hrs	/ /20		Exams Office Use Only

University of the Witwatersrand, Johannesburg

Course or topic No(s)	ELEN3015
Course or topic name(s) Paper Number & title	Data and Information Management
Examination/Test* to be held during month(s) of (*delete as applicable)	June 2009
Year of Study (Art & Sciences leave blank)	Third
Degrees/Diplomas for which this course is prescribed (BSc (Eng) should indicate which branch)	B.Sc (Eng) Elec.
Faculty/ies presenting candidates	Engineering
Internal examiners and telephone number(s)	Mr. D. J. J. Versfeld x7212 Prof. S HazelHurst
External examiner(s)	Prof. A.S.J. Helberg
Special materials required (graph/music/drawing paper) maps, diagrams, tables, computer cards, etc)	None
Time allowance	Course Nos: ELEN3015 — Hours: Three
Instructions to candidates (Examiners may wish to use this space to indicate, inter alia, the contribution made by this examination or test towards the year mark, if appropriate)	Answer *ALL* questions. Type '2' Examination. Total marks: 125 - Full marks: 110 Appendix included

Internal Examiners or Heads of Department are requested to sign the declaration overleaf

1. As the Internal Examiner/Head of Department, I certify that this question paper is in final form, as approved by the External Examiner, and is ready for reproduction.

2. As the Internal Examiner/Head of Department, I certify that this question paper is in final form and is ready for reproduction.

(1. is applicable to formal examinations as approved by an external examiner, while 2. is applicable to formal tests not requiring approval by an external examiner—Delete whichever is not applicable)

Name:______________________________ Signature:______________________________

(THIS PAGE NOT FOR REPRODUCTION)

Note: Show all workings, complete with the necessary comments. Marks will be allocated for all working and logical reasoning and not just for the correct answer.

Question 1

Consider the following cryptographic system. The elements $0, 1, \ldots, 25$ of $\mathbb{Z}_{29}$ represents the letters A, B, $\ldots$, Z in the usual order, and 26, 27 and 28 represent the comma, space and full stop, respectively. A key is a pair (a, b) with $a, b \in \mathbb{Z}_{29}$ and $a \neq 0$. Encryption consists of replacing each symbol x, by $e(x)$ defined as follows:

$$e_{(a,b)}(x) = (ax + b) \mod 29. \tag{1}$$

(a) Determine the size of the keyspace of the encryption function $e_{(a,b)}(x)$.

(3 marks)

(b) Determine the function $d(x)$ which will decrypt the ciphertext generated by $e(x)$. Verify that $d(x)$ is working as expected by encrypting the character B using the key $(a, b) = (5, 10)$ and then decrypting the resulting ciphertext with $d(x)$.

(8 marks)

(c) Show that the ciphers depicted by (1) form a group (the ciphers are closed).

(4 marks)

(Total 15 marks)

Question 2

Refer to the IDEA cryptographic algorithm depicted in Fig. 2 and the key schedule of Table 5 in the Appendix.

(a) Determine the output of the top left operation if $K_1 = AB78_{16}$ and the other input is $1AB8_{16}$.

(3 marks)

(b) Determine the output of the top left operation for the last round of decryption, given that $K_1^{(1)} = AB78_{16}$ and the input from the previous round is $5C5B_{16}$.

(8 marks)

(c) Given that $K_2^{(9)} = 2C1B_{16}$, determine the value of K_2 for the first round of decryption.

(2 marks)

(Total 13 marks)

Question 3

Consider a (very weak) hash function with hash size 2^5. Show that for a particular message, on average 22 messages are needed in order to obtain a better than even chance that another message will produce the same hash as the message in question. On the other hand, for the particular hash function, show that on average 7 messages are needed in order to obtain a better than even chance that any two messages will produce the same hash.

(Total 12 marks)

Question 4

Find the keyspace of the RSA cipher given by $n = 55$, by determining every possible set of keys. Do not count cases where the public and private keys are swapped as additional keys. Also be aware not to count degenerate keys such as $e = d$ and $e = 1$.

(Hint: Determine first if the candidate key is not its own inverse.)

(Total 15 marks)

Question 5

Consider the codebook $\mathcal{C}$ depicted in Table 1.

Table 1: Codebook $\mathcal{C}$

Codewords
(0 0 0 0 0 0 0)
(1 1 0 1 0 0 0)
(0 1 1 0 1 0 0)
(1 0 1 1 1 0 0)
(1 1 1 0 0 1 0)
(0 0 1 1 0 1 0)
(1 0 0 0 1 1 0)
(0 1 0 1 1 1 0)
(1 0 1 0 0 0 1)
(0 1 1 1 0 0 1)
(1 1 0 0 1 0 1)
(0 0 0 1 1 0 1)
(0 1 0 0 0 1 1)
(1 0 0 1 0 1 1)
(0 0 1 0 1 1 1)
(1 1 1 1 1 1 1)

(a) Determine the parameters n, k and d_{min} of this code.

(3 marks)

(b) Determine the generator matrix G in systematic form and show that the message (0 1 1 1) encodes to (0 0 1 0 1 1 1).

(5 marks)

(c) Determine the syndrome s of the received vector r = (1000101) and decode the vector.

(5 marks)

(Total 13 marks)

Question 6

Consider the polynomial $g(x)$ over $GF(2^3)$

$$g(x) = \alpha^3 + \alpha^1 x + x^2 + \alpha^3 x^3 + x^4,$$

and the particular Galois field depicted in Table 2.

Table 2: The Field $GF(2^3)$

0	0	(000)
1	1	(100)
α	α	(010)
α^2	α^2	(001)
α^3	$1+\alpha$	(110)
α^4	$\alpha+\alpha^2$	(011)
α^5	$1+\alpha+\alpha^2$	(111)
α^6	$1+\alpha^2$	(101)

(a) Determine the code length (n), the dimension of the code (k), the minimum distance d_{min} and the error-correcting capability t of the Reed-Solomon code $\mathcal{C}$ generated by $g(x)$.

(4 marks)

(b) The following polynomial is received and used as input to a Reed-Solomon decoder for code $\mathcal{C}$:

$$r(x) \quad = \quad \alpha^6 + \alpha^3 x + \alpha^5 x^2 + \alpha^6 x^3 + \alpha^6 x^4 + \alpha^2 x^5 + \alpha^2 x^6.$$

Table 3 lists a number of valid code polynomials for $\mathcal{C}$.

Table 3: Code polynomials

c_1	$=$	$\alpha^4 + \alpha^5 x + \alpha^6 x^2 + x^3 + \alpha^1 x^4 + \alpha^2 x^5 + \alpha^3 x^6$
c_2	$=$	$0 + \alpha^2 x + 0x^2 + \alpha^1 x^3 + \alpha^1 x^4 + \alpha^2 x^5 + \alpha^4 x^6$
c_3	$=$	$1 + x + \alpha^2 x^2 + \alpha^5 x^3 + \alpha^1 x^4 + \alpha^2 x^5 + \alpha^1 x^6$
c_4	$=$	$\alpha^6 + \alpha^3 x + \alpha^1 x^2 + \alpha^6 x^3 + \alpha^1 x^4 + \alpha^2 x^5 + \alpha^2 x^6$
c_5	$=$	$\alpha^4 + 0x + \alpha^2 x^2 + 0x^3 + \alpha^1 x^4 + \alpha^1 x^5 + \alpha^2 x^6$

From Table 3, find the most probable code polynomial that was transmitted.

(6 marks)

(c) Determine the error-locator polynomial $\sigma(x)$ for the received polynomial $r(x)$.

(5 marks)

(Total 15 marks)

Question 7

Consider a source with the distribution depicted in Table 4.

Table 4: Frequency distribution of source

Element	Frequency
m_0	0.4
m_1	0.18
m_2	0.14
m_3	0.16
m_4	0.12

(a) Determine the entropy of the source.

(3 marks)

(b) Create a Huffman tree for the source depicted in Table 4.

(5 marks)

(c) Decode the following Huffman encoded string, given the Huffman tree depicted in Fig. 1.

1100011010011110011100000101001101101011101101100110101110000111010

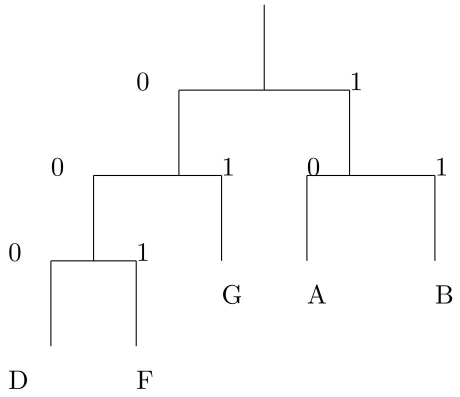

Figure 1: Huffman tree

(2 marks)

(Total 10 marks)

Question 8

The Running Club database keeps the list of all results from all races in one giant table called RESULTS. The columns are

- race: race name (varchar). Note that the race name, and date uniquely identify a race — often the same club organises races over different distance. Runners can choose which distance to run.
- distance: string since unit is kept with it.
- date: of race (note that the same race could could be run on different dates, typically the same time each year). The race name, distance and date uniquely identify a particular running of the race.
- posn: the position the runner came in the race;
- fn: the first name or initial of the runner
- surname
- gen: gender
- age: at date of race
- time taken

An extract of the table is shown — note that information has been truncated for the table to fit into a page width.

```
race             dista  date         posn    FN   SURNAME       gen  age   time
---------------  -----  -----------  ------  ---  ------------  ---  ----  ----------
Sportmans Wareh  15km   2008-10-26   73      B    Chamberlain   M    59    01:04:06
Sportmans Wareh  15km   2008-10-26   367     K    Chaba         F    43    01:31:19
Sportmans Wareh  15km   2008-10-26   608     J    Shillington   M    63    02:00:17
Maritimo Hunter  21km   2008-10-25   406     I    Burns         M    58    02:07:05
Maritimo Hunter  21km   2008-10-25   491     K    Roberts       F    56    02:14:31
Maritimo Hunter  10km   2008-10-25   83      N    Carminati     F    31    00:44:38
Woodlands        21km   2008-10-18   203     A    Campbell      M    33    01:58:34
Woodlands        21km   2008-10-18   277     I    Burns         M    58    02:04:06
Woodlands        21km   2008-10-18   315     A    Wainwright    M    58    02:07:33
Woodlands        21km   2008-10-18   383     S    Wainwright    M    24    02:14:21
```

There is another table called QUALIFIERS which lists all the races that are official qualifiers for Comrades marathon. It has 3 columns, all varchars:

- race name
- distance
- time: the qualifying time.

For a runner to qualify for the Comrades marathon they must complete one of the qualifiers in a time less than or equal to the qualifying time.

(a) Give the SQL code to find all the runners who came in the top 10 of a race in 2008.

(2 marks)

(b) Give the SQL code to find the fastest time for the marathon in the table (the marathon is 42km).

(1 marks)

(c) Give the SQL code to find the names and the times of the runners who ran the 5 fastest times for the marathon. (NB: the same runner could have run all 5 fastest times).

(2 marks)

(d) Give the SQL code to find the names and the times of the five fastest runner of the marathon. (NB: for example if Smith has run the 5 fastest times, then s/he appears only once).

(2 marks)

(e) Given the SQL code that produces a list of all the runners who have ever qualified for Comrades

(2 marks)

(f) Explain the difference between `inner` and `outer` join.

(2 marks)

(g) Discuss the normalisation of the RESULTS table. Suggest a better DB design.

(6 marks)

(Total 17 marks)

Question 9

Discuss the JPEG still image compression standard. (Hint: Sketch the block diagram and discuss why each step in the compression process is needed and how each step is performed.)

(Total 15 marks)

(Exam Total 125 marks)

(100%=110 marks)

Appendix

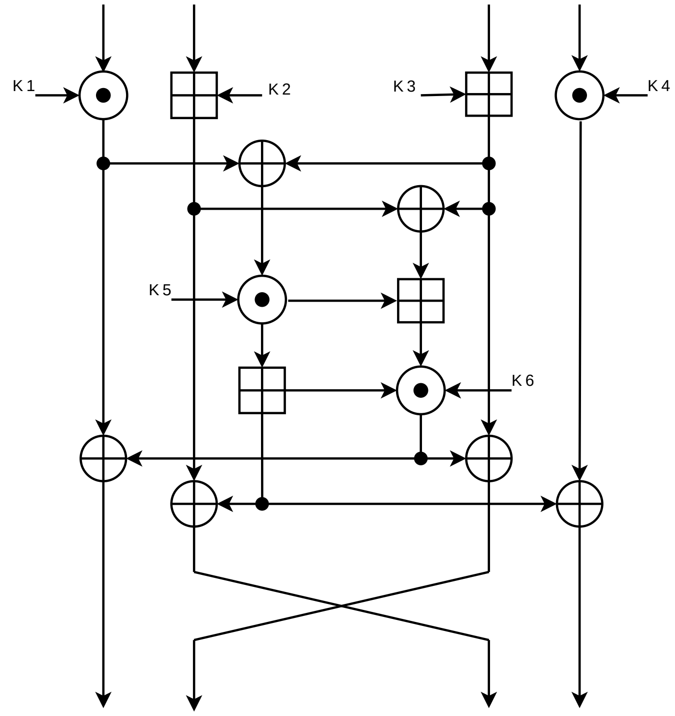

Figure 2: IDEA

Table 5: Key Schedule - Decryption

Round	Subkeys					
1st	$Z_1^{(9)-1}$	-$Z_2^{(9)}$	-$Z_3^{(9)}$	$Z_4^{(9)-1}$	$Z_5^{(8)}$	$Z_6^{(8)}$
2nd	$Z_1^{(8)-1}$	-$Z_3^{(8)}$	-$Z_2^{(8)}$	$Z_4^{(8)-1}$	$Z_5^{(7)}$	$Z_6^{(7)}$
3rd	$Z_1^{(7)-1}$	-$Z_3^{(7)}$	-$Z_2^{(7)}$	$Z_4^{(7)-1}$	$Z_5^{(6)}$	$Z_6^{(6)}$
4th	$Z_1^{(6)-1}$	-$Z_3^{(6)}$	-$Z_2^{(6)}$	$Z_4^{(6)-1}$	$Z_5^{(5)}$	$Z_6^{(5)}$
5th	$Z_1^{(5)-1}$	-$Z_3^{(5)}$	-$Z_2^{(5)}$	$Z_4^{(5)-1}$	$Z_5^{(4)}$	$Z_6^{(4)}$
6th	$Z_1^{(4)-1}$	-$Z_3^{(4)}$	-$Z_2^{(4)}$	$Z_4^{(4)-1}$	$Z_5^{(3)}$	$Z_6^{(3)}$
7th	$Z_1^{(3)-1}$	-$Z_3^{(3)}$	-$Z_2^{(3)}$	$Z_4^{(3)-1}$	$Z_5^{(2)}$	$Z_6^{(2)}$
8th	$Z_1^{(2)-1}$	-$Z_3^{(2)}$	-$Z_2^{(2)}$	$Z_4^{(2)-1}$	$Z_5^{(1)}$	$Z_6^{(1)}$
Last	$Z_1^{(1)-1}$	-$Z_2^{(1)}$	$-Z_3^{(1)}$	$Z_4^{(1)-1}$		