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 2008
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 Prof M.A. Van Wyk
External examiner(s)	Dr. W. A. Clarke
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.

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

Describe a combined strategy making use of the method of Kasiski and the Incidence of Coincidence to decipher polyalphabetic enciphered text where the number of alphabets are known to be less than 5.

(Total 5 marks)

Question 2

The following questions refer to the DES algorithm.

(a) When is a key a weak key, a semi-weak key, or a possibly weak key?

(3 marks)

(b) The following data stream is used as input to the S-Box Substitution of a specific round (In little-endian Format):

$$AE\ AA\ 77\ B8\ C8\ 50$$

Determine the output of the S-Box Substitution.

(10 marks)

(Total 13 marks)

Question 3

(a) What are the two most important properties of a secure hash function?

(2 marks)

(b) With reference to hash algorithms, explain what the birthday attack is and how the threat of this attack affects the choice of a hash algorithm's hash size. (Note: probability calculations are not required).

(6 marks)

(Total 8 marks)

Question 4

(a) Write a small pseudocode algorithm for determining the value of $a^b \mod p$. Assume that a is smaller than p. Further assume that the underlying architecture can handle numbers smaller or equal to p^2. (You can make use of the function $a \mod p$)

(3 marks)

(b) One of the steps in the Solovay-Strassen primality test is to compute the GCD of the candidate prime p and a witness a. Calculate the GCD of the following numbers and comment on the outcomes and how the outcomes effect the primality testing:

- p = 16100023 , a = 121003
- p = 700781 , a = 5003

(6 marks)

(c) Determine the Jacobi symbol J(1235/20003).

(5 marks)

(d) Encrypt the message 00054 using RSA encryption. For the particular encryption, assume $p = 109$, $q = 107$ and the decryption key $d = 8501$.
Hint 1:

$$\begin{aligned} 8501 \times e_1 &= 73 \times Q + r_1 \\ 8501 \times e_2 &= 74 \times Q + r_2 \\ 8501 \times e_3 &= 75 \times Q + r_3 \end{aligned}$$

Hint 2: $54^{99} \equiv 11042$, using the specified modular arithmetic.

(10 marks)

(Total 24 marks)

Question 5

Determine all the codewords of the (n, k) linear code C with parity-check matrix

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

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

(Total 10 marks)

Question 6

Do a systematic encoding of the following binary data using a Reed-Solomon code with n = 7:

011101110111001.

State the parameters (n,k, d_{min} and t) of the code and show all intermediate steps.

Hint: Table 1 lists some primitive polynomials.

Table 1: Primitive polynomials

$1 + x + x^3$
$1 + x + x^4$
$1 + x^2 + x^5$
$1 + x + x^6$

(22 marks)

Question 7

Consider the following string of data:

BDBGFBAGBDFGFABGGBGABFAABADBAA.

(a) Determine the entropy of the source based on the sample string.

(3 marks)

(b) Encode the text using a Huffman code.

(7 marks)

(Total 10 marks)

Question 8

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)

Question 9

(a) Suppose the student mark system data contains records like this.

ID number	Student Number	Course	Mark
861234123	06123456a	ELEN0001	80
861234123	06123456a	ELEN0003	70
861234123	06123456a	ELEN0004	60
867761123	06244191b	ELEN0001	53
867761123	06244191b	ELEN0002	66
867761123	06244191b	ELEN0004	58
867761123	06244191b	ELEN0003	91
865331726	06009345h	HART0001	72
865331726	06009345h	ENGL0001	78
865331726	06009345h	FREB0001	52

(i) State one problem with recording data like this.

(1 marks)

(ii) Normalise the data to at least second normal form.

(3 marks)

(b) A SNP is a position in a genetic sequence where some individuals in a population might have one character appearing and other individuals have another character (we always assume there are exactly two options). For example, I might have the character C appearing at a SNP and you might have a G. A SNP's location is given by the chromosome in which it appears and a location as measured from the beginning of the sequence.

The SNP table consists of a SNP identifier (text), the chromosome on which the SNP appears (a one or two letter code), its position (an integer), and a description of what the two alternatives for the SNP, which drawn from the set A, C, G, T, and `-`. There are 25 possible SNP alternatives: `A/C`, `A/G`, `A/T`, `A/-`, `C/C`, `C/G`, ...

The table below is an example. The first row says than SNP *rs15001218* appears in chromosome Z at position 22660621 and the two possible characters that can appear are G and T (these are always given in alphabetic order).

```
rs15001218,Z,22660621,G/T
rs15712282,10,22661114,C/T
rs15712280,10,22661127,A/G
rs13788418,10,22661285,C/G
```

In an experiment, we sample individuals from a population and record for each individual which SNPs were tested and what character was found. Our experiment table is shown below. The columns are the ID number of the individual sampled, the SNP tested, and which letter was found in the individual. For example, in the table A0001 was tested and two SNPs were found — in rs15001201, she had the symbol G, in rs15001202 she had the symbol C.

```
A0001,rs15001202,C
A0001,rs15001201,G
A0002,rs15712285,C
A0002,rs15001219,G
```

Give good SQL code that can be used for the following

(i) Give the SQL to create a table called *exps* which can store the necessary information for the experiments conducted (i.e. which experimental subjects had which SNP).

(2 marks)

(ii) Give the SQL that would find which SNPs were tested on experimental subject A1234.

(1 marks)

(iii) Give the SQL that would list the IDs for all the experimental candidates which were tested for an SNP on chromosome *Z*.

(2 marks)

(iv) Give the SQL that would find list the IDs for all the experimental candidates which were tested for an SNP on chromosome *Z* where one of the alternative symbols for the SNP is a *C*. You need only write what addition you would make to the previous query.

(1 marks)

(c) (i) Why are suffix arrays a useful technique for text databases?

(1 marks)

(ii) Give the suffix array for *orangesandlemonssay*.

(2 marks)

(Total 13 marks)

(Exam Total 120 marks)

(100%=110 marks)

Appendix A

Table 2: Key Permutation

57	49	41	33	25	17	9	1	58	50	42	34	26	18
10	2	59	51	43	35	27	19	11	3	60	52	44	36
63	55	47	39	31	23	15	7	62	54	46	38	30	22
14	6	61	53	45	37	29	21	13	5	28	20	12	4

Table 3: Number of Key Bits Shifted per Round

Round	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Shifts	1	1	2	2	2	2	2	2	1	2	2	2	2	2	2	1

Table 4: Compression Permutation

14	17	11	24	1	5	3	28	15	6	21	10
23	19	12	4	26	8	16	7	27	20	13	2
41	52	31	37	47	55	30	40	51	45	33	48
44	49	39	56	34	53	46	42	50	36	29	32

Table 5: S-Box 1

14	4	13	1	2	15	11	8	3	10	6	12	5	9	0	7
0	15	7	4	14	2	13	1	10	6	12	11	9	5	3	8
4	1	14	8	13	6	2	11	15	12	9	7	3	10	5	0
15	12	8	2	4	9	1	7	5	11	3	14	10	0	6	13

Table 6: S-Box 2

15	1	8	14	6	11	3	4	9	7	2	13	12	0	5	10
3	13	4	7	15	2	8	14	12	0	1	10	6	9	11	5
0	14	7	11	10	4	13	1	5	8	12	6	9	3	2	15
13	8	10	1	3	15	4	2	11	6	7	12	0	5	14	9

Table 7: S-Box 3

10	0	9	14	6	3	15	5	1	13	12	7	11	4	2	8
13	7	0	9	3	4	6	10	2	8	5	14	12	11	15	1
13	6	4	9	8	15	3	0	11	1	2	12	5	10	14	7
1	10	13	0	6	9	8	7	4	15	14	3	11	5	2	12

Table 8: S-Box 4

7	13	14	3	0	6	9	10	1	2	8	5	11	12	4	15
13	8	11	5	6	15	0	3	4	7	2	12	1	10	14	9
10	6	9	0	12	11	7	13	15	1	3	14	5	2	8	4
3	15	0	6	10	1	13	8	9	4	5	11	12	7	2	14

Table 9: S-Box 5

2	12	4	1	7	10	11	6	8	5	3	15	13	0	14	9
14	11	2	12	4	7	13	1	5	0	15	10	3	9	8	6
4	2	1	11	10	13	7	8	15	9	12	5	6	3	0	14
11	8	12	7	1	14	2	13	6	15	0	9	10	4	5	3

Table 10: S-Box 6

12	1	10	15	9	2	6	8	0	13	3	4	14	7	5	11
10	15	4	2	7	12	9	5	6	1	13	14	0	11	3	8
9	14	15	5	2	8	12	3	7	0	4	10	1	13	11	6
4	3	2	12	9	5	15	10	11	14	1	7	6	0	8	13

Table 11: S-Box 7

4	11	2	14	15	0	8	13	3	12	9	7	5	10	6	1
13	0	11	7	4	9	1	10	14	3	5	12	2	15	8	6
1	4	11	13	12	3	7	14	10	15	6	8	0	5	9	2
6	11	13	8	1	4	10	7	9	5	0	15	14	2	3	12

Table 12: S-Box 8

13	2	8	4	6	15	11	1	10	9	3	14	5	0	12	7
1	15	13	8	10	3	7	4	12	5	6	11	0	14	9	2
7	11	4	1	9	12	14	2	0	6	10	13	15	3	5	8
2	1	14	7	4	10	8	13	15	12	9	0	3	5	6	11