Breaking A Playfair Cipher Using Memetic Algorithm

Asst. Lecturer
Dalal Abdulmohsin Hammood
College of Elec. & Electronic Techniques
Foundation of Technical Education
Alsaady_Dalal@yahoo.com

Abstract :

The playfair cipher operates on pairs of letters. The cipher text is broken into group of blocks, each block contains two letters.

Encryption and decryption cipher depends on key search, which is a square matrix that contains 5x5 letters. The matrix contains every letter expect letter j.

In this paper the cryptanalyst a playfair is presented using search techniques. It is a Memetic Algorithm, another method is used to know a key through a correct Digram letters, which is used as an essential factor in the cryptanalyst cipher text.

Results without memtiec algorithm and with it have been compared, to see which one has a best solution.

The result clears that the memetic algorithm has a best solution more than a classical method, and faster than without used it, which has optimal solution.

The length of text is 1802 letter. The key size is 25 letters as a matrix 5x5.

All the algorithms were programmed using Matlab programming as M-file, and tested successfully

Keyword: Playfair Cipher, Frequency of Bi-Grams, cryptography, Memetic Aalgorithm (MA), And Key Space.

تحليل شفرة Playfair باستخدام خوارزمية البحث MA

م.م دلال عبد المحسن حمود
كلية التقنيات الكهربائية والالكترونية
هيئة التعليم التقني

الخلاصة:

تعمل شفرة الـ playfair على زوج من الأحرف. يقسم النص غالى مجموعة من البلوكات، يحتوي كل بلوك على زوج من الأحرف.

تعتمد عملية التشفير وفك التشفير على مفتاح البحث، حيث يكون على شكل مصفوفة مربعة من الأحرف 5x5.

في هذا البحث تم استخدام تحليل شفرة الـ playfair باستخدام تقنية من تقنيات خوارزميات البحث وهي memetic Algorithm، واستخدمت طريقة أخرى استندت على تكرار الحروف المزدوجة الأكثر شيوعا في اللغة الانكليزية، حيث يستخدم كعامل أساسي في تحليل الشفرة.

تم مقارنه النتائج باستخدام الخوارزمية memetic Algorithm وبدون استخدامها، حيث أظهرت النتائج سرعة استخدام الخوارزمية بدلا من الطريقة الكلاسيكية المعروفة. تهدف memetic Algorithm ، حيث تمتلك الحل الأمثل وبوقت قليل. طول النص 1802 حرف. طبقت الخوارزميات وتمت برمجتها ببرنامج MATLAB بنجاح

1. Introduction

Encryption is the study of mathematical techniques related to the aspects of information security such as confidentiality, data integrity, authentication and data origination [1-3]. It has been associated with the problem of designing and analysing encryption schemes. Schemes which provide secret communication over insecure communication media[4]. Cryptography was used as a tool to protect national secrets and strategies [1].

Decryption is the branch of cryptology concerned with solving the cryptographic systems used by others. The objects of cryptanalysts are to read the text of encrypted messages and to recover the cryptographic systems used. The text is recovered for its potential intelligence value. The systems are recovered for application to future [2,5].

Encryption and decryption algorithms are collectively called cryptographic algorithms[6].

Memetic algorithm is a search algorithm that is based on Evolutionary programming, it is a local improvement procedures for problem search[6-8].

The algorithm is discovered by (Darwin 1983), it depends on a genetic inheritance, then it could obtained a new generation. This generation represents a best solution[7,8].

The purplish papers in playfair cipher are: In 2009 M. Packirisamy & S. Gandhidss presented a new approach for secure transmission of message by modified version of Playfair cipher combining with Random number generator methods [9].

In 2011 S.S., Srivastava & G. Gupta presented a paper which dealt in with the security aspects of the proposed cipher and found that it is considerably secure against attacks. Moreover from the analysis made we conclude that the cipher is potentially a strong one[10].

In 2013 D. A. Hammood. Presented paper. It deals with cryptanalyst a playfair cipher using a classical and a certain methods to find the key[11].

In this paper, two methods are used. They are genetic and memetic algorithm to find optimal solution, and comparing between them to see which one has a best solution. It clears that the two methods have the same solution but the memetic algorithm clears the time less than genetic algorithm. So the mematic algorithm is faster than genetic algorithm. The length of text is 1802 letters.

2. The algorithms

2.1 Encryption Algorithm

Step 1: The length of text is 1802 letter.

Step 2: Removing all special characters (" ' = + * & ^ % # $ @ ! , ; : / ? > < , .), and blanks.

Step 3: Replacing all i with j.

Step 4: Encryption the plain text as the following:[12,13]

The key is: ilovemyworkcanuabdfghqstxz

i. The key must arrange as a square matrix 5x5, as shown below:

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

ii. The plain text is:

thefriendshipsididntwanttodepartmyhometownihavebeenheresinceiwasbornididn twanttoleavemyschoolandmostimportantlyididntwanttoleavesamthatussamandka herineourfriendshipstartedingradeschoolfourgradeidroppedmypenciandwasgoing togetitandofcoursesamleanedtogetittooourfriendshipstartedthatdaywebecamebest friendswewereinseparablewedideverythingtogetherhetoldmehelovedmelikeasiste rsomethingstartedchangingiwaslookingatsamdifferentlyiknewwhatthisfeelingwa sbutitcrushedmeilovedsambecauseiwasjustlikeasistersoitoldhimhewasjustlikeabr otherbutthenihadtomovetoawholeanewcountrymyimportantjournalwithconfessio nletterupupupsamhurryupkatherinesgannaleavehereyourchanceimessedupprealba dtellingkatherineilovedheronlylikeasisterbadideanowitstimetotellyouiloveherno wnowhafounditsbeentwoweekssincesamdiedheranstraightintoatruckcallingout mynamejusttogivemeanoteohyeahiamgladthatidroppedatpencilitletmemeetagreat personlikeyousamifyoucanhearmeimissyouandiloveyouwaitformedearkathguess whatiloveyouforeverandnomatterhowforwearejustrememberyouhavemesamthee ndmostofotherstoriesareaboutloveandhowpeopleenduphappymystoryisfarfromth

atmynameisjoshuebutmothercallsmeyouthisismysidinmystorybigbrotherwakeup
motheriscomingupherewiththestickgetupyoustupidfoolyesmomihadalwaysloved
mymotherbutshehadalwaysblamedmeformyfatherdeathsonowiamdoingeverythin
gtogainbackherloveevenifitmintkillingmyselfitgottothepointwhereiwastryingsoh
ardthatiwouldstarvefordayswasitryinghardenoughmotherdidntthinksowhatwasid
oingwrongsoiwalkeduptomysistersroommotherwhatwasidoingwrongwhydontyo
ulovemeanymoremotheriamsorryforthefirsttimeinalongtimeiactuallysmiledbutth
atnightiforgottocleanuponeoftheroomsmotherbealmeuntilicouldntistandupmothe
rdoyoulovemenowhaveimadeupforeverythingihavedonewhycouldntyouoseethati
lovedyouallalongmotherwhy

iii. Dividing the plain text into blocks, each block contains pairs of letters.

th ef ri en **ds** hi ps id id nt wa nt to de pa rt my ho me to wn ih av eb **ee nh** er es in ce iw as bo rni di dn tw an tt ol ea ve my sc h

Cases:

- If there is the same two letters (ee), the process to follow is by adding x instead of the second letter and moving letter to the next block, as the following block: ee **à** ex en.
- If there is one letter at the end of block, (x) is append as h (hx).
- Replacing j with i , because the matrix must be 5x5.
- In case the two letters are not appearing in the same row or column, the below letter that represent the same raw will take. And another letter will take above letter that represent in the same row.
  th à is ciphered FZ
- When the two letters are appearing in the same row, then it takes the right letter of the same row, and another letter takes next letter. If there is the last letter then it circulates to the first letter of the same row (wrapping back to start from end).
  pa à CN
  ds à SL

Finally: The cipher text is

ZFLQMVOBSIDVSTMSMSAXYNAXXLQITBYZYWGVKIXLNGVDULKQ
OPOBQVKVIMUAILYNPCVWCOSMGCXYNUXZXLOIULIKMTUDWOV
ONUSCIXSLKSVWLFAXYAMSMSAXYNAXXLOIULIPCYZFFLCZTCYC
CGYBQVMVBOVNYHMVOBSIDVSTLFYZIQOCHWCFIPUDWOVOGLH

UHWCFILHMEXEKSCKTOBDMNUGMCTXWOCFXWXLPLSNUGIDAV
NMZIPCYOINUIQXLQOSLXZXLWOVNYHMVOBSIDVSTLFYZIQZFFLF
CWRKQIBCYKQIPLTMVOBSIKOKOKVOCPITBYUAEOKIQMSIEVKAL
DVGXXLQOZFVKQVXLIFKIQVOVEISCIOEMLBIMTXVKXIKIZFOCDX
LFYZIQUDNUDOGXOMCTOVEWOCFNXTCYSMGTQLKVAXYAEMBO
NORGFLZFMIQLIOOCXNCTCBSLSAUHZDIQKILOVEIQTCKCIBNBPIO
MCTVCTXOLBKCTMIPLMZVLXLIFDVRDOKCTVCTXOLBKNCWVZFV
KCBXZZFOBVDCFXLWIEIXLNYGVOINUOKNIBUZYWYMLKSVWLFA
XLVHUUNOYLSDUWGQLTZIMWGOIXZPLUHZBZBSTCYZHWZKWBZ
YBZFVKOCIPFNGOUNOIULVQVKLKVNMUFUUAILKITZPIHCZKLBEA
CFPLOTOLGXYBZFVKOCILOVEIFQVKWGYAOLBKCTMIPLKUCFMSL
BGWMOXTSLKIXLPLOTYAVNLOVEVQVKGWNGWNFUGLBUSMSLX
TQKOBXYWNOPKBTZIMUAIPCYSMIQQVYUCXZYCLHQSLAXLNZYB
AMBFYOLGXVNSYWACYILCZXZXLDOEIKINULXIVFRLBDVCYFOCF
ZFFLMSWVSZEKFCXSOBDMOLLYLPKIKILPNFKVFLEKMZWGOLBK
WLCZCYLDWLBANUQVUYKIMCMITMVNNUSMOVEIWLNRCLLTVW
KIQIUYYBZFHNIPXMFUSLOVEIWLAHVWIEVKNUGCIWFLPLUZWNG
LKRLBKVVCTXKVKIKCVKWLHZULIKIPCYZFOPOBSCIXXLGLZFVKT
XVWLITCKVNCVNLYVELBCGGVKXIVTEOPOBHCZQBTTKYWTXVW
MLTDUYHYIWZFFLYWUNKIMILVZDBVCBSYLXQVMUFYITKIWLAZ
DVIMICMTMSOCYWTXVWKAODUKLXQVKRBYVBSKLXQVMVIDIW
OCHNZQVKOKLSFZQVTXMDWQLPBZWLCZZASEFGWOVOKLICIWV
DCFFYYNMTOVEISCWYLXQVKUAZZDVQCFFYYNMTAECYIQKIGLK
YATFLQVMHLBZFXIGWMOCYGIOCQOEIKWZFOCFXWXCLUCNARQ
VKOVEIIEOBLDLSCMAXMEOTOLGXYWPIYTLSXWXZXLZFKEVLAX
RGVKILYNTXKWOCDXVGUYFSFUSLNWAVSILFURLQVWFCMTYNI
MZYMLGXFUMHOBVNHQWIZFVKSMGCXZZFOCMPWNFUXYCTMSV
LGXRKWGDXVLYNEYIQBZXLYWIMTXVKZMWOIWWIZFVKRGFLYN
IMGIOCXNWVGXRGMFWGLAVNOVEIKINUWYVWIKLXQVMVCYXI
WZKWGLYZQVDLMZXZSLKIOCFYWGFXMCILNAZAFYYAICLOIQCB
XZZFFLCOHQSLGLWHLXXLAILBUBXEBOLGZFVKWOIWICLXQVKU
LBIYVBAXLOMDVNIFAXMILFCGBZWIZFVKGIWLAVVEIKOBWNFUE
IMCCFVBTQVWIEVKALDVGXVDULIQWGOKFRNIAVGCLAVNIXOPL
PFUSLOVEIFMVNFYYFOVGXWIZFVKRGWT

2.2 Decryption Algorithm

Step 1: Dividing a cipher text into blocks, each block contains two letters.

ZF LQ MV OB SI DV ST MS MS AX YN AX XL QI TB YZ YW GV KI XL NG VD UL KQ OP OB QV KV

- If there is two letters in the same row, left letter is taken.
- If two letters in the same column, above letter is taken
- If two letters are not in the same row or column, intersection points are taken. [12-15,17]

3. Memetic Algorithm[6-8]

Step (1): Specifying the number of population=pop, max number of generation=mx_gen
Step(2): population(pop); g=0; reproduction rate=0.4, mutation rate=0.2.
Step(3): initialization of population: size of population.
Step(4): testing: population(pop)
Step(5): if max of generation then go to step(15)
Step(6): pop=pop+1
Step(7): selection of parent randomly; population(pop)
Step(8): reproduction population(pop) according to rate (0.4) as shown in figure 3.
Step(9): mutate population(pop); according to rate (0.2) as shown in figure 4.
Step(10): testing: population(pop)
Step(11): apply local search
Step(12): survive population(pop)
Step(13): Applying final local search to find best child.
Step(14): go to step(4)
Step(15): best solution
Step(16): end

3.1 Fitness Value

3.1.1 Fitness Scoring

$$Fm = m \sum^{Q} fi \; Si \; /100 \qquad \mathbf{L}\,(1)$$

m is text length, F_m fitness value, fi is the percentage frequency of that bi- or tri-gram in the text, Si is the fitness score to the i^{th} bi- or tri-gram or four letter tested for, and the summation is over the Q bi- and tri-grams checked.
The fitness scoring table is shown in table 1

Table (1): Fitness Scoring

Bigram	score	Bi / trigram	Score	Four letter	score
TH	+3	ED	+2	THAT	+7
HE	+2	THE	+5	THEN	+7
IN	+2	ING	+5	THEM	+7
ER	+1	AND	+5	EEEE	-7
AN	+2	EEE	-5		

3.1.2 Fitness Diphthong(Di-Gram) Letters

$$F(K) = w_1 \sum_{i=1}^{N} (L_i^{(1)} - D_i^{(1)})^2 + w_2 \sum_{i,j=1}^{N} (L_{i,j}^{(2)} - D_{i,j}^{(2)})^2 \text{.....(2)}$$

Where N is the number of letters in the English language (A..Z), $L_i^{(1)}$, and $L_{i,j}^{(2)}$ are the known language unigram, bigram statistics, and $D_i^{(1)}$,and $D_{i,j}^{(2)}$ are the unigram, bigram statistics of the message decrypted with key *K*. The weights w_1,and w_2 can be varied to allow more or less emphasis on particular statistics. Fig. 1: shows the frequency of 15 Di-grams letters in English text[1,14-16].

Figure(2,.3, and 4) clear the algorithms of Memetic Algorithm [17].

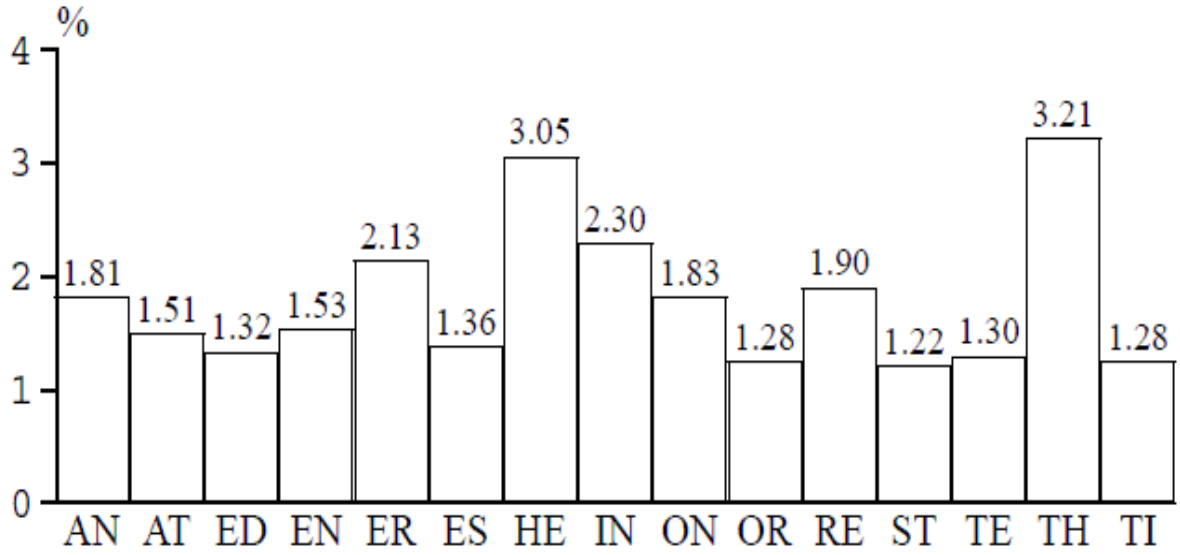

Fig. (1): Frequency Of 15 Common Di-grams In English Text

4. Flow Char Of Memetic Algorithm

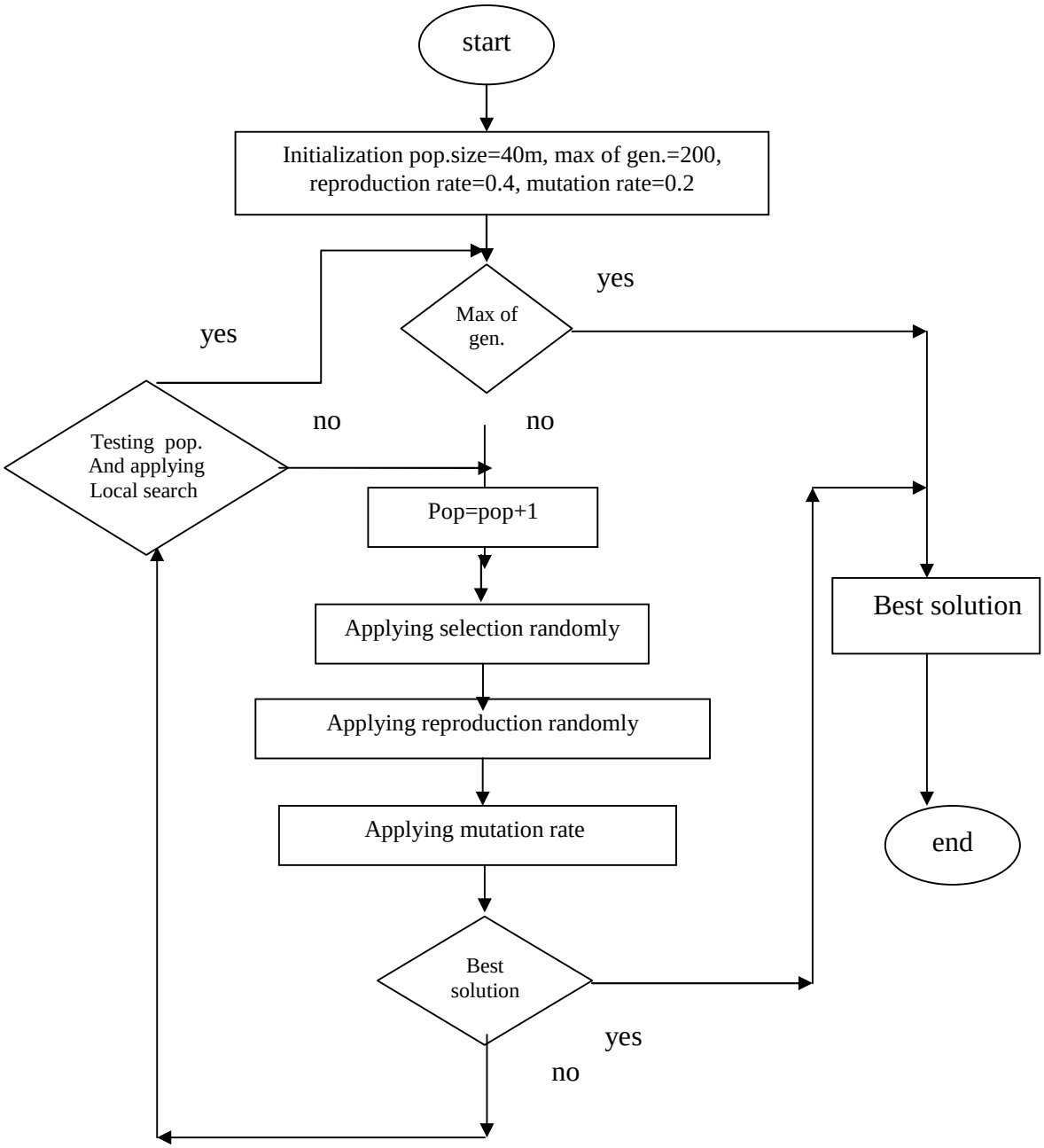

Fig. (2) : flow chart of MA

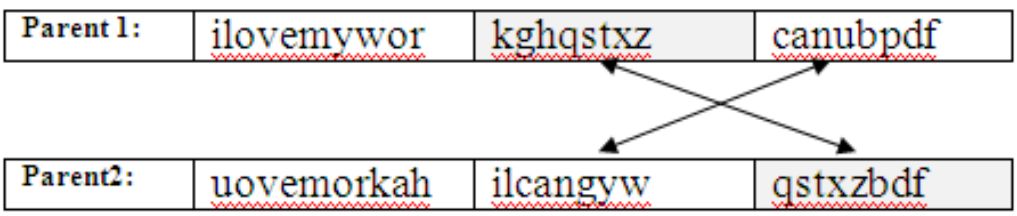

Fig.(3) : reproduction operation (multi point reproduction)

Child1:	I	L	O	V	M	Y	W	O	R	Q	S	T	X	Z	B	D	F	c	a	n	G	K	p	U	E

Child2:	U	O	V	E	M	W	R	K	A	H	Q	S	T	X	Z	B	D	F	G	S	C	I	L	N	P

Fig. (4) : mutation rate

5. Results

5.1 Breaking cipher text without knowing the key.

5.1.1 The first method(classical method):

Figure.(2) shows the frequency of most di-grams letters in the English language for 10000 letters. Comparison between figure.(2), and frequency of di-grams letters in cipher text. Trying to guess the plain text more time until the decipher text could be understand. As shown in decryption operation in section 2.2 and figure.(2).

5.1.2 The second method(MA):

Using a memetic algorithm to break a playfair cipher

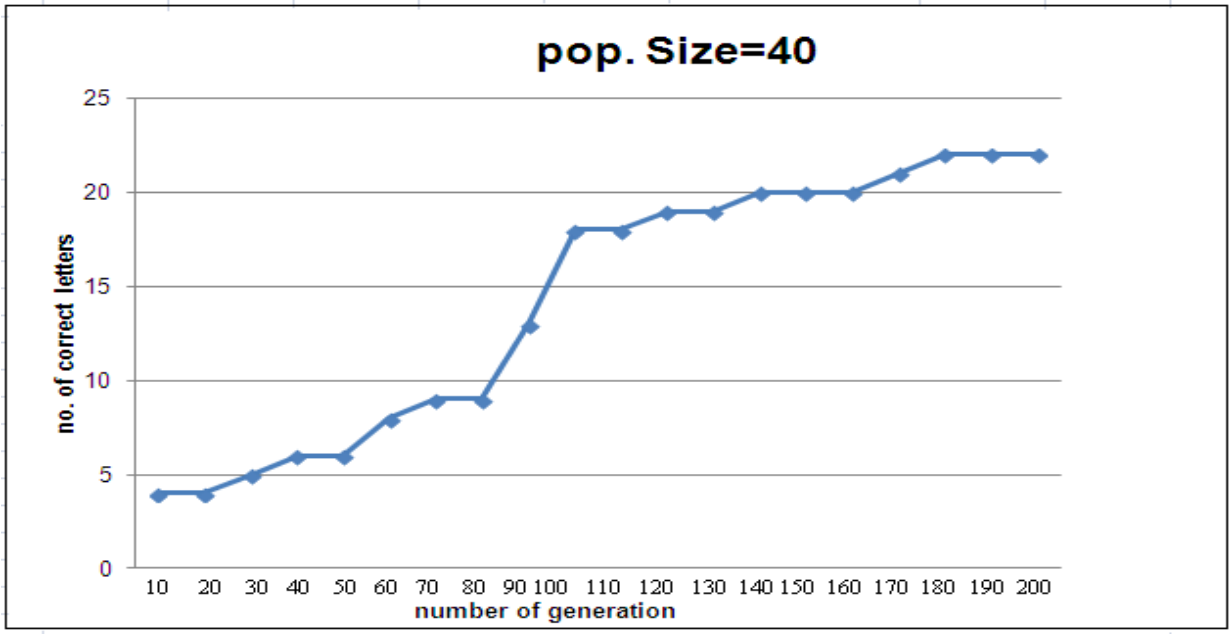

Fig. (5) : the relation between number of generation and no. of correct letter, for population 40, and mutation rate 0.2 with MA

Figure. (5) clears that the number of correct letters was 22 out of 25, when MA is used. The population size 40, mutation rate 0.2 , and cross over 0.4. the optimal solution was 22 out of 25. But when a classical method is used the number of correct letter is not appear. Because the length of text 1802. The classical method depending on frequency letters for English language for 10000 letters, in this paper the length of text 1802.

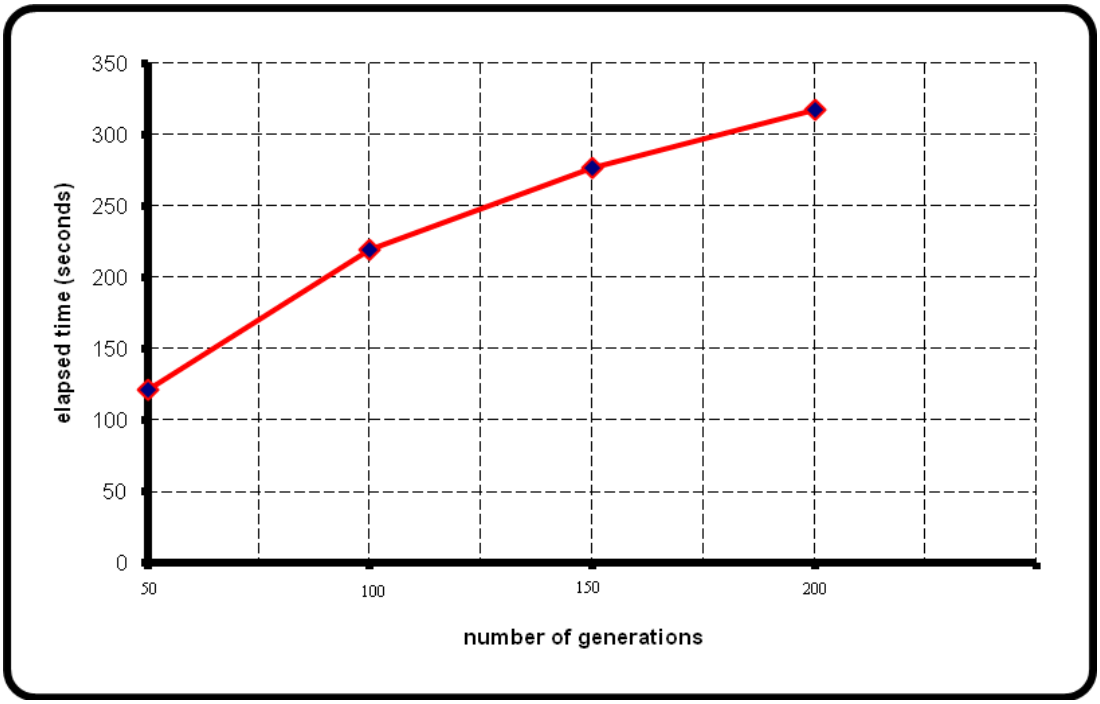

Fig. (6) : the elapsed time with number of generation

Figure.(6) clears that the number of generation and elapsed time to find the best solution. It clears that the elapsed time increases when the number of generation increases.

6. Comparing results

6.1 Using Memetic Algorithms and without used it

In 2008 G. Poonam presents a comparison of memetic & tabu search for the cryptanalysis of simplified data encryption standard algorithm. The methods were tested and various experimental results indicates that the proposed memetic algorithm is able to produce high quality solutions quickly and it also demonstrate that memetic algorithm performs better than the genetic algorithm for such type of NP-Hard combinatorial problem.[17]

In this paper, a cryptanalyst on playfair cipher using MA is presented. The methods were tested and various experimental results indicates that the proposed memetic algorithm is able to produce high speed solutions quickly and it also demonstrate that memetic algorithm performs better than the classical method that depend on frequency letter in English language.

Key search	MA	Classical method
Original key	After number of generation 200	Depending on frequency of letters in English language
ilovemyworkcanuabdfghqstxz	ilovemyworkcandabufghqztxs	ianualovemyorcbdfghqwkstxz:

7. CONCLUSIONS

1- In this paper a cryptanalysis on a playfair cipher, was implemented successfully. The algorithm was implemented using the MATLAB program.

2- Two methods are used as a classical and a certain method to find a key. It is apparent from the results that the number of correct letters has been discovered using MA.

3- The population size is 40, cross over rate 0.4, mutation rate 0.2, and number of generation 200.

4- MA faster than without using it.

5- The computer is a Pentium 4, processor is 2.7 GB, and RAM is 2 GB.

8. References

1. **Charles P. Pfleeger, "*Security In Computing*", Prentice Hall,2006.**
2. **Oppliger R. "*Contemporary Cryptography*", computer security series, ARTECH HOUSE, INC 2005.**
3. **Henk,G.A., *Fundamentals of cryptology, A professional References and interactive tutorial,* Kluwer Academic Publishers, New York, London, Moscow, 2002.**
4. **Wembo Mao., Modern *Cryptography: Theory And Practice*, prentice Hall , July 2003.**
5. **Goldreich. O, "*Foundations of Cryptography- Basic Tools*" 1st edition, Oded Goldreich, Cambridge, UK, 2004.**
6. **Jan Pelzl,C.P., *"Understanding Cryptography"*, Springer, 2010.**
7. **Jason Brownlee , *"Clever Algorithms, Nature-Inspired Programming Recipes"* , 1st edition, January 2011**
8. **Thomas, W., *"Global Optimization Algorithms, – Theory and Application –"* , 4th edition January 2008**
9. **Edmund k, B & Graham , K ., *"search methodologies, Introductory Tutorials in Optimization and Decision Support Techniques"*, Springer., 2005.**
10. **Packirisamy Murali & Gandhidss Senthikumar , *Modified Version of Playfair ipher using Linear Feedback Shift Register*. IEEE, International Conference on Information Management and Engineering,, p488-490, 2009.**

11. **Shiv Shakti Srivastava & Nitin Gupta , *Security aspects of the Extended Playfair cipher*. IEEE International Conference on Communication Systems and Network Technologies, p144-147,2011.**
12. **Trappe, W. and L. Washington, *"Introduction to Cryptography with Coding Theory"*, Pearson International 2nd Edition , India: Pearson Practice Hall, 2006.**
13. **Bauer, F.L., *"Decrypted Secrets, Methods and Maxims of Cryptology"*, 4th Edition, Springer 2007.**
14. **Churchhouse, R.F., *"Code and Cipher, Julius Caesar, the Enigma and the internet"*, Cambridge University Press, 2004.**
15. **Darel w. Hardy., Fred Richman., Carol l. Walk.er, *Applied Algebra , Codes, Ciphers, And Discrete Algorithms*, 2nd Edition, Hall CRC Press, by Taylor & Francis Group, LLC, 2009.**
16. **Stallings, W., "*Cryptography And Network Security, Principle And Practices*", 3rd Edition, Pearson Education, 2005.**
17. **Poonam. G, "*A comparison of memetic & tabu search for the cryptanalysis of simplified data encryption standard algorithm*", International Journal of Network Security & Its Applications (IJNSA), Vol.1, No 1,pp34-42,**