A Comparison between Single and Multi- Crossover Points to Break Hill Cipher Using Heuristic Search: MA & GA

Dalal A. Hammood
Computer Engineerin Department / College of Electrical & Electronic Techniques
Foundation of Technical Education/Baghdad
Email:ALsaady_Dalal@yahoo.com

Received on:22/12/2011 & Accepted on:6/12/2012

ABSTRACT

Hill cipher is a classical cipher which is based on linear algebra. In this method, matrices and matrix multiplication have been used to combine the plaintext.

Heuristic search is a search techniques. The methods of HS are: (GA, SE, EP, MA, TS). Genetic algorithms are one of Heuristic search, it is search techniques which is used natural selection. GAs select optimal solution through three operations, they are : selection, crossover and mutation. The parameters are kept in memory and the best values of fitness have been selected to represent next generation.

Memetic Algorithm is one of Heuristic search , a memetic algorithm is an extension of the traditional genetic algorithm. It uses a local search technique to reduce the likelihood of the convergence, to reach the best solution.

This paper focuses on using MA and GA to find optimal solution to cryptanalyse Hill cipher. Then comparing two methods of crossover to see which one has best solution, and comparing between GA and MA to see which one has best solution.

MATLAB is used as M-FILE.Theresults ofcryptanalysis cleared as following:-

1- Without genetic algorithms: The number of correct letters for the key was 1 out of 9.
2- Using genetic algorithms: two methods are used, and they have been compared of crossover, they are single and multi- crossover points randomly. After (250) generation, the number of correct letters was 4 out of 9 when single crossover point is used. The number of correct letters was 8 out of 9 when multi crossover point are used. So multi crossover point have best solution. Genetic algorithms are applied successfully.
3- Using Memetic Algorithms. After (100) generation, the number of correct letters was 8 out of 9. So MA is better than Genetic algorithms.
4- the number of correct letter was 9 out of 9 when the MA is used.

Keywords: Hill cipher, Memetic Algorithm, Genetic Algorithms, Cryptanalyse, Key Search, single & multi crossover point.

مقارنة بين نقاط التزاوج المفردة والمتعددة لتحليل شفرة Hill باستخدام خوارزميات البحث العشوائي: الميميائية والجينية

الخلاصة

شفرة هيل هي شفرة تقليدية (كلاسيكية) حيث تستند على الجبر الخطي. في هذه الطريقة، تم استخدام ضرب المصفوفات لتكوين النص المشفر.

البحث الاستدلائي هي طريقة بحث. طرق البحث الاستلائي هي : (الخوارزميات الجينية، خوارزميات الانصهار والتبريد، البرمجة التطويرية، الخوارزمية الميميائية، بحث التابو). الخوارزميات الجينية هي احدى طرق البحث الاستدلائي، هي تقنيات بحث حيث تستخدم الانتقاء الطبيعي. الخوارزميات الجينية تختار الحل الامثل من خلال ثلاث عمليات : الانتقاء (الاختيار) ، التزاوج والطفرة. الباراميترات تحفظ في الذاكرة ويتم اختيار افضل قيمة للفتنس لتمثل بالجيل القادم.

خوارزمية MA هي طرق البحث الاستدلائي، وهي امتداد للخوارزمية الجينية التقليدية. ويستخدم تقنية البحث المحلي إلتى يقلل من احتمالات التقارب للوصول الى الحل الامثل.

يركز هذا البحث على استخدام الخوارزميات الجينية وخوارزمية memetic لايجاد الحل الامثل لتحليل شفرة Hill. ثم مقارنة طريقتين لعملية التزاوج لرؤية ايهما يمتلك الحل الافضل. ومقارنة بين خوارزمية GA وMA وملاحظة اي منهما لها افضل الحلول.

استخدم الماتلاب كـ M-File، اوضحت النتائج مايلي:

1- بدون استخدام الخوارزميات الجينية: عدد الحروف الصحيحة للمفتاح كان 1 من اصل 9 حروف.
2- باستخدام الخوارزميات الجينية: استخدمت طريقتين وقورنت نقاط التزاوج، وهي النقطة المفردة للتزاوج والنقاط المتعددة (نقطتي تزاوج)عشوائياً . بعد 250 جيل عدد الحروف الصحيحة كان 4 من اصل 9 عند استخدام النقطة المفردة للتزاوج. عدد الحروف الصحيحة كان 8 من اصل 9 عندما استخدم النقاط المتعددة. لذا الحل الافضل عند استخدام التزاوج المتعدد. طبقت الخوارزميات الجينية بنجاح.
3- باستخدام Memetic Algorithm: استخدمت طريقة النقاط المزدوجة، بعد 100 جيل عدد الحروف الصحيحة كان 8 من اصل 9. وقورنت مع الخوارزمية الجينية حيث اوضحت النتائج ان الخوارزمية MA اعطت نفس النتائج لعدد من الاجيال 100 جيل بينما اعطت الخوارزمية الجينية بعد 250 جيل.
4- كان عدد الحروف الصحيحة 9 من اصل 9 حروف بعد 150 جيل عندما استخدمت خوارزمية MA.

الكلمات المرشدة: شفرة الهيل، الخوارزمية الميميائية، الخوارزمية الجينية، تحليل الشفرة، المفتاح، نقاط التزاوج المفردة والمزدوجة.

INTRODUCTION

Cryptography means hidden and *secrecy in writing*. The aim of cryptography to render a message incomprehensible to an unauthorized reader[1].

Cryptanalysis a set of rules which are in turn intended to make that unauthorized decryption more difficult (encryption security)[1]. Plaintext is the message that will be encrypted[2].

There are many papers that have been published about hill cipher, but these papers have not used genetic algorithms.

In 2007 A.Mahapatra & R. Dash presented data encryption and decryption by Using hill cipher technique and self Repetitive matrix[3].

In 2008 B. Acharya, D. Jena, S. Kumar Patra , & G. Panda, presented Invertible, Involutory and Permutation Matrix Generation Methods for Hill Cipher System[4].

In 2009 M. Toorani, & A. Falahati presented A Secure Variant of the Hill Cipher, that overcomes all the security drawbacks of the Hill cipher. The proposed scheme includes an encryption algorithm that is a variant of the Affine Hill cipher for which asecure cryptographic protocol is introduced[5].

In this paper two methods of crossover are used, and then comparing with them to see which one has the best solution. The two methods are: single crossover point, and multi crossover point.

Two algorithms are used Genetic and Memetic Algorthim, and comparing between them.

HILL CIPHER

Hill cipher is a classical cipher which is based on linear algebra. In this method, matrices and matrix multiplication have been used to combine the plaintext.

This cipher was invented in 1929 by Laster S. Hill . Let *m* be a positive integer. The idea is to take *m* linear combinations of the m alphabetic characters in one plaintext element, thus producing the m alphabetic characters in one ciphertext element.

For example, if *m*=3, the plaintext element is written as $x=(x_1, x_2, x_3)$ and a ciphertext element as $y=(y_1, y_2, y_3)$. Here, y_1 would be a linear combination of x_1, x_2, and x_3, as would y_2 and y_3. as shown below[1,2,6-8]:

$$(y_1, y_2, y_3) = (x_1, x_2, x_3).\begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix}$$

Where $x_1 \ldots x_m$ are plaintext, $k=(k_{ij})$ is a key to encrypt or decrypt text, and $y_1 \ldots y_m$ are ciphertext. The matrix must be square (m*m).

The plaintext is X, which X is :

algorithmsarequitegeneraldefinitionsofarithmeticprocesses

The Key cipher is r r f v s v c c t

Table 1 shows the number of each alphabetic letters[8-10].
Now, to encrypt plaintext, the key is following:

$$k = \begin{pmatrix} 17 & 17 & 5 \\ 21 & 18 & 21 \\ 2 & 2 & 19 \end{pmatrix}$$

Plaintext= x =(algorithmsarequitegeneral definitionsofarit hmeticprocesses).
Equation (1) represents the encryption process.

$$E=k \cdot x \mod 26 \qquad \text{.....(1)}$$

Where k is a key, x is a plaintext, and E is a ciphertext.
The message is broken into blocks, each block must be 3 letters. and then multiplying each block with key[2].
alg=k.(0 11 6) mod 26 =J , rit=k.(17 8 19) mod 26 =M, and so on.
Then , the ciphertext is

JMGVOGIXUBHVYMELWABDHTAQYFAACNFCYDWGLUMAQVFFK BNOQFSKYQWWW

To decrypt ciphertext, equation (2) shows the decryption process.

$$D=X^{-1} \cdot K \qquad \text{.....(2)}$$

Where k is a key , x is the inverse ciphertext, and D is a decipher.

GENETIC ALGORITHMS

The Genetic Algorithm (GA) is a stochastic optimization strategy. It derives its behavior from a nature selection, and works by the creation a population of individuals represented by chromosomes. The individuals in the population then go through a process of evolution"[11].

Figure (1) shows the structure of a simple genetic algorithm. Genetic algorithms work on populations of individuals instead of single solutions. In this way the search is performed in a parallel manner.

At the beginning of the computation a number of individuals (the population) are randomly initialized. The objective function is then evaluated for these individuals. The first/initial generation is produced.

If the optimization criteria are not met with the creation of a new generation starts. Individuals are selected according to their fitness for the production of offspring.

Parents are recombined to produce offspring. All offsprings will be mutated with a certain probability. The fitness of the offspring is then computed. The offsprings are inserted into the population replacing the parents, producing a new generation.

In practice, the genetic model of computation can be implemented by having arrays of bits or characters to represent the chromosomes. Simple bit manipulation operations allow the implementation of selection, crossover, and mutation[11,12].

When the genetic algorithm is implemented it is usually done in a manner that involves the following cycle: Evaluate the fitness of all of the individuals in the population**.** Create a new population by performing operations such as crossover, fitness-proportionate reproduction and mutation on the individuals whose fitness has just been measured. Discard the old population and iterate using the new population[11].

One iteration of this loop is referred to as a generation. There is no theoretical reason for this as an implementation model. Indeed, behavior in populations in nature is not found as a whole, but it is a convenient implementation model[10-12].

The first generation (generation 0) of this process operates on a population of randomly generated individuals. From there on, the genetic operations, in concert with the fitness measure, operate to improve the population[9-13].

PROPOSED ALGORITHM OF GA

t=0;
initialize population P(t);
evaluate P(t);
until (done)
t=t+1;
parent selection P(t);
crossover P(t)

1- single point crossover
2- multi point crossover

mutate P(t);
swapping between position randomly
evaluate P(t);
survive P(t);

MEMETIC ALGORITHM

The genetic algorithm is a global search, this is not well suited to find optimal solution[14].

The memetic algorithms can be viewed as a marriage between a population-based global technique and a local search made by each of the individuals. They are a special kind of genetic algorithms with a local hill climbing. Like genetic algorithms, memetic Algorithms are a population-based approach. They have shown that they are orders of magnitude faster than traditional genetic Algorithms for some problem domains[15].In a memetic algorithm the population is initialized at random or using a heuristic. Then, each individual makes local search to improve its fitness.To form a new population for the next generation, higher quality individuals are selected. The selection phase is identical inform to that used in the classical genetic algorithm

selection phase. Once two parents have been selected, their chromosomes are combined and the classical operators of crossover are applied to generate new individuals. The latter are enhanced using a local search technique. The role of local search in memetic algorithms is to locate the local optimum more efficiently then the genetic algorithm[15].

THE ALGORITHM OF MA[16]

t=0;
set of population P(t), max_gen, gen=0;
cross_rate=0.2,mut_rate=0.2;
initialize population
evaluate P(t);
until (done)
t=t+1;
parent selection P(t);
crossover P(t)
multi point crossover
mutate P(t);
swapping between position randomly
evaluate P(t);
apply local search
survive P(t);
applying final local search to best offspring

FITNESS MEASURE

The technique used to compare candidate keys is to compare uni-gram statistics of the decrypted message with those of the language (which are assumed known)[13].

$$Fitness = a\left(1-\sum_{i=1}^{26}\left\{\left|SF[i]-DF[i]\right|\right\}/4\right)^{8}+$$

$$b\left(1-\sum_{i=1}^{26}\left\{\left|SDF[i][j]-DDF[i][j]\right|\right\}/4\right)^{8}\ \mathbf{L}\ (3)$$

where SF and SDF denote the relative frequencies of single characters and digrams in the English language (respectively), and DF and DDF denote the relative frequencies of single characters and digrams in the message decrypted using *key*. Varying α and β allows the weighing in favour of either the single character frequencies of the di-gram frequencies[13].

USING GENETIC ALGORITHMS TO ATTACK A HILL CIPHER

In the first phase, an initial population, describing representatives of the potential solution, is created to initiate the search process. The elements of the population are encoded into bit-strings, called chromosomes[17] .

Another process is a selection. It is supposed to be able to compare each individual (chromosomes) in the population. Selection is done by using a fitness function.

After selection, two methods have been used to apply crossover. They are single point and two points crossover, to see which else has optimal solution.

1-Single point crossover

The traditional genetic algorithm uses single point crossover, where the two mating chromosomes are cut once at corresponding points and the sections after the cuts exchanged. Here, a cross-site or crossover point is selected randomly along the length of the mated strings and bits next to the cross-sites are exchanged. If appropriate site is chosen, better children can be obtained by combining good parents else it severely hampers string quality[12,17-19].

2- Two point crossover

In two-point crossover, two crossover points are chosen and the contents between these points are exchanged between two mated parents, as shown in Figure (2) [12,17-19].

After crossover, some keys are subjected to **mutation.** Mutation prevents the algorithm to be trapped in a local minimum. The mutation operation is used in this cipher. randomly select two elements in the child and swap those elements as shows in Figure (3) [12,17-19].

USING MEMETIC ALGORITHMS TO ATTACK A HILL CIPHER

For the memetic algorithm, the population size was set to 20; the probabilities for crossover and mutation were both 0.2 for all the test problems because it was the best configuration found empirically for the memetic algorithm. Stochastic selection is used.multi crossover is used to reproducing child. A marriage between a population-based global and local search is used to reproducing a new generation[16].

RESULTS

The attack to hill cipher was implemented for population size 20, different numbers of generations and mutation rate 0.2.

1- Using Genetic Algorithm.

Figure (4) shows the single point crossover process. It shows the relation between the number of correct letters and different number of generations for population 20. It is clear that the number of correct letters is 4 out of 9 after 250 generation when the population size 20. This solution is not enough.

Figure (5) shows the two point crossover process. It shows the relation between number of correct letters and different number of generations for population 20. It is

clear that the number of correct letters is 8 out of 9 after 250 generation when the population size 20.

Figure (6) shows the elapsed time between the number of correct letters and different number of generations for population 20. It is clear that the elapsed time is increased when the number of generations increase for two point crossover.

2- Using Memetic Algorithm

Figure 7 shows the two point crossover process. It shows the relation between number of correct letters and different number of generations for population 20. It is clear that the number of correct letters is 8 out of 9 after 100 generation when the population size 20, and 9 out of 9 after 150 generation.

Figure (8) shows a comparistion between GA and MA. It is clear that the MA is better than GA to find a good sloution.

COMPARISION RESULTS

The true key is:

R R F V S V C C T

Table (2) shows decryption ciphertext without genetic algorithm. It is clear that the number of correct letter is 1 out off 9.

Table (3) shows decryption ciphertext with genetic algorithms using two point crossover. It is clear that the number of correct letter is 8 out off 9.

Table (4) shows decryption ciphertext with memetic algorithm using two point crossover. It is clear that the number of correct letter is 9 out off 9.

The best key is: $\begin{Bmatrix} R & U & F \\ V & S & V \\ C & C & T \end{Bmatrix}$

CONCLUSIONS

In this paper memetic and genetic algorithms, were implemented successfully to break a hill cipher.

This work has used single letters frequency. Using different number of generations, and population 20 . the text length is 59 letters and mutation rate is 0.2. key size is 9 letters.

The algorithms were implemented using the MATLAB program. Different parameters were tested such as the number of population and the time required finishing the algorithm for different number of generations.

Two algorithms are used to see which one has a best key.

Using Genetic Akgorithms: The best number of correct letters was 4 out of 9 letters after 250. generation for single point crossover. When Multi points crossover is used, the number of correct letters was 8 out of 9, after 250 generation.

Using Memetic Algorithm: The best number of correct letters was 9 out of 9 letters after 100 generation for multi (two) point crossover which represent the best solution.

So memetic algorithms gives best solution.

The algorithms were run in MATLAB for the processor 2.70 GHz and RAM 512MB.

REFERENCES

[1].Bauer, F.L., "Decrypted Secrets, Methods and Maxims of Cryptology", 4th Edition, Springer 2007.

[2].Erdogmus, P., A. Ozturk, and S. Tosun, Continious Optimization Problem Solution with Simulated Annealing and Genetic Algorithm. 5th International Advanced Technologies Symposium (IATS, 09) 2009. May: p. 13-15.

[3]. Mahapatra & R. A. "Dash presented data encryption and decryption by Using hill cipher technique and self Repetitive matrix"., Bachelor of Technology Thesis in Electronics & Instrumentation Engineering, Department of Electronics & Instrumentation Engineering National Institute of Technology in Rourkela, 2007.

[4]. Acharya, B. D. Jena, S. Kumar Patra , & G. Panda, "Invertible, Involutory and Permutation Matrix Generation Methods for Hill Cipher System ", International Conference on Advanced Computer Control, IEEE, p410-414, 2008.

[5]. Toorani,M. and A. Falahati "A Secure Variant of the Hill Cipher", IEEE, p313-316,2009

[6].Morelli, R., R. Walde, and W. Servos, A Study Of Heuristic Approaches for Breaking Short Cryptograms. International Journal on Artificial Intelligence Tools 2004. **13**(1): p. 45-64.

[7].Schneier, B., Applied Cryptography: protocol, algorithms and source code in C 2nd Edition 1996, New York: Jhon Wiley & Sons, Inc.

[8].Stinson, D.R, Cryptography, Theory and Practice. 3rd ed. 2006. Chapman&Hall/CRC, Taylor & Francis Group.

[9].Stallings, W., "Cryptography And Network Security, Principle And Practices", 3rd Edition, Pearson Education, 2005.

[10].Menezes, A., P.V. Oorschot, and S.Vanstone, "Handbook of Applied Cryptography",1996, Boca Raton: CRC Press, 1996.

[11].Sumathi,S , Hamsapriya, T & P. Surekha, Evolutionary Intelligence, An Introduction to Theory and Applications with Matlab, 2008 Springer.

[12].Genetic Algorithms and Direct Search Toolbox™ 2, User's Guide , MATLAB. 2009.

[13].Spillman, R., et al., Use of A genetic Algorithm in the Cryptanalysis of simple substitution ciphers. Cryptologia, 1993. **17**(1): p. 31-44.

[14].Goldberg, D.E., "Genetic Algorithms in Search, Optimization and Machine Learning", Addison Wesley, Reading, 1989.

[15]. Moscato,P, On evolution, scorch, optimization, genetic algorithms and martial art toward memetic algorithms", Technical report, California,1989.

[16].Garg. P , A Comparison Betweem Memetic Algorithm and Genetic Algorithm for the cryptanalysis of Simplified Data Encryption Standard Algorithm, IJNSA, April 2009,1(1), p.34-42.

[17].Sivanandam, S.N. and S.N. Deepa, Introduction to Genetic Algorithms. 2008: Springer.

[18].Haupt, R.L. and S.E. Haupt, "Practical genetic algorithms", 2nd Edition, John wiley & Sons, INC, 2004.
[19].Melanie, M., "An Introduction to Genetic Algorithms". 5th Edition, Cambridge, Massachusetts • London, England: The MIT Press, 1999.

Table (1) numbers of each alphabetic letters.

a	b	c	d	E	F	g	h	i	j	k	l	m
0	1	2	3	4	5	6	7	8	9	10	11	12
n	o	p	q	R	S	t	u	v	w	x	y	z
13	14	15	16	17	18	19	20	21	22	23	24	25

Table (2) Number of correct letters Without using Gas.

population	Keys	Number of correct letters
1	HIEFBCDGA	0
2	BFAHGIEDC	0
3	BCEHIFGDA	0
4	GCEDFAHIB	0
5	CGBFEDHIA	0
6	AECBFGHDI	0
7	FDHAEGBCI	0
8	HEGFIACBD	0
9	CDEBFHIGA	0
10	DB**F**EGAHIC	1
11	HH**F**BRDSSB	1
12	FFRRBDTUR	0
13	UGGYTREWY	0
14	FHHHDTTRR	0
15	VT**F**HRETTT	1
16	BAEWTRETO	0
17	HTREWQIUY	0
18	**R**EERIITURE	1
19	KJUYTTEWQ	0
20	ACTRVAAU**T**	1

Table (3)Number of correct letters after 250 generation for two point crossover using GA.

population	Keys	Number of correct letters
1	HH**F**BRDSSB	1
2	TUTRV**VCCT**	4
3	ACTR**V**VFU**T**	2
4	HEGFIACBD	1
5	F**R**RHDTFHA	1
6	**RUFVSVCCT**	8
7	HH**F**BRDSSB	1
8	**RRF**FI**VCCT**	7
9	**R**EER IITURE	1
10	**RRF**FIA**CCT**	6
11	KJUYTTEWQ	0
12	**RF**FHGI**VC**C	4
13	AECBFGHDI	0
14	**RRF**FI**VCC**A	6
15	UGGYTREWY	0
16	**RR**RTVV**CCT**	5
17	HTREWQIUY	0
18	**RRF**T**V**C**VCT**	7
19	**RRF**FI**VCCT**	7
20	GCEDFAHIB	0

Table (4) Number of correct letters after 100 generation for twopoint crossover Using MA.

Population	Keys	Number of correct letter
1	**R**UTRS**VCC**A	4
2	**RRF**FI**VCCT**	7
3	**RUFVSVCCT**	8
4	**RRF**FI**VCCT**	7
5	**RRF**FIA**CCT**	6
6	**RF**FHGI**VC**C	4
7	**RRF**FI**VCC**A	6
8	**RR**RTVV**CCT**	5
9	**RRF**TVC**VCT**	7
10	**RRF**FI**VCCT**	7
11	**RRF**FI**VCC**A	7
12	**RRFVSVCCT**	9
13	**RRF**FI**VCCT**	7
14	**RR**RTVV**CCT**	5
15	**RRFVSVCC**U	8
16	**RRFVSVCC**U	8
17	**RRF**FI**VCC**A	6
18	**RRF**FI**VCCT**	7
19	**RRF**FI**VCC**A	6
20	**RRFVSVCCT**	9

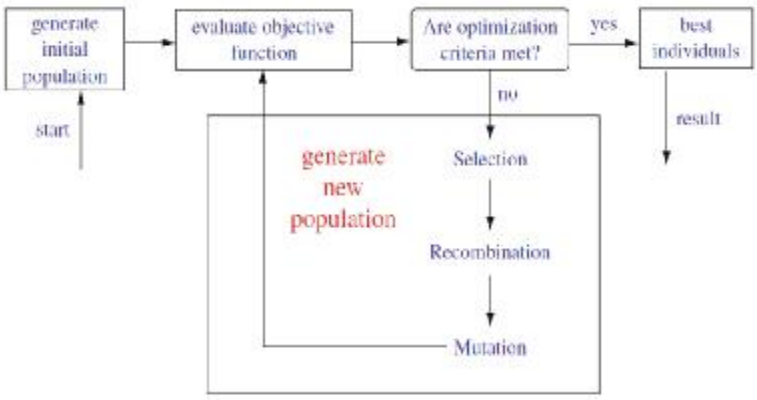

Figure (1) Cycle Of Gas.

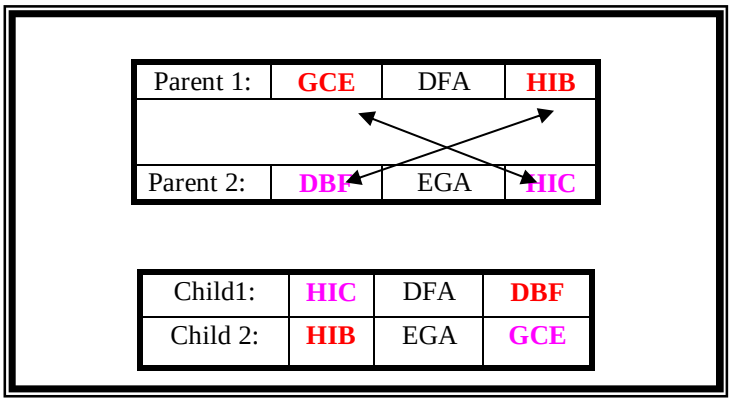

Figure (2) Applying Crossover.

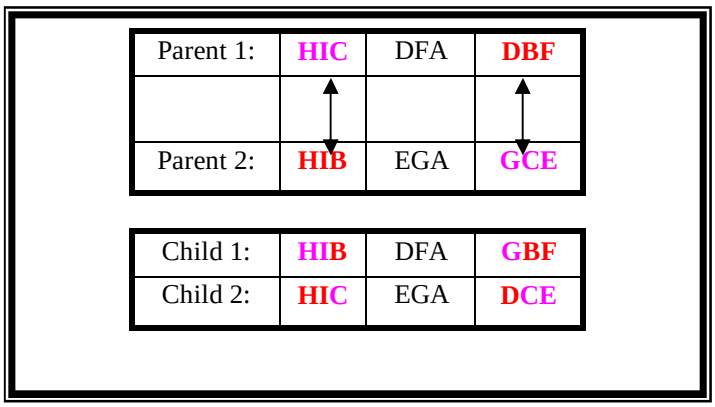

Figure (3) Applying Mutation.

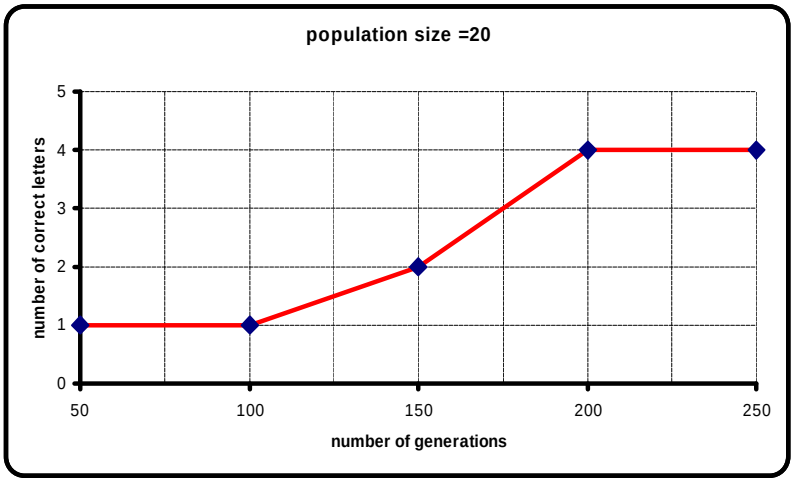

Figure (4) Single point crossover, pop. 20, No.of correct letters & diffrenet No. of generations.

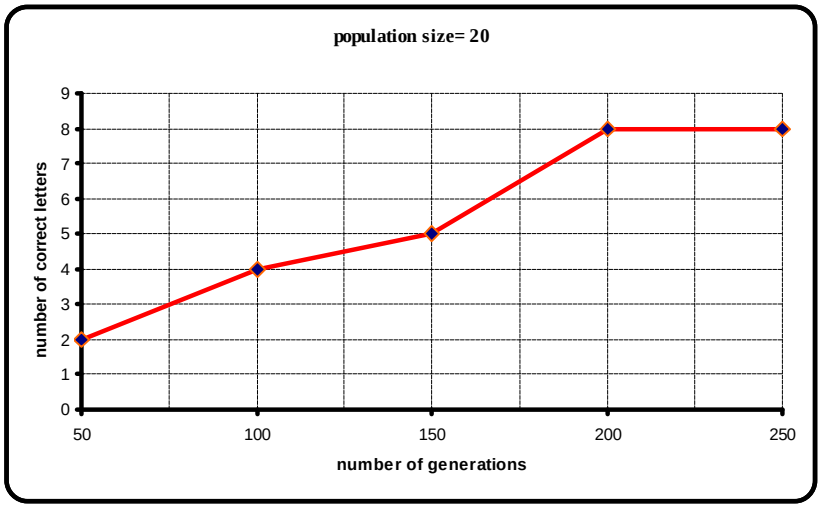

Figure (5) Two point crossover, pop. 20, no. of correct letters & diffrenet no. of generations.

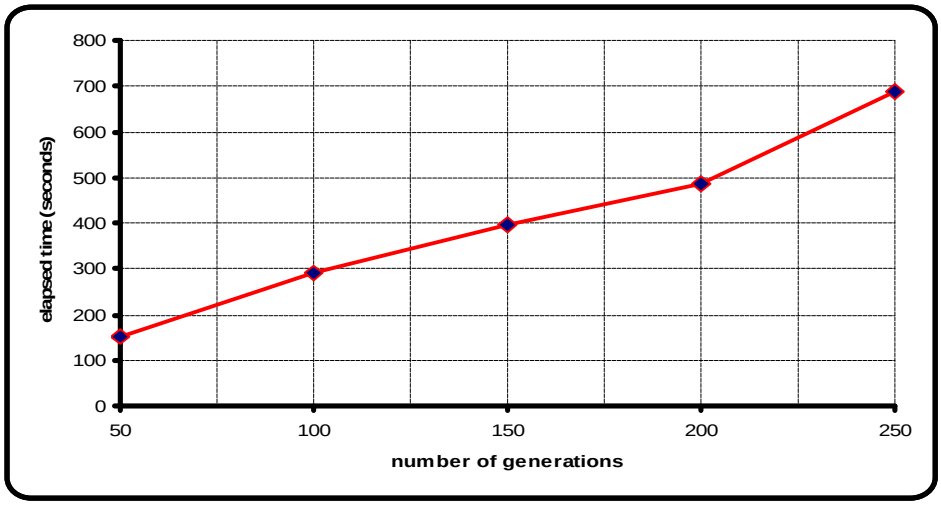

Figure (6) Elapsed time between no. of correct letters & different no. of generation.

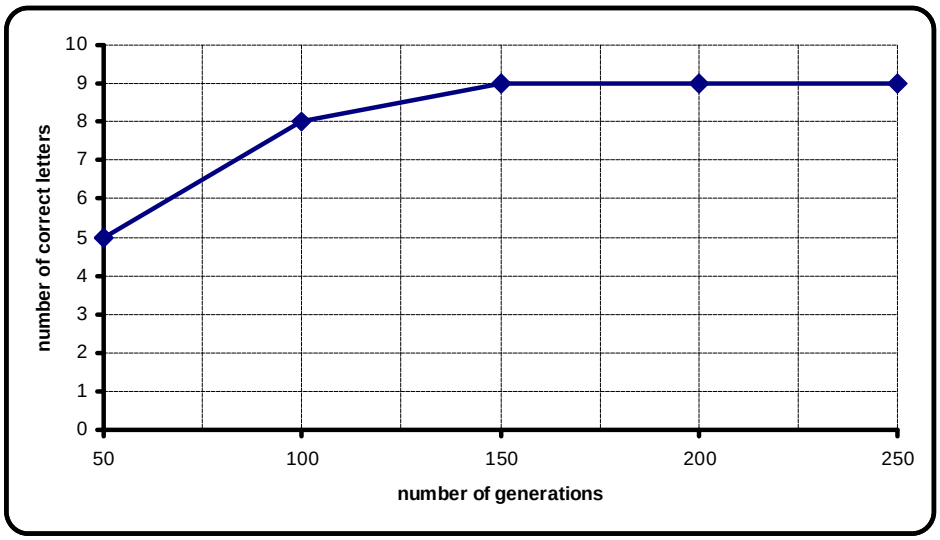

Figure (7) Two point crossover, population 20, number of correct letters and diffrenet number of generations using MA.

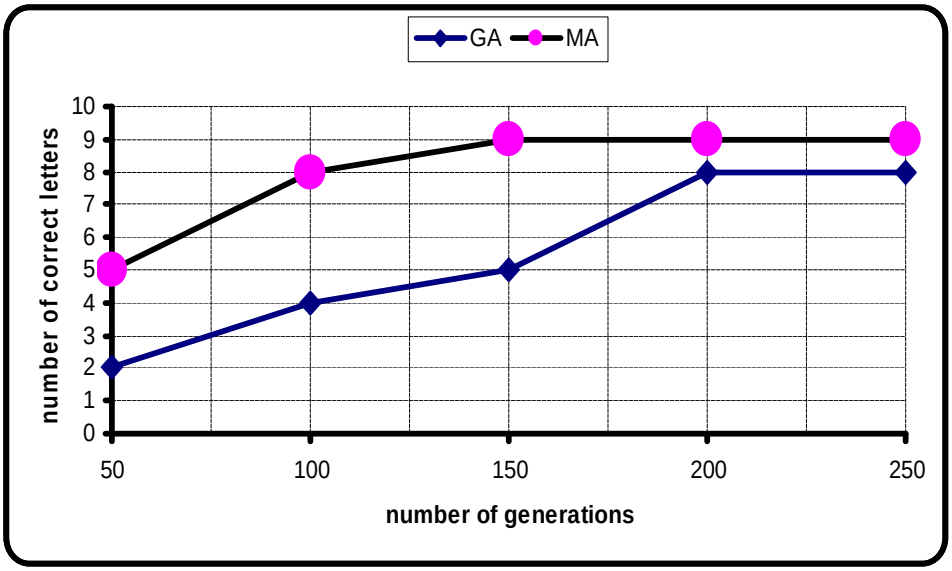

Figure (8) A comparistion between GA and MA.