IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 1, Ver. IV (Jan.-Feb. 2017), PP 78-84
www.iosrjournals.org

Cryptanalysis of Stream Cipher Cryptosystem Based on Soft Computing Techniques

Prof .DrSalim Ali Abbas Al-Ageelee[1],RiyamNoori Jwad[2]
[1]*(department of ComputerScience, college of Education/Al-MustansiriyaUniversity,Iraq)*
[2]*(department of ComputerScience, college of Education/ Al-Mustansiriya University,Iraq)*

***Abstract**:This paper presents a new investigation for cryptanalysis stream cipher based on Genetic Algorithm(GAs), Particle Swarm Optimization (PSO).GA and PSO utilized for the automatic recovery of the key, and hence the plaintext. Based on a mathematical model, it is shown that such algorithms can be used to reduce the number of trials which are needed to determine the initial state of the attacked generator using ciphertext only attack. Thesealgorithms have been shown to be effective at finding optimal solutions. Experimental results show the ability of GA and PSO in finding the correct secret key which is used to recover the plaintext.*

***Keywords**: Cryptanalysis, Particle Swarm Optimization, Genetic Algorithms.*

I. Introduction

Cryptanalysis is the science of recovering the plaintext of a message without access to the key. It is a method of transforming cipher text into a plaintext without knowing the key or algorithm [1][2].However the cryptanalysis of stream ciphers through soft computing techniques as Particle Swarm Optimization (PSOs), Genetic Algorithms (GAs)is still an emerging issue.GA is based on the evolutionary ideas of Natural selection and genetics [3].GAis a good condidate for the optimal solutions to optimization and search problems. The algorithm have been successfully applied to Vertex-Cover problem [4][5], Maximum-Clique problem [6][7], Regression testing [8], N-puzzle problem [9], Traveling Salesman Problem [10].

PSO was originally developed by a social-psychologist J. Kennedy and an electrical engineer R.Eberhart in 1995 and emerged from earlier experiments with algorithms that modeled the "flocking behavior" seen in many species of birds. Where birds are attracted to a roosting area in simulations they would begin by flying around with no particular destination and in spontaneously formed flocks until one of the birds flew over the roosting area [11]. PSO has been an increasingly hot topic in the area of computational intelligence. It is yet another optimization algorithm that falls under the soft computing umbrella that covers genetic and evolutionary computing algorithms as well [12].

There are many researches has been written on using soft computing techniques to cryptanalysis different types of encryption systems some of these: A.J.Clark, in his Thesis uses various optimization heuristics in the fields of automated cryptanalysis and automated cryptographic function generation[13], M.F. Uddin in his paper focused on using of PSO in cryptanalysis of simple substitution ciphers using ciphertext only attack[14],R.R.Yako In her research, an optimization approach such as GAs is considered to improve the cryptanalysis problem[15], S. M. Hameedin her work used PSO to cryptanalysis transposition cipher, PSO used ciphertext only attack to recover the secret [16],H.A.MAl_Sharifi,in his research focused on using of PSO algorithm to cryptanalysis stream cipher using plaintext attack choosing one Linear Feedback Shift Register (LFSR) [17], B.N.Ferriman, in his Thesis focused on the RC4 algorithm and present a new approach for cryptanalysis of the cipher by attacking RC4s state register[18], Ali A.Abd in his research is considered a new approach to cryptanalysis stream cipher systems based on genetic algorithm(GA)[19].The present work explores the related work done and applicability of GAs and PSOs in a field of cryptanalysis.The rest of this paper has been organized as follows: Section 2 presents a brief overview of GAs. Section 3 presents a brief overview of PSOs. Section 4 presents designing cryptanalysis systems for stream cipher.Section 5 presents acomparisonresults of cryptanalysis system between GA and PSO. The last section explains the conclusion

II. Genetic Algorithm

GAs is the search heuristic that mimics the process of natural evolution [3]. It is based on the Darwin's principle of Natural selection. According to this theory the chromosomes with the best fitness function should survive and create new offspring (survival of the fittest). GAs gives useful solution to optimization and search problem. It is a rapidly growing area of Artificial Intelligence. The GAs starts with the population which is nothing but chromosomes which can be decimal or binary or even hexadecimal. The GAs operator is applied to population in order to optimize the results [20]. The new population is formed from the old population with better fitness value. The population can be crafted using following operators:

Population size: The population size is generally taken 5 to 100 [21].
But in this paper we took it about 20,100 and 200.Numerical experiments show that too large and too small number of chromosomes in the population can lead to poor solutions [22].

Fitness Function: The fitness function decides whether the given solution is achieving the aims [23]. The proper crafting of the fitness function is a crux of the solutions.
In this paper we will use the T.SiegenthalerMethodas a Fitness Function[24].

Selection: The chromosomes are selected from the population for reproduction. The chromosomes with higher fitness value are more likely to be selected. [23]. There are many ways to do selection process some of them are Tournament selection, Roulette wheel selection, Stochastic universal selection, Truncation selection etcetera.

Crossover: The operator created a new offspring from the population by exchanging subsequences of two chromosomes to create two offspring. There are many types of crossovers for the binary chromosomes. Some of them are Single-point crossover, Two-point crossover,Multi-point crossover, Uniform crossover. In this work we will use the Single-point crossover which is One crossover point is selected, binary string from beginning of chromosome to the crossover point is copied from one parent, the rest is copied from the second parent [25] .

Mutation: Mutation operator flips the bit in chromosomes. The purpose of mutation is to maintain the diversity within the population.

III.Particle Swarm Optimization

Swarm Intelligent is a kind of Artificial Intelligence based on the behavior of animals living in groups and having some ability to interact with another and with the environment in which they are inserted. Every particle in the swarm acts in a distributed manner using the intelligence of its own and the group intelligence. Every particle has two features: a position and a velocity. The particles exchange the information to correct their positions and velocities by using the received information [26].
In this work, GA and PSO is used as soft computing algorithm employed in cryptanalysis system to find initial values of the key stream and hence, to obtains plaintext.

3.1 Basic Elements of the PSO Technique [27][28]

The basic elements of PSO technique are briefly stated and defined as Follows:

1. **Particle, X^i**: It is a candidate solution represented by an m-dimensional vector, where m is thenumber of optimized parameters.
2. **Population, pop(t)** : It is a set of n particles at time i, i.e. pop(i) = $[X^i_1,..., X^i_n]$T. The number of particles in population would be between 20to 30.
3. **Swarm**: It is an unsystematic moving particles population, which Band together and at the same time every particle moves in aUnorganized direction.
4. **Particle velocity, V_i**: It is the speed of the moving particles which canbe characterized by k-dimensional vector.
5. **Inertia weight, w_i**: It is a control factor used to control the effect of the preceding velocities on the present velocity.
6. **Individual best, p_i**: it is the composition of the particle fitness value at the present position to the best fitness value it has ever reached.
7. **Global best, p_j^g**: It is the best location obtained in all individual locations.
8. **Stopping criteria**: it is the terms which finish search process.

3.2 PSO Methods

There are several methods of PSO depending on the shape of
Updated velocity equation of the particle those are:

- **Original PSO:** Basic algorithm introduced to calculate the velocityand position of each particle and it is used to find the optimal solution**.**

velocity of individual particles updated as follows:
$V^i_{j+1}=V^i_j+c_1r_1(P^i_j- X^i_j) +c2r2 (P^g_j- X^i_j)$ j=1,…,n(1)
with the position calculated as follows:
$X^i_{j+1}= X^i_j+V^i_{j+1}$ j=1,….,n(2)
Where
$\mathbf{X^i_j}$: Particle position
$\mathbf{V^i_j}$: Particle velocity

$\mathbf{P^i_j}$: Best position found by ith particle (personal best)
$\mathbf{P^g_j}$: Best position found by swarm (global best, best of personal bests)
$c_1$and c_2 : are the cognitive (individual) and social, (group) learning, rates, respectively, The values of c_1 and $c_2$are usually assumed to be 2. r_1 and r_2 : are uniformly distributed random numbers in the range 0 and 1.

- **Inertia Weighted PSO:** The inertia weighted PSO is added to decrease the velocity. Its value varies from 0.9 to 0.4.The value of the jth particle velocity can be formulatedas:

$V^j_{j+1} = w_i v_j^i + c_1 r_1 (p_j^i - x_j^i) + c_2 r_2 (p_j^g - x_j^i)$ j=1,…,n(3)

Then the value of the inertia weight can be calculated:
$w_i = W_{max} - (W_{max} - W_{min}/imax) * i$ ... (4)
where
$W_{max}$is the initial value of the weight.
W_{min} :is the final value of the inertia weight.
i_{max}:is the maximum number of iterations.
In this work, Inertia Weighted PSO type is used.

IV. Designing Cryptanalysis Systems for Stream Cipher

In this work we used soft computing techniques as GA and PSOwhich could be implemented and applied easily to solve various optimization problems. These techniques employed for the purpose of Cryptanalysis. Wesuggest main steps todesigning cryptanalysis systems for stream cipher, we select the Geffe generator to be attacked .these steps of the analysis and procedure can be summarized as follows:

Step1: select plaintext.
Step2: generate the key stream from Geffe generator system.
Step3: generate the ciphertext calculated as follows:
P_t XOR $K_i=C_t$ when P_t=plaintext,K_i=keystream,C_t=ciphertext.
Step4: design fitness function.
Step5: apply soft computing techniques.
Step6: select the optimal solution.

4.1 Fitness Function Calculation:

the main goal of cryptanalysis is to get the key in order to obtain the plaintext. Cryptanalysis stream cipher should get the correct key. to decrypt the ciphertext.Using soft computing techniques to cryptanalysis stream cipher needs fitness to determine the best new generation. In this work, fitness function based on correlation between C_n and X^i_n.Cryptanalysis of stream cipher based onstatistical model is used to find the Linear Feedback Shift Regester(LFSRi)part of the key, i.e., the initial of theLFSRi, i €{1, …, s}. Further, the number of tests to find the LFSRi-part of the key is determined as a function of the number of ciphertext digits used in the correlation attack. Let the inputs
$x^1_n, x^2_n,, x^s_n$ of the functionf be generated by independent and identically distributed (i.i.d) random variables (r.v.) Xnwithprobability distribution P_x such that $P(X^i_n = 0) = P(X^i_n = 1)$ for all i and n. The function generates i.i.d. r.v. $Zn = f(X^1_n, X^2_n, X^3_n, \ldots, X^s_n)$ with probability distribution P_z where
$P(Z_n = 0) = P(Z_n = 1)$(5)
Where Z_n: output key of random generator number n
The plaintext is assumed to be the output of a binary memorylesssource (BMS) with
$P(Y_n = 0) = P0$. (6)
The r. v. a as a measure for the correlation between C_n, and X^i_n is where:
C_n:cipher text digits.
X^i_n:initial state of linearfeedback shift regester
N:ciphertext length
definedas:

$$\alpha = \sum^N_{n=1} (1 - 2(C_n \text{ XOR } X^j_n)) = N - 2 \sum^N_{n=1}(C_n \text{XOR } X^j_n) \quad (7)$$

where j €{O,1,.. ,s}

Here, the best fitness>= 0.60, this rate considered the threshold in our work, it change according to different plaintext size.
We will attack the Geffe generator, nonlinear combining function which it consist of 3 LFSR in different Length:7,9,11.
The algebraic normal form is:

$$f(x1, x2, x3) = x1x2 \text{ XOR } x2x3 \text{ XOR } x3. \quad (8)$$

Here, the number of ciphertext symbols is determined to perform a ciphertext-only attack on the Geffe Cipher using the correlation attack. Our conclusion from the analysis is that the pseudonoise generator's output sequence and the sequences generated by the linear feedback shift registers should be uncorrelated. This leads to constraints for the nonlinear combining function to be used.

4.2Using Genetic Algorithm (GA) to cryptanalysis stream cipher systems

GA has been successfully applied to numerous applications in the field of search and optimization. It is recursive procedure that consists of a fixed population size of chromosomes. These chromosomes are created randomly or heuristically which represent the initial population. The population evolves by applying three basic operations: selection crossover and mutation with probability. For the Initial Population, the cryptanalysis process begins with randomly generated numbers between {0, 1} as the key size for n chromosomes and sorting these numbers in ascending order. The sequence of these numbers represents the candidate keys (chromosomes). Each chromosome represents the candidate key which it uses to decrypt the ciphertext and then calculate the fitness value to determine the best chromosome (candidate key).
For the Selection operator, selects chromosomes in the population for reproduction. The better chromosome has the opportunity to select more timed to reproduce. Many selection procedures have been proposed, this paper used Roulette-wheel selection (that are described in section 2) to attack nonlinear stream cipher systems, which it used to selecting potentially useful solutions for reproduction. The chromosome (sequence) with high fitness has a higher probability of participate one or more offspring to the next generation. For the Crossover operator, two chromosomes are combining to produce a new generation that possesses both their characteristic. There are several crossover techniques, this thesis used single-point crossover (that are described in section 2) with probability of crossover (p_c)equal to 0.7 to attack streamcipher. For the Mutation operator, this process is used to maintain diversity in population from one generation to the next generation in order to obviate local minima .For this paper, a simple two point mutation is used. This process uses to select two individuals randomly on population in chromosome and swap between them. If the random number that generated to represent the candidate key is equal to 0.1.
For the Fitness Function calculation in this paper, equation (7) is used to calculate the fitness function of GA to attacks stream cipher. For the GA parameters there are a set of values which are considered as the most appropriate to attacks stream cipher by GA. Table (1) shows the different parameters of GA to cryptanalysis stream cipher Systems.

Table (1): GA parameters to attack stream cipher

Parameters	Symbol	Value
Key Length	KeyLen	[24]
Text Length	TxtLen	[10-100]
Number of chromosomes	Popsize	[20,100,200]
Maximum number of Iteration	MaxIter	[100-300]
Probability of crossover	P_c	0.7
Probability of mutation	P_m	0.1

4.3Using PSO algorithm to cryptanalysis stream Cipher

In Evaluation For the initial population, the cryptanalysis process begins with randomly generated numbers between {-1, 1} as the key size for n particles .The sequence of these numbers represents the candidate keys (particles). Randomly generates Velocity for each particle which it's bounded to some minimum and maximum values [Vmax, Vmin] where Vmin= -Vmax and it uses to reinforces the local search reconnoitering of the problem space. Each particle represents the candidate key and use to decrypt the ciphertext and then calculate the fitness value to determine the best particle (key).For the evaluation process, the fitness value for each particle (candidate key) must be calculated for each generation.

Table (2) shows the most parameters of PSO that preferred to be used to decrypt stream cipher.

Table (2): PSO parameters to attack stream cipher

Parameters	Symbol	Value
Number of particles in the swarm	Popsize	[20-200]
Number of Key	KeyLen	[12-24]
Length of text	TxtLen	[10-100]
The maximum number ofIteration	MaxIter	[100-300]
The maximum of velocity	V_{max}	4
The minimum of velocity	V_{min}	$-V_{max}$
Inertia Weight	W	[0.4- 0.9]
Acceleration parameter	C_1,C_2	[0.5-2]
Random number between [0,1]	r_1,r_2	[0-1]

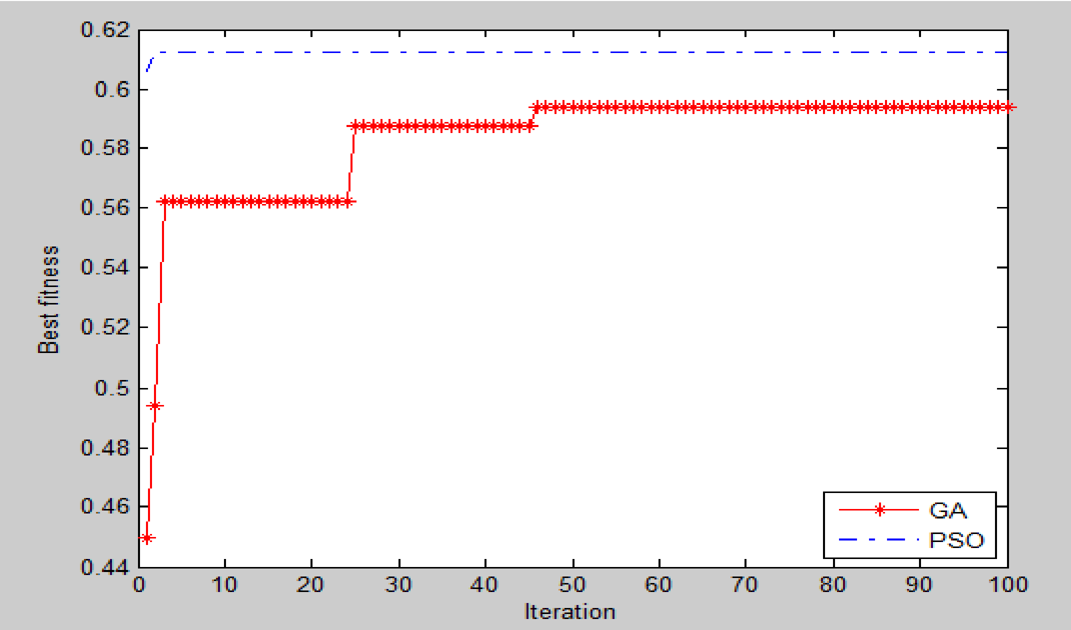
V. Comparison Results of cryptanalysis system between GA and PSO

The following Tables (3,4,5) shows the results of applying proposed cryptanalysis system GA and PSO For Popsize(20,100,200) and Maxiter(10,300) For TxtLen=100,40,10 characters ,the following notations are used:
Popsize = Population size
MaxIter= Maximum Iteration
BF=Best Fitness
T/sec=Time/second
T.T/sec=Total Time/second
Iter_Num= Iteration_Number

Table (3) results of applying GA and PSO For Popsize(20,100,200) and MaxIter(100,300) For TxtLen=100 characters

Popsize	MaxIter	GA				PSO			
		BF	T/sec	T.T/sec	Iter_ Num	BF	T	T.T/sec	Iter_ Num
20	100	0.5869	2.09	15.91	13	0.6188	.23	16.31	1
	300	0.5496	0.67	48.30	4	0.5520	0.58	47.20	1
100	100	0.5896	3.21	79.33	4	0.5656	1.21	82.32	2
	300	0.6023	4.84	238.36	6	0.5556	0.82	240.98	1
	100	0.6048	4.86	163.88	3	0.5850	3.29	160.47	2
200	300	0.5694	5.62	476.26	4	0.5760	1.58	460.32	1

Fig. (1) shows the comparison between GA and PSO in cryptanalysis system for Popsize (20) and MaxIter(100) for TxtLen=100

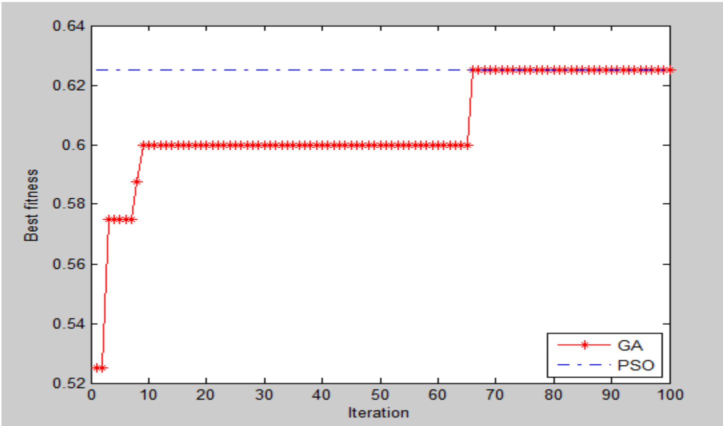


Figure(1) comparison between GA and PSO in cryptanalysis system for Popsize (20) and MaxIter(100) for TxtLen=100

Table (4) shows the results of GA For Popsize(20,100,200) and Maxiter(100,300) For TxtLen=40 characters.**Table (4)** results of applying GA and PSO For Popsize(20,100,200) and MaxIter(100) For TxtLen=40 characters.

Popsize	MaxIter	GA				PSO			
		BF	T/sec	T.T/sec	Iter_ Num	BF	T/sec	T.T/sec	Iter_ Num
	100	0.5906	2.11	6.98	31	0.5656	0.17	6.97	2
20	300	0.5938	0.21	19.81	3	0.5875	0.18	22.41	2
	100	0.6225	25.48	33.99	75	0.6225	0.70	32.75	2
100	300	0.6262	20.30	101.30	35	0.6262	0.41	97.51	1
	100	0.6061	30.10	65.50	35	0.5844	0.66	64.53	2
200	300	0.6023	5.60	465.30	4	0.5906	1.60	463.10	1

Fig. (2) shows the comparison between GA and PSO in cryptanalysis system for Popsize (100) and MaxIter(100) for TxtLen=40

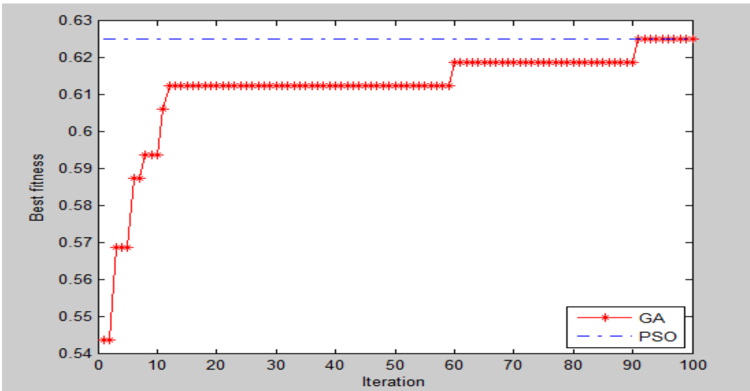


Figure(2) comparison between GA and PSO in cryptanalysis system for Popsize (100) and MaxIter(100) for TxtLen=40

Table (5) shows the results of GA For Popsize(20,100,200) and Maxiter(100,300) For TxtLen=10 characters
Table (5) results of applying GA and PSO For Popsize(20,100,200) and MaxIter(100,300) For TxtLen=10 characters

Popsize	MaxIter	GA				PSO			
		BF	T/sec	T.T/sec	Iter_ Num	BF	T	T.T/sec	Iter_ Num
	100	0.6250	0.11	2.02	5	0.6250	0.05	2.10	1
20	300	0.6750	0.83	5.87	42	0.6625	0.04	5.60	2
100	100	0.6250	1.96	10.99	18	0.6625	0.09	8.50	1
	300	0.6625	27.25	33.36	250	0.6625	0.81	24.40	1
200	100	0.6750	12.25	27.34	45	0.6625	0.18	16.80	2
	300	0.6625	13.92	82.10	51	0.6625	0.35	50.47	2

Fig. (3) shows the comparison between GA and PSO in cryptanalysis system for Popsize (20) and MaxIter(100) for TxtLen=10



Figure(3) comparison between GA and PSO in cryptanalysis system for Popsize (20) and MaxIter(100) for TxtLen=10

VI. Conclusion

1) The cryptanalysis using GA and PSO can find the optimal solution for text withlengths with 10 characters as shown in Tables 3, 4 and 5.
2) We conclude that the PSO cryptanalysis system performs better than GA in term of time as shown in Tables 3, 4 and 5.
3) As shown in Tables 3, 4 and 5the results of applying GA, PSO to cryptanalysis stream cipher, we notice that 5 iterations are enough to find the best solution for the PSO but this number of iterations are not enough for GA to find the best solution.
4) As shown inTable(3) we conclude that the best results of GA and PSO inTxtLen=10 characters is Popsize=20 and MaxIter=100.
5) As shown in Table (4) the best results of GA and PSOinTxtLen=40 characters isPopsize=100 and MaxIter=100.
6) As shown in Table (5) we conclude that the best results of GA and PSO in TxtLen=10characterswhenPopsize=20 and MaxIter=100.
7) From a sequent 4,5 and 6 ,we conclude that Popsize=20 and MaxIter=100 is enough to find the optimal key.

Acknowledgements

I would like to thank my advisorProf.Dr.Salim AL-Ageeleefor motivation and support in presenting this paper.

References

[1]. Schneier, B.1996, Applied Cryptography, Second Edition: Protocols,AlgorithmsandSource Code in C.
[2]. Delman,B.2004,Genetic Algorithm in Cryptography,Doctorol thesis, Rochester Institute of Technology.
[3]. Holland,J.H.1992,Adaptation in Natural and Artificial Systems.
[4]. M.Milanovic,"Solving the generalised vertex cover problem byGenetic Algorithm ", Computing and Informatics,2010.
[5]. H.Bhasin,M.Amini,"The applicability of Genetic Algorithm to Vertex cover", International Journal of ComputerApplication,2015.
[6]. Bazgan,C.,Luchian,H.1995. A genetic Algorithm for maximal Clique Problem. Inproceeding of the International ConferenceinAles,France.
[7]. H.Bhasin et al,"Hybrid Genetic algorithm for Maximum Clique Problem",International Journal of Application of Innovation inEngineering &Management,2013.
[8]. H.Bhasin,Manoj,"Regression testing using Coupling and GeneticAlgorithms",International Journal of Computer Science andInformation Technologies,2012.
[9]. H.Bhasin,N.Singla,"Genetic based algorithm for N-Puzzle problem", InternationalJournal of Computer Application,2012.
[10]. Y.Liao et al,"Evolutionary algorithm to Traveling Salesman Problems", Computer& Mathematics with Applications,2012.
[11]. Papoulis, A. “Probability Random Variables, and StochasticProcess”, McGraw-Hill College,October, 2001.
[12]. Parsopoulos K. E. and Vrahatis M.N., “Recent Approaches to Global optimization Problems through Particle Swarm Optimization”, Kluwer Academic Publishers, Netherlands, Natural Computing 1, pp 235–306, 2002.
[13]. A.J.Clark," Optimisation Heuristics for Cryptology",InformationSecurity Research Centre Faculty of Information TechnologyQueensland University of Technology,1998.
[14]. M. F Uddin and Amr M. Youssef," Cryptanalysis of Simple Substitution CiphersUsing Particle Swarm Optimization".IEEE, Congress on EvolutionaryComputation, Canada,2006.
[15]. RajaaR.Yako,"Decrypting A Class Of Stream Cipher UsingCiphertext Only, Comparative Study", Master Thesis, Higher Academy for Scientific andHumanistic Studies, Department of Computer Science,2007.
[16]. Sarab M. Hameed and Dalal N. Hmood, " particles swarmoptimization for the cryptanalysis of transposition cipher " .Journalof Al-NahrainUniversity , Vol.13(4), pp.211-215, 2010.
[17]. Hussein Ali Mohammed Al_Sharifi ," Cryptanalysis of StreamCipher System Using Particle Swarm OptimizationAlgorithm ".Journal of KerbalaUniversity , Vol. 8 No.4 Scientific , 2010.
[18]. Benjamin Nicholas Ferriman," Cryptanalysis of the RC4 Stream Cipher usingEvolutionary Computation Methods",MasterThesis,University of Guelph,Guelph, Canada,2013.
[19]. Ali A. Abd ,Hameed A. Younis, and Wasan S. Awad," Attacking ofstream Cipher Systems Using a Genetic Algorithm". Journal ofUnivesity of Thi-Qar, ISSN: 66291818,Vol. 8, Issue. 3, 2013.
[20]. Goldberg,D.E.1989. Genetic Algorithm in search, optimization andmachine learning.
[21]. Alander.1992. On optimal population size of genetic algorithm.In Proceedings of theIEEE computer systems andsoftware engineering.
[22]. Goldberg,D.E et al.2000. Bayesuan Optimization Algorithm, population sizing and timeto convergence, University ofIllinois,USA.
[23]. Melanie,M.1996. An introduction to a Genetic Algorithm:MITpresspaperback edition.
[24]. T. Siegenthaler," Decrypting a Class of Stream Ciphers UsingCiphertextOnly",IEEE,1985.
[25]. Bhasin,H.2015.Algorithms: Design and Analaysis.
[26]. Singiresu S. Rao "Engineering Optimization Theory and Practice" Book, by John Wiley& Sons, Inc.2009.
[27]. James Kennedy and Russell Eberhart "Particle SwarmOptimization", Book, IEEE 1995.
[28]. James Kennedy "The Particle Swarm: Social Adaptation of Knowledge" IEEE, 1997.