prius</Model>
    <HorsePower>120</HorsePower>
  </Car>
</Vehicles>
```

What are the results of the following XPath expressions?
(a)   /Vehicles/Car[@manf="Toyota"]/HorsePower
(b)   /Vehicles/*[HorsePower>200]/Model
(c)   //*[@manf="Toyota"]/@manf
(d)   /child::*/*/parent::node()/parent::node()

(6 Marks)

7c   Your uncle runs a distribution company for computer parts. He needs to design a new way to store the information about his customers, suppliers and parts. He wants to use text files to store this information.
- List 4 reasons why XML would be better for this task.
- List 4 reasons why text files would be better than XML.

(8 Marks)
(Total 20 marks)

Question 8
8a

(i) Explain why *indexing* a document or web page is necessary in information retrieval.

(2 Marks)

(ii) Giving examples, outline how the processes known as *tokenisation, stopword removal* and *stemming* are involved in indexing.

(3 Marks)

(iii) Define and explain the index term weights *tf* and *idf* and explain what these weights tell us about the importance of an index term.

(5 Marks)

8b A collection of web pages *"c"* contains 10 individual web pages. The table below shows the search result rankings obtained from running query *"q"* on collection *"c"* using two different web retrieval systems.

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	wp10	wp10
9	wp5	wp1
10	wp9	wp7

(i) Assuming web pages wp1, wp7 and wp8 are the only relevant pages in collection *"c"* for query *"q"*, calculate the recall and precision values at ranking positions 1, 4 and 8 for both retrieval systems.

(5 Marks)

(ii) Why is Average Precision an important measure in Web-based Information Retrieval? Calculate the Average Precision Value for both systems in the example above.

(5 Marks)
(Total 20 marks)