An Efficient Modification to Playfair Cipher

[1]Md. Ahnaf Tahmid Shakil and [2]Md. Rabiul Islam
[1] Department of Computer Science & Engineering, University of Information Technology and Sciences, Bangladesh, at.shakil@yahoo.com
[2] Department of Computer Science & Engineering, Rajshahi University of Engineering & Technology, Rajshahi-6204, Bangladesh, rabiul_cse@yahoo.com

Abstract— Playfair is one of the best-known traditional ciphers but it is limited from different aspects. This paper deals with some of its limitations and extensibilities. Proposed modification uses a 7 × 7 matrix with a matrix randomization algorithm to extend the data holding capability and security at the same time. Some limitations like I/J inconsistency and padding character ambiguity is eliminated. According to the performed cryptanalysis, this modification is stronger than playfair.

Keywords— Algorithm Enhancement and Optimization, Classical Cryptography, Computational Algorithm, Cryptography, Polyalphabetic Cipher, Private-key Encryption

Manuscript received on 19 July 2014 and accepted for publication on 24 August 2014.

1 Introduction

CRYPTOGRAPHY is the study of mathematical techniques related to aspects of information security such as confidentiality, data integrity, entity authentication, and data origin authentication [1]. It discusses about a set of techniques, Encryption is one of them. One of the primitive purposes of data and information is to interact with it via various communication channels. These channels are not always authentic. Information or data must be masked before the communication is initiated to assure confidentiality. The process of masking data before transmission through communication channel is encryption, though purposes of encryption may differ. Most of the cryptosystem follows a generic structure to encipher and decipher data. It involves plaintext, ciphertext, encryption algorithm, decryption algorithm and key [2].

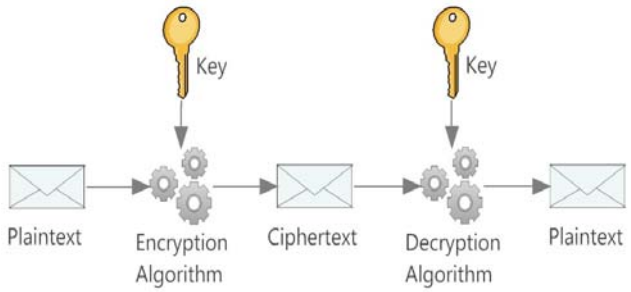


Figure 1: General structure of cryptography.

According to the structure, encryption and decryption uses two different algorithms which may use either identical or different keys. Based on the usage of key, encryption may be categorized into two distinct sections – symmetric or private key encryption and asymmetric or public key encryption [3]. Symmetric encryptions can be block ciphers or stream ciphers.

2 The Playfair Cipher

Playfair is a symmetric polyalphabetic encryption system that uses block substitution. It was invented by Charles Wheatstone in 1954 but implementation was popularized by Lord Playfair [4], [5]. This cipher was also used as a British field cipher [6]. Playfair cipher uses a 5 × 5 matrix which is shown in table 1.

TABLE 1
A PLAYFAIR MATRIX

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

The matrix is constructed by choosing a keyword from which duplicate characters are removed and placed in the matrix. Then the rest of the empty spaces are filled with remaining characters by following an alphabetic order. Consistency with English alphabet is kept by putting any two characters in a single entry (Traditionally, these characters are I and J). Then plaintext is considered as a construction of two character blocks. A plaintext with odd length is normalized by appending a padding character at the end. Each block is substituted by following the rules below:

- If both characters are same, a filler character e.g., x is added after the first character.

- If both characters are on the same row of the matrix, they are replaced by their immediate next with the first element of the row circularly following the last.

- Two characters that are on the same column are replaced by the character beneath them with the top element of the row circularly following the bottom.

- Two characters when neither on the same column or on the same row, replaced by the character on its row that intersects another character by column.

For every possible key there is different number of matrix arrangements. So, for 25 letters, a permutation of 25 (which is approximately 10^{25}) number of possible matrix can be generated [7]. Also, with 26 letters there is a possibility of 676 digrams, which was considerably secure for the time when playfair invented. But, with the change of time, different cracking method arisen, some of which doesn't even require technical device and can be solved by pencil and paper [8].

3 PROPOSED MODEL

In the proposed model, a 7 × 7 matrix is considered for extended character support and additional features. Primarily, the matrix supports 49 characters. But, the model uses 47 of them for general purpose and 2 for special purpose. The character set includes 26 lower-case letters, 10 numerals, 10 most frequently used punctuation marks and a whitespace character. The two remaining characters serve exclusively as a filler character and a padding character. This two particular character are not eligible to participate in plaintext or keyword. During decryption they are omitted. They eliminate the existing ambiguity in playfair that couldn't resolve the following scenarios:

- *Scenario 1:* A substitution pair includes identical characters and each character in the pair is filler character. For example, if 'X' is a filler character, then according to conventional playfair algorithm, pair 'XX' will be replaced by 'XXX', which, in turn, will create ambiguity. And according to the cryptanalysis by Michael J. Cowan, this is a potential source of exposure of plaintext structure [8].

- *Scenario 2:* Plaintext has odd number of characters. A plaintext with odd length is processed by appending a padding character. But decryption algorithm has no clue, whether that particular last pair uses a padding character or not.

Similar algorithm exists that uses dedicated characters to reduce these ambiguity [9].

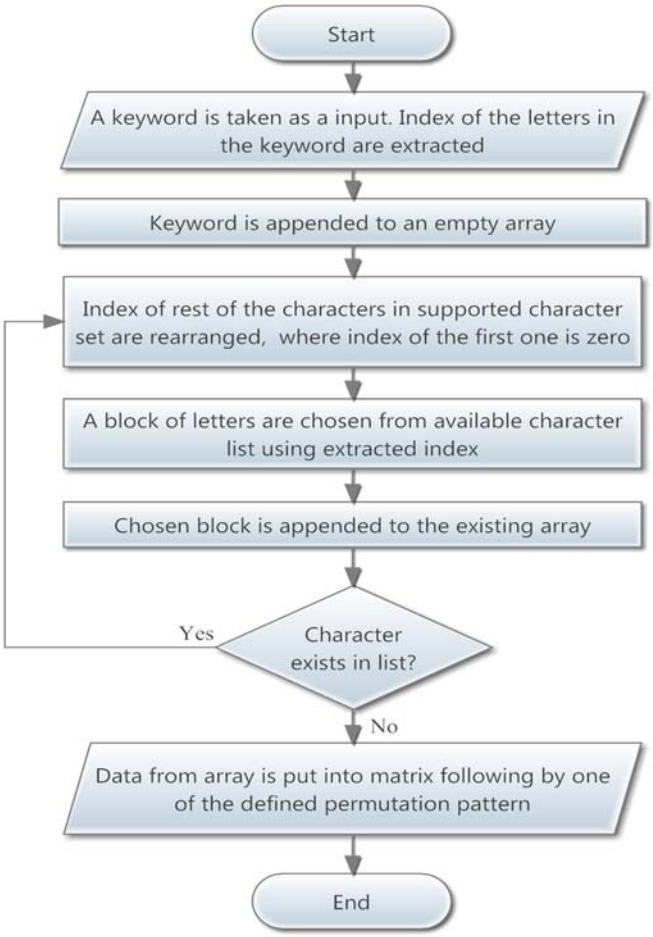


Figure 2: Data flow diagram of matrix construction.

In the matrix construction phase, following steps are applied:

- A character set (C.S.) is considered which is composed of all 49 supported characters.

- Every character in C.S. possesses a temporary index number. And, the character groups follow an indexing hierarchy. In the primary state, it appears like as table 2.

TABLE 2
PRIMARY INDEXING OF CHARACTER GROUPS

Groups	Characters	Index Range
Alphabets	26	0 – 25
Numerals	10	26 – 35
Punctuations	10	36 – 45
Whitespaces	1	46 – 46
Filler and Padding	2	47 – 48

- An array is considered where characters are temporarily stored before putting in the matrix. It is primarily empty.
- First, a keyword is chosen, which is a composition of valid letters in C.S. (excluding padding and filler character).
- Index list of keyword characters (I.K.) is calculat-

ed. Then, Keyword is placed in the empty array.
- Then, C.S. is rearranged by removing characters that are already in the array. C.S. is also re-indexed in a way that, index of the first character is 0; later one is 1 and so on.
- Now, a block of characters is extracted from C.S. by using I.K. If any character of referred index is not available, it is simply ignored.
- The extracted characters are appended to the array.
- This extraction and appending process iterates until there is no character left in C.S. (Fig. 2 provides an explicit view on the process).
- Finally, data from array is placed in the matrix by following a matrix permutation pattern (3.2).

Once the matrix construction is complete, plaintext data blocks are substituted using the same principle as 5 × 5 playfair algorithm.

3.1 Example

Consider a keyword K = "ace". K contains 3 characters. Also, consider C.S. which consists of punctuations in the list ['(', ')', '$', '&', '+', ',', '/', ':', ';', '='], C.S. is in primary state and using '!' as filler character and '~' as padding character. Table 3 shows the indexing of C.S. for primary state.

TABLE 3
INDEXING OF C.S. IN PRIMARY STATE

Char.	a	b	c	d	e	...	!	~
Index	0	1	2	3	4	…	47	48

First, index of K is calculated. So, index(K) = [0, 2, 4]. Let, A is an empty array. After appending characters from K, A = ['a', 'c', 'e'].

Now, C.S. is rearranged by removing characters in A and re-indexed, which is shown in table 4.

TABLE 4
INDEXING OF C.S. AFTER A REARRANGE AND RE-INDEXING OPERATION

Char.	b	d	f	g	h	...	!	~
Index	0	1	2	3	4	…	44	45

Then, a block of characters ['b', 'f', 'h'] is extracted from C.S. by using index(K). Characters are appended to array. Now, the array, A = [a, c, e, b, f, h]. This way, all the characters are extracted. Finally, A = ['a', 'c', 'e', 'b', 'f', 'h', 'd', 'i', 'k', 'g', 'l', 'n', 'j', 'o', 'q', 'm', 'r', 't', 'p', 'u', 'w', 's', 'x', 'z', 'v', '0', '2', 'y', '3', '5', '1', '6', '8', '4', '9', ')', '7', '$', '+', '(', ',', ':', '&', ';', ' ', '/', '!', '=', '~']

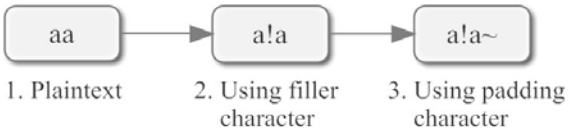
Now, applying a spiral pattern (direction: clockwise, starting point: upper-left edge) on data in A, we get the matrix which is shown in table 5.

TABLE 5
GENERATED MATRIX FOR KEYWORD "ace" IN A SPIRAL PATTERN CONFIGURATION

a	c	e	b	f	h	d
z	v	0	2	y	3	i
x	(	,	:	&	5	k
s	+	=	~	;	1	g
w	$	!	/		6	1
u	7	)	9	4	8	n
p	t	r	m	q	o	j

Encryption: Plaintext "aa"

Step1: Plaintext processing –



Step 2: Block substitution -

a! ⟶ *ew*

a~ ⟶ *bs*

Decryption: Ciphertext "ewbs"

Step 1: Block substitution -

ew ⟶ *a!*

bs ⟶ *a~*

Step 2: Omitting padding and filler character –

Ignoring filler and padding characters, retrieved plaintext is "aa"

The screenshot in figure 3 gives an example of another plaintext encryption, which is a software implementation of the explained algorithm of modified playfair cipher.

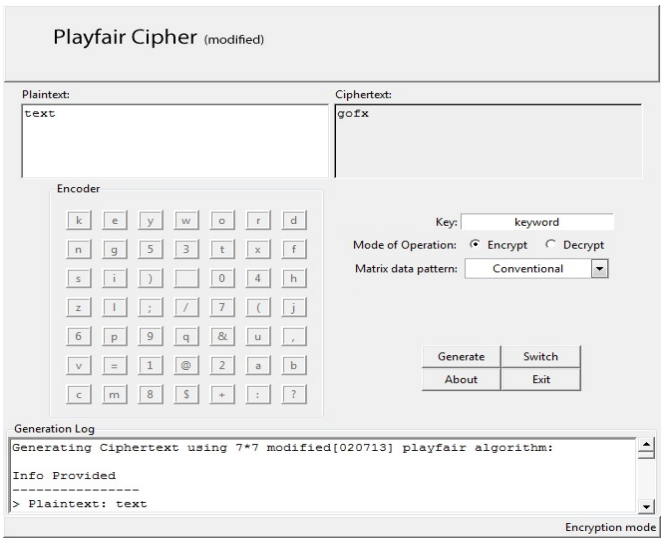


Figure 3: A software implementation of modified playfair cipher algorithm [10].

3.2 Matrix Permutation Patterns

Matrix permutation patterns define how data is to be arranged in matrix. For example, the traditional playfair used a left to right and top to bottom order which is re

ferred in this paper as conventional pattern. Unlike a single pattern, this model uses multiple permutation patterns to choose from. Some of the permutation patterns:

Spiral Pattern

This pattern takes any of the four edges as starting point and consumes the matrix at a spiral concentric fashion. Or, starts from the center and expands through the matrix at a spiral expanding fashion. A total possible variation is 16. A clockwise spiral pattern using upper-left edge as starting point is given in table 6.

TABLE 6
A SPIRAL PATTERN CONFIGURATION USING 7 × 7 MATRIX

1	2	3	4	5	6	7
24	25	26	27	28	29	8
23	40	41	42	43	30	9
22	39	48	49	44	31	10
21	38	47	46	45	32	11
20	37	36	35	34	33	12
19	18	17	16	15	14	13

Diagonal Pattern

As the name suggests, diagonal pattern follows a diagonal route to consume the matrix. A total possible variation is 8. A diagonal pattern using upper-left edge as starting point is shown in table 7.

TABLE 7
A DIAGONAL PATTERN CONFIGURATION USING 7 × 7 MATRIX

1	3	6	10	15	21	28
2	5	9	14	20	27	34
4	8	13	19	26	33	39
7	12	18	25	32	38	43
11	17	24	31	37	42	46
16	23	30	36	41	45	48
22	29	35	40	44	47	49

J-Pattern

J pattern uses a matrix consumption path that is composed of multiple J shaped route. A total possible variation is 8. A J-pattern using horizontal configuration with upper-left edge is shown in table 8.

TABLE 8
A J-PATTERN CONFIGURATION USING 7 × 7 MATRIX

1	2	3	4	5	6	7
14	13	12	11	10	9	8
15	16	17	18	19	20	21
28	27	26	25	24	23	22
29	30	31	32	33	34	35
42	41	40	39	38	37	36
43	44	45	46	47	48	49

It is also possible to generate user defined patterns. The sole purpose of multiple patterns is to scramble/permutate the matrix. Patterns can be changed with every data exchange session, which generates a random behavior that makes it difficult for attacker to decide what pattern is used. At the same time, number of possible structure rises dramatically. For example, if there are m patterns and n possible structures for each pattern, then, the total structures will be m × n. This makes cryptanalysis more difficult.

4 ANALYSIS OF PROPOSED ALGORITHM

Playfair was considered safe at the beginning of 20th century, because of the effort it takes to break the cipher manually. But, after invention of computers, this became a trivial problem [7]. The first known solution to this digram cipher is given by J. Mauborgne in 1913 [11]. After this, different methods are discovered to effectively crack this substitution cipher [7], [8], [12]. There are several common attacks on ciphers, which are – ciphertext only attack, known plaintext attack and chosen plaintext attack [13]. The proposed algorithm tends to increase the security by character set extension, generation of random key and a further permutation in the arrangement pattern of matrix. This creates confusion for attacker that makes the algorithm stronger. One way to exploit the security of algorithm like this, attacker needs to know the nature of the language. This is because the frequency of a letter in a language is always the same. And, this may lead an attacker to expose the plaintext structure. So, this possibility must be eliminated / minimized. Hence, a pre-encryption and post-encryption frequency analysis is required to show the effectiveness of an algorithm. To generate frequency distribution graph, number of occurrences of each letter in character set are counted and divided by occurrence of e (the letter in English with highest frequency). As a result, a relative frequency in range 0 and 1 is gained. The points on the horizontal axis correspond to the letters in order of decreasing frequency. More flat the relative frequencies are, more concealed the information is [14].

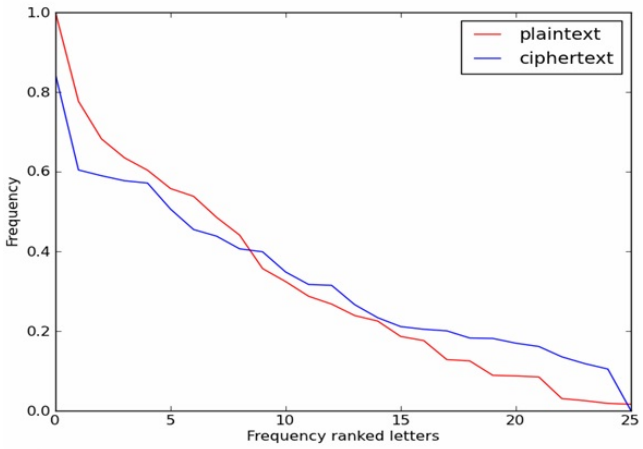


Figure 4: Frequency distribution in plaintext vs. conventional playfair ciphertext.

A frequency distribution analysis is performed on 55,900 popular words in English consists of 419,968 letters. Figure 4 shows the change in frequency distribution between a plaintext and conventional playfair ciphertext. The ciphertext curve is slightly flatter than plaintext, which denotes that some frequency information has been concealed.

On the other hand, figure 5 shows a relative comparison between plaintext, conventional playfair and modified playfair ciphertext. Modified playfair cipher curve shows some significant improvement over conventional playfair. Because, it provides a more flat curve then that of conventional playfair, which is better security. But yet, like the conventional playfair, it can be broken by following the identical principles, except, the modified one requires harder effort

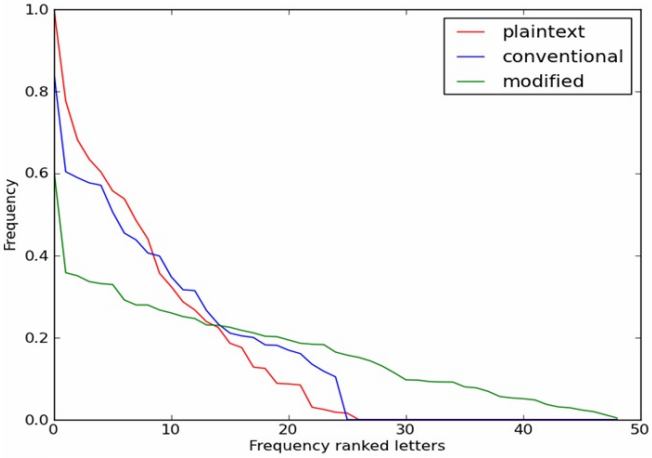


Figure 5: Relative frequency distribution between plaintext, conventional playfair ciphertext and modified playfair ciphertext.

5 Conclusion

This paper attempts to modify and extend the existing 5 × 5 playfair cipher in different ways, by extended character set, I and J ambiguity reduction, matrix modification and one to one ciphertext generation. The original playfair uses a single pattern to generate matrix in a left to right and top to bottom order. But instead, in proposed model, multiple matrix generation patterns are introduced. Selection of these patterns is driven by user which acts as initialization vector. This randomized behavior generates confusion for attacker that increases the security. At the same time, matrix extension to 7 × 7 produces more possible structures than original playfair. This paper can be used as a learning resource that will help to understand playfair, its vulnerabilities and will show an effective way to improve it, thus, helping students in understanding cryptography, algorithm enhancement and cryptanalysis in an easier way, which would be otherwise difficult with more advanced ciphers like DES or AES. It is possible to encrypt any binary data by using modified playfair cipher along with base 32 encoding. Wholly, the algorithm is unique, unambiguous and simple that leaves a lot of possibilities to be a useful learning resource and possibilities to implement it as a low-security protocol in a wide range of devices, including low powered embedded ones.

References

[1] A. Menezes, , A. J. Menezes, P. van Oorschot, S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996, pp. 4.

[2] William Stallings, *Cryptography and Network Security,* Fifth Edition, Pearson Education, 2011, pp. 33

[3] William Stallings and Lawrie Brown, *Computer Security - Principles And Practice*, Second Edition, Pearson Education, 2011, pp. 39-62.

[4] Simon Singh, *The Code Book – The Science of Secrecy from Ancient Egypt to Quantum Cryptography*, Anchor Books, 1999.

[5] William Stallings, *Cryptography and Network Security - Principles And Practice,* Fourth Edition, Prentice Hall, 2005, pp. 40.

[6] A. Menezes, , A. J. Menezes, P. van Oorschot, S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1996, pp. 274.

[7] Keith M. Martin, *Everyday Cryptography – Fundamental Principles & Applications*, Oxford University Press, 2012, pp. 62.

[8] Michael J. Cowan (2008): *Breaking Short Playfair Ciphers with the Simulated Annealing Algorithm*, "*Cryptologia*", 32:1, pp. 71-83.

[9] Aftab Alam, Sehat Ullah, Ishtiaq Wahid & Shah Khalid, *Universal Playfair Cipher Using MXN Matrix, "International Journal of Advanced Computer Science"*, Vol. 1, No. 3, pp. 113-117, Sep. 2011.

[10] A. T. Shakil, *A demonstration project on conventional and modified Playfair cipher algorithms.* Available online: http://sourceforge.net/projects/cryptographytools/files/Playfair%20Cipher/

[11] Joseph O. Mauborgne, *An advanced problem in cryptography and its solution*, Army Service Schools Press, 1918.

[12] Dorothy L. Sayers, *Have His Carcase,* Victor Gollancz, 1932.

[13] Mark Stamp, Richard M. Low, *Applied Cryptanalysis - Breaking Ciphers in the Real World*, Wiley Publication, pp. 2.

[14] William Stallings, *Cryptography and Network Security - Principles And Practice,* Fifth Edition, Pearson Education, pp. 45.