

# 1.3.1

## **COMPRESSION, ENCRYPTION AND HASHING**

### **TOPIC WISE EXAM QUESTIONS**

**A-LEVEL**

**OCR**

- a) Lossy vs Lossless compression.
- b) Run length encoding and dictionary coding for lossless compression.
- c) Symmetric and asymmetric encryption.
- d) Different uses of hashing.

Candidates need to understand the need for compression, especially when transferring data via the Internet. Candidates need to understand the difference between lossy and lossless compression, and the benefits and drawbacks of each type. Candidates need to be able to recommend a type of compression for a given scenario.

Candidates need to understand how run-length encoding can reduce the size of a file for example with a text file or image. Candidates should understand how dictionary coding works by substituting entries with a unique code. Candidates should have practical experience of using these algorithms with small example files.

Candidates should understand the need for encryption. Candidates should understand how symmetric and asymmetric encryption work to encrypt and decrypt data.

Candidates should understand the need for and purpose of using hashing algorithms to store data.

Candidates should be aware of different uses for hashing, such as the storing of passwords.







2 A survey is carried out to look at the types of vehicle that travel down a stretch of motorway.

For each vehicle that passes by, a letter is entered into the system.

For a car 'C' is entered.

For a motorbike 'M' is entered.

For a lorry 'L' is entered.

For any other vehicle 'O' is entered.

It is decided to compress the data that has been generated.

(a) State what is meant by the term 'compression'.

.....  
..... [1]

It is decided that Run Length Encoding will be used.

(b) The following sequence has been compressed using Run Length Encoding.

3C3M4C

Show the result of decompressing the sequence.

.....  
..... [2]

(c) Show the result of using Run Length Encoding to compress the sequence:

CCCCOLLCCCCMOCCCC

.....  
.....  
.....  
..... [3]

The survey takers want to find out the longest continuous sequence of cars in any given chunk of data. For example, in the data

CCMCCCCLLCCC

the longest sequence would be 4.

- (d) Write the function `longest` which takes in a string of characters as an argument and returns an integer representing the longest continuous sequence of Cs. [5]

- 4 Traditionally films have been distributed on optical media such as DVDs.

Being able to stream high resolution films is only possible due to improvements in compression.

- (e) Explain why compression is important for the streaming of high resolution films.

.....  
.....  
.....  
.....  
.....  
.....  
..... [3]

- 8 A student writes a program to apply a symmetric encryption algorithm to work on messages of up to 25 ASCII characters.

The encryption algorithm works in the following way.

A message of up to 25 characters (spaces and punctuation are not included) is placed in a 5×5 array. Any leftover spaces are filled with random letters. The message I LOVE COMPUTER SCIENCE becomes:

I	L	O	V	E
C	O	M	P	U
T	E	R	S	C
I	E	N	C	E
T	O	W	R	M

The key is a sequence of ten numbers.

In this example we will use 1 2 3 4 5 1 2 3 4 5. The first 5 numbers state how many spaces the rows 0 to 4 must be rotated right.

A key with the first 5 digits 1 2 3 4 5 would result in

E	I	L	O	V
P	U	C	O	M
R	S	C	T	E
E	N	C	E	I
T	O	W	R	M

The next 5 digits state how many spaces down the columns 0 to 4 should be rotated.

Applying the last 5 digits 1 2 3 4 5 to the grid above would give

T	N	C	O	V
E	O	C	T	M
P	I	W	E	E
R	U	L	R	I
E	S	C	O	M

Part of the pseudocode for the algorithm is written below.

```
global array grid[5,5]
addMessage()
// letters and random letters have been entered
// into the 2D array, grid

for i = 0 to 4
    x = getNextDigitInKey()
    shiftRow(i,x)
next i

for i = 0 to 4
    x = getNextDigitInKey()
    shiftColumn(i,x)
next i

//Now reassemble array back into string.
```

(b) Show the result of running the algorithm on the grid and key below.

[2]

**KEY: 3 3 3 3 3 1 1 1 1 1**

T	O	P	S	E
C	R	E	T	M
E	S	S	A	G
E	Y	R	P	L
U	O	G	G	Q

Grid after only the rows are shifted:


Grid after columns have also been shifted:




3 A charitable organisation is trying to make the works of William Shakespeare available to more people.

(a) The organisation decides to make a copy of Shakespeare's entire works available as a downloadable text file from its website. It further decides to compress the file before making it available to download.

(i) State an advantage to the website's visitors of the file being compressed.

.....  
.....  
..... [1]

(ii) Explain why the company should use lossless and not lossy compression.

.....  
.....  
.....  
.....  
.....  
..... [3]

(b)\* The organisation looks at using either run length encoding or dictionary encoding to compress the file described in part (a).

Discuss the **two** compression methods and justify which you would recommend. You may refer to the extract of text below to illustrate your argument.

*What's in a name? that which we call a rose  
By any other name would smell as sweet;  
So Romeo would, were he not Romeo call'd,*

[12]

.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....  
.....



1. A software company decides to release a duplicate file finder which it has named "De-Duplicator". Duplicate files are files that are exactly the same (bit for bit identical). Space is often wasted on computers by having multiple versions of the same file. Duplicate file finders are programs that find and identify duplicate files on a hard drive so that they can be removed.

Every time the program encounters a file it takes a hash of the file and checks it against a list. If the hash exists in the list, the file is marked to be deleted. If the hash does not exist it is added to the list.

- (i) Explain **two** characteristics you would look for in a hashing algorithm for this purpose.

1 -----  
-----  
-----  
-----

2 -----  
-----  
-----  
-----

[4]

- (ii) After running the program a user finds that they still have apparent duplicates of some of their images. Explain why these apparent duplicates might still be present.

-----  
-----  
-----  
-----

[2]



**If you found this  
useful, drop a follow  
to help me out!**

**THANK YOU!**

**GCST**