A Framework for the Development Playfair Cipher Considering Probability of Occurrence of Characters in English Literature

Uttam Kr. Mondal[1] , Satyendra Nath Mandal[2] and J. PalChoudhury[3]
um@cse.cemk.ac.in, satyen_kgec@rediffmail.com, jnpc193@yahoo.com

[1]Lecturer,Dept. of CSE & IT/College of Engg. & Management, Kolaghat , Midnapur(W.B), India
[2]Lecturer, Dept. of IT/Kalyani Govt. Engg. College,Kalyani, Nadia(W.B),India
[3]Asst. Professor, Dept. of IT/Kalyani Govt. Engg. College,Kalyani, Nadia(W.B),India

Abstract
The Playfair cipher was the first practical digraph substitution cipher. The scheme was invented in 1854 by Charless Wheatstone, but was the named after Lord Playfair who promoted the use of the cipher. In this paper, two algorithms have been proposed to written the Playfair cipher. The first effort is being made to write the playfair cipher in a new way by considering the probability of the occurrence of english character in context to english literature. In the proposed algorithm at each and every time, the algorithm will check whether, the plain text contains odd number of words or not. If the plain text is of odd number of characters, the algorithm will concatenate a dummy character at the end with lowest probability to form even length word in the cipher text and then search is being made in the dictionary to find out whether the word with dummy character is present in the dictionary or not. If that word is present in the dictionary, delete that dummy character of that word and the same effort is being made with another dummy character with next lowest probability occurrence. The second method add another new approach, it replace all blanks between two consecutive word by some character which has least probability to appear at that position in the sentence and search the dictionary to check whether this new word has different meaning . If new word if present, replace blank with next higher probability character for this blank. Finally, time comparison has made of encryption and decryption of some sentences using original with two algorithms and it has been found that the decryption time are much better than original play fair algorithm and that will produce the original message from encrypted message directly.

Keywords: Plain text, Cipher text, Substitute cipher, dummy character, Playfair cipher

1. Introduction

The Playfair cipher was the first practical digraph substitution cipher. The scheme was invented in 1854 by Charless Wheatstone, but was named after Lord Playfair who promoted the use of the cipher. The technique encrypts pairs of letters, instead of single letter as in the simple substitution cipher. The Playfair is significantly harder to break since the frequency analysis used to simple substitution does not work with it. Frequency analysis can still be undertaken, but on the ~600 possible digraphs rather than the 26 possible monographs. Frequency analysis thus requires much more cipher text in order to work.It was used for the practical purpose by British forces in the Second Boer War and in World War I and the for the same purpose by the Australians during World War II . This was because Playfair is reasonably fast to use and requires no special equipment. A typical scenario for Playfair use would be to protect the important but non-critical secrets during actual combat.

In this paper, a first effort is being made to write the Playfair cipher in a new way by considering the probability of the occurrence of English character in context to English literature. In the proposed algorithm at each and every time, the algorithm will check whether, the plain text contains odd number of words or not. If the plain text is of odd number of characters, the algorithm will concatenate a dummy character at the end with lowest probability to form even length word in the cipher text and then search is being made in the dictionary to find out whether the word with dummy character is present in the dictionary or not. If that word is present in the dictionary, delete that dummy character of that word and the same effort is being made with another dummy character with next lowest probability occurrence. If the word in the source text is not present in the dictionary, that word has to be incorporated in the dictionary. If one word(with same spelling) occurs more than one time in the source text, a left rotation will make in the next word after each occurrence to get the revised word formation. Now the

same effort is being made for the revised word. Other conditions are same as existing Playfair cipher algorithm. The entire steps discussed earlier have been repeated for the entire words of the source text.

The 'key' for the play fair cipher is generally a 5X5 matrix of character. The idea was to place word in the 5X5 matrix and rest of the matrix filled up by the characters which are not in the word otherwise the matrix can be formed in different ways. The characters i and j are place in the same place of the matrix as there is no j combined with i. Then encrypt the word according the Playfair cipher algorithm. But problem of this algorithm is, after encryption of odd length word with dummy character , it may happen that it has different meaning (e.g the word quart means two points , after addition z , it will be quartz and meaning is rock crystal). The another example, in original Playfair cipher is very strong but in addition of dummy character the new character may have the different meaning (e.g The word "Ira" is odd length ,if we add a dummy character "q" usually used in Playfair cipher, the new word will be "Iraq" which has totally different meaning. This is the reason to rewrite the algorithm in this paper. This is the reason to rewrite the first algorithm (Algorithm A) in this paper.

In the second algorithm (Algorithm B), an effort has been made to fill the blanks among the word in the line so that at the time of decryption from cipher text, the algorithm recovers the original plane text without any confusion. It is done by considering rear frequency set with respect to the position of the letter that can be appear after a word and replace the blanks with all these characters so that at time of decryption from cipher text, the algorithm directly get the original plane text. But in original Playfair algorithm, only one dummy character is added to make the length of string even. The algorithm is good but problem is at time of getting original text from cipher text gives more than one option. This is the reason to rewrite the algorithm.

2. Cryptographic Techniques

Cryptography, a word with Greek origins, means "secret writing". However, we use the term to the science and art of transforming messages to make them secure and immune to attacks. There are several techniques to implement cryptography. These methods will be described in the following sections.

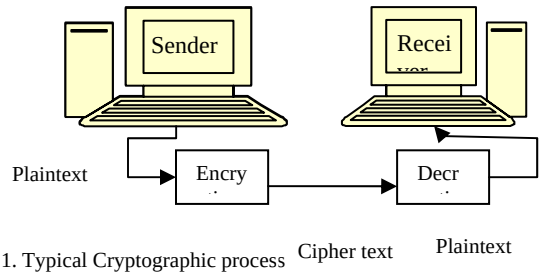

Figure 1. Typical Cryptographic process

2. 1 Plain text And Cipher text

The original message, before transformed, is called plaintext. After the message is transformed, it is called cipher text. An encryption algorithm transforms the plaintext into cipher text; a decryption algorithm transforms the cipher text back into plaintext. The sender uses an encryption algorithm, and the receiver uses a decryption algorithm.

2.2 Technique of cryptography

Types of cipher are furnished in Tree structure

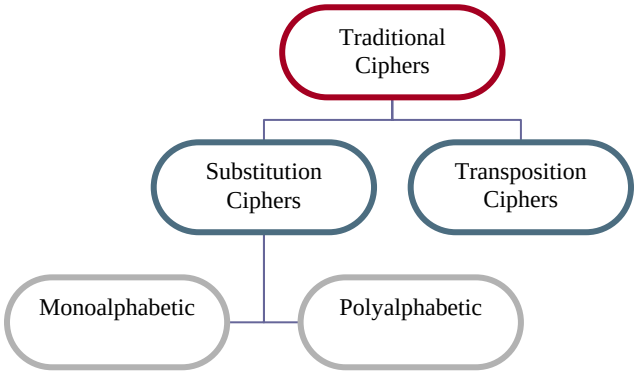

Figure 2. Different Traditional Cipher

2.3 Substitution Techniques

The substitution techniques are –

2.3.1 Caesar Cipher

It is a special case of substitution techniques where each alphabet in a message is replaced by an alphabet three places down the line.

Example= plain text 'ATUL' will become cipher text 'DWXQ'.

2.3.2 Modified Version of Caesar Cipher

In this scheme the cipher text alphabets corresponding to the original plain text alphabet is not necessarily be three places down the order, but instead, can be any places down the order.

Example= the plain text alphabet 'A' can be replaced by any alphabet i.e.,

from 'b' to 'z'.

So this scheme has 25 possibilities of replacement.

This method is susceptible to "BRUT FORCE ATTACK".

2.3.3 Mono-alphabetic cipher

Rather than using a uniform scheme for all the alphabets in a given plain text message, we decide to use random substitution.

This means 'A' can be replaced by any other alphabet ('B' through 'Z'),

Each 'B' can also be replaced by any other random alphabet ('A' or 'C' through 'Z').

This is also susceptible to attack.

2.3.4 Homophonic Substitution Cipher

It is similar to monoalphabatic cipher. Like a plain substitution cipher technique we replace one alphabet with another in this scheme. However the difference the two techniques is that the replacement alphabet set in the case of simple substitution technique is fixed, where as in the case of this cipher one plain text alphabet can map to more than one cipher text alphabet. For instance, 'A' can be replaced by 'D', 'H', 'P'. 'B' can be replaced by 'E', 'I', 'Q'.

2.3.5 Polygram Substitution Cipher

Polygram substitution cipher technique replaces one block of a plain text with a block of cipher text- it does not work on a character by character basis. Playfair is one type of polygram substitution cipher. For instance 'HELLO' could be replaced by 'YUQQW'. But 'HELL' could be replaced by totally different cipher text block 'TEUI'.

2.3.6 Polyalphabatic Substitution Cipher-

This cipher uses multiple one character key. Each of the keys encrypts one plain text character. The first key encrypts the first plain text character; and the second key encrypts the second plain text character, and so on. After all the keys are used they are recycled. Thus if we have 30 one letter keys, every 30th character in the plain text would be replaced with the same key. This number is called as the period of the cipher.

2.4 Transposition Technique

Transposition techniques differ from substitution techniques in the way that they do not simply replace one alphabet with another: they also perform some permutation over plain text alphabets.

In this category there are the following types of techniques:

Rail Fence Technique
Simple Columnar Transposition Technique
Vernam Cipher (One-Time Pad)
Book Cipher / Running Key Cipher

2.4 Encryption and Decryption:

The process of encoding plain text messages into cipher text messages is called encryption.

The reverse process of transforming cipher text messages back to plain text is called decryption.

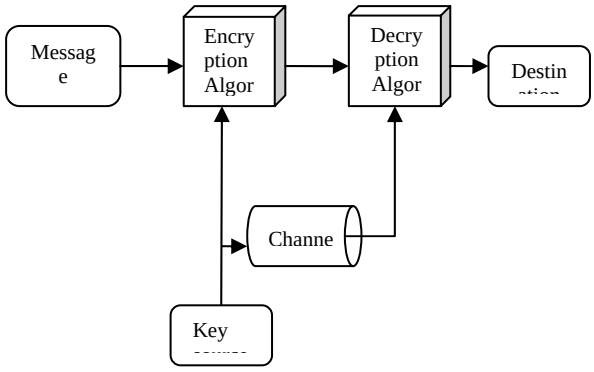

Figure 3. Model of Conventional Cryptosystem

2.6 Symmetric Key Cryptography

If the same key is used for encryption and decryption, we call the mechanism as Symmetric Key Cryptography.

Cipher text = encrypt (plaintext, key)
Plaintext = decrypt (cipher text, key)

Symmetric key cryptography is useful if you want to encrypt files on your computer, and you intend to decrypt them yourself. (But if you intend to send them to someone else to be decrypted, then you have a "key distribution problem".) In security, we assume the encryption algorithms that we have chosen to use are publicly known; only the key is secret to the participants.

2.7 Asymmetric Key Cryptography

If two different keys are used in a cryptographic mechanism, wherein one key is used for encryption, and another, different keys are used for decryption; we call the mechanism as Asymmetric Key Cryptography.

2. Proposed Modified Playfair Cipher

Algorithm A.

The 'key' for a Playfair cipher is generally 5X5 matrix of character. e.g

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

We can apply the encryption rules to encrypt the plaintext.

Remove any punctuation or characters that are not present in the key square except blank character (this may mean spelling out numbers, punctuation etc)
Identify any double letters in the plaintext and replace the second occurrence with an 'X' e.g 'hammer' -> 'hamxmer'.
If plaintext contain has an odd number of characters, append a letter which has lowest probability of occurrence in the English literature to make it even. After search the dictionary to check whether this new even length word present or not. If it is present, replace the append character with next lowest probability character . repeat 3 until a new word is formed which has no other meaning in the dictionary. e. g 'quart' is odd length and z has lowest probability in the English literature but if you append z , the new word will be 'quartz' which has totally different meaning)
If same word present more than one times, rotate this word one character after each occurrence. e.g pussy will be after first occurrence ussyp).
Break the plaintext into pairs of letters e.g. 'hamxer' ->'ha mx er'
The algorithm now works on each of the letter pairs.
Locate the letters in the key square(the examples given are using the key square above)
a. If the letters are different rows and columns, replace the pair with the letters on the same row respectively but at the other pair of corners of the rectangle defined by the original pair. The order is important – the first encrypted letter of the pair is the one that lies on the same row as first of the plaintext letter. 'HA'->'CD'

b. If the letters appear on the same row of the table, replace them with the letters to their immediate right respectively.'LB' ->'GA'.

c. If the letters appear on the same column of the table, replace them with the letters immediately below respectively.'LX'->'QZ'.

Algorithm B.

Step1 : Rules for finding RFS

I. Find out the letters with frequency of occurrences for i^{th} position of RFS
from the English dictionary .
II. Choose less frequency used letter for i^{th} position for RFS.

Step 2: Rules for Removing Space:
. For space at i^{th} position ,choose i^{th} letter from RFS , if it again form a new word choose $(i+1)^{th}$ letter, so on ….(*i* is counted from previous next of the space position, and from the first position at the time of beginning)

II. For the last position add RFS letter if last position is odd , otherwise do nothing.

III. For non-dictionary word add ZQ in the space position.

For compound word like " without " after "with" if o 's position is RFS position then add QZ after "with " for showing continuation to next word ,here "out".
For non-dictionary word if ZQ is present , then add QZ then continue ZQ and corresponding letter.
(because in English dictionary there is not a single occurrence of ZQ & QZ in
any position of word)
VI. For non-dictionary last odd position first add QZ then add i^{th} element (the space position after the word)from the RFS.

VII. Put x between two same consecutive letter.
VIII. If after adding x total length becomes odd add 'z' at the end. if 'z' is present (previous character) put 'q' .

4. Result

An example encryption of plaintext "pussy cat pussy cat where had you been, I had been to London to see the queen".

The idea in this paper for algorithm A is frequency of occurrence of letters in English text. The frequency of occurrence of characters of English text of 100 characters is furnished in table 1.

Table 1: Frequency of occurrence of letters in an English text.

Letter	Freq Uency	Letter	Frequency	Letter	Frequency	Letter	Frequency
E	12.7	H	6.1	W	2.3	K	0.08
T	9.1	R	6.0	F	2.2	J	0.02
A	8.2	D	4.3	G	2.0	Q	0.01
O	7.5	L	4.0	Y	2.0	X	0.01
I	7.0	C	2.8	P	1.9	Z	0.01
N	6.7	U	2.8	B	1.5		
S	6.3	M	2.4	V	1.0		

Each and every times, the dummy character has been added depending on lowest frequency character so that new word after adding dummy character has not been different meaning. If it has been occurred, the next lowest character has been chosen.
After the rules 3. the string will be "pusxsy catz usxsyp atcz wherez hadz youz bexenz iz adhz exenbz to london ot sexe thez quexen"
After the execution of the algorithm the cipher text will be "ZFOVVW FCPY NXVOWZ BPQP DNMIQT CDLW TVXL EIOQUW UT LBQW QOHILT BT BXSHSI TB OHOQ WEQT UXQOHI"

For the algorithm B. the rear frequency set (RFS) of English character has been considered.
The set is {{X, Z},{J, Z},{Q,Z},{ J ,X},{Q ,J},{J,Q},{J, Q},{J,Q},J ,J,Q,J,F ,H ,P ,B ,K ,A ,M ,O ,L ,C ,W ,G,K ,V ,U ,T}. In this set , the character X and Z has least number of probability to appear at first position and other characters are arranged in ascending order is furnished in table 2.

Table 2. Rear Frequency Set (RFS)

Position	Letter	Frequency	Position	Letter	Frequency
1	X, Z	12,60	15	P	0
2	J,Z	14, 27	16	B	0
3	Q, Z	122,130	17	K	0
4	J, X	78,85	18	A	0
5	Q, J	26,13	19	M	0
6	J, Q	9,46	20	O	0
7	J,Q	1,10	21	L	0
8	J,Q	2,9	22	C	0
9	J	0	23	W	0
10	J	0	24	G	0
11	Q	0	25	K	0
12	J	0	26	V	0
13	F	0	27	U	0
14	H	0	28	T	0

*Length of longest word is 28 in English dictionary.
(antidisestablishmentarianism , considering total no. of dictionary words: 45428) . The position column in this table indicate the probability of appearance of character of character column.

Now again considering the same line “pussy cat pussy cat where had you been
I have been to london to see the queen” the algorithm B will as follows

After applying the above rules.

1)
PUSQZSYJCATQZJPUSQZSYJCATQZJWHEREQZJH
AQZDJYOUQZJBEQZENQZQIZQHAQZVEQBEQZEN
QZQTOQZQLONDONZQTOQZQSEXEQZJTHEQZJQ
UEXENQZJQZXZ
(extra letter (Z) is added at end because total length is odd.)

2) now apply playfair cipher rules:
taking the following KEY

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

We get the following Cipher text …

zfxhwxtrfczetuzfxhwxtrfczetudnmicmtucdulbntvxutueiulh
iulullucdulomelcmtquhlubtuluqsibsuwezxeluohoqulobecul
ueiqoquhtuulzl

after decoding.....

PUSQZSYJCATQZJPUSQZSYJCATQZJWHEREQZJH
AQZDJYOUQZJBEQZENQZQIZQHAQZVEQBEQZEN
QZQTOQZQLONDONZQTOQZQSEXEQZJTHEQZJQ
UEXENQZJQZXZ
putting space & generating original message..

pussy cat pussy cat where had you been
I have been to london to see the queen

5.Conclusion and future work

The original playfair cipher is very strong but in addition of dummy character the new character may have the different meaning (e.g The word “Ira” is odd length ,if we add a dummy character “q” usually used in playfair cipher, the new word will be “Iraq” which has totally different meaning. In this algorithm A, the blank character is allowed, result is easy cryptanalysis. In this algorithm B, the blank character is replaced by character which has least value in rear frequency set. So, It is easy to recover plain text from cipher text and the time comparison of encryption and decryption of some sentences are furnished in Table 3 and Table 4.Finally, It has been observed that the decryption time is always less than from original play fair algorithm.

Table 3: Time comparison for encryption among three algorithms

SL No.	Sentences	Original Algo(s)	First Algo(s)	Second Algo(s)
01	Pussy cat pussy cat where had you been.	0.057	0.103	0.841
02	Twinkle twinkle little star.	0.057	0.070	0.306
03	An error doesn’t become a mistake until you refuse to correct it.	0.083	0.095	0.553
04	The best part of beauty is that which no picture can express.	0.097	0.096	0.506
05	Freedom rings wherever opinions clash.	0.043	0.082	0.351
06	Sincerity and truth are the basis of every virtue.	0.103	0.116	0.432
07	A thing is valuable where it born.	0.056	0.070	0.356
08	A good face is the best letter of recommendation.	0.072	0.092	0.609
09	Necessity is the mother of invention.	0.043	0.059	0.405
10	Ability is poor man’s wealth.	0.043	0.058	0.263

Table 4: Time comparison for decryption among three algorithms

SL No.	Sentences	Original Algo(s)	First Algo(s)	Second Algo(s)
01	Pussy cat pussy cat where had you been.	0.919	0.143	1.640
02	Twinkle twinkle little star.	0.706	0.140	0.805
03	An error doesn't become a mistake until you refuse to correct it.	1.482	0.228	2.036
04	The best part of beauty is that which no picture can express.	1.393	0.230	2.588
05	Freedom rings wherever opinions clash.	1.430	0.095	0.987
06	Sincerity and truth are the basis of every virtue.	1.145	0.201	0.335
07	A thing is valuable where it born.	0.773	0.158	1.087
08	A good face is the best letter of recommendation.	1.097	0.160	1.569
09	Necessity is the mother of invention.	2.469	0.127	0.334
10	Ability is poor man's wealth.	0.841	0.135	0.803

References.

[1] Wikipedia has a good description of the encryption/decryption process, history and cryptanalysis of play fair algorithm.

[2] Basic Cryptanalysis, FM 34-40-2,FIELD MANUAL, DEPARTMENT OF THE ARMY http://www.umich.edu/~umich/fm-34-40-2/

[3]Lecturer 3 " Cryptanalysis of the Classical Ciphers" by Dr. AlexBiryukov. Available at http://www.wisdom.weizmann.ac.il/~albi/cryptanalysis/lect3.htm.

[4]" The Code Book" by Simon Singh's

[5] Singh, Simon(200)."The Science of Secrecy from Ancient Egypt to Quantum Cryptography. ISBN 0-385-49532-3.

[6] "Cryptography and Network Security" by Atul Kahate

[7]" Introduction to Cryptography with Coding Theory" by Wade Trappe and Lawrence C. Washington.

[8] "Playfair Cipher- Description and Cryptanalysis" by James Lyons at http://www.dsplabs.com.au-2007/

[9]" Cryptography & Network Security" by Behrouz A. Forouzan



Uttam Kr. Mondal(05/03/1982) has received his Bachelor of Engineering(B.E) degree in InformationTechnology in 2004 and Master ofTechnology (M.Tech) in Information Technology in 2006 from University of Calcutta , India. He has now working as a lecturer in department of Computer science & Engineering and Information Technology in College of Engineering & Management, Kolaghat,West Bengal, India. His research areas include cryptography & network Security. He has one research paper in national conference.



Satyendra Nath Mandal (23/10/1975) has received his B.Tech & M.Tech in Computer Science & Engineering from university of Calcutta, West Bengal India. He is now working as a lecturer in department of Information Technology at Kalyani Govt. Engg. College , Kalyani , Nadia, West Bengal, India. His field of research areas include cryptography & network Security, fuzzy logic, Artificial Neural Network, Genetic Algorithm etc. He has about 15 research papers in national and International conferences. His one research paper is published in International journal of computer science and Engineering published by WASET.



Dr. J. Paul Choudhury(Jagannibas Paul Choudhury) completed Bachelor of Electronics and Tele-Communication Engineering(Hons) from Jadavpur University, Kolkata M. Tech in Electronics and Electrical Engineering under the specialization of Control and Automation Engineering from Indian Institute of Technology, Kharagpur and. thereafter completed PhD(Engg) from Jadavpur University, Kolkata. At present Dr. Paul Choudhury is with Information Technology Department, Kalyani Government Engineering College, Dt Nadia, West Bengal, India, and he has 60 publications in National and International Journals and in Conference Proceedings. His field of interest is Soft Computing, Data Base, Object Oriented Methodology, etc.