GENETIC ALGORITHMS, TABU SEARCH AND SIMULATED ANNEALING: A COMPARISON BETWEEN THREE APPROACHES FOR THE CRYPTANALYSIS OF TRANSPOSITION CIPHER

POONAM GARG

Institute of Management Technology, INDIA

E-mail: pgarg@imt.edu

ABSTRACT:

Due to increasing incidents of cyber attacks, the demand for effective internet security is increasing. Cryptology is the science and study of systems for secret communication. It consists of two complementary fields of study: cryptography and cryptanalysis. In this paper, we propose a cryptanalysis method based on genetic algorithm, tabu search & simulated annealing to break a transposition cipher. We will also compare and analyze the performance of these algorithms in automated attacks on a transposition cipher. A generalized version of these algorithms can be used for attacking other cipher as well.

Keywords: *Transposition Cipher, Genetic Algorithm, Tabu Search, Simulated Annealing, Key search.*

1. INTRODUCTION

The demand for effective internet security is increasing exponentially day by day. Businesses have an obligation to protect sensitive data from loss or theft. Such sensitive data can be potentially damaging if it is altered, destroyed, or if it falls into the wrong hands. So they need to develop a scheme that guarantees to protect the information from the attacker.

Cryptology is at the heart of providing such guarantee. Cryptology is the science of building and analyzing different encryption and decryption methods. Cryptology consists of two subfields; Cryptography & Cryptanalysis. Cryptography is the science of building new powerful and efficient encryption and decryption methods. It deals with the techniques for conveying information securely. The basic aim of cryptography is to allow the intended recipients of a message to receive the message properly while preventing eavesdroppers from understanding the message. Cryptanalysis is the science and study of method of breaking cryptographic techniques i.e. ciphers. In other words it can be described as the process of searching for flaws or oversights in the design of ciphers.

Classical ciphers fall into one of the two broad categories: substitution cipher & transposition cipher. Modern crypto-systems have now supplanted the classical ciphers but cryptanalysis of classical ciphers is most popular crypto logical application for meta-heuristic search research. The basic concepts of substitution and transposition are still widely used today in the Advanced Encryption Standard(AES). Advanced Encryption Standards(AES) & International Data Encryption Algorithm (IDEA) are widely used encryption algorithm which use only three very simple operators, namely substitution, permutation(transposition) and bit-wise exclusive-OR operator. Since the operations of the classical cipher are the building blocks of modern ciphers, so the classical ciphers are usually the first ones considered when researching new attacks.

Transposition cipher simply performs some permutation over the plain text alphabet. Transposition cipher works by breaking a message into fixed size blocks, and then permuting the characters within each block according to a fixed permutation; say $\prod$ the key to the transposition cipher is simply the permutation $\prod$. The transposition cipher has the property that the encrypted message (i.e., the ciphertext) contains all the characters that were in the plaintext message, albeit in a different (and hopefully meaningless) order. Example of the transposition cipher is presented in Table 1.

The size of the permutation is known as the period. In this example a transposition cipher with a period of six is used. Let $\prod$ = {4, 2, 1, 5, 6, 3}. Then the message is broken into blocks of six characters. Upon encryption the fourth character in the block will be moved to position 1, the second remains in position 2, the first is moved to position 3, the fifth to position 4, the sixth to position 5 and the third to position 6. The message used in the previous two examples is now used to illustrate this process on a number of blocks.

KEY:

Plaintext: 123456	
Ciphertext	: 421563

ENCRYPTION:

Position:	123456 123456 123456
Plaintext:	HELP_I_AM_LOST_SIR
Ciphertext:	PEH_IL_A_LOMSTSIR_

Table 1 : *Example transposition cipher key and encryption.*

Giddy and Safavi-Naini[12] have published an attack on the transposition cipher using simulated annealing and Methew[8] presented an attack on transposition cipher using genetic algorithm. Werner R. Grundlingh[14] presented an attack on the simple cryptographic cipher using genetic algorithm. Russell [11] presented a attack on transposition cipher using ant colony. R. Toemeh[13] presented an attack on transposition cipher using genetic algorithm.

This paper introduces an attack on the transposition cipher using genetic algorithm, tabu search & simulated annealing. The previously published attacks were enhanced and modified in order that an accurate comparison of three techniques could be obtained.

All experiments presented in this paper were performed on text using 27 character alphabet, i.e A-Z, and the _ character. All punctuation and structure (sentences/paragraphs) has been removed from the text before encryption.

2. METHODOLOGY: GENETIC ALGORITHM , TABU SEARCH & SIMULATED ANNEALING ATTACK ON TRANSPOSITION CIPHER

The technique used to compare candidate key is to compare 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 ,K is known as language Statistics i.e for English, [A,…….,Z_], D is the decrypted message statistics, and u/b/t are 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

$$C_k \approx \alpha.\sum_{i\in A}|K^u_{(i)} - D^u_{(i)}| + \beta.\sum_{i,j\in A}|K^b_{(i,j)} - D^b_{(i,j)}| + \gamma.\sum_{i,j,k\in A}|K^t_{(i,j,k)} - D^t_{(i,j,k)}| \quad (1)$$

Since the unigram frequencies for a message are unchanged during the encryption process of a transposition cipher and so, they are ignored when evaluating a key. And it is also an expensive task to calculate the trigram statistics. Hence we will use assessment function based on bigram statistics only. In the process of determining the cost associated with a transposition cipher key the proposed key is used to decrypt the ciphertext and then the statistics of the decrypted message are then compared with statistics of the language.

2.1 Genetic algorithm attack

The genetic algorithm is based upon Darwinian evolution theory. The genetic algorithm is modeled on a relatively simple interpretation of the evolutionary process, however, it has proven to a reliable and powerful optimization technique in a wide variety of applications [2]. Holland [7] in 1975, was first proposed the use of genetic algorithms for problem solving. Goldberg and Dejong[1] were also pioneers in the area of applying genetic processes to optimization. Over the past twenty years numerous application and adaptation of genetic algorithms have appeared in the literature.

An attack on the transposition cipher using Genetic Algorithm is presented here. The mating i.e. reproduction technique used here for creating the two children is given in Figure 1.

1. Notation: p1 and p2 are parents, c1 and c2 are the children, $p_i(j)$ denotes the element j in parent i, $c_i(j)$ denotes element j in child i, $\{C_i^{j,k}\}$ denotes the set of elements in child i from position j to k with the limitation that if k =0 or j = P+1 then $\{C_i^{j,k}\}=\{\phi\}$,
   2. Child 1 :

```
        a)    Choose a random number $r \in [1,P]$
        b)    $c_i(j)= p_i(j)$ for j=1,……,r
        c)    for i=1..........P-r and k=1,......,P
              if $p_2(k) \notin \{C^{l,i+r-1}\}$
                then
                  $c_i(i+r)=p_2(k)$
          else
              k=k+1
  2.  Child 2 :
        a)    Choose a random number $r \in [1,P]$
        b)    $C_2(j)= p_i(j)$ for j=P,……,r
        c)    for i=1..........r and k=P,…….,1
                if $p_2(k) \notin \{C^{r-i+1,P}\}$then
                   $c_2(r-i)=p_2(k)$
              else
                 k=k-1
```

Figure 1 : *A mating process*

A mutation operation randomly selects two elements in the child and swaps those elements. Figure 2 describes the attack on Transposition cipher using genetic algorithm.

1. Input to the algorithm given : The ciphertext (and its length), the statistics of the language (unigrams, bigrams and trigrams).
2. Initialize the algorithm parameters. The *solution* pool size M, and the maximum number of iterations MAX.
3. Generate an initial pool of solutions randomly P^{CURR}, and calculate the cost of each of the solutions in the pool using Equation 1
4. For i = 1,.........., MAX, do
   (a) Select $M/2$ pairs of keys from $P^{(CURR)}$ to be the parents of the new generation.
   (b) Perform the mating operation described in Figure 1 on each of the pairs of parents to produce a new pool of solution $P^{(NEW)}$.
   (c) For each of the M children perform a mutation operation described above.
   (d) Calculate the suitability of each of the children in $P^{(NEW)}$ using Equation 1
   (e) Sort $P^{(NEW)}$ from most suitable (least cost) to lease suitable (most cost).
   (f) Merge $P^{(CURR)}$ with $P^{(NEW)}$ to give a list of sorted solutions (discard duplicates) the size of this list will be between M and 2M. Choose the M best solutions from the merged list to become the new $P^{(CURR)}$.
5. Output the best solution from $P^{(CURR)}$.

Figure 2: *A genetic algorithm attack on transposition cipher*

2.2 Tabu search attack

The basic concept of Tabu Search is described by Glover [5- 6] in 1989 for solving combinatorial optimization problem. It is kind of iterative search and is characterized by the use of a flexible memory. It is able to eliminate local minima and to search beyond the local minimum. Therefore it has the ability to find the global minimum multimodal search space.

Transposition cipher can also be attacked using a tabu search. The overall algorithm is described in Figure 3.

1. The inputs to the algorithm are the known (intercepted) ciphertext , key size, P.
2. Set maximum number of iterations MAX, the size of the tabu list S_TABU, and S_POSS, the size of the possibilities list. Initialize the tabu list with a list of random and distinct keys and calculate the cost associated with each of the keys in the tabu list.
3. For iteration i =1……………. MAX, do
   (a) Find the best key in the tabu list which has the lowest cost associated with it. Call this key K^{BEST}.
   (b) For j = 1………, S_POSS, do
      i. Choose $n1,\ n2 \in [1,N]\ n1 \neq n2$.
      ii. Create a possible new key K^{NEW} by swapping the elements n1 and $n2$ in K^{BEST}.
      iii. Check that K^{NEW} is not already in the list of possibilities for this iteration or the tabu list. If it is return to Step 3(b)i.
      iv. Add K^{NEW} to the list of possibilities for this iteration and determine its cost.
   (c) From the list of possibilities for this iteration find the key with the lowest cost – call this key P^{BEST}.
   (d) From the tabu list find the key with the highest cost – call this key T^{WORST}.
   (e) While the cost of PBEST is less than the cost of T^{WORST}
      i. Replace T^{WORST} with P^{BEST}.
      ii. Find the new P^{BEST}.
      iii. Find the new T^{WORST}.

4. Output the best solution (i.e., the one with the least cost) from the tabu list.

Figure 3 : *A tabu search attack on transposition cipher*

2.3 Simulated annealing attack

In 1983 Kirkpatrick [3] proposed an algorithm which is based on the analogy between the annealing of solids and the problem of solving combinatorial optimization problems.

An attack on the transposition cipher using simulated annealing is described in Figure 4.

1 . Set the initial temperature, $T^{(0)}$.

2. Generate an initial solution - arbitrarily set to the identity transformation (could be randomly generated or otherwise).

3. Evaluate the cost function for the initial solution. Call this $C^{(0)}$.

4. For temperate T do many (eg., 100 $\times M$) times:

- Generate a new solution by modifying the current one in some manner
- Evaluate the cost function for the newly proposed solution.
- Consult the Metropolis function to decide whether or not the newly proposed solution will be accepted.
- If accepted, update the current solution and its associated cost.
- If the number of accepted transitions for temperature T exceeds some limit (eg. 10 $\times M$) then jump to Step 5.

5. If the number of accepted transitions for temperature T was zero then stop (return the current solution as the best), otherwise reduce the temperature (eg. $T^{(i+1)} = T^{i} \times 0.95$ and return to step 4

Figure 4 : *A tabu search attack on transposition cipher*

3. RESULTS AND DISCUSSIONS

Experimental results for the three techniques were generated with 100 runs per data point using ‘C’ language. 10 different messages were created for each run and each attack run 10 times per message. The best result for each message was averaged to produce the data point. We are comparing each of the three techniques on the basis of two criteria.

The first criterion is made upon the amount of ciphertext provided to attack. These results are presented in Table 1. Here each algorithm was run on different amount of ciphertext. Table 1 represents the average number of key elements correctly recovered for a transposition size 15. Note that because a transposition cipher key, which is rotated by one place, will still properly decrypt a large amount of the message, a key element is said to be correctly placed if its neighbors are the same as the neighbors for the correct key(except for end position). In that case, the message will almost readable, especially if the period of the transposition cipher is large. It can be seen from figure 5 the result that each of the three algorithms performed roughly equally well when the comparison is made based upon the amount of known ciphertext available to the attack.

Amount of Ciphertext	GA	TS	SA
200	7.5	6.75	7
400	12.5	9.25	10.5
600	13.5	11.6	12
800	14	12.25	13.25
1000	14.25	14	14.1

Table 1: *The amount of keys recovered versus available ciphertext, key size is 27*

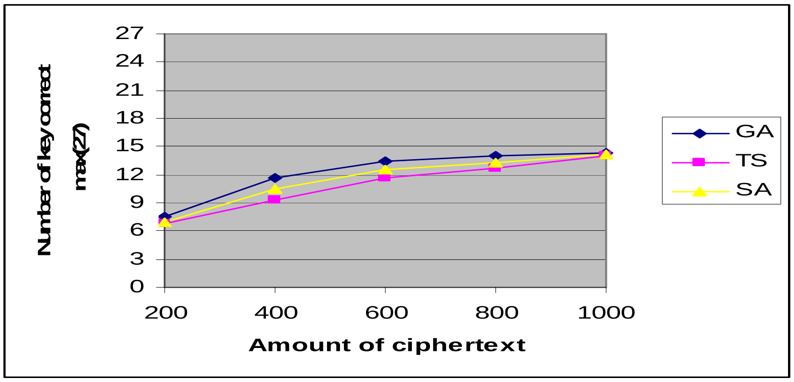

Figure 5: *The amount of keys recovered versus available ciphertext, key size is 27*

The second criterion is made upon the period of transposition cipher. It should be noted from Table 2 that for the period less than fifteen, with one thousand available ciphertext characters, each of the algorithm could successfully recover the key all the time.

Transposition Size	GA	TS	SA
15	14.25	14	14.1
20	16.50	16.75	16.25
25	20.75	21.5	20.75
30	25.25	26	25.5

Table 2: *The amount of keys recovered versus transposition size, 1000 known ciphertext character*

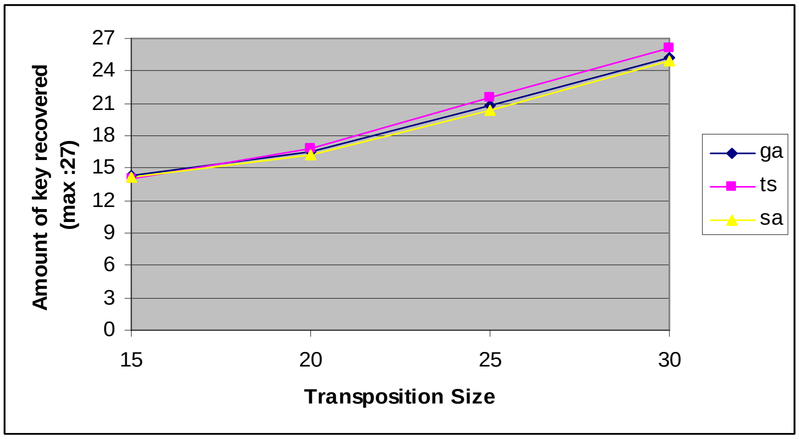

Figure 6: *The amount of keys recovered versus transposition size, 1000 known ciphertext character*

The Figure 6 shows that the tabu search attack is most powerful. For a transposition cipher of period 30 the tabu search attack is able to correctly place 26 of the key element, on an average.

4. CONCLUSION

The paper explains genetic algorithm, tabu search & simulated annealing attack on the transposition cipher. This paper has developed the theory and presented a number of automated attacks against transposition ciphers. The principles used in transposition ciphers form the foundation for many of the modern cryptosystems. The first performance comparison was made on the average number of key elements (out of 27) correctly recorded versus the amount of ciphertext which is assumed to known in the attack for a transposition size 15. It was found that for the transposition cipher all the three algorithms performed equally w.r.t amount of known cipher text available to attack. The second comparison was made upon the period of transposition cipher. It was found that the tabu search is most powerful to find the correct solution. Result indicates that tabu search is extremely powerful technique for attack on transposition cipher.

REFERENCES

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

[2] Davis, L. , "Handbook of Genetic Algorithms", Van Nostrand Reinhold, New York ,1991.

[3] Kirkpatrick S., Gelatt C. D., Jr., and Vecchi M. P., "Optimization by simulated annealing", *Science,* 220(4598):671–680, 1983.

[4] Fred Glover, "Tabu search: A tutorial". Interfaces, 20(4): 74-94, July 1990.

[5] Glover Fred. "Tabu search part II", ORSA Journal on computing, Vol 2. ,pp. 14-32,1990

[6] Glover F., Taillard E., and Werra D.D., "A user's guide to tabu search", *Annals of Operations Research*, 41:3–28, 1993.

[7] Holland, J., "Adaptation in Natural and Artificial Systems", University of Michigan Press, Ann Arbor, 1975.

[8] Methew, R.A.J. (1993, April), "The use of genetic algorithms in cryptanalysis", *Cryptologia, 7*(4), 187-201.

[9] Poonam Garg , "Genetic Algorithm & Tabu Search attack on monoalphabetic cipher", BIMC *proceeding of Business Information Management Conference - 2005,* 322-328

[10] Richard Spillman et. al., "Use of a genetic algorithm in the cryptanalysis of simple substitution ciphers", Cryptologia, 17(1):187-201, April 1993.

[11] Russell, M.D.; Clark, J.A.; Stepney, S. "Making the most of two heuristics: breaking transposition ciphers with ants Evolutionary Computation, 2003. CEC '03. The 2003 Congress on Volume 4, 8-12 ,Page(s):2653 - 2658 Vol.4, Dec. 2003

[12] Giddy J. P and Safavi-Naini R. ,"Automated cryptanalysis of transposition ciphers", *The Computer Journal,* Vol 37, No. 5, 1994.

[13] R. Toemeh, S. Arumugam. Breaking Transposition Cipher with Genetic Algorithm // Electronics and Electrical Engineering. – Kaunas: Technologija, 2007. – No. 7(79). – P. 75–78

[14] Werner R. Grundlingh and Jan H Van Vuuren, "Using Genetic Algorithm to Break a Simple Cryptographic Cipher", Article, http://www.apprendre-en-ligne.net/crypto/bibliotheque/