Journal of Theoretical and Applied Information Technology
10th October 2013. Vol. 56 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

USAGE OF FIREFLY ALGORITHM IN VIGNERE CIPHER TO REDUCE VARIABLE LENGTH KEY SEARCH TIME

[1]V.RAJENDRAN,[2]DR.T.PURUSOTHAMAN
[1]Research Scholar, Anna university, Coimbatore, Tamilnadu, India.
[2]Faculty Of Computer Science And Engineering, Government College Of Technology, Coimbatore, India.

E-mail: [1]dhanaraja2006@yahoo.com, [2]purusothaman.t@gct.ac.in

ABSTRACT

Various cryptographic methods rely on optimization techniques to find the key of a given cipher text. A fitness function is defined based on frequency analysis or any other known information. Vignere cipher was crypt analyzed using GA, PSO and other optimization techniques. Firefly algorithm is an optimization algorithm based on the brightness and attraction among fireflies. A fitness function is defined based on frequency of occurrence monograms and bigrams in English. Firefly algorithm is tried on 1 KB of Vignere cipher and it was seen that keyword length ranging from 5 to 20 were found without error and also with less time complexity than GA and PSO.

Keywords: *Cryptology, Vignere cipher, Network Security, PSO, Firefly Algorithm.*

1. INTRODUCTION

Cryptology deals with the different methods of encryption and decryption of the information or data. Cryptography deals with the new methods of information or data encryption . Cryptanalysis means the art of breaking of the new methods of encryption by finding the weakness of the encryption methods [1].In symmetric cipher, same password is used for encryption and decryption. Cryptanalysis of Vignere cipher using Genetic Algorithm was performed by [2] and it involves two operations such as mutation followed by crossover tries to find the better solution to the problem.

Particle Swarm Optimization is an optimization technique and was also used to find the key of vignere cipher in less time than with the GA method [3].

Naturally fireflies are having blinking characteristics. Based on these blinking characteristics of fireflies, firefly algorithm was developed. This new nature inspired evolutionary algorithm was utilized for many optimization problems related to multimodal , continuous constraint , stochastic test functions and optimization problems related to design.[1][4].

This article brings the usage of FA in vignere cipher to reduce the variable length key search time. This algorithm is utilized for finding the variable length key used in vignere cipher and it is proved that this firefly algorithm performs well when compared with the GA and PSO.

2.0 VIGNERE CIPHER

The main drawback of simple substitution cipher is that they are easily breakable using frequency analysis of each alphabet in the cipher text. Hence a technique was developed by Blaise de Vignere in the year 1553 and so it was named as the Vignere cipher. The vignere cipher is a basic form of substitution cipher and the type of substitution is polyalphabetic.

In Vignere cipher tabula recta is used for both encryption and decryption. This table consists of 26 alphabets written in 26 different rows, in which each and every alphabet is left shifted with reference to the last alphabet .Consider key letter in X axis and plaintext letter in Y axis. At the intersections of key letter labeled X axis and plaintext labeled Y axis, results in cipher text letter.

For example, if the plaintext word to be encrypted is ‘UNCONDITIONALLY’, the

keyword used is RIGHT and this keyword is repeated until all the letters in the plaintext are encrypted.
Plaintext : UNCONDITIONALLY
Key Word: RIGHTRIGHTRIGHT
Cipher text: LVIVGUQZPHEIRSR

In the plaintext, the alphabet N occurs three times, the alphabets O,I and L occurs for two times. In this the first alphabet N is encrypted using the key letter I to give the cipher text letter V. Likewise for the second and third occurrences of letter N in the plaintext, the cipher text letters are G and E. Hence it is observed that the strength of this cipher is, for the same letter N, three different enciphered letters are produced for each unique letter of the keyword. Similarly for the two time occurrences of the alphabets O,I and L, six different enciphered letters are produced.

2.1 Particle Swarm Optimization

Optimization is a process of computing a parameter of a function to reach a optimum solution. PSO is an optimization technique derived by Kennedy and Eberhert in the year 1995.PSO depends upon the nature swarm behavior of some animals called swarm intelligence [3].Each and every particles present in the swarm population is looking for a better solution by the way of flying across a search space which may be a multidimensional. The location of each and every particle in the swarm is changing in accordance with the flying experience of its own. The location of the i^{th} particle is written as Xi = (Xi_1, Xi_2,…,Xi_N) in a N dimensional search space. The previous best location of the every particle in the swarm is stored in a memory is written as P_{besti}= ($p_{i1,}p_{i2}\dots p_{iN}$).The global best in the swarm is written as P_{gbest}= ($p_{g1},p_{g2}\dots p_{gN}$).The velocity of each and every particles in the population is written as V_i= ($v_{i1,}v_{i2},v_{i3},\dots v_{in}$).The location of each particle is decided by the current velocity during that iteration. Velocity vector and position vector are two important fundamental equations which will decide the function of PSO. The velocity vector is written as

$$v_{in}^{t+1} = wv_{in}^{t+1} + c_1r_1(p_{in}^t - x_{in}^t) + c_2r_2(p_{gn}^t - x_{gn}^t)$$
(1)

wv_{in}^{t+1} → previous velocity inertia

$c_1r_1(p_{in}^t - x_{in}^t)$ → Cognition component

$c_2r_2(p_{gn}^t - x_{gn}^t)$→ Social component

The position vector is written as

$$x_{in}^{t+1} = x_{in}^t + v_{in}^{t+1}$$
(2)

c_1 and c_2 are the acceleration constants (Kennedy,1997; Eberhart and Shi,2001) and inertia weight w(Shi and Eberhart,1998) are predefined by the user and the random numbers r1,r2 are generated uniformly in the range of (0,1).In each dimension all the particle's velocities are clamped to V_{max}. As V_{max} is the limited velocity in each dimension, all the accelerations in a dimension are added together will cause the velocity to exceed V_{max}. V_{max} refers to the English language which has 26 alphabets. To reach a better fitness value this same procedure is followed for a pre decided times of iterations. Hence PSO was used to find out the keyword of Vignere cipher in less time.

2.2 Firefly Algorithm

FA was derived by [5][6],deals with the blinking characteristics of the fireflies in nature. The blinking light from a firefly is considered as a signal system, in which each firefly is attracted by the neighboring fireflies. According to the FA algorithm, initially each firefly is given a random position in the search space. All the fireflies have its own flashing light property. The attracting capability of each firefly depends upon its brightness. This will fix the present position in the search space depends upon the computed fitness value. The attractiveness of a firefly depends upon its brightness. Each firefly will be attracted by a nearby firefly which is having more brightness in the search space. Thus each firefly moves towards the brighter firefly and then its fitness function changes accordingly based on its current position. Thus the moving of fireflies towards brighter firefly results in a better position in the search space. [6].The light intensity from a source at a distance r is based on the inverse square law. This law states that the intensity of the light I reduces as the distance r increases in terms of

$$I \propto \frac{1}{r^2}$$
(3)

In addition to this, air in the atmosphere absorbs the light and so the brightness of the light decreases as the distances increases. Hence it is observed that for the above reasons many of the fireflies visible only to a short distance [6].

The flashing characteristics of this firefly algorithm are clearly stated as three rules.

1. Irrespective of the sex of the fireflies each firefly is attracted by another one because they are all are unisex.
2. Due to the blinking nature of the fireflies, those fireflies which is having less brightness will be attracted by the fireflies which is having more brightness and so the Attractiveness is directly proportional to their brightness, and they both decreases as their distance increases. A firefly can move randomly, If no one is brighter than a particular firefly.
3. The brightness of a firefly is calculated by the visual features of the fitness function to be optimized [7][8].

The pseudo code can be briefly stated as [6]:

```
Begin

1) The fitness function f (X),  X = $(x_1,….,x_n)^T$

2) Population of fireflies are initialized as $x_i$ (i = 1,2,…,m) .

3) Compute intensity of light Ii at $X_i$ associated with f(x).

4) Consider Light absorption coefficient $\gamma$
While (t < Max Generation)
For i =1:m (all n fireflies)
For j=1:i (all n fireflies)
                    if ($I_j$> $I_i$)
                    move firefly i towards j;
                    end if
    Vary attractiveness with distance r via exp[-$\gamma r$];
compute new solutions and the  light intensity is updated ;
end for j
end for i
 Rank fireflies and find the current best;
end while
Post-processing the results and visualization;
End
```

3 PROPOSED TECHNIQUE

It is proposed to utilize firefly algorithm to search the variable length key used in Vignere cipher. To analyze this plaintext size is limited to 1KB.Keywords of various lengths such as 5, 6, 7 etc. are used to convert plaintexts into cipher texts using vignere encryption. It is known that the keyword length is the first requirement to crack a vignere cipher. With the help of William Friedman’s coincidence test, the length of the key word used can be found. With the help of frequency analysis method, every character in the password will be identified. With the help of the English language statistics such as one and two alphabet word frequencies, and also based on the frequency of the occurrences of the each alphabet, a suitable fitness function was arrived. The above technique is implemented using MAT lab R2008 and it was executed using Intel’s Core *i3* PC.

3.1 Coincidence Test

Index of coincidence test is a technique formulated by William F. Friedman, can be used to calculate the absolute length of the keyword used in Vignere cipher. In this test, placing two texts nearer and then the number of similar English letters present in the same position was counted in both the texts. In a random source model, index of coincidence is stated as a ratio of the normalized counting to the expected counting.

Coincidence counting is used when both the plaintext and cipher text written in the same language. Similar language has a higher coincidence counting than dissimilar language. For example English language has 26 letters, so 1/26 is the probability for randomly selection of any given alphabet. Thus the sum of squares of each letter probabilities was arrived for plaintext is 0.0683 and for random text it is 0.0385.With these values we can easily estimate the coincident count. In order to find out the interval of the repeating keys, it is necessary to pick two texts which contain intelligence information of the similar language. There are two forms of index of coincidences are available, they are
1.Delta IC Test
2. Kappa IC Test.
Delta IC Test deals with the autocorrelation of a single distribution. Kappa IC deals with the matching of two text strings. In this paper Kappa IC Test is utilized to search out the length of the keyword. Hence a practical way to compute the observed number of coincidences divided by the occurrences of the total number of cases occurs will give the final count of superimposed alphabet comparisons. If the ratio is very near to the value

0.0667, we can find out the correct superimposition has been found and if the ratio is 0.0385, then it will results in incorrect superimposition.

Frequency distribution for a given letter can be computed mathematically using the technique IC is written as

$$IC = \frac{\sum_1^X n_i(n_i-1)}{N(N-1)/x}$$

(4)

N → length of the text
n_i→ frequency of occurrence of X characters of the alphabet
X→ letters of the alphabet
n (n-1) → number of combinations of elements taken two at a time.
N (N-1) → count the number occurrences of English alphabet pairs in the entire text.

Hence the above formula will give the ratio of the observed total number of coincidences to the total number of coincidences that are possible. From the relative alphabets frequencies of English language, the expected average value for the IC can be computed.

The expected IC for plaintext will be 0.0683, if all the letters of the English language are equally distributed. So the required IC can be estimated for every length of the password or keyword. If the coincidences between the length of the keyword and the cipher text are 0.06 then it will leads to the exact length of the password. In this work, the length of the password was varied from five to twenty ,hence the IC value has to be 0.06 within this range.Hence the maximum length can be safely assumed to be the length of the keyword [3].

3.2 Fitness Function

Each letter frequencies present in the cipher text is estimated to derive the fitness function. This estimated frequency of all the letters present in the cipher text is normalized by dividing the frequency of occurrence of each letter by the total number of the letters present in the cipher text file. Then the difference between computed normalized frequency and the expected frequency of the letter is found. The absolute value of the differences of all the characters present in the cipher text is computed and the final normalized value lies between normally from 0 to 1.Finally from the analysis of monogram and bigram, a fitness function can be written as

$$fitness = \sum_{i=1}^{20} a * \left|\mathrm{SF(i)} - \mathrm{DF(i)}\right| + \sum_{i=1}^{20}\sum_{j=2}^{20} b * \left|\mathrm{SDF(i,j)} - \mathrm{DDF(i,j)}\right| \quad (5)$$

SF(i)→ Standard monogram frequency
DF(i)→ Decoded monogram frequency
SDF→ Standard bigram frequency
DDF→ Decoded bigram frequency
a → Weightage for monogram difference
b→ Weightage for Bigram difference

The difference between the SDF and DDF can be calculated using the above fitness function and the error can be reduced by using an optimization technique called Firefly Algorithm [3].

4 EXPERIMENTAL RESULTS

Plain texts up to 1 KB were converted to cipher text with varying keyword lengths from 5 to 20 in vignere cipher. These cipher texts were given to PSO and the time taken for finding the key was noted. Firefly algorithm was also applied to these cipher texts and the keyword used for encryption was found. The time taken to find the keyword was noted. It was seen that with the firefly algorithm the search time was very less when compared with PSO algorithm. Three sets of password (key) lengths from 5 to 20 were tested and the corresponding time taken to find the exact key was tabulated in table (1) and table (2). It is observed clearly from the figure(1) that the time required for finding the key using FA is very less when compared with PSO.

5 CONCLUSION

In this paper, an approach was presented using Firefly Algorithm to find out the variable length keyword used in Vignere cipher. Many experiments were conducted for key sizes varying from 5 to 20 characters for the cipher text length of 1Kb.It was observed that firefly algorithm requires less time for finding the key than with the PSO algorithm. If the cipher text is around 2 KB size then the difference in time would still be larger and it would be possible to find the keyword of lengths around 40, which is more than sufficient for any practical applications. In this paper, the performance of Firefly algorithm was compared with the performance of PSO in obtaining the keyword of Vigenere cipher. Hence, with new blooming optimization algorithms, it might be possible to reduce the time taken to find the keyword further. 1Kb of cipher text was used in this paper to find the keyword and hence experiments can be done to find the keyword with lesser cipher text .

Since frequency analysis of the English alphabets is used to find the keyword, this method can be applied to plaintexts of normal English language. Passages like LJYDE JTTD YKKUAHYDRAKA cannot be used as plaintext since this passage doesn't follow the frequency pattern of English alphabets.

REFERENCES

[1]. JitinLuthra, Saibal K.Pal,2011,"A Hybrid Firefly Algorithm using Genetic Operators for the Cryptanalysis of a Monoalphabetic Substitution Cipher"

[2]. .Purusothaman,T.,V.Gopalakrishnan, S.Arumugam, V.Palanisamy and S.Balraja et al., "Cryptanalysis of Vignere cipher using genetic algorithm and dictionary analysis",The IASTED international Conference on Technology for Education 2009

[3]. Ganapathy Sivagurunathan and Dr.T.Purusothaman " Reduction of Key Search Space of Vignere Cipher Using Particle Swarm Optimization". Journal of computer science 7(11):1633-1638,2011.

[4]. Shadi MashhadiFarahani, Azam Amin Abshouri, BabakNasiri, Mohammad Reza Meybodi "An Improved Firefly Algorithm With Directed Movement"in Proceedings of 4th IEEE International Conference on computer Science and Information Technology, 248-251 Chengdu, China.

[5]. Yang, X. S., "Nature-Inspired Metaheuristic Algorithms", Luniver press, 2008.

[6]. Yang, X. S.,"Firefly Algorithms for Multimodal Optimization" in: Stochastic Algorithms: Foundations and Applications, SAGA 2009, Lecture Notes in Computer Sciences,Vol. 5792, pp. 169-178 (2009).

[7]. Yang, X. S.,"Firefly Algorithm, Stochastic Test Functions and Design Optimisation", Int. J.Bio-Inspired Computation, Vol. 2, No. 2, 2010,pp.78—84

[8]. X.-S. Yang, "Firefly Algorithm Levy Flights and Global Optimization", Research and Development in Intelligent Systems XXVI (Eds M. Bramer, R. Ellis, Petridis), Springer London, , 2010, pp. 209-218.

Table 1: Elapsed Time To Find The Keys Of Different Length And Average Time Taken For Threesets Of Keywords Of Same Length Using PSO .

Password Size(Key)	Password	PSO	
		Elapsed Time (insecs.)	Average Time (insecs.)
5	SAIRA	160	185.33
	MUSIC	185	
	RAJEN	211	
6	SAIRAM	183	205.66
	SAVICH	200	
	CIPHER	234	
7	EDRFTGY	206	230
	CHANNEL	257	
	BCDAXYZ	227	
8	AFEQMOST	247	252
	STRENGTH	280	
	ELEPHANT	229	
9	RAJENDRAN	251	274.44
	PRESIDENT	302	
	RSPOQNYCD	271	
10	GODISLOVEE	278	298.33
	ACDINTOPLV	323	
	DQMEFABNOU	294	
11	ACEGIKMOQSU	311	324.66
	FOUNDATIONS	345	
	MOSHIPRSTVY	318	
12	INTOXICATING	334	348.66
	NEUROSCIENCE	367	
	VSTMPBCDAHNI	345	
13	SURIYAPRAVINK	354	370.33
	INTERNATIONAL	388	
	OQSRTLUVWYZEG	369	
14	NOTESAVINXZYPO	380	397
	UNDERPREVILEGE	416	
	CRUVFQLMNACDWI	395	
15	BOXIAYGCFMNSHEW	404	419
	VIRTUALINTERACT	423	
	ADHCOPEGHZXYUVN	430	
16	AISYOPQRTYZXGHMU	426	441.66
	ENTREPRENEURSHIP	436	
	EACDHNROVUISXYLB	463	
17	CSKBNRDHOIYFLQTAV	454	465.33
	THERMALPOWERPLANT	462	
	POMNSHQRSTUVACDBL	480	
18	ACDEMFGOSTNXUVREGY	491	489.66
	AABBCCDDEEFFGGHHII	478	
	RDHASKOPNIQRSTUVWX	473	
19	SHEMALEMECALUOVPZQS	495	499
	ASDFGHJKLOIUYTREWQZ	502	
	IMCBADHNGFEOSTRVUNZ	500	
20	IVEYQRTSMNOPCABDFEZY	520	520.66
	QWERTYUIOPLKJHGFDSAZ	522	
	NPRSTUVWABNDHCTWYLMG	520	

Table 2: Elapsed Time to find the keys of different length and Average time taken for threesets of keywords of same length usingFA.

Password Size(Key)	Password	Fire Fly Algorithm (FA)	
		Elapsed Time (insecs.)	Average Time (insecs.)
5	SAIRA	38.28	37.81
	MUSIC	35.80	
	RAJEN	39.37	
6	SAIRAM	43.15	42.35
	SAVICH	39.97	
	CIPHER	43.95	
7	EDRFTGY	46.89	46.38
	CHANNEL	47.66	
	BCDAXYZ	44.59	
8	AFEQMOST	50.41	52.62
	STRENGTH	53.66	
	ELEPHANT	53.80	
9	RAJENDRAN	57.40	58.26
	PRESIDENT	61.51	
	RSPOQNYCD	55.87	
10	GODISLOVEE	63.49	61.68
	ACDINTOPLV	62.01	
	DQMEFABNOU	59.54	
11	ACEGIKMOQSU	69.12	67.68
	FOUNDATIONS	69.46	
	MOSHIPRSTVY	64.46	
12	INTOXICATING	73.93	72.61
	NEUROSCIENCE	74.78	
	VSTMPBCDAHNI	69.14	
13	SURIYAPRAVINK	78.05	77.14
	INTERNATIONAL	79.11	
	OQSRTLUVWYZEG	74.26	
14	NOTESAVINXZYPO	83.12	82.19
	UNDERPREVILEGE	84.11	
	CRUVFQLMNACDWI	79.34	
15	BOXIAYGCFMNSHEW	89.46	88.18
	VIRTUALINTERACT	88.71	
	ADHCOPEGHZXYUVN	86.38	
16	AISYOPQRTYZXGHMU	93.06	92.84
	ENTREPRENEURSHIP	95.24	
	EACDHNROVUISXYLB	90.22	
17	CSKBNRDHOIYFLQTAV	98.97	98.47
	THERMALPOWERPLANT	102.30	
	POMNSHQRSTUVACDBL	94.16	
18	ACDEMFGOSTNXUVREGY	97.96	101.84
	AABBCCDDEEFFGGHHII	105.90	
	RDHASKOPNIQRSTUVWX	101.68	
19	SHEMALEMECALUOVPZQS	108.19	107.75
	ASDFGHJKLOIUYTREWQZ	109.50	
	IMCBADHNGFEOSTRVUNZ	105.57	
20	IVEYQRTSMNOPCABDFEZY	113.96	113.59
	QWERTYUIOPLKJHGFDSAZ	113.10	
	NPRSTUVWABNDHCTWYLMG	113.73	

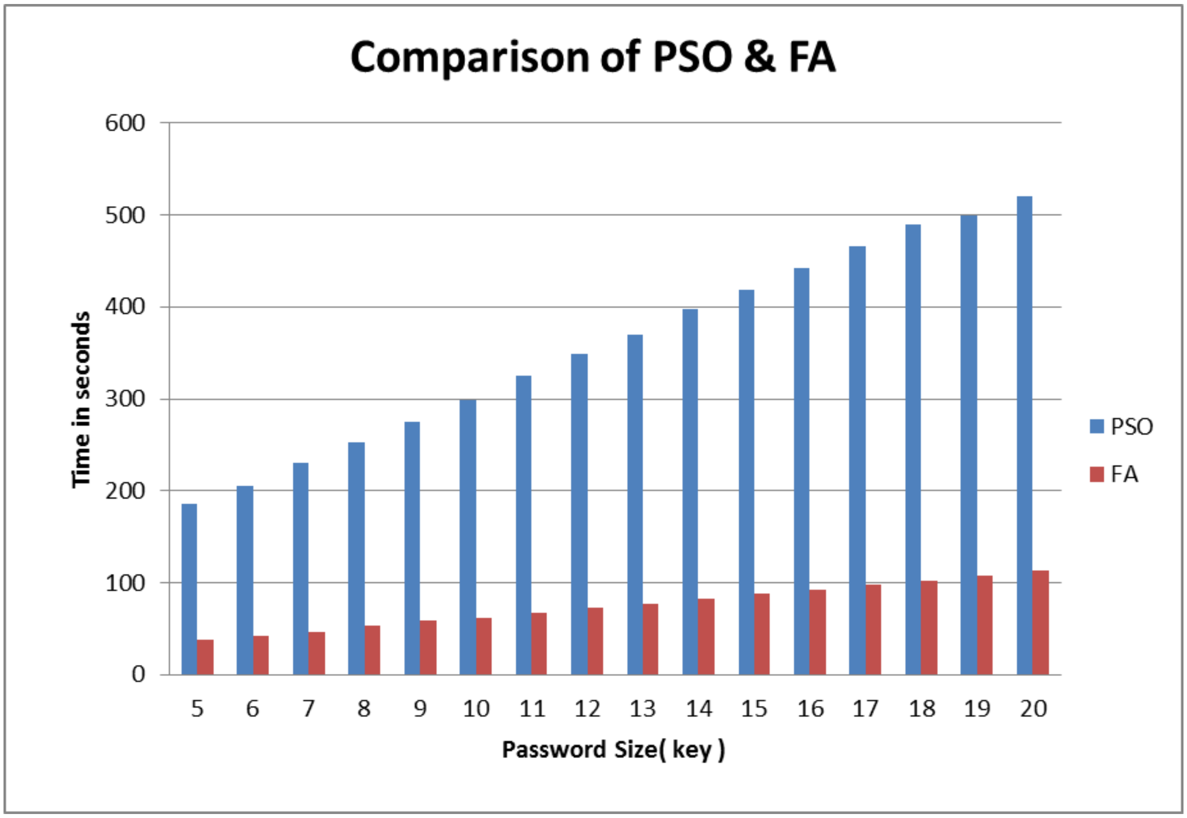

Figure 1 :Comparison Of Time Taken For Finding The Keyword Using PSO And FA