Cryptanalysis of ADFGX using Genetic Algorithm

Ali.S.Alkhalid, and Alaa.O.Alkhfagi

***Abstract*—** A Genetic algorithm is a search tool that's used to insure high probability of finding a solution by decreasing the amount of time in key space searching. In this paper the focus will be on the cryptanalysis of a ADFGX cipher by using genetic algorithm and study the algorithm factors that affect finding solution taking in consideration the time and efficiency of the cryptanalysis process.

***Keywords*—** substitution cipher, transposition cipher, genetic algorithm, ADFGX cipher.

I. Introduction

CRYPTOGRAPHY means secrecy in writing. The aim of cryptography is to render a message Mysterious to an unauthorized reader (1). While the cryptanalysis is the attempt to break that cryptosystem to test its strength and study its weaknesses. There are different types of encryption technique can be used to encrypt text, But this paper will be focused on the polyalphabetic substitution cipher, taking the Hill cipher as case study, then attempts to break this cipher using genetic algorithm which is a search tool that's usable to insure high probability of finding a solution by decreasing the amount of time in key space searching.

In March 1918, the Germans implemented a now famous cipher - the ADFGX cipher - to encrypt communications between their corps and division level field headquarters. This new cipher used only the letters A, D, F, G and X and resisted all attempts at cryptanalysis. This cipher contributed to the tactical 206 surprise achieved by the Germans during the initial spring offensive. A French cryptanalyst, Georges Panvin, attacked the new cipher using frequency analysis. By the end of April he was able to read some of the messages protected by this cipher, but he never achieved a general solution. On June 1, Panvin noticed that the Germans had slightly changed the cipher by including the letter V. Panvin used frequency analysis once more and was able to decipher some messages protected by this new cipher by June 2. He disseminated the key he uncovered to the other French cryptanalysts. On June 3, one of the French cryptanalysts, using the key provided by Panvin, deciphered an intercepted German message that provided the first hint as to the location and timing of the anticipated German offensive. When the new attack came, there was no surprise and the Germans were not as successful as in their previous attempts. Georges Panvin and the method of frequency analysis helped save France. [2]

Central Technical University, College of Elec. & Electronic Techniques, (Baghdad-IRAQ)

II. The Genetic Algorithm

Genetic Algorithms (GA) deal with the science of computational methods over the mechanisms of natural selection and genetics [4, 5]. These computational methods increase based on the key space size and some of the algorithm parameters that's will be further studied later in this paper. Normally GAs comprised of set of population(s) which might represent one or more groups. These populations are composed of feasible solutions for any given problem to solve. The evolution of such population is based on biological metaphors when are subjected to probabilistic operators. GAs show least processing time compared to other methods.

The application of genetic algorithm to cryptanalysis is as follows:

2.1. Individual Description

Each individual in the population represent a guess of the optimal key so large population size will insure get to the result in less generation number but increase the time required for the single generation.

2.2. Fitness Function

The fitness function is used to rate the quality of each population which mean the nearest guess to the optimal solution. It depends on letter frequency occurrence of the cipher with corresponding to the letter frequency in English language as shown in the following equation [6].

$$\textbf{Fitness}= 1 - \sum_{i=1}^{26} |sf(i) - df(i)| + \sum_{j=1}^{26} (|sdf(i,j) - ddf(i,j)|)/4$$

SF [*i]* is the standard frequency of character *i* in English plain text, *DF* [*i*] is the measured frequency of the decoded character *i* in the cipher text, *SDF* [*i,j*] is the standard frequency for a diagram and *DDF*[*i,j*] is the measured frequency for the diagram. When the measured and standard frequencies are the same, the summation term equals to zero, making the fitness value equal to one. Larger values of fitness function represent smaller errors. Sensitivity to small values of the difference is increased by raising the result to 8th power, while sensitivity to large values of the difference is decreased by the division of the summation terms by 4. The best keys at the last generation are used to decipher the cipher text.

2.3. Crossover and Mutation

2.3.1. Crossover: This operator randomly chooses a locus and exchanges the subsequences before and after that locus between two chromosomes to create two offspring. For example, the strings 10000100 and 11111111 could be crossed over after the third locus in each to produce the two offspring 10011111 and 11100100. The crossover operator roughly mimics biological recombination between two single−chromosome (haploid) organisms.

2.3.2. Mutation: This operator randomly flips some of the bits in a chromosome. For example, the string 00000100 might be mutated in its second position to yield 01000100. Mutation can occur at each bit position in a string with some probability, usually very small (e.g., 0.001)[7].

III. ENCRYPTION DECRYPTION PROCESS

The encryption process consist of two parts substitution and transposition

A.Substitution part

This method makes use of a 5×5 table containing the letters of the alphabet, treating i and j as a single letter. However, the alphabet may be entered in any order into the table, as for example.

	A	D	F	G	X
A		b	c	d	e
D	F	g	h	i/j	k
F	L	M	n	o	p
G	Q	r	s	t	U
X	V	w	x	y	z

Each plaintext letter is replaced by the label of its row and column. For example, s becomes FG , and z becomes XX. Suppose the plaintext is "Kaiser Wilkelm" The result of this initial step is XD AA GD FG XA DG DX GD AF XD XA AF DF

B. Transposition part

In the second step, we employ a keyword, like charm, to effect a transposition (that is, a permutation) of the fractionated text: Begin by writing the fractionated text in rows under the characters of the keyword:

C	H	A	R	M
X	D	A	A	G
D	F	G	X	A
D	G	D	X	G
D	A	F	X	D
X	A	A	F	D
F				

Then, by rearranging the columns of this array by alphabetizing the letters of the keyword, we obtain

A	C	H	M	R
A	X	D	G	A
G	D	F	A	X
D	D	G	G	X
F	D	A	D	X
A	X	A	D	F
	F			

Finally, the cipher text is obtained by reading off the columns of the last array: AGDFAXDDDXFDFGAAGAGDDAXXXF

3.1. Cryptanalysis

To cryptanalysis such a cipher it's not possible to encrypt it by genetic directly because it has two stage in the encryption step (substitution and transposition) so it should be separated.

So by assuming that no transposition permutation occur in the encryption so only substitution ciphering occur. Then arrange the alphabet letter according to their frequency in descending order.

Optimal order="etoianshrludymwgfcbpkvxzq"

Then use a counter loop to find the most occurrence letter and assign it to its corresponding letter in English frequency letters order (optimal order) so as example if 'FA' is the most occurrence letter then its mean the letter "e" will assign to "FA" and so on.

Then genetic algorithm take place by its effort to change the order of the optimal order to get the best fitness value by taking in consideration diagram and bigram should be used in the fitness function.

The 1st generation should give fitness around 0.74 which mean 7 of 25 letter is correct and by swap the letter position many time by using genetic algorithm for 300-500 generation will lead to get fitness around 0.89 which mean 18 of 25 correct letter which take around 3-6 minute.

For the transposition brute force should be use and at each guess perform the previous step and calculate fitness.

Transposition key length	Key space	Time to find the transposition key	Time to find plain text in minute
1	1	None	3-6
2	2	40 second	3.6-6.6
3	6	1 minute	4-6
4	24	4 minute	7-10
5	96	20 minute	19-22
6	576	2 hours	123-126
7	4032	14 hours	843-846
8	32256	112 hours	6723-6726

From the above table it clear that when long key length used it's hard to get the plain and take very long time so this technique cannot be used for key length is exceed 7.

Plain=['thefriendshipsididntwanttodepartmyhometownihave beenheresinceiwasbornididntwanttoleavemyschoolandmostim portantlyididntwanttoleavesamthatussamandkaherineourfriend shipstartedingradeschoolfourgradeidroppedmypenciandwasgoi ngtogetitandofcoursesamleanedtogetittooourfriendshipstartedt hatdaywebecamebestfriendswewereinseparablewedideverythin gtogetherhetoldmehelovedmelikeasistersomethingstartedchang ingiwaslookingatsamdifferentlyiknewwhatthisfeelingwasbutitc rushedmeilovedsambecauseiwasjustlikeasistersoitoldhimhewa sjustlikeabrotherbutthenihadtomovetoawholeanewcountrymyi mportantjournalwithconfessionletterupupupsamhurryupkatheri nesgannaleavehereyourchanceimesseduprealbadtellingkatherin eilovedheronlylikeasisterbadideanowitstimetotellyouilovehern ownowhafounditsbeentwoweekssincesamdiedheranstraightin toatruckcallingoutmynamejusttogivemeanoteohyeahiamgladth atidroppedatpencilitletmemeetagreatpersonlikeyousamifyouca nhearmeimissyouandiloveyouwaitformedearkathguesswhatilov eyouforeverandnomatterhowforwearejustrememberyouhaveme samtheendmostofotherstoriesareaboutloveandhowpeopleendu phappymystoryisfarfromthatmynameisjoshuebutmothercallsm

eyouthisismysidinmystorybigbrotherwakeupmotheriscomingu
pherewiththestickgetupyoustupidfoolyesmomihadalwaysloved
mymotherbutshehadalwaysblamedmeformyfatherdeathsonowi
amdoingeverythingtogainbackherloveevenifitmintkillingmysel
fitgottothepointwhereiwastryingsohardthatiwouldstarvefordays
wasitryinghardenoughmotherdidntthinksowhatwasidoingwron
gsoiwalkeduptomysistersroommotherwhatwasidoingwrongwh
ydontyoulovemeanymoremotheriamsorryforthefirsttimeinalon
gtimeiactuallysmiledbutthatnightiforgottocleanuponeoftheroo
msmotherbealmeuntilicouldntistandupmotherdoyoulovemeno
whaveimadeupforeverythingihavedonewhycouldntyouoseethat
ilovedyouallalongmotherwhy ’]

The above plain with length equal to 1769

Cipher=[‘GGDFAXDAGDDGAXFFAGGFDFDGFXGFDGA
GDGAGFFGGXDAAFFGGGGFGAGAXFXAAGDGGFDX
GDFFGFDAXGGFGXDFFDGDFAAXAAXADAXAXFFDF
AXGDAXGFDGFFAFAXDGXDAAGFADFGGDFFDGAG
DGAGFFGGXDAAFFGGGGFGFAAXAAXAAXFDXGGF
AFDFFGFGFAAAFFAGFDFGGFGGDGFDFXFGGDGGA
AFFGGFAXGDGAGDGAGFFGGXDAAFFGGGGFGFAA
XAAXAAXGFAAFDGGDFAAGGGXGFGFAAFDAAFFA
GDXAADFAXGDDGFFAXFGGXGDDAGDDGAXFFAGG
FDFDGFXGFGGAAGDGGAXAGDGFFDDGDAAAGAXG
FAFDFFGFGFADAFGGXGDDDGDAAAGAXDGAGGDF
GFXFXAXAGFDXGFXAXFFAFDGAAFFAGXDAAGFDD
FGDGFFDDGGFGDDAXGGDGGGAAFFAGFGDAAFFG
GXGDGFAXGFAAFDFAAXAAFFAXAGGGFGDDAXGG
DGGGGGFGFGFGGXGDDAGDDGAXFFAGGFDFDGFX
GFGGAAGDGGAXAGGGDFAAGGAGAAXGXDAXADA
XAFAAFDAXADAXGFGGDAGDDGAXFFAGGFXDAXX
DAXGDAXDGFFGFAXFXAAGDAAADFAAXXDAXAGD
GAGAXXAAXGDXGGGDFDGFFDDGGFGDDAXGGDFA
XGDDFAXGGFGFAAGFDAXDFAXFAFGXAAXAGFDA
XFADGDXAXAAGFDGGFGGAXGDGFFGFDAXGGDFD
GFFDDGFGGAAGDGGAXAGAFDFAAFFDDDGFFDDD
GXDAAGFFAFGFGDXDGFFDDAAGGGFAAFDAGDGD
ADAAXGDAXFFGGFAXGDGDXFFAXXDXDDFAAGGG
GDFDGGFDAAXAXFADGFFDDXDAAGFADGXGGDGG
GAFGDGXGFDFAXAGFDAXDGFAFGXAAXAGGFAAF
DADAXAFAAGXGFAXDGXDAAGFDGGXGFGGFADGD
XAXAAGFDGGFGGAXGDGFFGDGGGFGFAAGDFDGF
DDFAXXDAAGFDGGXGFGGFADGDXAXAAADGDFGG
GDFAXGDADGXGGGGDFAXFFDGDFAAAGGGFGFDF
GXAAXGGFGAAXDDFFGFAAXAAFFAXXDAFFGGXFF
GGGDXGFDXGDGFDFXFGGDGGAAFFGGDGFGGXGD
FFAAFAXDDGGGDFAFFGFFDAAXGFGFDGFGFFFAAX
GGGGAXGDGXFXGXFXGXFXGFAAFDDFGXGDGDXG
GXFXDXAAGGDFAXGDDGFFAXGFDDAAFFFFAAFAA
XAAXAAXDFAXGDAXXGFGGXGDAFDFAAFFAFAXD
GFDAXGFGFAXAGGXFXGDAXAAFAADAAAGGGAXF
AFADGFFDDDXAAGGDFAXGDDGFFAXDGFAFGXAA
XAGDFAXGDFGFFFAXGFADGDXAXAAGFDGGFGGA
XGDADAAAGDGAGAXAAFFFGXDDGGGGFGGDGFDA
XGGFGGGAXFAFAXGFGGXDGFAFGXAAXDFAXGDF
FFGXDFFFGXDDFAADAFGGXFFAGDGGGDGGGGFAD
AXAXFFGGXDFGXDAXAXDXGFGFDGFFAFAXGFAAF
DAGDGAXAGDFAXGDAAFFGFGGGDAADGDDDFGGD
GFFGGFGAAGGGDGXAFDXAFAAFAFADGFFDDFGGX
GGFDXGFFAAFDAXDGGXGFGGGGFGDDDGXAAXFD
AXAAFFFGGGAXFGDFXGAXAADFDGAAFDDDFAAA
AGGGDFAAGGDGAGGDFGFXFXAXAGAAGGFXAXFF
AFDGFADGGGFAAXGGFDAXFDAXAXGGAADDGDAX
AAGGFXAXGDGFFGFFFADGDXAXXGFGGXGFAAFD
DGDAXGFGGXAFAAFFDFAXAAGDFDAXDGFDDGGF
GFXGFGGXAAFFAGDGFAFGXAAXXGFGGXXDAADG
GGDAFGGDFDAXAGAXAAGDDXAAGGDFDDGXAXG
FGFXDDFAAGGDGFAFGXAAXXGFGGXDAFGGDAXX
AAXGDAAFFAGFFFGFDAAGGGGAXGDDFFGXDDAFG
GDXDAXAAGDAXDGGXGFGGGDAXFDAXFDADAXG
DXGFGGXDFAAXAAXFDAXGFAAFDGGDFAXAXFFA
GFDFGGFGGFGDAFGGGDFAXGDGFGGFGGDDGAXG
FAAGDAXAAADFGGXGGFAFGXAAXAAFFAGDFFGX
DFXAXFGFXFAAXAXFFAGGXFXDFAAFXFXXGFDXG
GFGGFGGDXGDGGFDAAAGDDAGDFGFDGGDFAAGG
FDXGFFAAFDAXDGGFDGFGGFDFGXAXADGXGGFDF
GGGDFAXGDAFAAFAFAGFFDAXXGFGGXGGDFDGG
FDGGFFDXGGFDGAGDGFFFDXGGFGGFGGDXGADD
GDDADGDFGGGDFAXGDXDAADXAXGXFXFDFGGGD
FAXGDDGGFAFFGFDDGFFDDGXFXDFAXGDAXXDD
GGGDFGGDFAXGFGGDGAFDXDDAXGGGXFXXGFGG
XGFGGGXFXDGAGDAFGFGFAXGAXGFFDFGFDDGDF
AAAGAAFAXDAAXGGFFAFGXAAXAGFDXGFDFGGG
DFAXGDADGXGGGFDFAXDFAAAGAAFAXDAAXGGF
ADFAAAFDAXAGFDAXDAFGGDFDXGDAAAGGDFAX
GDAGAXAAGGDFGFFGFFFGXDDGAAFDAGFGDGFFD
DAXXAAXGDXGGGDFDGFFDDGGFGDDAADGFFADA
AAFDXDFAXGDFAFGXAAXAXXAAXFFDGDADGGGF
DDGFFGGDXDGFAFADGFFDDFDXGGFAXFADADGG
GDDFGGGGGFGGGDFAXFXFGDGFFGGXDDFAXGDA
XDGXDAAGFGGGDXGDGFFDDGFFGDFAAGDAGGGD
FAAGGDGXDFGGXFAAGGFGGAAGDXAAXDAFGGDA
GAAXGGFXDAAGFDGGGGDXGDGFFDDDFAAGDAGA
XFFFGGXDDDFFDFGGGDFAXGDAGDGAGFFGGGGDF
DGFFDXGFFGXDDFAAGGXDAAGFDGAGFGDGFFDD
XDGDFGFFDDGFFGDGXDAAFADXAXAGGXFXGGFG
FDXGGFDGGFGGAXGDGFGDFGFGFDFDFGGGDFAXG
DXDDFAAGGXDAAGFDGAGFGDGFFDDXDGDFGFFD
DXDDFXGAGFGFFGGXGFGGXFAFGXAAXFDAXAAFF
XGFDFGGDAXFDFGGGDFAXGDDGAAFDGFFGGDGD
XGDAFGGDGGDFAXDADGGDGFGGGGDGFDAXDGFF
AAFAFGFFDDGGDGFDAXDGAAAFGGGXAAFAFAXG
GFFDDGFAAXAGADGXGGGGDFAAGGFFDGDDDFGG
DGDAFGGDDDFGGGGGFGAFFAAXAAFFGXFXFGFFA
XFGDAGGDFAXGDFGFGFDGFFDFGGGDFAXGDADA
XAAFAFDAXGXFFGGDGFADGAFFGGXFAAGFFGGDG
GFGGAAFFAGGXFXFDFGGGDFAXGDAGFGXGFGGXF
AFGXAAXFDAXFFFGXDDFAAXAAXDGFDAAAGAXG
XFXDAFGGDAXXAAXGDXGGGDFDGFFDDDGDFAAX
AAXAGFGFFAXXDDFXGAFFGGXFAAGFFGGXGFGGX
FGGFAXAXGGDFAAGGDGFAFGXAAXAGXGFGGXAA
FAFAAAFAFGFFDDFDFGGGDFAXGDXDDFXG’]

Optimal order 1st generation [**et**o**a**i**n**shrludymw**g**fcbpkv**xzq**] fitness 0.74

Optimal order 10th generation [**et**o**a**i**n**s**hr**ludymw**g**fcbpkv**xzq**] fitness 0.0.792

Optimal order 23th generation [**etoians****hr**ludymw**g**fcbpkv**xzq**] fitness 0.0.841
Optimal order 41th generation [**etoians****hr**ludy**mwg**fc**bp**kv**xzq**] fitness 0.0.89

REFERENCES

[1] LEASTER s. HILL, Cryptography in an algebraic alphabet, American
[2] Kahn, David, the Code Breakers: The Story of Secret Writing, the McMillian Company, 1967, pp. 333-347.
[3] *Singh*, *Simon*: "Geheime Botschaften", *Carl Hanser Verlag*, 1999, P.132
[4] David E. Goldberg , Genetic Algorithms in Search, Optimization, and Machine Learning , The University of Alabama,1989
[5] Holland J.H, Adaption in natural and artificial systems, Univ. of Michigan Press, 1975.
[6] Richard Spillman, Mark Janssen, Bob Nelson, Martin Kepner, Use of a Genetic Algorithm in the Cryptanalysis of Simple Substitution Ciphers, Cryptologia, Vol. 17, Issue 1, pp. 31 – 44, 1993. http://dx.doi.org/10.1080/0161-119391867746
[7] Mitchill Melanie, an introduction to genetic algorithms, First MIT Press paperback edition, 1998.