Accredited

AS and A LEVEL

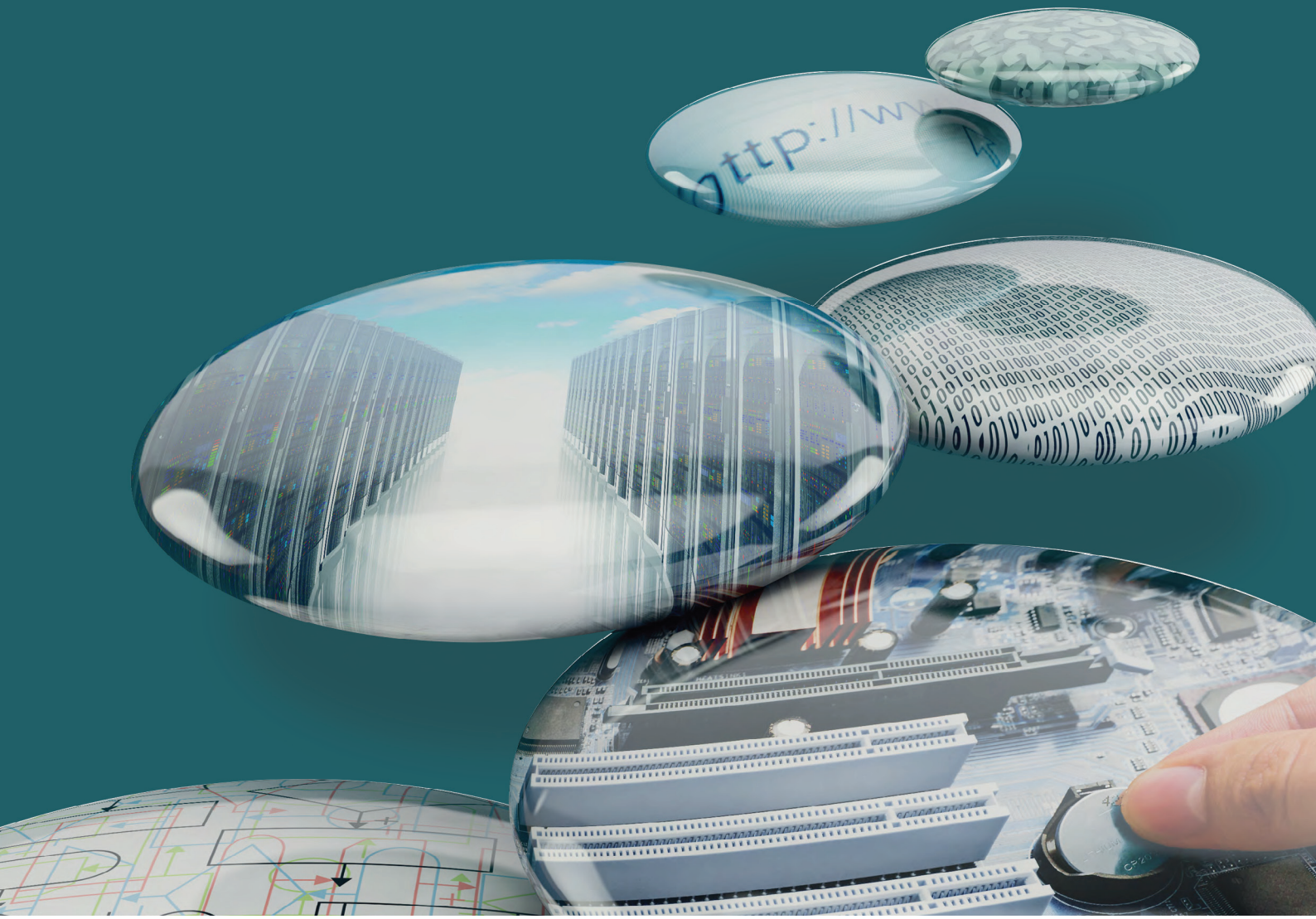
Topic Exploration Pack

H046/H446

COMPUTER SCIENCE

Theme: Compression Encryption Hashing

September 2015



CAMBRIDGE UNIVERSITY PRESS
www.cambridge.org

OCR
Oxford Cambridge and RSA

We will inform centres about any changes to the specification. We will also publish changes on our website. The latest version of our specification will always be the one on our website (www.ocr.org.uk) and this may differ from printed versions.

Oxford Cambridge and RSA Examinations is a Company Limited by Guarantee.
Registered in England. Registered company number 3484466.

Registered office: 1 Hills Road
Cambridge
CB1 2EU

OCR is an exempt charity.

Contents

This activity offers an opportunity for English skills development.

Compression, Encryption and Hashing

Compression

To speed up the transmission of data it is often compressed so that it can be transmitted faster over slower Internet connections. To do this the redundant information in the file is removed. One frequently used example is John F. Kennedy's inaugural address 'Ask not what your country can do for you – ask what you can do for your country'.

This phrase contains 17 words, 61 letters, a dash and a full stop. With spaces this sentence takes up 79 units of memory. But many of the words appear twice and by simply listing the words used and the order in which they are used we can reduce the size of the file significantly.

By listing the words used in a dictionary

1. ask
2. not
3. what
4. your
5. country
6. can
7. do
8. for
9. you

And the sentence becomes 1 2 3 4 5 6 7 8 9 – 1 3 9 6 7 8 4 5

The dictionary now takes up 41 units and the sentence 35 units. Not a lot less but none the less it has reduced the size of the file. By using pattern recognition within the structure of this sentence it is possible to create a different dictionary that will further reduce the file size. For example the phrase 'can do for' is a repeated phrase; 'you' and 'r' can be used to make 'you' and 'your'. Imagine how a much longer piece of text could be reduced using this method.

This is an example of lossless compression but with images and sound we can afford to lose some of the data and still make sense of the image or sound. Compression applied to images can reduce the resolution and therefore the number of bits per pixel, and similarly sound can be sampled at less frequent intervals and at lower resolution and still be useable, particularly speech where the human brain is capable of filling in the gaps.

Encryption:

To avoid data being accessible if intercepted it is normal to encrypt data that is being transmitted. There are examples of encryption techniques dating back centuries, for example the simple Caesar cipher that uses a simple substitution technique. Variations on this include key words to provide the offset rather than a single value. More complex codes, such as the Enigma code used during the Second World War, used varying offsets for each letter in the message, and a good introduction to codes and code breaking should include discussion of Bletchley Park's role in the Second World War. http://www.bletchleypark.org.uk/content/hist/worldwartwo/enigma.rhtm

Bletchley Park's work deciphering coded messages played a key role in the development of the computer.

The ability to decipher codes using computational techniques leads to ever more complex encryption techniques, including those based on quantum mechanics. There is an introduction to this topic on how stuff works: http://science.howstuffworks.com/science-vs-myth/everyday-myths/quantum-cryptology.htm

Hashing

Hashing is a technique to encrypt data such as passwords so that the password is stored in an encrypted form to reduce the risk of it being deciphered and used fraudulently. The password when set is subjected to a hashing algorithm and stored. When users type in their password the hashing algorithm is applied and compared to the stored value. This eliminates the problems associated with storing passwords as clear text but does mean passwords cannot be retrieved and need to be reset if forgotten.

Encryption is a two-way process using a key; if the key is known the encrypted data can be deciphered. Hashing is a one-way process that scrambles the data. One common approach is to add a random number to the submitted message and apply the hashing algorithm. If the message was encrypted using a key, common passwords would produce the same encrypted text and through comparison processes, these passwords and the key could be discovered; with the addition of a random number, common passwords will not be identifiable and the process of deciphering them becomes far more complex.

A good article about this topic is published on the GCN website: http://gcn.com/articles/2013/12/02/hashing-vs-encryption.aspx

Suggested activity

Children's storybooks often use the same words over and over again. Find a suitable example and create the dictionary and story to compare how much the data can be compressed for the storybook.

Use software such as IrfanView or Photoshop to reduce the resolution of a high-resolution image to the point where there is a noticeable lack of clarity/detail for use illustrating an article (half A4 width for example), then as an image to be viewed on a mobile phone screen. What decrease in file size can be achieved for these purposes? What would be the typical transmission times on a 3G network for each of these?

Using software such as Audacity re-sample a music track until the loss in quality is noticeable. Record or re-sample a spoken track until it is no longer possible to clearly make out what is being said. What sample rates were necessary for this and what effect did this have on the file size?

Simple substitution ciphers and substitution ciphers using keys can be experimented on using spreadsheets. Using the ASCII value for an alphabetic character and modular arithmetic a key can be used to shift the cipher text alphabet a number of places to the right or left to create a simple substitution. It is normal for the shifts to wrap around so that the later letters are mapped on to the earlier letters in the alphabet after a shift to the right for example.

Using a keyword and repeating the keyword characters to match the length of the original message, the ASCII values of the letters in the keyword can be used to create a more complex substitution cipher.

There is an excellent resource to cover these techniques available from Digital Schoolhouse, a UK-based resource for computer science teachers, http://freepdfs.net/using-spreadsheets-to-teach-data-encryption-techniques-digital/07f9f95eef6abfcede11c21445688a9b/

Use a spreadsheet to encrypt a message using initially a single offset and then a keyword.

Consider more complex approaches to this using multiple keywords and keys to create more secure encryption techniques.

The students will need to be aware of how to use VLOOKUP to lookup the cipher text for each character from a table, the use of CODE function to find the ASCII value for a character and CHAR to look up the character for an ASCII value. They will also need a working knowledge of modular arithmetic to make sure the cipher text codes wrap around when shifted.

Suggested activities

[The resources provided in the links identified above provide sufficient support for these activities. Worksheets could be created with the task headings on but not sure this is worthwhile. Eg]

Children's storybooks often use the same words over and over again. Find a suitable example and create the dictionary and story to compare how much the data can be compressed for the storybook.

Use software such as IrfanView or Photoshop to reduce the resolution of a high-resolution image to the point where there is a noticeable lack of clarity/detail for use illustrating an article (half A4 width for example), then as an image to be viewed on a mobile phone screen. What decrease in file size can be achieved for these purposes? What would be the typical transmission times on a 3G network for each of these?

Using software such as Audacity re-sample a music track until the loss in quality is noticeable. Record or re-sample a spoken track until it is no longer possible to clearly make out what is being said. What sample rates were necessary for this and what effect did this have on the file size?

Use a spreadsheet to encrypt a message using initially a single offset then a key word. You will need to be aware of how to use VLOOKUP to lookup the cipher text for each character from a table, the use of CODE function to find the ASCII value for a character and CHAR to look up the character for an ASCII value. You will also need to use modular arithmetic to ensure the cipher codes wrap around when a shift is applied.

We'd like to know your view on the resources we produce. By clicking on the 'Like' or 'Dislike' button you can help us to ensure that our resources work for you. When the email template pops up please add additional comments if you wish and then just click 'Send'. Thank you.

If you do not currently offer this OCR qualification but would like to do so, please complete the Expression of Interest Form which can be found here: www.ocr.org.uk/expression-of-interest

OCR customer contact centre
General qualifications
Telephone 01223 553998
Facsimile 01223 552627
Email general.qualifications@ocr.org.uk