Hybrid Bees Algorithm With Simulated Annealing for Cryptanalysis of Simple Substitution Cipher

Ismail Khalil Ali
Alma'mon University College
Dr_ismail_cs@yahoo.com
Aseel Ghazi Mahmod
College of Nursing- University of Nursing
Aseel_bhl11@yahoo.com

Abstract

This research is concerned with the development of a hybridization technique for cryptanalysis a simple substitution ciphers. The exploration/exploitation balancing strategy of Simulated Annealing is incorporated into the original Bees Algorithm to improve its search efficiency and reduce its computational cost. Experimental results demonstrate the applicability of algorithm for the cryptanalysis a simple substitution ciphers.
Keywords: Cryptanalysis, Substitution Cipher, Bees Algorithm, Simulated Annealing

الخلاصة

هذا البحث يتعلق بتطوير أسلوب هجين جديد لتحليل شفرة النظام التعويضي البسيط . طريقة التهجين عبارة عن تركيبة من خوارزمية النحل وخوارزمية التلدين المقلد. الأسلوب المقترح يحسن من سرعة التقارب والتوازن في استكشاف واستثمار الحلول المرشحة. النتائج التجريبية تظهر قابلية التطبيق للخوارزمية على تحليل شفرة النظام التعويضي البسيط.
الكلمات المفتاحية : تهجين خوارزمية النحل مع التلدين المقلد لتحليل شفرة النظام التعويضي البسيط

1. Introduction

The whole point of cryptography is to keep the plaintext (or the key, or both) secret from eavesdroppers (also called attackers, interceptors, or simply the enemy). Cryptanalysis is the science of recovering the plaintext or the key. An attempted cryptanalysis is called an attack. The goal of cryptanalysis is to find some weakness or insecurity in a cryptographic scheme, thus permitting its subversion or evasion. The use of automated techniques in the cryptanalysis of cryptosystems is desirable as it removes the need for time-consuming (human) interaction with a search process. Making use of computing technology also allows the inclusion of complex analysis techniques, which can quickly be applied to a large number of potential solutions in order to weed out unworthy candidates. Two fundamental goals in computer science are finding algorithms with provably good run times and with provably good or optimal solution quality (Clark, 1998).

There are two basic types of encryption ciphers: substitution and transposition (permutation). The *substitution cipher* replaces bits, characters, or blocks of characters with different bits, characters, or blocks. The *transposition cipher* does not replace the original text with different text, but moves the original text around, it rearranges the bits, characters, or blocks of characters to hide the original meaning (http://en.wikipedia.org/wiki/Classical_Cipher. Download in 10/1/2013).

Their importance stems from the fact that most of the ciphers in common use today utilize the operations of the classical ciphers as their building blocks. For example, the Data Encryption Standard (DES), an encryption algorithm used widely in the finance community throughout the world, uses only three very simple operators, namely substitution, permutation (transposition)

and bit-wise exclusive-or (admittedly, in a complicated fashion) (U.S. Department of Commerce National Bureau of Standard, 1988).

Swarm Intelligence (SI) is a part of Artificial Intelligence based on the socio cooperative behavioral pattern displayed by various species like birds, bees, termites, ants etc. During the past decade, algorithms based on SI have emerged as potential candidates for solving complex and intricate global optimization problems which are otherwise difficult to solve by traditional methods. Some popular SI based techniques for global optimization include Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Bees Algorithm (BA), Firefly Algorithm (FFA), Fish School Algorithm (FSA), Bacteria Foraging Optimization (BFO) algorithm etc.(Kusum. and Millie.,2013)

This work introduce a hybrid technique of BA with Simulated Annealing (SA) in the cryptanalysis of simple substitution cipher. The rest of the paper is organized as follows: Section 3 presents a brief overview of the substitution cipher. The underlying principles of Bees Algorithm and Simulated Annealing are presented in Sections 4 and 5 respectively. Section 6 describes the algorithm proposed. Experimental results of computational tests to evaluate the performance of the proposed algorithm are reported in section 7.

2. Literature Review

(Uddin. and Youssef.,2006) proposed cryptanalysis of simple substitution ciphers using Particle Swarm Optimization (PSO). They showed that PSO provides a very powerful tool for the cryptanalysis of simple substitution ciphers using a ciphertext only attack. Also, Uddin M.

(Uddin. and Youssef,. 2006b) investigate the use of Ant Colony Optimization (ACO) for automated cryptanalysis of simple substitution ciphers using the ciphertext only attack. It given the noticeable accuracy gain of the bi-gram based attack as compared to the unigram based one. Hilal H. (Hilal. Salih, Ahmed. Sadiq, Ismail. Ali,*at al*., 2010) have carried out interesting studies on the use of modification particle swarm optimization with 2-opt technique for the cryptanalysis of mono-alphabetic substitution cipher. Ahmed T.(Ahmed T.,2012) present a benefit developed PSO using the mutation operator, so it called Mutation PSO (MPSO). The benefit of mutation in PSO is use as momentum and diversity tool in the population. MPSO used to attack the two types of classical cryptography (substitution and transposition. Aditi,

(Aditi Bhateja , Shailender Kumar, Ashok. Bhateja,*at al,*. 20013), In this paper they have investigated the use of PSO for the cryptanalysis of vigenere cipher and proposed PSO with Markov chain random walk in which some of the worst particles are replaced with new better random particles to enhance the efficiency of PSO algorithm.

3. Simple Substitution Cipher

Simple substitution cipher is sometimes referred to as the monoalphabetic substitution cipher to distinguish it from the polyalphabetic substitution cipher. Each symbol in the plaintext maps to a (usually different) symbol in the ciphertext. If the plaintext is English, then the simple substitution cipher consists of 26! $\approx 2^{88}$ possible keys. On an average, an attacker has to try 2^{87} to break the cipher using the brute force approach. Suppose the attacker can test 240 keys per second, then the above key can be exhausted in $2^{87}/2^{40} = 2^{47}$ seconds, or about 4.4 million years. This implies that if the key space is big, then the brute force approach is highly impractical on a simple substitution cipher (http://en.wikipedia.org/wiki/Classical Cipher. download in 10/1/2013). This number is far too large to allow a brute force attack even on the fastest of today's computers. However, because of the properties of the simple substitution cipher they are relatively easy to cryptanalysis (Henery and Piper 1982; Douglas, 1995). For example, the string

"THE MONEY IS IN THE BAG" is encrypted using a simple substitution cipher key and encryption operation is shown in Table 1.

Table 1: Example of Substitution Cipher

KEY	
Plaintext	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Ciphertext	PQOWIEURYTLAKSJDHFGMZNXBCV
ENCRYPTION	
Plaintext	T H E M O N E Y I S I N T H E B A G
Ciphertext	M R I K J S IC Y G Y S M R I Q P U

4. Bees Algorithm (BA)

Bees Algorithm (BA) is a population-based search algorithm first developed in 2005(Pham *et al*, 2005). It mimics the food foraging behavior of swarms of honeybees. In its basic version, the algorithm performs a kind of neighborhood search combined with random search and can be used for both combinatorial optimization and functional optimization. Then the fitness's of the sites visited by the scout bees are evaluated and Bees that have the highest fitness's are chosen as "selected bees" and sites visited by them are chosen for neighborhood search. Then, the algorithm conducts searches in the neighborhood of the selected sites, assigning more bees to search near to the best e sites. Searches in the neighborhood of the best e sites are made more detailed by recruiting more bees to follow them than the other selected bees. Together with scouting, this differential recruitment is a key operation of the BA. The remaining bees in the population are assigned randomly around the search space scouting for new potential solutions. These steps are repeated until a stopping criterion is met. At the end of each iteration, the colony will have two parts, those that were the fittest representatives from a patch and those that have been sent out randomly. The algorithm performs a kind of neighborhood search combined with random search and can be used for both combinatorial and functional optimization (Pham *et al.,* 2006).

The BA requires a number of parameters to be set, namely: the number of scout bees (*n*), the number of patches selected out of *n* visited points (*m*), the number of elite patches out of *m* selected patches (*e*), the number of bees recruited for the best e patches (*nep*), the number of bees recruited for the other (*m-e*) selected patches (*nsp*) and the size of patches (*ngh*) including termination criterion. The pseudo code for the bee's algorithm in its simplest form (Pham *et al.,* 2006):

1. Initialize population with random solutions.
2. Evaluate fitness of the population.
3. While (stopping criterion not met) // Forming new population
4. Select sites for neighborhood search.
5. Recruit bees for selected sites (more bees for best e sites) and evaluate fitness's.
6. Select the fittest bee from each patch.
7. Assign remaining bees to search randomly and evaluate their fitness's.
8. End While.

5. Simulated Annealing (SA)

Simulated Annealing (SA) is a combinatorial optimization search technique first introduced in 1983 by Kirkpatrick, (Kirkpatrick S., Gelatt C.D. and Vecchi M.P, 1983). This technique inspired by the cooling processes of molten metal's. It merges hill climbing with the probabilistic acceptance of non-improving moves. The search starts at some initial state $S := S^0$. There is a control parameter T known as the temperature. This starts 'high' at T^0 and is gradually lowered. At each temperature, a number *Mil* (Moves in Inner Loop) of moves to new states are attempted. A candidate state Y is randomly selected from the neighborhood $N(S)$ of the current state. The change in value, ΔE, of f is calculated. If it improves the value of $f(S)$ (i.e., if $\Delta E < 0$ for a minimization problem) then a move to that state is taken is taken ($S = Y$); if not, then it is taken with some probability. The worse a move is, the less likely it is to be accepted. The lower the temperature T, the less likely is a worsening move to be accepted. Probabilistic acceptance is determined by generating a random value U in the range [0, 1] and performing the indicated comparison. Initially the temperature is high and virtually any move is accepted.

As the temperature is lowered, it becomes ever more difficult to accept worsening moves. Eventually, only improving moves are allowed and the process becomes 'frozen'. The algorithm terminates when the stopping criterion is met. Generally, the best state achieved so far is also recorded (since the search may actually move out of it and subsequently be unable to find a state of similar quality). At the end of each inner loop, the temperature is lowered. The simplest way of lowering the temperature is to multiply by a constant cooling factor μ in the range (0..1); this is known as geometric cooling. Figure 1 shows outline of the basic simulated annealing algorithm (Kirkpatrick S., Gelatt C.D. and Vecchi M.P,1983):

```
S: = S^0 ; T: = T^0 ;
Repeat
{
   For (int i = 0; i < Mil; i + +)
  {
      Select Y € N(S);
      ΔE:= f(Y) − f(S);
     If (ΔE < 0) then S: = Y
   Else
       Generate U: = rand [0, 1];
      If (U < exp (- ΔE /T)) then S: = Y;
   }
  T = T × μ;
}
Until stopping criterion is met
```

Figure 1: The Basic Simulated Annealing Algorithm

6. Proposal Hybrid BA-SA for Cryptanalysis Simple Substitution Cipher

The implementation part of our cryptanalysis of a simple substitution cipher involves a different phases that is will be described below:

6.1 Initial Population

Population initialization is a crucial task in evolutionary algorithms, which affects the convergence speed as well as the quality of the final solution. When no information about the solution is available, then random initialization is the commonly used method to generate candidate initial population. For the initialization process either use some heuristics different alphabetic strings, or initialize the swarm by a random sample of permutation of {A, B, ... , Z}.

6.2 Fitness Function Calculation

The fitness function is the main factor of the algorithm. The choice of fitness measure depends entirely on the language characteristics must be known. The technique used by Nalini (Nalini, 2006) to compare candidate key n-gram statistics of the decrypted message with those of the language (which are assumed known). Equation 1 is a general formula used to determine the suitability of a proposed key (k). Here, Ã denotes the language alphabet (i.e., for English alphabet [A... Z]. *K* and *D* denote known language statistics and decrypted message statistics, respectively, and u/b/t is the unigram, bigram and trigram statistics. The values of α, β and γ allow assigning of different weights to each of the three n-gram types where $\alpha + \beta + \gamma = 1$.

$$\textbf{\textit{Fitness}} = \alpha \sum_{i \in A} \left| K_i^u - D_i^u \right| + \beta \sum_{i,j \in A} \left| K_{i,j}^b - D_{i,j}^b \right| + \gamma \sum_{i,j,k \in A} \left| K_{i,j,k}^t - D_{i,j,k}^t \right| \qquad \textbf{(1)}$$

When trigram statistics are used, the complexity of Equation 1 is O (N^3) where *N* is the alphabet size. Therefore, it is an expensive task to calculate the trigram statistics. Hence, we will use assessment function based on unigram and bigram statistics only. Equation 2 is a formula that used as fitness function for this work.

Where according to this work the best values of α and β are 0.7 and 0.3 respectively.

$$\textbf{\textit{Fitness}} = \textbf{1-} \left(\alpha \sum_{i \in A} \left| K_i^u - D_i^u \right| + \beta \sum_{i,j \in A} \left| K_{i,j}^b - D_{i,j}^b \right| \right) \qquad \textbf{(2)}$$

6.3 Hybrid BA-SA in the Cryptanalysis Substitution Cipher

Exploration and exploitation are the two important aspects in evolutionary computing paradigms. Exploration is the ability to search the solution space to find promising new solutions, and exploitation is the ability to find the optimum solution in the neighborhood of a good solution. Generally, evolutionary algorithms have two general drawbacks. One is that have problem dependent performances. This drawback is usually tied to internal parameters controlling the behavior of the chosen technique. No optimal parameter setting applies to all problems and tuning these parameter settings can result in large performance variances. The other major drawback is premature convergence, as individuals in the population becomes too similar and loses population diversity. The problem however is that all individuals can exchange information and good solutions (possible local optima) thus possibly attract attention too fast directing the search away from even better solutions (a possible global optimum), in other words the population converges prematurely. Keeping population diversity can thus be vital for searching the possible solutions. Too much population diversity can however also be a problem

resulting in individuals wasting time investigating poor solutions or individuals unable to reach a fine-grained result.

In this paper, a Hybrid technique is proposed, which combines the merits of BA and SA to breaking the key of simple substitution ciphers; this integration is (BA-SA). The purpose of BA-SA is to use a conventional annealing approach during the chosen decision to guide the search process towards a more optimal solution space, in other words, to allow for search intensification to occur by manipulating the cooling schedule of simulated annealing, BA-SA practitioner can exercise control over convergence. On the other hand, local search algorithms suffer from the problem of finding a good starting solution; the artificial BA provides these solutions, since simulated annealing is a time-consuming process to wait during any cooling step until the equilibrium distribution of states is reached. In particular, SA knows little about whether a region in the configuration space has been explored or whether a region is a good place to search.

6.4 Proposal Hybrid Algorithm (BA-SA)

The following is an algorithmic description of the attack on a simple substitution cipher using BA-SA:

Input: The cipher text, the statistics of the language (unigrams, bigrams), the algorithm parameters (n, m, e, nep, nsp, Max_Iter,T,Mil, μ).

Output: The key having the highest fitness as fond by BA-SA.

Step 1: Randomly generate the initial bees (keys of the simple substitution cipher) to form a population.
Step 2: Calculate the fitness function of each of the bees (keys) using equation 2.

$$\textit{Fitness} = 1 - \left(\alpha \sum_{i \in A} \left| K_i^u - D_i^u \right| + \beta \sum_{i,j \in A} \left| K_{i,j}^b - D_{i,j}^b \right| \right)$$

Step 3: Repeat
Step 4: Select sites for neighborhood search.
Step 5: Perform Improved SA algorithm.
Step 6: Recruit bees for selected sites (more bees for best e sites) and evaluate fitness's.
Step 7: Select the fittest bee from each patch.
Step 8: Assign remaining bees to search randomly and evaluate their fitness's.
Step 9: Until stopping criterion is met.
Step10: Copy the best key obtained so far in the output key variable and exit.

6.5 Parameters Selection

The appropriate values of the BA and SA parameters are obtained by examining different values of these parameters and making many trial and error runs. The algorithm tested includes a substantial number of settings where it would be difficult to treat them simultaneously. Table 2 shows the values of the parameters adopted for BA, where the values decided empirically. The parameters of SA algorithm are summarized in Table 3.

Table 2: Parameters selection of BA

Parameters	Symbol	Value
Number of scout bees	*n*	30 – 50
Number of sites selected out of n visited sites	*m*	4 – 7
Number of best sites out of m selected sites	*e*	1
Number of bees recruited for best e sites	*nep*	7
Number of bees recruited for the other(*m*-*e*) selected sites	*nsp*	3
Number of iterations	*Iter*	500-1000

Table 3: Parameters selection of SA algorithm

Parameters	Symbol	Value
Max temperature	*T*	1000
Cooling factor	μ	0.95
Moves in inner loop	*Mil*	10

7. Experimental Results

BA-SA was used for cryptanalysis of simple substitution ciphers. We summarize the results of these experiments in this section. When constructing a key for a simple substitution cipher, there are 26 choices of letters to substitute for a, then 25 remaining letters that can be substituted for b, then 24 remaining letters that can be substituted for c, etc. This results in 26×25×24×…….×3×2×1=26! possible keys. In fact, there are 26! = 403,291,461,126,605,635,584,000,000 possible keys.

The experiments of proposed hybrid algorithm were operated on diverse texts encrypted using simple substitution ciphers with three kinds of keys: simple, as Cesar's keys or average as keyword key and more difficult as that of mixed key. The attack was run a number of times with a variety of parameter values. In general it was found that (500) iterations were usually enough to break the ciphertext and the algorithm was fast enough that this took a few seconds. However, it is very difficult to ensure that trajectories of the swarms do not intersect during the execution. The experiment shown in Figure 2 was performed on simple substitution cipher using proposed algorithm and the X-axis here shows the number of iteration. The value of fitness function used has started from about 0.20 and come up to 0.72. In addition, we can see a rapid increase in fitness function at the beginning and the rate of increases as we run the experiments for longer period. This is because as the fitness value increases, it becomes more and more difficult to find a better key.

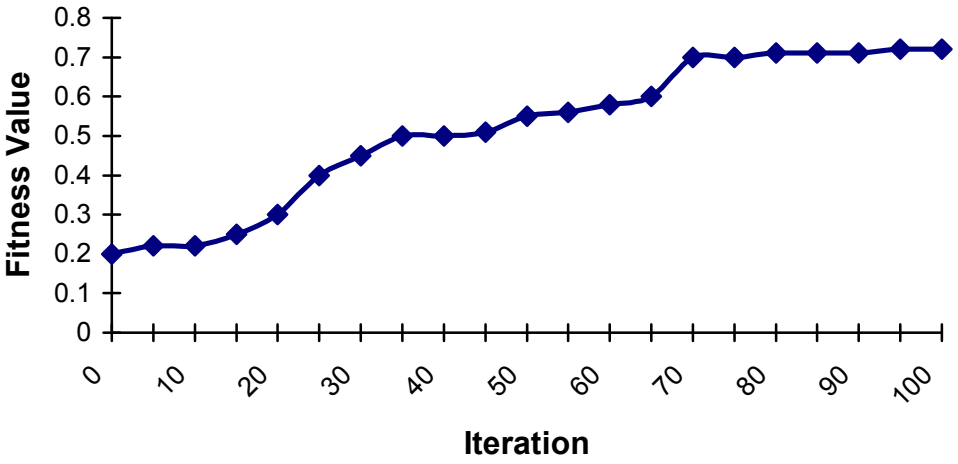

Figure 2: Performance of BA-SA

Table 4 shows the compares the average number of key elements correctly recovered for a key size of 26 versus the amount of ciphertext, which is assumed known in the attack. Due to limited length of ciphertext none of the key is true key. Still a large portion of the cipher text be decrypted correctly, the message was almost readable.

Table 4: The amount of key recovered versus available ciphertext substitution key size of 26

Amount of Ciphertext Length	Average Number of Key Recovered
200	11.50
400	14.70
600	18.55
800	20.48
1000	21.67
1500	22.75
2000	24.13

Figure 3 shows the performance of BA-SA. Thus, BA-SA is a very promising approach for solving the problem of cryptanalysis of simple substitution cipher and any discrete optimization problem in general.

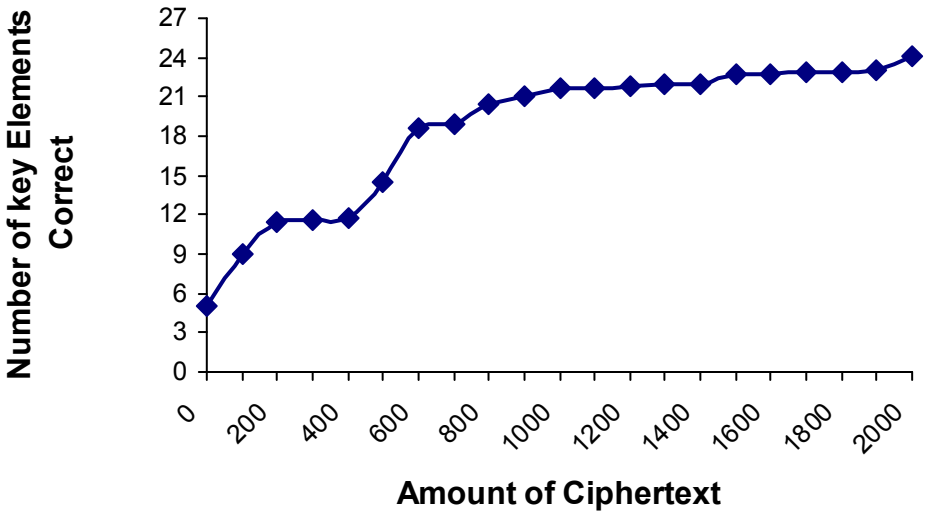

Figure 3: A Comparison Based on the Amount of Ciphertext

8. Conclusions

In this study, we have proposed an optimization technique based on the hybridization of the Bees Algorithm with Simulated Annealing to cryptanalysis of simple substitution ciphers. The proposed algorithm focuses mostly on exploration in its early search stages and allows higher levels of exploitation only when promising areas of the search space have been well identified. The exploration capability using a central temperature is adopted from SA to control the exploration and exploitation capability of the algorithm in different stages of the search process for improving its performance. Experimental results tested on several ciphertexts demonstrate that BA-SA has good performance, few parameters need to be tuned for the best possible performance, advantage of easy implementation, and it is very efficient in finding optimal or near optimal solutions performance. For future research, the proposed HBA can be applied for cryptanalysis other more complicated cryptosystems like knapsack ciphers, block ciphers or design efficient cryptographic Boolean functions.

References

"A Classical Cipher, Substitution Ciphers", retrieved from http://en.wikipedia.org/wiki/Classical_cipher. download in 10/1/2013

Aditi Bhateja , Shailender Kumar, and Ashok K. Bhateja, 2013 ***"Cryptanalysis of Vigenere Cipher Using Particle Swarm Optimization with Markov Chain Random Walk"*** Aditi Bhateja et.al / International Journal on Computer Science and Engineering (IJCSE) ISSN : 0975-3397 Vol. 5 No. 05 May.

Ahmed Tariq, 2012. ***"Mutation-Based Particle Swarm Optimization (MPSO) to Attack Classical Cryptography Methods"*** Journal of Advanced Computer Science and Technology Research 2, pp: 50-65,

Clark. A., 1998. ***"Optimization Heuristics for Cryptology"***, PhD Thesis, Security Research Centre, Faculty of Information Technology, Queensland University of Technology, Australia, February

Douglas R., 1995. ***"Cryptography: Theory and Practice"***, CRC Press, Boca Raton, Florida, USA,

Henry B. and Fred P., 1982. ***"Cipher Systems: the Protection of Communications"***, Wiley-Inter science, London, UK,

Hilal H. Salih, Ahmed T. Sadiq, and Ismail K. Ali, 2010. ***"Attack on the Simple Substitution Ciphers Using Particle Swarm Optimization"***. Engineering & Technology Journal – University of Technology, Vol.28, No.11,

Kirkpatrick S., Gelatt C.D. and Vecchi M.P., 1983. ***"Optimization by Simulated Annealing"***, Science Vol: 220, No: 4598, pp: 671-680,

Kusum D. and Millie P. 2013.,***"Swarm Intelligence for Global Optimization"***, IEEE Symposium Series on Computational Intelligence, Singapore, 15-19 April.

Nalini, 2006 ***"Cryptanalysis of Simplified Data Encryption Standard via Optimization Heuristics"***, International Journal of Computer Sciences and network security, vol 6, No 1B, Jan.

Pham D.T., Ghanbarzadeh A., Koc E., Otri S., and Zaidi M., 2006. ***"The Bees Algorithm–A Novel Tool for Complex Optimization Problems"***, in second Virtual International Conference on Intelligent Production Machines and Systems (IPROMS 2006), Elsevier. Cardiff, UK. pp: 454-459.,

Pham D.T., Ghanbarzadeh A., Koc E., Otri S., Rahim S. and Zaidi, 2005. ***"The Bees Algorithm"***, Technical Note, Manufacturing Engineering Centre, Cardiff University, UK,

U.S. Department of Commerce/National Bureau of Standard**s** ***"Data Encryption Standard"***, Federal Information Processing Standards Publication 46-1,

Uddin, M. F. and Youssef, A. M., 2006 ***"An Artificial Life Technique for the Cryptanalysis of Simple Substitution Ciphers"***, to appear in Proc. Of IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2006) Ottawa, May.

Uddin, M. F. and Youssef, A. M., 2006. "***Cryptanalysis of Simple Substitution Ciphers Using Particle Swarm Optimization"***, IEEE Congress on Evolutionary Computation, Vancouver, BC,Canada, July 16-21,