

A Variation in the Working of Playfair Cipher

Pathikrit Pal*, Thejas G. S. †, Sanjeev Kaushik Ramani†, S. S. Iyengar† and N. R. Sunitha*

†Discovery Lab, School of Computing and Information Sciences, Florida International University, Miami, FL, USA

Email: [tgs001, skaus004]@fiu.edu, iyengar@cis.fiu.edu

*Computer Science and Engineering, Siddaganga Institute of Technology, Tumakuru, Karnataka, India

Email: [pathikrit.pal95, nrsunithasit]@gmail.com

Abstract—Cryptography has decidedly been in the field of research for decades with the motif of enhancing the security of information exchange. This paper exhibits a variation in the implementation of a classical cipher technique, the Playfair cipher. The motive is to make the ciphertext produced less vulnerable to attacks; we have tested the same with a common attack, the brute-force attack. The proposed model is also statistically analyzed for vulnerability against the performance of the classical encryption technique.

Keywords - Cryptography, ciphers, cryptanalysis, attacks, brute-force, hill climbing, Fibonacci sequence, golden ratio, generator, vulnerability, plaintext, ciphertext

1. Introduction

Cryptography is a fine blend of mathematics and computing and is known to have been used by people even before the advent of the computer era. Early traces of cryptography are seen in the Egyptian practice of hieroglyphics, which was then known only to the elite few.

Prior to the 20th century, cryptography was approached in the conventional manner, using pen and paper, or often with simple mechanical aids. 20th century embarked on a revelation in the field with the invention of Enigma, a complex electro-mechanical machine engineered by Arthur Scherbius [1] at the end of World War I.

The era of computers has provided unprecedented freedom for the cryptographers to come up with robust algorithms to generate ciphertexts. These algorithms would be highly prone to errors when approached by classical means of pen and paper or are far too costly to be practically implemented using electro-mechanical machines.

The parallel development of cryptanalysis, the art of breaking of the ciphers has not gone unnoticed.

Cryptography, in its early days was extensively deployed in war-zones where it was utilized in breaking the secret messages of the opponent army. Today, cryptography prevails and is majorly used in hiding personal data or classified

credentials and also in securing the social media accounts, bank details and even e-mails.

Sections II and III discuss the classical Playfair cipher technique and other related works in this domain. Sections IV and V concentrate on the proposed model and the crypt-analytic ways. We also discuss the outcome of some of the common attacks we tried on the proposed model.

2. The Playfair cipher

The Playfair cipher [2], [3] or the Wheatstone cipher was invented by Charles Wheatstone in 1854 but is well known by the name Playfair since Lord Playfair was at the forefront of the promotion of the cipher.

The cipher technique belongs to the family of symmetric cipher which dominantly uses a single key for both encryption and decryption. The key in this case is a word or a phrase.

The cipher technique also uses a matrix formed with the elements in the alphabet set of the language of the original message. Playfair cipher is the first digram substitution technique.

It is assumed in this paper that the language of encryption and decryption is English with the alphabet set as follows :

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

There have been attempts to implement Playfair Cipher in other languages as well.

2.1. The Cipher Technique

The method is explained along with an example in the description below.

The chosen key phrase or word (here, HELLO WORLD) is manipulated to have no spaces and only the distinct characters (HELOWRD) and arranged in a matrix (Fig. 2).

The matrix, as mentioned, is a 5x5 matrix with 25

```

plaintext : LET US MEET
augmented plaintext : LETUSMEXET
grouped in pairs : LE TU SM EX ET

```

Figure 1. Plaintext augmented and grouped in digrams

H	E	L	O	W
R	D	A	B	C
F	G	I	J	K
M	N	P	S	T
U	V	X	Y	Z

Figure 2. Matrix formed with key phrase 'HELLO WORLD' as per Playfair cipher

letters from the alphabet (Fig. 2). 'Q' is opted out since the frequency of usage of 'Q' is pretty less (0.095%).

The plaintext is rewritten to lose the blank spaces and 'X' is inserted in between repeated characters (Fig. 1). The augmented plaintext is then arranged into pairs and padded with 'Z' at the end when necessary, to have a pair.

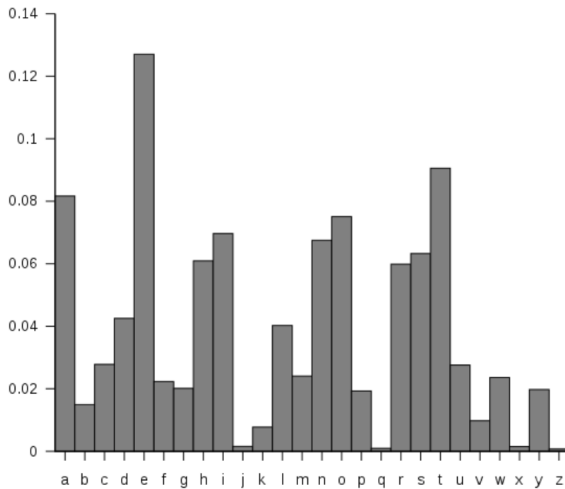


Figure 3. A histogram showing the relative frequency of occurrences of the letters in English alphabet [4], [5], [6]

In Fig. 1, padding with 'Z' is not necessary since the number of characters is even and thus satisfies the requirements of a digram.

The substitution is carried out as follows

The Playfair cipher was predominantly used during the World Wars. The frequency of occurrence of 'Q' as a monogram in other languages are much lesser than that in English language and it has been ever since the mode of opting out 'Q' in the cipher table. Another approach to make matrix in the cipher technique involves keeping 'Q' and keeping 'I' and 'J' together in the same cell. This approach is practiced less.

A : 6.34	K : 1.50	U : 3.76
B : 2.21	L : 3.72	V : 0.94
C : 2.71	M : 2.75	W : 1.40
D : 4.92	N : 9.59	X : 0.07
E : 15.99	O : 2.75	Y : 0.13
F : 1.80	P : 1.06	Z : 1.22
G : 3.02	Q : 0.04	Ä : 0.54
H : 4.11	R : 7.71	Ö : 0.24
I : 7.60	S : 6.41	Û : 0.63
J : 0.27	T : 6.43	ß : 0.15

Figure 4. Frequency (in percent) of occurrence of monograms in Deutsch [7]

If both the letters in the digram are in the same row, substitute each letter with the letter to their right, with wraparound.

LE would substitute to OL

If both the letters in the digram are in the same column, substitute each letter with the letter below them, with wraparound.

SM would substitute to TN

If the letters in the digram are in different columns and different rows, substitute each letter with a letter in their row (horizontally) such that a rectangle is formed with the 4 letters.

ET would substitute to WN

thus, following the procedure mentioned above, the ciphertext is as follows

```

plaintext : LE TU SM EX ET
cipher text : OL MZ TN LV WN

```

Figure 5. Plaintext enciphered to ciphertext as per Playfair cipher with assumed inputs

3. Related Work

Several modifications to the Playfair cipher have been proposed over the years. While few notable changes include modifications in building the matrix, others portray a change in the method of encryption of the plaintext.

Table 1 shows an overview of some of the different proposals.

4. The proposed variation

The procedure of grouping of the augmented plaintext into digrams remain the same as that in the classical Playfair cipher. The proposed model uses a 6x6 matrix instead of a 5x5 one which includes elements from the decimal digit set and the special character underscore ('_').

The decimal digit set is

0, 1, 2, 3, 4, 5, 6, 7, 8, 9

Wraparound is a situation when the letter that is to be enciphered is at the either end columns of the matrix, in this case the letter is ciphered based on the mathematical modulo operation

TABLE 1. TABLE SHOWING A FEW OF THE MODIFICATIONS AND EXTENSIONS MADE ON PLAYFAIR CIPHER

Title	Authors	Year of Publication	Contributions
Modified Block Playfair Cipher using Random Shift Key Generation [8]	Arvind Kumar, Gagan Gupta et al.	November 2012, International Journal of Computer Applications (0975 – 8887) Volume 58– No.5	1. SHA(Secure Hash Algorithm) 2. Random numbers 3. Matrix shifts with random numbers
3D - Playfair Cipher with additional Bitwise Operation [9]	Versha Verma, Dilpreet Kaur et al.	2013 International Conference on Control, Computing, Communication and Materials (ICCCCM)	1. Trigrams 2. Random numbers 3. 4x4x4 3d matrix 4. 26 letters, 10 digits, 28 special characters 5. XOR
An Extended PlayFair Cipher using Rotation and Random Swap patterns [10]	Swati Hans, Rahul Johari et al.	2014, 5th, International Conference on Computer and Communication Technology	1. Two 4x4 matrices 2. Dummy row 3. Matrix rotation 4. Frequency of letters increased on cipher 5. Digrams map to separate digrams, more secure against frequency analysis attack

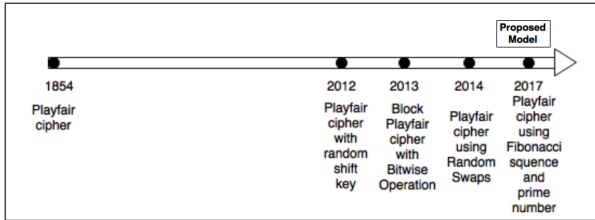


Figure 6. Timeline showing a few modifications made on Playfair cipher

The underscore (‘_’) is a fairly frequently used special character, not far behind the comma (‘,’), the most frequently used special character as shown in Fig. 7.

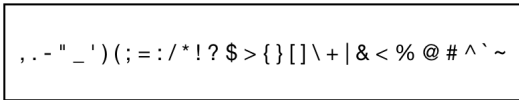


Figure 7. Special characters arranged as per their frequency of use. [11]

The rules for substitution has been changed with the inclusion of prime number, Fibonacci sequence and the golden ratio.

4.1. The Fibonacci Sequence

The Fibonacci sequence [12] is one where a term is the sum of its preceding 2 terms, the first two terms being 0 and 1.

A sequence is an ordered list of numbers and a series is the sum of the terms in the sequence.

The Fibonacci sequence is one with the fixed generators 0 and 1.

The sequence is as follows :

0 1 1 2 3 5 8 13 21 34 55 89 144 233

The mathematical formula to generate the series is

$$F(n) = F(n-1) + F(n-2) , F(0) = 0 \text{ and } F(1) = 1; n \geq 2 \quad (1)$$

4.2. The Golden Ratio

In mathematics, two numbers are said to be in the golden ratio [13] if their ratio (larger to smaller) is equal to the ratio of their sum to the larger value.

$$\frac{a+b}{a} = \frac{a}{b} , a \text{ and } b \text{ are the two numbers.} \quad (2)$$

The golden ratio, resembled by the greek symbol Phi(ϕ), is practically the most irrational number. The beauty lies in the fact that ϕ can be represented using itself as shown in equation 3. It is a recurring fraction.

$$\phi = 1 + \frac{1}{\phi} \quad (3)$$

$$\phi = \frac{1}{2} + \frac{\sqrt{5}}{2} = \frac{\sqrt{(1+5)}}{2} \approx 1.6180339887 \quad (4)$$

The relationship between the golden ratio and the Fibonacci sequence is not unknown to us, the golden ratio is the limit of the ratio of a term in Fibonacci sequence to the one preceding it as in equation 5.

$$\lim_{n \rightarrow \infty} \frac{F(n)}{F(n-1)} = \phi \quad (5)$$

A small part of this concept has been used to modify the encryption technique in the existing Playfair cipher.

5. The proposed cipher technique

5.1. The model

5.1.1. Encryption . The rules are as follows :

1. Select a prime number (represented by the greek symbol Nu, ν) under 36 (since 36 is the maximum number of cells in the 6x6 cipher table).

2. Select a keyword or a keyphrase. The symmetric key (represented by K) for the cipher technique is the prime number appended to the keyword or the keyphrase and then reduced to its distinct characters.

3. Construction of the matrix :

a. The cells of the matrix are first filled with the characters of the key (K) in row major order starting with the first row.

b. The remaining empty cells of the matrix are first filled alternately with the remaining elements from the decimal digit set and the special character underscore ('_') in its (matrix) alternate positions.

c. The remaining cells of the matrix are then filled with the remaining elements from the alphabet set. In this case also, 'Q' is opted out due to the same reason as in Playfair cipher.

We could, however, include 'Q' and consider 'I' and 'J' to exist together in one cell; performance will not vary noticeably.

4. Augment, pad and rewrite the plaintext as necessary as per Playfair cipher.

5. Generate the first ν terms in the Fibonacci sequence and keep only the distinct prime numbers.

6. Calculate the ratio of one term to its preceding term among the remaining terms of the sequence. It is represented by the Greek symbol Rho, ρ .

$$\rho = \frac{F(n)}{F(n-1)} \quad (6)$$

7. Calculate the offset, the difference of ρ from ϕ . It is represented by the greek symbol Theta, θ .

$$\theta = \phi - \rho \quad (7)$$

For steps 6 and 7, follow Table 2.

8. Keep only the sign (+ or -) from the resulting series. The first two terms, 0 and 1 have no sign since they are the generators. here,

~ ~ - + - - - , ~ is used to represent no sign

9. Arranging the series of signs :

1. Arrange the sign series, as obtained, along with the augmented and padded plaintext repetitively until it matches the length of the plaintext and group into pairs as necessary.

2. Scan throughout the sign series thus formed and consider where there is no sign, as represented by '~'.

TABLE 2. TABLE SHOWING THE CALCULATIONS OF θ AND ρ

Serial Number	Resulting Fibonacci number	Ratio of each number to the one before it (this estimates phi) ρ	Difference from Phi; Offset $\theta = \phi - \rho$
0	0	-	-
1	1	-	-
2	2	2.0000000000	-0.3819660113
3	3	1.5000000000	+0.1180339887
4	5	1.6666666667	-0.04863267797
5	13	2.6000000000	-0.9819660113
6	89	6.84615384615	-5.22811985745

Alternatively, change pairs of '~' to '+' and '-' starting with the first pair being changed to '+'.

3. The plaintext now has a sign assigned to every character and has been grouped into pairs. If in the digram, the signs are opposite, i.e one is '+' and the other is '-', only the position of the signs are interchanged.

10. The digraph substitution is performed as follows:

a. if the associated sign is '+', trace forward with wraparound.

b. if the associated sign is '-', trace backward with wraparound.

5.1.2. Decryption . The decryption procedure requires the listener (one to whom the message is sent) to have prior knowledge about the chosen key phrase and the chosen prime number.

With the help of the key combination and the skill set to construct the required matrix and generate the list of symbols with the golden ratio and Fibonacci sequence, the following rules are used to decrypt the message.

1. Rewrite the ciphertext devoid of spaces and arrange as digrams.

2. Arranging the series of signs :

a. Arrange the sign series, as obtained after constructing the table as in Table 2, along with the ciphertext repetitively until it matches the length of the ciphertext and group into pairs as necessary.

b. Scan throughout the sign series thus formed and consider where there is no sign, as represented by '~'. Alternatively, change pairs of '~' to '+' and '-' starting with the first pair being changed to '-'.

c. If in the digrams the signs are the opposite to one another, i.e. one is '+' and the other is a '-', do not change the signs; in all other cases, change '-' to '+' and vice versa.

3. The digraph substitution is performed as follows:

a. if the associated sign is '+', trace forward with wraparound.

b. if the associated sign is '-', trace backward with wraparound.

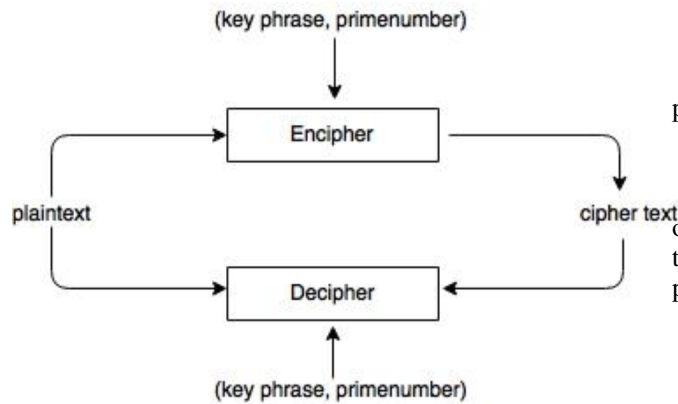


Figure 8. Flowchart showing the overview of the proposed model

5.2. An example showing the working of the proposed model.

Assumed inputs :

plaintext : HIDE THE DEAD BODY
 key word : BRITISH
 prime number : 11

Figure 9. Assumed inputs for the proposed model

key : BRITSH1

Figure 10. Final key(K) with the assumed inputs

The cipher table created with the distinct letters of the key is shown in Figure 11.

Generate first 11 Fibonacci terms.

0 1 1 2 3 5 8 13 21 34 55

Keep only the prime numbers from the terms.

0 1 2 3 5 13

The signs are assigned to the terms as per Table 2.

B	R	I	T	S	H
1	A	0	C	2	D
3	E	4	F	5	G
6	J	7	K	8	L
9	M	-	N	O	P
U	V	W	X	Y	Z

Figure 11. Matrix formed with key word 'BRITISH' and prime number 11 as per proposed cipher model

0 1 2 3 5 13
 ~ ~ - + - -

Formation of the digrams in the plaintext along with padding.

H I D E T H E D E A D B O D Y Z
 ~ ~ - + - - ~ ~ - + - - ~ ~ - +

Since in the digrams DE, EA and YZ the signs are opposite, we interchange the position of the signs and obtain the following. Also, the pairs of '~' have been changed to pairs of '+' and '-' alternatively, starting with a pair of '+'.

H I D E T H E D E A D B O D Y Z
 + + + - - - - + - - - + + + -

As per the substitution rule,

H I becomes B T

D E becomes 1 3

T H becomes I S

E D becomes 3 2

E A becomes 4 1

D B becomes 2 H

O D becomes P 1

Y Z becomes Z Y

Hence,

the ciphertext is **BT13IS32412HP1ZY**.

6. Cryptanalysis of the proposed model and comparison with Playfair cipher

6.1. Analysis of Playfair cipher

Playfair cipher can be easily broken if sufficient text is available. Obtaining the key is comparatively easy if both the plaintext and the ciphertext are known.

When only the cipher text is known, the primary attempt to crack the code is a brute-force [14] approach. In this case, the frequency of occurrence of digrams in the cipher text is tallied against the frequency of occurrence of digrams in the assumed language. One thing to note is that a digram and its reverse (e.g. XY and YX) will always decrypt to the same letter pattern (e.g. RP and PR). Identifying repetition in digrams and their near about reversed patterns and matching them against a list of known plaintext words is a probable beginning for the construction of the key.

Another approach to crack a Playfair cipher is the hill climbing [15] method where a random combination of the letters of the alphabet set is assumed to be the cipher table. The substitutions in the cipher are then performed to come up with a possible plaintext. A few minor changes are then made to the square to form a child combination. The substitutions are again performed on the ciphertext, following the same rules, to come up with a better candidate for the plaintext. The two candidates are then duelled for a better match for the plaintext and the corresponding square combination is then updated as the parent square.

In the Playfair cipher, a plaintext digram always enciphers to the same ciphertext digram and this is a major drawback of the technique.

6.2. Analysis of the proposed model

At the very first sight of the proposed changes, the newer model has a combination of a word or a phrase along with a number as compared to the single element in the key for the classical Playfair cipher.

This combination has a prime number and this itself increases the strength of the cipher technique.

Cipher techniques are nowadays implemented in the social media and the keyword need not be a combination of letters only. The keyword which is a combination of only letters of the alphabet set in the classical approach is now upgraded to bear the same properties as that of an identifier in a few programming language, i.e. it can have decimal digits and an underscore (‘_’) as well.

This increases the maximum number of elements of the matrix to 36 which is assumed to be a 6x6 matrix for simplicity of computation.

The classical approach has 625 possible diagrams on each substitution out of which one is the true one with a probability of 0.0016. The proposed approach has 1296 choices for each diagram and there are 12 prime numbers under 36 (considering 1 as prime) as follows,

1, 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31

Hence, in the proposed model, the probability of guessing a correct combination during an attack is $\frac{1}{12}$ times lessers than getting a correct combination in the classical approach. A combination in this discussion is a choice of key word or a key phrase. In both the approaches, even if the choice for the key word or the key phrase is same, the proposed model has a prime number attached to it during the formation of the matrix and thus the chances decrease by 12 for brute-force attack where the approach is guessing a key word.

In the proposed model the same digram in the plain-text does not always encipher to the same digram in the ciphertext. This is due to the pseudo-random sign series as generated in Table 2. Hence, chosen-ciphertext attack approach would not generate the correct plaintext and cannot be easily cracked.

The statistics shown in Figures 14 and 15 are as tested on passcode.org.

7. Conclusion

The Playfair Cipher is easily breakable when sufficient ciphertext is available since a digram always enciphers to the same pair of characters. This problem has been solved where the user chooses a prime number, almost randomly. The proposed model is less vulnerable to attacks than the classical Playfair Cipher. Brute force attacks on the proposed model will take more time than that on the Playfair Cipher. Further research and inclusion of modern cyptographic techniques on this proposed model with strengthen the algorithm and portray for practical usage.

. In both the approaches, a key word selected in either, having the same length, has equal probability of being chosen, essentially because they are selected from the same alphabet set.

. This is not completely random since there is a limitation to the choice of the prime number and the sign series does repeat after certain length.

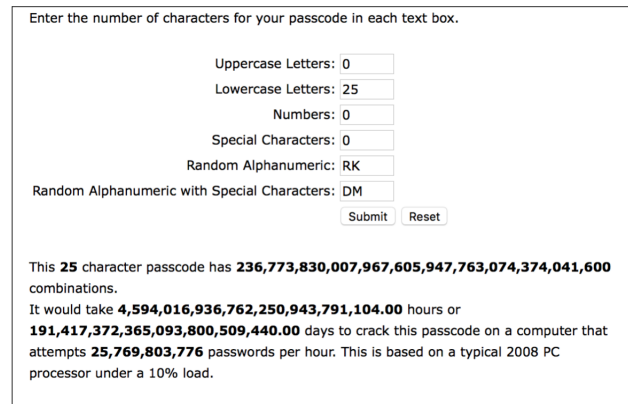


Figure 12. Statistics showing the vulnerability of classical Playfair cipher [16]

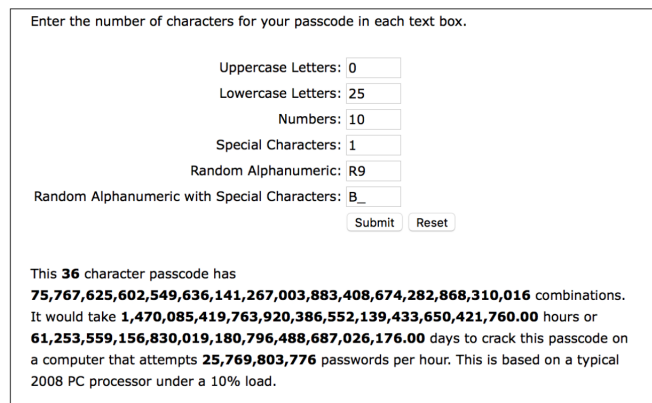


Figure 13. Statistics showing the vulnerability of the proposed model [16]

Acknowledgment. This work was supported by Florida International University Graduate School Dissertation Year Fellowship award received by the author Thejas G.S.

References

- [1] S. Arthur, “Electric ciphering apparatus,” Oct. 13 1925, uS Patent 1,556,964.
- [2] H. F. Gaines, *Cryptanalysis: A study of ciphers and their solution*. Courier Corporation, 2014.
- [3] M. Smith, *Station X: the codebreakers of Bletchley Park*. Pan Macmillan, 2004.
- [4] O. Dictionaries, “What is the frequency of the letters of the alphabet in english,” 2010.
- [5] N. Phando, “Statistical distributions of english text,” 2007.
- [6] E. S. Lee, “Essays about computer security,” *University of Cambridge Computer Laboratory*, p. 181, 1999.
- [7] P. Cryptography. (2017) German letter frequencies. [Online]. Available: <http://practicalcryptography.com/cryptanalysis/letter-frequencies-various-languages/german-letter-frequencies/>
- [8] A. Kumar, P. S. Mehra, G. Gupta, and A. Jamshed, “Modified block playfair cipher using random shift key generation,” *International Journal of Computer Applications*, vol. 975, p. 8887, 2012.

- [9] V. Verma, D. Kaur, R. K. Singh, and A. Kaur, "3d-playfair cipher with additional bitwise operation," in *2013 International Conference on Control, Computing, Communication and Materials (ICCCCM)*. IEEE, 2013, pp. 1–6.
- [10] S. Hans, R. Johari, and V. Gautam, "An extended playfair cipher using rotation and random swap patterns," in *2014 International Conference on Computer and Communication Technology (ICCT)*. IEEE, 2014, pp. 157–160.
- [11] S. Arbesman. (2017) The rarity of the ampersand: Frequencies of special characters. [Online]. Available: <https://www.wired.com/2013/08/the-rarity-of-the-ampersand/>
- [12] M. Beck and R. Geoghegan, *The Art of Proof: basic training for deeper mathematics*, 2010.
- [13] E. W. Weisstein, "Golden ratio," 2002.
- [14] J. Owens and J. Matthews, "A study of passwords and methods used in brute-force ssh attacks," in *USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*, 2008.
- [15] Y. Izbenko, V. Kovtun, and A. Kuznetsov, "The design of boolean functions by modified hill climbing method," in *2009 Sixth International Conference on Information Technology: New Generations*. IEEE, 2009, pp. 356–361.
- [16] Passcodes.org. (2017) Brute force attack calculator | secure passcode generators. [Online]. Available: <http://passcodes.org/>

Appendix

C

Computer era : Computer era in cryptography dates back to the time of WWII when Colossus was engineered by Tommy Flowers, an electronics engineer at the Post Office Research Station(a part of GPO) at Dollis Hill, UK. It is the world's first fully digital programmable computer that was used to decrypt ciphers generate by German Army's Lorenz SZ40/42 cipher machine.

Ciphertext : A cipher text is the encrypted plaintext received by the receiver.

D

Digram : A digram or bigram is a sequence of two letters, syllables or words.

G

Generator : A generator of a sequence is the term(s) that can used to generate each term of the sequence following certain rules.

P

Plaintext : A plaintext is the original message sent by the sender.

S

Special character : Any printable ASCII characters with ASCII codes between 33 and 126, both inclusive except the white space, the alphabet set, both uppercase and lower case, and the decimal digit set.

The characters in the set are

, . - " _ ') (; = / * ! ? \$ > { } [] \ + | & < % @ # ^ ` ~