CHAPTER 17 – INFORMATION SCIENCE

Binary and decimal numbers – a short review:

base 10

For decimal numbers we have 10 digits available (0, 1, 2, 3, … 9)

4731 = $\frac{4}{1000}$ $\frac{7}{100}$ $\frac{3}{10}$ $\frac{1}{1}$ ← place values

$4(1000) + 7(100) + 3(10) + 1(1)$

For binary numbers we have 2 digits available, 0 and 1.

___	___	___	___	___	___	___
64	32	16	8	4	2	1
2^6	2^5	2^4	2^3	2^2	2^1	2^0

Express the following binary numbers as decimal numbers:

10101 = $1(16) + 0(8) + 1(4) + 0(2) + 1(1) = 16 + 4 + 1 = 21$

16 8 4 2 1

11100010 = $1(128) + 1(64) + 1(32) + 0(16) + 0(8) + 0(4) + 1(2) + 0(1)$

128 16 8 4 2 1 32 64

$= 128 + 64 + 32 + 2 = 226$

Express the following decimal numbers as binary numbers:

55 = $\frac{1}{32}$ $\frac{1}{16}$ $\frac{0}{8}$ $\frac{1}{4}$ $\frac{1}{2}$ $\frac{1}{1}$ $= 110111_{Two}$ ← place value

(64)

88 = $\frac{1}{64}$ $\frac{0}{32}$ $\frac{1}{16}$ $\frac{1}{8}$ $\frac{0}{4}$ $\frac{0}{2}$ $\frac{0}{1}$ $= 1011000_{Two}$ ← place value

$55 - 32 = 23$, $23 - 16 = 7$, $7 - 4 = 3$, $3 - 2 = 1$, $1 - 1 = 0$

$88 - 64 = 24$, $24 - 16 = 8$, $8 - 8 = 0$

An orbiting satellite can follow 16 different directions that are labeled 0 to 15 in binary (0000 to 1111). Each message is sent as the command along with 3 check digits. The check digits are arranged so that certain sums have even parity. These are called ***parity-check sums*** where the parity of a number refers to whether a number is even or odd. Even numbers have ***even parity*** and odd numbers have ***odd parity***.

For our satellite, the following sums must be even (0 mod 2)

$a_1 + a_2 + a_3 + c_1$, $a_1 + a_3 + a_4 + c_2$, and $a_2 + a_3 + a_4 + c_3$

What are the check digits for command 9? $\underset{8}{\underline{1}}$ $\underset{4}{\underline{0}}$ $\underset{2}{\underline{0}}$ $\underset{1}{\underline{1}}$ ← place value

$a_1 + a_2 + a_3 + c_1 = 1 + 0 + 0 + \underline{\quad}$ $c_1 = 1$ so sum is even

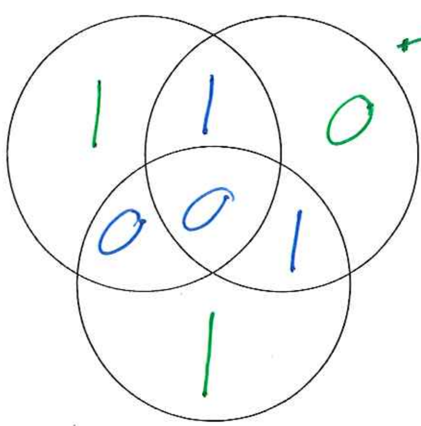
$a_1 + a_3 + a_4 + c_2 = 1 + 0 + 1 + \underline{\quad}$ $c_2 = 0$ so sum is even

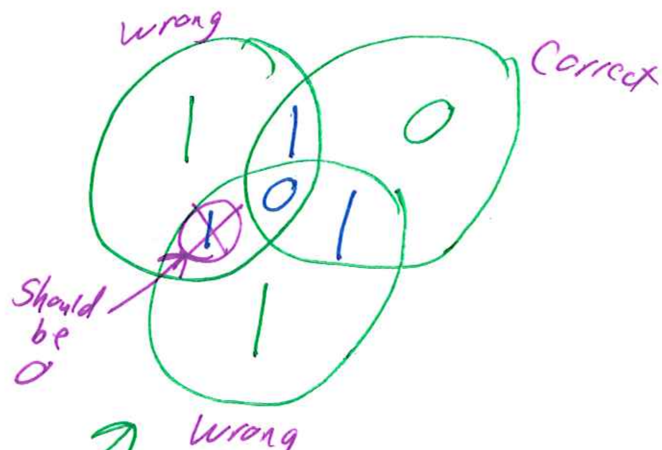
$a_2 + a_3 + a_4 + c_3 = 0 + 0 + 1 + \underline{\quad}$ $c_3 = 1$ so sum is even

Check digits are 101

full code word is 1001101

For this type of parity-check sum, we can use a Venn diagram to help find the check digits or find errors.





Fix the error in the code 1101101 if it is known only one digit has an error.

1001 101

A set of words composed of 0's and 1's that has a message and parity check sums appended to the message is called a ***binary linear code***. The resulting strings are called ***code words***.

The process of determining the message you were sent is called ***decoding***. If you are sent a message x and receive the message as y, how can it be decoded?

The ***distance between two strings*** of equal length is the number of positions in which the strings differ.

(a) 10101 and 11101

11101

distance of 1

(b) 111111 and 000000

000000

distance of 6

The ***nearest neighbor decoding method*** decodes a message as the code word that agrees with the message in the most positions provided there is only one such message.

How good a code is at detecting and correcting errors is determined by the weight of the code. The ***weight of a binary code*** is the minimum number of 1's that occur among all non-zero code words of that code.

Consider a code of weight t,

- The code can detect $t-1$ or fewer errors.
- If t is odd, the code will correct $\frac{t-1}{2}$ or fewer errors.
- If t is even, the code will correct any $\frac{t-2}{2}$ or fewer errors.

Consider the code C = {0000000, 0001111, 1111000, 1111111}

4 4 7

(a) What is the weight of the code? 4

(b) How many errors can this code detect? weight − 1
4 − 1 = $\boxed{3}$

(c) How many errors can this code correct? $\frac{4-2}{2} = \frac{2}{2} = \boxed{1}$

(d) Decode the message received as 0001101.

0001111

A ***compression algorithm*** converts data from an easy-to-use format to one that is more compact. jpg photo files use data compression as do most video and audio files.

Delta function encoding uses the differences in one value to the next to encode the data.

The data below is the closing price of the Dow Jones on Oct. 1, 2012 – Oct 5, 2012. Compress the data using delta function encoding and determine how much the data is compressed.

13610 13575 13495 13482 13515 25 characters

13610 −35 −80 −13 33 16 characters

25 − 16 = 9 characters saved

$\frac{9}{25}$ = 36% compression

Binary codes can also be compressed by assigning short codes to characters that occur frequently and longer codes to characters that occur rarely.

We have 5 symbols, π, μ, σ, δ, and α. If we give all the symbols a code of the same length, we would need 3 binary digits (000 to 101). So a string of 6 symbols would be 6 x 3 = 18 characters long. Can we devise a different binary code if we knew how often each character occurred?

Use ***Huffman coding*** is a way to assign shorter code words to those characters that occur more often.

π ✓	μ ✓	σ ✓	δ ✓	α ✓
0.16	0.19	0.23	0.17	0.25

π	δ	μ	σ	α
0.16	0.17	0.19	0.23	0.25

$\underbrace{0.16 \quad 0.17}_{0.33}$

μ	σ	α	$\pi\delta$
0.19	0.23	0.25	0.33

$\underbrace{0.19 \quad 0.23}_{0.42}$

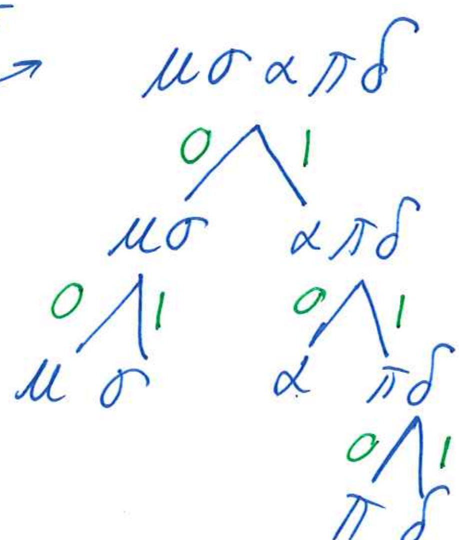
α	$\pi\delta$	$\mu\sigma$
0.25	0.33	.42

$\underbrace{0.25 \quad 0.33}_{0.58}$

$\mu\sigma$	$\alpha\pi\delta$
0.42	0.58

$\mu\sigma\alpha\pi\delta$

1



00 = μ
01 = σ
10 = α
110 = π
111 = δ

to decode

10|00|01|00|110|00

α μ σ μ π μ

The process of disguising data is called encryption. Cryptology is the study of making and breaking secret codes.

A ***Caesar cipher*** shifts the letters of the alphabet by fixed amount.

EXAMPLE

Create a Caesar cipher that shifts the alphabet by 10 letters and use it to encrypt the message THANKS.

DRKXUC

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J

	T	H	A	N	K	S
Position	19	7	0	13	10	18
Add 10	29	17	10	23	20	28
Mod 26	3	17	10	23	20	2
Translate	D	R	K	X	U	C

← Two different methods. You do not need to do both. They give the same code.

The message below was created with a Caesar cipher with a shift of 14. What is the original message?

HSZSDVCBS

TELEPHONE

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N

	H	S	Z	S	D	V	C	B	S
Position	7	18	25	18	3	21	2	1	18
Sub 14	-7	4	11	4	-11	7	-12	-13	4
Mod 26	19	4	11	4	15	7	14	13	4
Translate	T	E	L	E	P	H	O	N	E

← ~~Two~~ different methods for same thing.

A ***decimation cipher*** multiplies the position of each letter by a fixed number k (called the ***key***) and then uses modular arithmetic. To use a decimation cipher,

1. Assign the letters A – Z to the numbers 0 – 25.
2. Choose a value for the key, k, that is an odd integer from 3 to 25 but not 13 (why not?)
3. Multiply the value of each letter (i) by the key (k) and find the remainder when divided by 26.
4. To decrypt a message, the encrypted value x needs to be multiplied by the decryption letter j and then the remainder mod 26 is the original letter.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Use a decimation cipher with a key of 11 to encrypt THANKS

	T	H	A	N	K	S
Position	19	7	0	13	10	18
☆ Mul by key	209	77	0	143	110	198
Mod 26	1	25	0	13	6	16
Translate	B	Z	A	N	G	Q

$$\begin{array}{r} 8 \\ 26\overline{)209} \\ -208 \\ \hline 1 \end{array}$$

The message below was encrypted with a key of 21. The decryption key is 5. Decode the message.

	Q	R	G	S	M	O	J	T	K
Position	16	17	6	18	12	14	9	19	10
☆ Mul by decryp key	80	85	30	90	60	70	45	95	50
Mod 26	2	7	4	12	8	18	19	17	24
Translate	C	H	E	M	I	S	T	R	Y

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

A ***Vigenère cipher*** uses a ***key word*** to encode the characters.

Use a ***Vigenère cipher*** with a key word of MINT to encode the message
12 8 13 19

	N	E	W	P	R	I	N	T	E	R
Position	13	4	22	15	17	8	13	19	4	17
	M	I	N	T	M	I	N	T	M	I
Add Key word	12	8	13	19	12	8	13	19	12	8
Sum	25	12	35	34	29	16	26	38	16	25
Mod 26	25	12	9	8	3	16	0	12	16	25
Translate	Z	M	J	I	D	Q	A	M	Q	Z

A ***Vigenère cipher*** with a key word of LEX was used to encode the message below. Decode it.
11 4 23

	D	Y	M	P	V	J	L	R
Position	3	24	12	15	21	9	11	17
	L	E	X	L	E	X	L	E
Sub Key word	11	4	23	11	4	23	11	4
Difference	-8	20	-11	4	17	-14	0	13
Mod 26	18	20	15	4	17	12	0	13
Translate	S	U	P	E	R	M	A	N

To increase security, codes can be added together. Find 10110 + 00111 using binary addition. (mod 2) In ***binary addition***, if the sum is even, enter a 0.

$$\begin{array}{r} 10110 \\ +\ 00111 \\ \hline 10001 \end{array} \quad (\text{mod } 2)$$

SAMPLE EXAM QUESTIONS FROM CHAPTER 17

1. Convert the binary number 11001 to a decimal number.

16 8 4 2 1

(A) 3 (B) 25

(C) 6 (D) 31

16+8+1 = 25

2. What is the distance between received words 1100101 and 1010111?

x x x

1010111

(A) 1 (B) 2 (C) 3 (D) 4

(E) more than 4

3. Add the binary sequences 1100101 and 1110001. How many 1s digits are in the sum?

mod 2

+1110001

0010100

(A) 1 (B) 2 (C) 3 (D) 4

(E) more than 4

4. Use delta encoding to compress the data

1834 1831 1831 1825 1850.

1834 −3 0 −6 25

By how many characters is the data compressed?

(A) 9 (B) 10

(C) 11 (D) 13

orig code	20 characters
− compressed data	11 characters
Saved	9 characters

Data compressed by 9 char

5. Use the ~~nearest-neighbor~~ Venn diagram method to decode the received word 1101101.

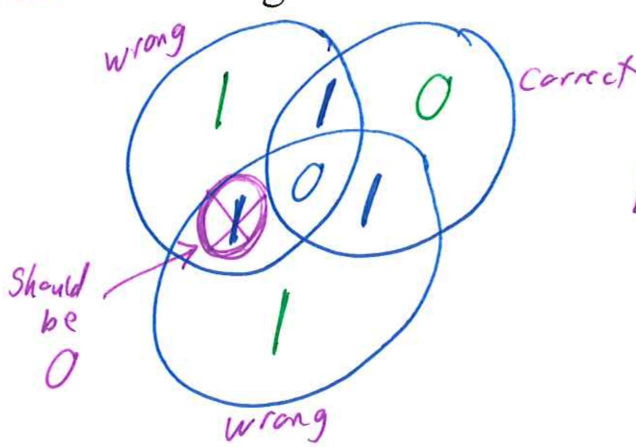
(A) 1001

(B) 0100

(C) 1101

(D) 1011

(E) None of these



1001 101

Questions 6 and 7 use the code {1100, 1010, 1001, 0110, 0101, 0011}.

2 2 2 2 2 2

Min # 1s in non-zero code words

6. What is the weight of this code?

(A) 0 (B) 1 (C) 2 (D) 3 (E) 4

Detect $2-1=1$

Correct $\frac{2-2}{2}=\frac{0}{2}=0$

7. Which one of the following is a true statement about this code?

(A) This code can detect and correct two errors

(B) This code can detect two errors and correct 1 error

(C) This code can detect and correct one error.

(D) This code can detect one error and correct 0 errors

(E) None of these

Question 8 (6 points)

Given binary codes A → 0, C → 10, I → 110, S → 1110, B → 11110.

(a) Encode the message CASSI

10 0 1110 1110 110 ← spaces are not needed so

1001110 1110110 ←

(b) Decode the message 1111001110110101110

B A S I C S

Question 9 (5 points) What is the code word for the message 110, if the code word is the message appended with three check digits found using the parity-check sums $a_1 + a_2 + a_3$, $a_1 + a_3$ and $a_2 + a_3$?

$a_1 a_2 a_3$

$\underbrace{1+1+0}_{\text{even so}}$ $c_1 = 0$

$\underbrace{1+0}_{\text{odd so}}$ $c_2 = 1$

$\underbrace{1+0}_{\text{odd so}}$ $c_3 = 1$

so code word is 110011

Question 10 (7 points)

Use a Huffman code to assign binary codes to the directions that occur with the probabilities given below. already in increasing order

Up	Down	Left	Right
0.12	0.20	0.25	0.43

.32

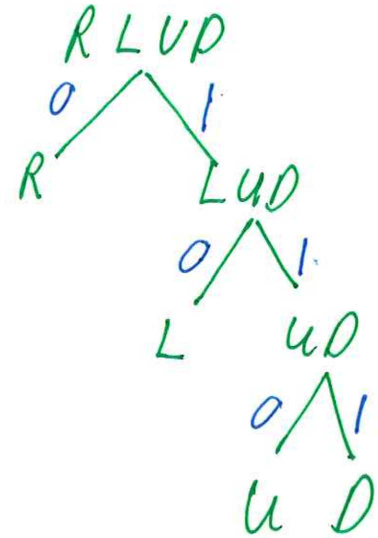
L	UD	R
.25	.32	.43

.57

R	LUD
.43	.57

RLUP

1



R = 0
L = 10
U = 110
D = 111

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Question 11 (5 points)

Use a decimation cipher with key 9 to encode the word CABLE.

	C	A	B	L	E
Position	2	0	1	11	4
Mul by Key	18	0	9	99	36
Mod 26	18	0	9	21	10
Translate	S	A	J	V	K

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

Question 12 (5 points) Use a Caesar cipher with a shift of 3 to encode the word BINARY.

	B	I	N	A	R	Y
Position ~~Translate~~	1	8	13	0	17	24
Add shift	4	11	16	3	20	27
Mod 26	4	11	16	3	20	1
Translate	E	L	Q	D	U	B

Two different methods for same thing

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Question 13 (6 points)

Use the Vigenere cipher with the key word PEN (15 4 13) to encode BASEBALL.

	B	A	S	E	B	A	L	L
Position	1	0	18	4	1	0	11	11
Add code word	P	E	N	P	E	N	P	E
	15	4	13	15	4	13	15	4
Sum	16	4	31	19	5	13	26	15
Mod 26	16	4	5	19	5	13	0	15
Translate	Q	E	F	T	F	N	A	P