Book Ciphers

- Book cipher is a cryptographic method in which the **key** is a commonly available **book** or some piece of **text**.

- The book selected is typically one that would not arouse any suspicions if found to be in the possession of one of the communicating parties.

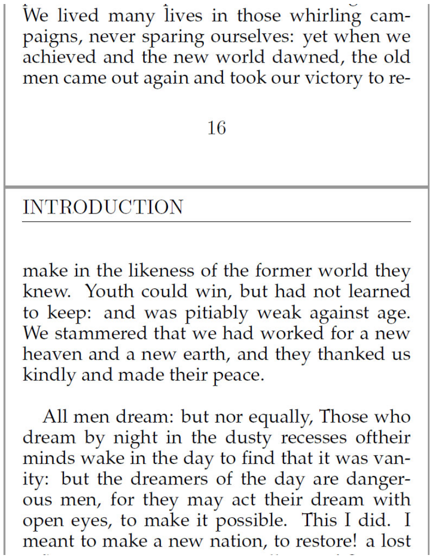
- It is typically essential however that both correspondents not only have the same book, but the **exact same edition**.

- The encryption method generates the ciphertext by simply replacing words in the plaintext of a message with the location of words from the book (key) being used.

- The **ciphertext** is then transmitted as a **string of numbers**, which are, in effect, the coordinates that point to the location of specific words in the book.

- For example, each coordinate might consist of a group of three numbers (22,12,8), which are interpreted as page 22, line 12, word 8.

- Consider the following screenshot, which shows the partial contents of two pages of a text, which are used to encrypt the string: "**victory achieved new peace possible**".



We lived many lives in those whirling campaigns, never sparing ourselves: yet when we achieved and the new world dawned, the old men came out again and took our victory to re-

16

INTRODUCTION

make in the likeness of the former world they knew. Youth could win, but had not learned to keep: and was pitiably weak against age. We stammered that we had worked for a new heaven and a new earth, and they thanked us kindly and made their peace.

All men dream: but nor equally, Those who dream by night in the dusty recesses oftheir minds wake in the day to find that it was vanity: but the dreamers of the day are dangerous men, for they may act their dream with open eyes, to make it possible. This I did. I meant to make a new nation, to restore! a lost

- The ciphertext is: "**16 19 8 16 18 1 17 6 5 17 12 6**"; decoded in 3-number groups as follows:

(16, 19, 8): page 16, line 19, word 8 => victory
(16, 18, 1): page 16, line 18, word 1 => achieved
(16, 18, 4): page 16, line 18, word 4 => new
(17, 6, 5): page 17, line 6, word 5 => peace
(17, 12, 6): page 17, line 12, word 6 => possible

- One of the problems with this technique is that the text to be used as the key may not contain the required plaintext word. For example, Shakespeare's works do not contain words such as computer, semiconductors, or missiles.

- Therefore, an alternative approach would be to provide coordinates to individual letters (syllables) rather than words.

- For example, using the above screenshot again, (**17 7 3 5**) would be decrypted as:

  Page 17, line 7, word 3, syllable 5 => 'm'

- We could also design a cipher that uses multiple syllable mapping. The following example maps two syllables at a time, (**17 7 2 2 3**):

  Page 17, line 7, word 2, syllables 2 and 3 => 'me'

- This technique however, results in generating a larger ciphertext (typically 4 to 6 digits being required to encipher each letter or syllables) and increases the time and effort required to encode and decode the messages.

- Book ciphers in general significantly increase the difficulty of frequency analysis attacks by disguising plaintext letter frequencies by using a technique called **homophony**.

- In these ciphers, plaintext letters map to more than one ciphertext symbol. Usually, the highest-frequency plaintext symbols are given more equivalents than lower frequency letters, which flattens the **frequency distribution is flattened**, thus making analysis more difficult.

- The famous **Beale ciphers** are a very good example of a **homophonic cipher**. The ciphertext, which supposedly points to a location of buried treasure somewhere in Bedford county, was coded in 1885, but to date no one has been able to decode them and locate the treasure.

- This secret (or hoax) has occupied some of the best cryptanalytic minds. The enciphered text was keyed to the Declaration of Independence.

- Here each ciphertext number was generated by taking the plaintext character and finding a word in the Declaration of Independence that started with that character and using the numerical position of that word in the Declaration of Independence as the encrypted form of that letter.

- Since many words in the Declaration of Independence start with the same letter, the encryption of that character could be any of the numbers associated with the words in the Declaration of Independence that start with that letter.

- The following links provide additional information on this ongoing challenge:

  https://en.wikipedia.org/wiki/Beale_ciphers
  unmuseum.org/beal.htm

- Likewise, when Simon Singh provided a set of 10 problems in the appendix of ***The Code Book***, stage 5, which was a Book Cipher, was the most difficult one for the Swedish team that won the £10,000 prize (https://simonsingh.net/cryptography/cipher-challenge/).

- The main strength of a book cipher is the key. The sender and receiver of encoded messages can agree to use any book or other publication available to both of them as the key to their cipher.

- The cryptanalysis process entails somehow identifying the key (book) from a massive number of possibilities available.

- In the context of espionage, a book cipher has a considerable advantage for a spy in enemy territory. A conventional codebook, if discovered by the local authorities, instantly incriminates the holder as a spy.

- It also provides the authorities with a significant counter-espionage advantage by deciphering the code and sending false messages impersonating the agent (i.e., disinformation).

- On the other hand, a book, if chosen carefully to fit with the spy's cover story, would seem entirely innocuous.

- The drawback to a book cipher is that both parties have to possess an identical copy of the key. The book must not be of the sort that would look out of place in the possession of those using it and it must be of a type likely to contain any words required.

- Even better if a copy of the book is not always in the possession of the operatives. In other words, it can be readily downloaded from a website (https://www.gutenberg.org/), used to encrypt or decrypt the messages, and then deleted.

- Thus, for example, a spy wishing to send information about troop movements and numbers of armaments would be unlikely to find a cookbook or romance novel useful keys.

- Another application for a book cipher (besides encryption large amounts of data) would be to use a book cipher to deliver a key phrase that is to be used for a modern cipher.

- One of the most important operational rules for secure communications is to frequently change passwords or pass phrases. One of the problems of frequent changes in a pass phrase is to ensure secure delivery of pass phrases to all of the communicating parties including those who are remotely located.

- Using a book cipher the pass phrase can easily be delivered remotely, and then used in a modern cipher such as AES, 3DES, etc.

Software Implementation of Book Ciphers

- Using the book as a key is relatively similar to one-time pad, insofar as the book can be considered to be a random stream of characters.

- The simplest implementation, called the “running key cipher”, uses letters of subsequent words in some text or book as a key to encode a message.

- Figure 1 is the simplest form, usually called the "**running key cipher**." In this case, text (usually from a book) provides a very long key stream, **equal to the size of the plaintext**.

- The choice of the book used is known is agreed upon in advance, while the passage used is selected randomly for each message and secretly indicated somewhere in a previous message.

- Consider the following example that uses George Orwell’s “Nineteen Eighty-Four” (Everyman’s Library Edition). The key stream starts on page 300, section VI, line 3: “**it was the lonely hour of fifteen**”.

- The plaintext will be: “**meet me at 3 capilano bridge**”.

- The plaintext and the running key stream are converted to their ASCII equivalents and then XOR’d to generate the ciphertext.

- The process is illustrated below (note that spaces will be ignored):

Plaintext:	**m e e t m e a t 3 c a p i l a n o b r i d g e**
ASCII:	6d 65 65 74 6d 65 61 74 33 63 61 70 69 6c 61 6e 6f 62 72 69 64 67 65
Key:	**i t w a s t h e l o n e l y h o u r o f f i f**
ASCII:	69 74 77 61 73 74 68 65 6c 6f 6e 65 6c 79 68 6f 75 72 6f 66 66 69 66

Ciphertext (hex):	**04 11 12 15 1e 11 09 11 5f 0c 0f 15 05 15 09 01 1a 10 1d 0f 02 0e 03**
Ciphertext (dec):	**04 11 18 21 30 17 09 17 95 12 15 21 05 21 09 01 26 16 29 15 02 14 03**

- The running key cipher is preferable to the **Vigenère cipher** (discussed later) since the same key is not used to encrypt subsequent strings of plaintext.

- However, security is still poor because the entropy per character of both the plaintext and the running key is low, and the combining operation can be inverted relatively easily.

- In addition, by guessing probable plaintexts along the ciphertext, a skilled cryptanalyst will eventually recognize the book and break the code.

- Thus, we go back to our earlier conclusion that replacing words in the plaintext using the coordinates of the words in a book is a much more secure solution.

- An even better solution is to replace individual letters rather than words as in our earlier example.

- The Book cipher is straightforward to implement and relatively secure, but using it in practical situations can be problematic. First example, we can only encrypt only letters, so encrypting white spaces or symbols such as “$”, “#”, tends to give away a lot of information to a cryptanalyst.

- One way around this is to pre-encode the plaintext using an algorithm similar to ***uuencode*** (install **sharutils** in Linux), which produces uppercase letters only.

- Another problem arises when encoding very large amounts of plain text using a Book Cipher. To encode a 5000-character plaintext, at least a 5000-word book will be required. In fact, it should be much longer since we cannot expect the frequency of all letters in the text to match the frequency of initial letters of words in the book.

- Compression would be very helpful to address the need for long keys (books), since due to the disproportion in frequencies of all letters and initial letters, the coding process may become impossible, and the process could run out of some letters even before the encryption has started.

- The number of random keys, i.e. the number of distinct usable books, is not as large as one might think. Especially in the modern era of computers, it would not be hard to write a program to try to decrypt a message repeatedly using each of a collection of digitized books as keys, and to apply a statistical test to the output to see how random it looks.

- The major weakness of a book cipher is that one must carry the book around as a key (or otherwise obtain repeated access to it, e.g. by regularly visiting a library or downloading it).

- If a cryptanalyst had captured some messages that are suspected of being encrypted with a book cipher, and had a hunch about the identity of the sender, some simple old-fashioned surveillance (and perhaps burglary) would narrow down the possible keyspace significantly.