The Book Cipher Optimised Method To Implement Encryption And Decryption

Rasika Lele, Rohit Jainani, Vihang Mikhelkar, Aniket Nade, Mrs. V. Meshram

Abstract: We plan to re innovate the age old method of Encryption namely The Book Cipher, which can be done by removing the constraints that made it obsolete in the first place. This can be done by providing a computerized touch to the Book Cipher, reducing the size of the formed Cipher Text after encoding, and creating an Encrypted Key for this Cipher. This system also provides an Infinite Key Space for the formed encryption. To introduce another level of security to this Cipher text, we plan to further Hash the encrypted key, and introduce an Auto-Destruct functionality to the Cipher Text.

Index Terms: Book Cipher, Encrypted Key, Infinite Key Space, Hashing, Auto- Destruct.

I) Introduction

"Encryption is basically an indication of users' distrust of the security of the system, the owner or operator of the system, or law enforcement authorities." Encryption transforms original information, called plaintext or clear text, into transformed information, called cipher text, code text or simply cipher, which usually has the appearance of random, unintelligible data. The transformed information, in its encrypted form, is called the cryptogram. Cryptography is a sophisticated, creative and mathematically challenging field of study. The basis of Cryptography involves receiving some data that is readily readable and to encrypt this data into a new format, that is gibberish to all but those for whom it is intended. This data can be mostly decoded only by those who contain relevant information needed for the decryption. The initially received data is referred to as the Plain Text and the data after the encryption is known as the Cipher Text. The major basis of different Cryptography Techniques is divided into two generic types: symmetric-key and public-key. Both of these types have their own advantages, and are used by the cryptographic community to exploit the strengths of each. However each of these techniques have a constraint, that being that the encrypted key can be breached easily by using an exhaustive key search. Recent research shows that the keys can be easily breached within the span of 24 hours. The means to overcome this constraint would be to increase the uncertainty factor of the generated key, such that it rivals that of the plaintext. Although traditional one time pad method provides infinite key space' to increase the uncertainty factor this method has drawback that length of key should be as long as the plaintext. This problem has also been overcome by using this novel method of cipher generation. Our team plans to introduce a modernistic approach to the Book Cipher encryption method in this paper.

This new method also overcomes the above mentioned drawback of the onetime pad method, because although the key space is infinite, the key used in this project will be small compared to the data that it protects. The organization of information within this paper is as follows: In the second section, we talk about the history of the Book Cipher, its current constraints and our plan to overcome them. The third section contains the basic work through of the entire project, and how each problem is overcome. This is followed by the Algorithms of Encryption and Decryption and the flowchart of the same. Then comes our Case Studies in the next section followed by the in detail description of Hashing and the AutoDestruct functions we shall be implementing in this project.

II) INFRASTRUCTURE

The Book Cipher encryption is done by using a particular key to encrypt and then later decrypt a particular piece of information. This particular key can be a book, or any other piece of text. The basis being that the words present in the information to be encoded is replaced by the location of the similar words present in the key. It is essential however, that the exact same key be present with both the sender and the receiver. It was used by George Scovell for the Duke of Wellington's army in some campaigns of the Peninsular War (A.D. 1807-1814). Even in Cold War, we can saw it served as a secure communication method. For instance, in 1970s, Taiwan broadcast book cipher codes via radio to instruct some of their espionages in mainland China. The strength of the Book Cipher lies in it being homophonic in nature, which allows a single word to be replaced by the locations of an infinite number of the same word used in another book, project or paper. However this comes at the expense of a large cipher text expansion. The main reasons for the decline of usage of the book cipher encryption was due to the following main reasons:

(1) The reference books had to be digitalized, uploaded and stored in a database on the computer, which seemed redundant at the time.
(2) If the hacker knew the language of the reference book, then it became easier to find the book in question.
(3) The Cipher Text becomes huge and generally has a larger size than the actual clear text.
(4) The entire text has to be converted to Upper Case before it could be encoded.

- *Rasika Lele : rasika.lele@gmail.com*
- *Rohit Jainani : rohitjainani@ymail.com*
- *Vihang Mikhelkar : vihangmirkhelkar@gmail.com*
- *Aniket Nade : aniketnade@gmail.com*
- *Prof. Mrs. V. Meshram : vidulakulkarni@viit.ac.in*

The above drawbacks and a few others are overcome by the novel modernistic approach of the book cipher method.

III) WORKING

The modernistic approach for book cipher takes into consideration all the constraints of book cipher and overcomes them. It introduces a novel idea of implementing age old encryption technique. The reference keys are easily available online nowadays, as there is an infinite supply of ebooks and pdf's available easily. These references are used as the basis for the encryption of the clear text that shall be provided as the input. The words present in the clear text are replaced by the page no, line no and word no of their respective counterparts present in the reference key. This process is further explained thoroughly in the next section. This encryption gives rise to the cipher text. The cipher text being extremely large in size is compressed using the data compression algorithm. After the encoding, the reference key is hashed using hashing (SHA) to create the hashing signature. Hashing generates a unique signature for each and every data, which is extremely difficult and time consuming to break. This helps give an extra layer of security to the cipher text. Both the cipher text and the hashing signature are then transmitted via the internet using an SMTP protocol. Also included with the cipher text is an Auto-Destruct functionality to prevent any external party that gains access to the sent files any time to decode them. This functionality compares the hashing key present with the cipher text with the hashing signature used as an attempt through the brute force attack of an exhaustive key search. If the keys do not match, then it automatically self- destructs the entire file, to prevent these tactics from working. These features not only overcome all the disadvantages previously associated with the book cipher, but also make it a reliable point to point means of information transfer. On a successful transfer, where the hashing signature matches with the receiver's, the decryption starts (further explained in the next section). After decryption is complete, the clear text can be easily readable by the receiver.

IV) ENCRYPTION AND DECRYPTION ALGORITHM

The encryption and decryption of this modernistic approach retains all the advantages of the original book cipher algorithm and removes almost all of the other constraints, to introduce a novel algorithm with a high efficiency.

1) Encryption

a) After the clear text to be encoded and the reference key (.pdf) have been entered into the system, the encryption process will START. During this process, each word present in the clear text shall individually be searched for its corresponding counterpart, and when found shall be replaced with its words page number, line number and word number. Till here falls under the general book cipher algorithm.

b) However, this still does not address any of the following constraints: what if the word doesn't exist in the reference, the addressing of spaces, the repetition of words, and the total size of the code generated being huge. These constraints are addressed as follows. If the word is not present in the reference, then that particular word is broken down into its syllables, and those corresponding syllables are searched. If this search comes up empty as well the word will be further broken down into its corresponding alphabets, and these are then mapped according to alphabet mapping.

c) Also, if multiple instances of the word exist in the clear text and reference, and only one instance of the word is chosen from the reference, then the code becomes vulnerable due to redundancy. To overcome this drawback, a randomization function is used. This function will search for multiple instances of the same word present in the reference, and will then randomly select any one of these instances to replace the word in the clear text. This method overcomes the drawback of redundancy.

d) Due to the above methods, the cipher text size becomes quite huge. This results In a slow transmission and more time for third parties to gain access to the data. To prevent this scenario, we use a lossless data compression algorithm. This algorithm is known as the LZF lossless data algorithm. This gives a 45% compression rate, which reduces the size of the encrypted text to an acceptable amount.

e) After all these processes, the cipher text is ready to be transmitted.

2) Decryption

a) After the cipher text and the hashing signature have been successfully transferred to the receiver's side, and the hashing signature matches, then decryption process begins. In this process, the steps are done in the reverse order of encryption.

b) Firstly the LZF decompression algorithm is used to return the code to its original size.

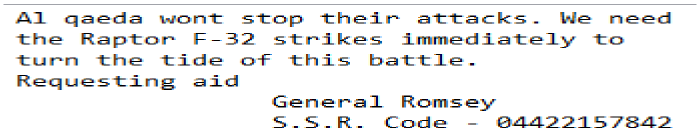
c) Then that code is decoded, by searching for the corresponding word present in the reference key and replacing the locator numbers with the appropriate word.

d) If syllable or character mapping has been done, then the corresponding words shall be joined via character foundation and then decoded to get the original word.

e) After the cipher text has been decoded back to clear text, it is then shown to the receiver. This ends the decryption algorithm of the book cipher.

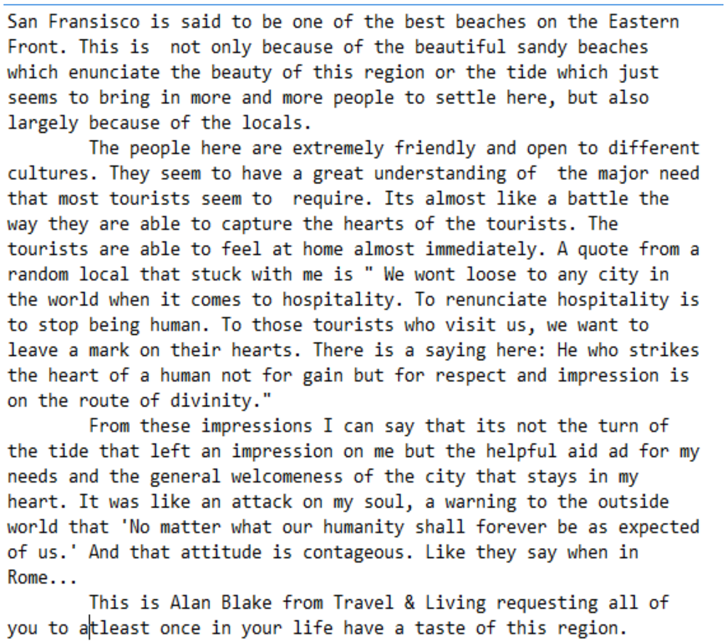
V) CASE STUDY

To clarify the project and to show the workings of this novel book cipher clearly, the following case study has been shown: The Text below is the clear text to be encoded into cipher text



```
Al qaeda wont stop their attacks. We need
the Raptor F-32 strikes immediately to
turn the tide of this battle.
Requesting aid
                   General Romsey
                   S.S.R. Code - 04422157842
```

And the following text is the reference key. For this case study, we shall be using a single page reference; however, the project can accept pdfs with pages up to 999.

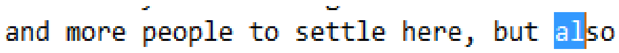


```
San Fransisco is said to be one of the best beaches on the Eastern
Front. This is  not only because of the beautiful sandy beaches
which enunciate the beauty of this region or the tide which just
seems to bring in more and more people to settle here, but also
largely because of the locals.
        The people here are extremely friendly and open to different
cultures. They seem to have a great understanding of  the major need
that most tourists seem to  require. Its almost like a battle the
way they are able to capture the hearts of the tourists. The
tourists are able to feel at home almost immediately. A quote from a
random local that stuck with me is " We wont loose to any city in
the world when it comes to hospitality. To renunciate hospitality is
to stop being human. To those tourists who visit us, we want to
leave a mark on their hearts. There is a saying here: He who strikes
the heart of a human not for gain but for respect and impression is
on the route of divinity."
        From these impressions I can say that its not the turn of
the tide that left an impression on me but the helpful aid ad for my
needs and the general welcomeness of the city that stays in my
heart. It was like an attack on my soul, a warning to the outside
world that 'No matter what our humanity shall forever be as expected
of us.' And that attitude is contageous. Like they say when in
Rome...
        This is Alan Blake from Travel & Living requesting all of
you to atleast once in your life have a taste of this region.
```

Here on encryption start

- Al doesn't exist in the reference key. Thus we start with syllable mapping

  Al →

  and more people to settle here, but also

  Al is thus found. Now it will be replaced with the following numbers

  1,4 ,8 [0-1]|[0]

  Where:
  1: Page no
  4: Line no
  8: Word no (excluding special characters)
  [0-1]: The Syllables mapping of the word.
  [0]: indicates that this is the solo syllable mapping. If there are more than 1 required to form the word, then syllables with the same number in this bracket will be combined to from the word. Also it is to be noted that the first 3 numbers are compulsory, and the other 2 are dependent on the situation.

- Qaeda here doesn't exist in the reference key. Syllable mapping on 'QA' comes up negative.

  This leaves us with character mapping.

  Thus Q is replaced by

  It is to be noticed that the characters can occur multiple times in the reference. To overcome this, a randomization function is used. This function Is shown clearly for words like the to make it easier to understand. Therefore the first reference of these letters found is applied here.

  Q→ 1,7,5 [2][1]

  Where:
  1,7,5 → Page no, line no, Word no Respectively
  [2] → Character no in word
  [1]→ Identification for word

AE Syllable mapping was not found as well. Thus character mapping of A is as follows:

A→ 1,1,2 [1][1]

Where
1,1,2 → Page no, line no, Word no Respectively
[1] → Character no in word
[1]→ Identification for word

ED syllable mapping was found. However EDA mapping was not. Thus we get:

ED→ 1,6,12 [2-3][1]

Where:
1,6,12 → Page no, line no, Word no Respectively
[2-3] → Syllables location in word
[1]→ Identification for word

Similarly A is again replaced by

A→ 1,1,2 [1][1]

Where:
1,7,5 → Page no, line no, Word no Respectively
[2] → Character no in word
[1]→ Identification for word

Here on decryption, the [1] is common and unique to these syllables and characters. The decryption algorithm recognizes this and thus is able to reform the word. Most of the words here exist in the reference. For multiple words like the, there exist multiple locations in the reference

The→ 1,1,9
→1,1,13
→1,2,7
.
.
.
.
→1,18,10

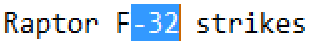
The randomization function will choose one at random, so as to prevent redundancy in the code.

The→1,5,7

This covers up the entirety of the message, except for the spaces , numbers and special characters. Those numbers and special characters if not found in reference, are then replaced by unique keys present in the application database.

Eg.

F-32



Raptor F-32 strikes

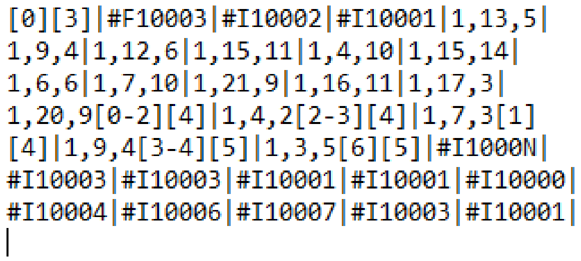
Here F is character mapped.
However '-32' is not present in the reference in any way.

They are replaced by the unique keys in the database

'-' → #F10003
'3'→ #I10002
'2'→#I10001

These unique keys are a combination of numbers and characters, to distinguish while decoding.

The Final Output of the Initial Clear Text to Cipher Text is :



```
[0][3]|#F10003|#I10002|#I10001|1,13,5|
1,9,4|1,12,6|1,15,11|1,4,10|1,15,14|
1,6,6|1,7,10|1,21,9|1,16,11|1,17,3|
1,20,9[0-2][4]|1,4,2[2-3][4]|1,7,3[1]
[4]|1,9,4[3-4][5]|1,3,5[6][5]|#I1000N|
#I10003|#I10003|#I10001|#I10001|#I10000|
#I10004|#I10006|#I10007|#I10003|#I10001|
|
```

This represents the Final Code of our project.

VI) HASHING

Hashing refers to Within this project hashing is done as stated before for the reference key, to provide a layer of security during transfer. The hashing used here will be the SHA512 method. The SHA-512 method is a novel hash functions computed with 32- and 64-bit words, respectively. It uses different shift amounts and additive constants, from other hashing methods, but their structures are otherwise virtually identical, differing only in the number of rounds.

VI) AUTODESTRUCT

Auto-destructing of the data is a method to vanish or self-destruct the message immediately after the user has read it once and it is no longer useful. We plan to use the auto-destruct sequence on the encrypted file that will be sent to the user. Along with the file the reference key hash of the encrypted text file will be sent. When the message is received by the client it will generate an automatic search to try and match the hash. If the user does not have the matching reference key file then the encrypted message will be destroyed automatically from user's system without being decrypted.

VII)CONCLUSION

This project idea was constructed by keeping in mind a secure point to point information transfer application. The major reason for choosing this encryption method is that it will provide an infinite key space and is a pure stand-alone method in its field.

VIII) REFERENCES

[1]. Changda Wang and Shiguang Ju " A Novel Method to Implement Book Cipher ",Journal Of Computer ,Vol 5,No 11,November 2010

[2]. Craig Smith "Basic Cryptanalysis Techniques " November 17th 2001

[3]. Roxna Geambasu ,Tadayoshi Kohno ,Amit A. Levy and Henry M. Levy "Vanish : Increasing Data Privacy with Self-Destructing Data ,University of Washington