UNIVERSITY OF DUBLIN

TRINITY COLLEGE

Faculty of Engineering, Mathematics & Science
School of Computer Science & Statistics

B.A.(Mod.) Computer Science
Senior Freshman Examination

Trinity Term 2009

2BA5

Telecommunications and Information Management

Thursday 21st May 2009　　Exam Hall　　09:30 – 12:30

Dr Declan O'Sullivan, Dr Stefan Weber

Instructions to Candidates:

□ Answer FIVE questions, at least TWO from each section.
□ Use separate answer books for each question.
□ All questions carry equal marks (20 marks)

Materials permitted for this examination:

□ Use of non-programmable calculators is permitted.
□ Please note the make and model of your calculator on your answer book.

Section A

Question 1

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) Dictionary-based approaches represent a common class of compression algorithms. Explain the central idea of this type of compression algorithm and give an 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.

(5 marks)

(1b) Assume you have connections with the following characteristics:

   i) A connection that exhibits sporadic bursts of errors. Transmissions over this connection are not costly and the transmission delay and round-trip-time are considered to be short.

   ii) A connection that exhibits few errors which are generally limited to single bits. Transmissions over this connection are costly and the transmission delay and round-trip-time are considered to be short.

Propose an error detection approach for the connections in i) and an error detection approach for connections in ii) and explain the details of each approach. Justify your choice and contrast it against an alternative that would not be as suitable as your choice.

(10 marks)

(1c) Assume you have a connection between two stations that are limited in processing power and storage capacity. Suggest a flow control mechanism that would be suitable for this connection, explain the details of this mechanism and justify your choice by contrasting the mechanism against an alternative mechanism.

(5 marks)

(Total 20 marks)

Question 2

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) A computer L sends a UDP packet of 10000 bytes – including the UDP header - to a computer M. Computer L is connected to network N1 and Computer M is connected to network N3. Network N1 has a maximum transfer unit (MTU) of 4000 bytes and is connected by a router R1 to network N2. Network N2 has an MTU of 600 bytes and is connected by a router R2 to network N3. Network N3 has an MTU of 1500 bytes. All networks and computers use IPv4 to transfer the data.

Determine how many IP packets computer M will receive and discuss the progress of the IP packets from the source to the destination. Assume that the length of an IP header is 20 bytes. Your analysis should be accompanied by a diagram that visualises the infrastructure and the transmission of the data between computer L and computer M.

(10 marks)

(2b) Describe and contrast the two concepts of distance vector routing and link state routing on the example of the network of Trinity College. Assume that schools such as "Computer Science&Statistics (CSS)", "Electronic Engineering (EE)", "Genetics (G)", etc have individual routers and that these routers are partially connected to one another e.g. the router from CSS may be connected to the router from EE but not to the router from G. The description should include diagrams that visualise the process that each concept follows to establish routing tables.

(10 marks)

(Total 20 marks)

Question 3

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) The transfer of a webpage consists of the transfer of a request from host A to host B, a response that contains the text of the webpage from host B to host A, two additional requests for two images from host A to host B, and two responses that contain the requested images. All requests are 300 bytes long; the webpage is 4000 bytes long and the images are 2000 and 3000 bytes long. A segment in this transfer can hold at maximum of 1300 byte payload.

Describe the transfer of the webpage in as much detail as possible including such details as the control overhead in the communication and the reaction to the loss of segments. A figure of the TCP header is given below in Figure 1.
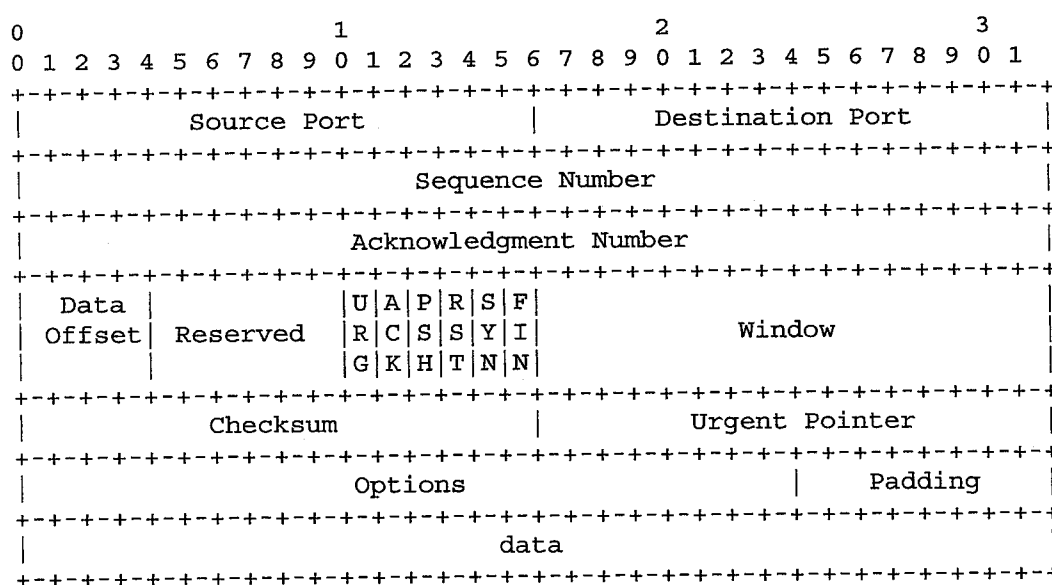
```
0                   1                   2                   3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Source Port          |       Destination Port        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Sequence Number                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Acknowledgment Number                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Data |           |U|A|P|R|S|F|                               |
| Offset| Reserved  |R|C|S|S|Y|I|            Window             |
|       |           |G|K|H|T|N|N|                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Checksum            |         Urgent Pointer        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    Options                    |    Padding    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             data                              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
Figure 1: TCP Header Format

(15 marks)

(3b) Describe the term congestion control in your own words and explain the motivation for the implementation of congestion control. As part of the description of the term congestion control, discuss the behaviour of additive increase/multiplicative decrease (AIMD) and its advantages and disadvantages.

(5 marks)

(Total 20 marks)

Question 4

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) Assume that eight stations are connected by a bus and each station wants to communicate with other stations on the bus. Suggest 3 schemes to allocate the medium to individual stations and discuss the advantages and disadvantages of each scheme with respect to the scenario described above. At least one scheme should allocate the access to the medium dynamically and at least one scheme should use a fixed allocation of the medium.

Your analysis should be accompanied by diagrams that visualize the behaviour of each of the three schemes.

(10 marks)

(4b) Explain the terms "hidden terminal problem" and "exposed terminal problem". Each explanation should be accompanied by a diagram that visualises the characteristics that the term describes.

(4 marks)

(4c) IEEE 802.11 defines two coordination functions: the distributed coordination function (DCF) and the point coordination function (PCF). Explain the details of the two coordination functions and how they are used in IEEE 802.11.

(6 marks)

(Total 20 marks)

Section B

Question 5

5a   Describe what is meant by a "Solid State Drive" and name two advantages they have over "Hard Disk Drives".

(3 Marks)

5b   Describe the parts (hardware and software) of a computer involved in moving a piece of data from an application and encoding it onto a disk. In other words describe "the journey of a byte" during the writing process.

(7 Marks)

5c   Consider a disk with the following characteristics: block size B = 512 bytes, interblock gap size G = 128 bytes, number of blocks per track=25, number of tracks per surface=400. A disk pack consists of 15 double sided disks.

   i.   What is the total capacity of a track and what is its useful capacity (excluding interblock gaps)?

   ii.   What is the total capacity and useful capacity of a cylinder?

   iii.   What is the total capacity and the useful capacity of a disk pack?

   iv.   Suppose an average seek time is 20 msec, average rotational delay of 10msec, and block transfer time of 1 msec. How much time does it take (on average) in msec to locate and transfer a single block given its block address?

   v.   Calculate the average time it would take to transfer 20 random blocks and compare it with the time it would take to transfer 20 consecutive blocks using double buffering to save seek time and rotational delay.

   vi.   Assuming 500 contiguous blocks make up a sequential file. Calculate the amount of time needed to do a binary search on the file

(10 Marks)

(Total 20 marks)

Question 6

6a Describe why hashing techniques have limitations with respect to indexing.

(3 Marks)

6b Describe what is a B-Tree and explain the rules involved in **inserting** a value into a node, including rules for splitting nodes. Use diagrams to illustrate.

(7 Marks)

6c Given the B-Tree in Figure A (with m=5), show the growth of the B-Tree as you add the keys in the following order 16, 18, 21, 3, 22, 25, 14, 30, 35, 42, 53, 23, 55, 67, 70
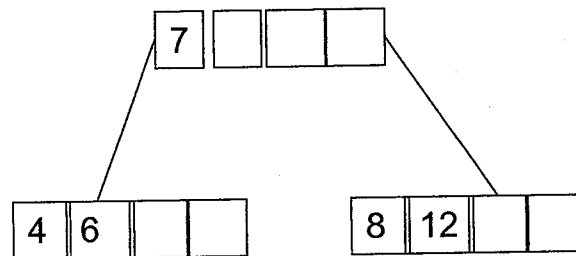
Figure A

i. Show the tree at each stage and provide an explanation for any splits

ii. What is the average number of node splits per new key inserted?

iii. What is the average search time?

(10 Marks)

(Total 20 marks)

Question 7

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<news>
<news_item timestamp='1'>
  <title> Gorilla Corporation acquires YouNameItWeIntegrateIt.com </title>
  <content>
    <par> Today, Gorilla Corporation announced that it will purchase YouNameItWeIntegrateIt.com. The shares of
YouNameItWeIntegrateIt.com dropped $3.00 as a result of this announcement.
    </par>
    <par> As a result of this acquisition, the CEO of YouNameItWeIntegrateIt.com Bill Smarts resigned. He did not
announce what he will do next.  Sources close to YouNameItWeIntegrateIt.com hint that Bill Smarts might be taking a
position in Foobar Corporation. </par>
    <par>YouNameItWeIntegrateIt.com is a leading systems integrator that enables <quote>brick and mortar</quote>
companies to have a presence on the web. </par>
  </content>
  <date>1-20-2000</date>
  <author>Mark Davis</author>
  <author>John Smith</author>
  <news_agent>News Online</news_agent>
</news_item>
<news_item timestamp='2'>
  <title>Foobar Corporation releases its new line of Foo products today</title>
  <content>
    <par> Foobar Corporation releases the 20.9 version of its Foo products.  The new version of Foo products solve
known performance problems which existed in 20.8 line and increases the speed of Foo based products tenfold. It also
allows wireless clients to be connected to the Foobar servers.
    </par>
    <par> The President of Foobar Corporation announced that they were proud to release 20.9 version of Foo products
and they will upgrade existing customers <footnote>where service agreements exist</footnote> promptly.
    </par>
    <figure>
      <title>Presidents of Foobar Corporation and TheAppCompany
      Inc. Shake Hands</title> <image source='handshake.jpg'/>
    </figure>
  </content>
  <date>1-20-2000</date>
  <news_agent>Foobar Corporation</news_agent>
</news_item>
<news_item >
<title>Foobar Corporation is suing Gorilla Corporation for patent infringement </title>
  <content>
    <par> In surprising developments today, Foobar Corporation announced that it is suing Gorilla Corporation for patent
infringement. The patents that were mentioned as part of the lawsuit are considered to be the basis of Foobar
Corporation's <quote>Wireless Foo</quote> line of products.
    </par>
    <par>The tension between Foobar and Gorilla Corporations has been increasing ever since the Gorilla Corporation
acquired more than 40 engineers who have left Foobar Corporation, TheAppCompany Inc. and
YouNameItWeIntegrateIt.com over the past 3 months. Most of the engineers have relocated to Hawaii where the Gorilla
Corporation's server development is located. </par>
  </content>
  <date>1-20-2000</date>
</news_item>
</news>
```

Figure B

7a   What is XML Schema and what is it used for? Describe **two** of the statements in
XML Schema that has no equivalence in DTDs

(3 Marks)

Question 7 continued from the previous page

7b   Describe the relationship between the "XML tree model" and XPath expressions. List and describe the different types of nodes (e.g. comment node) possible in the tree model.   Draw a diagram showing the nodes representing the subtree of element **<news_item timestamp='2'>** from the document shown in Figure B.

(7 Marks)

7c   Describe what is meant when describing Xquery as consisting of "FLWOR" expressions. In addition, define and explain XQuery statements for each of the following queries posed over the document in Figure B. Also show expected results and explain your design decisions.

   i.   Return the second <par> element for each news_item

   ii.   Return in one element called <summary>, summary information about each individual news_item which has a named author(s). The summary should include a news_item element with timestamp, title, date and news_agent subelements for each individual news_item

   iii.   For each news item that is relevant to the Gorilla Corporation, create an "item summary" element. The content of the item summary is the content of the title, date, and first paragraph of the news item, separated by periods. A news item is relevant if the name of the company is mentioned anywhere within the content of the news item.

(10 Marks)

(Total 20 marks)

Question 8

8   This question is about Information Retrieval and the role of Indexing.

Indexing of resources is key to the success of Information Retrieval systems.

Discuss the above statement and illustrate appropriately with examples throughout.

Include in your answer at least the following: A discussion why Information Retrieval is hard; Why information retrieval is considered different to data access; A description and diagram of the structure and main components of an Information Retrieval system; What are the main parameters of indexing effectiveness; A brief description of the main components of object indexing (e.g. term normalisation); A detailed description and illustration of **three** of the main techniques (e.g. stop word removal); and outline what is considered a basic indexing strategy.

(Total 20 marks)

© UNIVERSITY OF DUBLIN 2009