Cryptanalysis of Simple Substitution Ciphers Using Particle Swarm Optimization

Name: Nguyễn Bình Long - 20176807

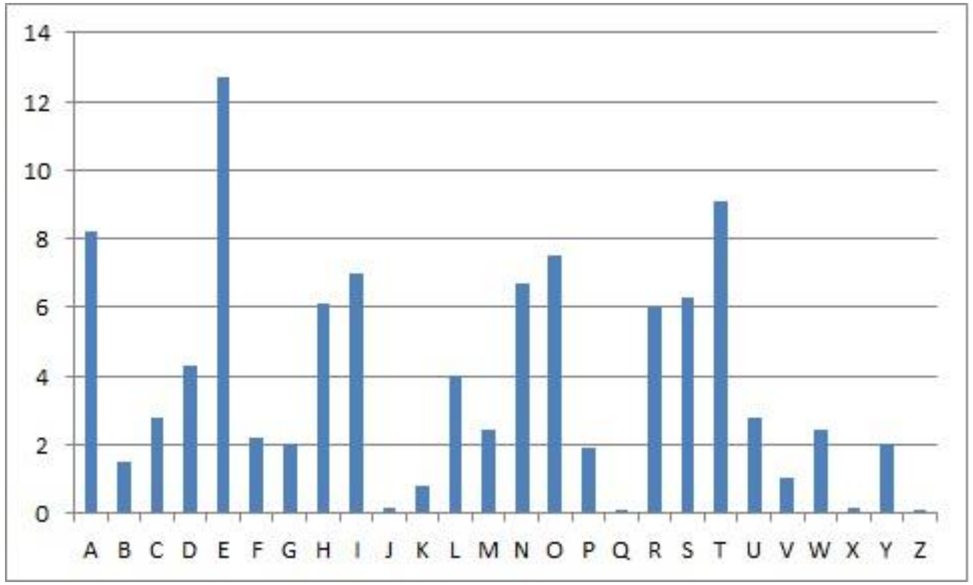
Lê Minh Hải Phong - 20170221

1. Mono-alphabetical Substitution Cipher

- The key space: all permutations of Σ = {A, B, C, …, Z}
- Encryption given a key π:
  - each letter X in the plaintext P is replaced with π(X)
- Decryption given a key π:
  - each letter Y in the ciphertext C is replaced with $\pi^{-1}(Y)$
- Example:
  - A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
  - π = B A D C Z H W Y G O Q X S V T R N M S K J I P F E U
  - SUBSTITUTION → SJASKGKJKGTV

2. Cryptanalysis of Substitution Ciphers

- Exhaustive search is infeasible since key space size is 26! ≈ $4*10^{26}$
- However, substitution ciphers preserve the language features: frequency of letters, or of groups of two or more letters.
  - vulnerable to frequency analysis attacks



2. Cryptanalysis of Substitution Ciphers

- The letters in the English alphabet can be divided into 5 groups of similar frequencies:
  - I. e
  - II. t, a, o, i, n, s, h, r
  - III. d, l
  - VI. c, u, m, w, f, g, y, p, b
  - V. v, k, j, x, q, z
- Some frequently appearing bigrams or trigrams
  - Th, he, in, an, re, ed, on, es, st, en at, to
  - The, ing, and, hex, ent, tha, nth, was eth, for, dth.

3. PSO to cryptanalyze Substitution Ciphers

- Particle swarm optimization is a population based, self-adaptive search optimization technique inspired by social behavior of bird flocking or fish schooling.
- In PSO every potential solution are called particles. Each particle keeps track of its coordinates in the search space which are associated with the best solution it has achieved so far. This value is called particle best or *pbest*.
- Another value is the best value achieved so far by any particle in the population. This value is called global best or *gbest* .

3. PSO to cryptanalyze Substitution Ciphers

- After finding the two best values each particle updates its velocity ($v_{i,j}$) and position ($P_{i,j}$) towards its *pbest* and *gbest* locations as follows:
  - *Particle velocity update*:

$$v_{i,j} = c_o v_{i,j} + c_1 r_1 \left(Ppbest_{i,j} - p_{i,j}\right) + c_2 r_2 \left(Pgbest_{i,j} - p_{i,j}\right)$$

  - *Particle position update*:

$$p_{i,j} = p_{i,j} + v_{i,j}$$

where, $Ppbest_{i,j}$ and $Pgbest_{i,j}$ are the particle best and global best position of the particles respectively.

$0 \le r_1, r_2 \le 1$ are uniformly distributed random variables

c_0, c_1, c_2 are learning factors

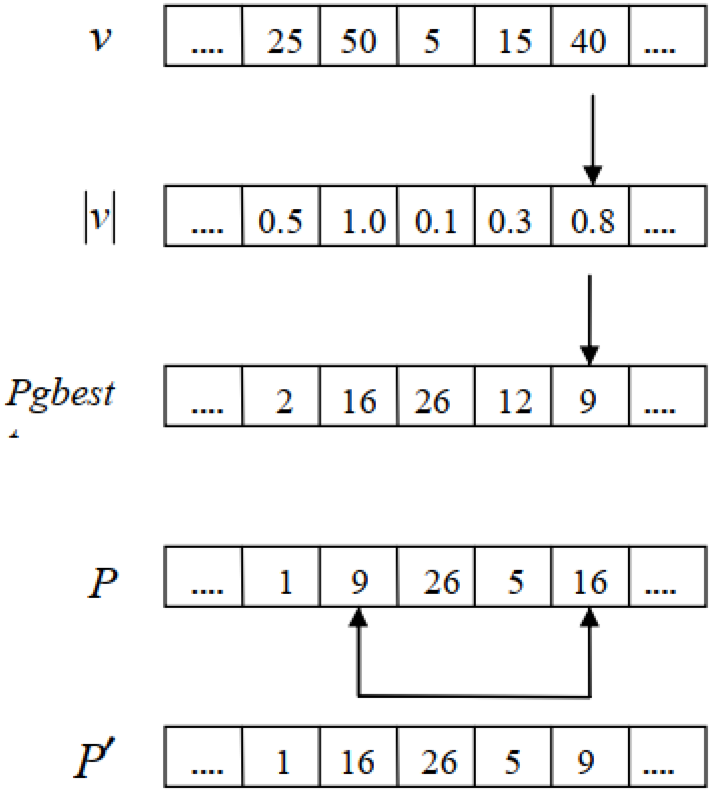
3. PSO to cryptanalyze Substitution Ciphers

- Let R^U, R^B be the reference language unigram and bigram statistics and DK^U, DK^B be the decrypted message unigram and bigram statistics (using a key K) respectively.
- Then, our cryptanalysis problem corresponds to finding a decryption key K such that minimizes the following weighted objective function:

$$Cost(K) = \alpha_1 \sum_{c \in \{A,B,..,Z\}} \left| R^U_{(c)} - DK^U_{(c)} \right| + \alpha_2 \sum_{c_1,c_2 \in \{A,B,..,Z\}} \left| R^B_{(c_1,c_2)} - DK^B_{(c_1,c_2)} \right|$$

3. PSO to cryptanalyze Substitution Ciphers

- New updating position method: the velocity is added to the particle on each dimension. When the velocity is larger, the particle is more likely to change to a new permutation sequence.



v	....	25	50	5	15	40	....
$\|v\|$	....	0.5	1.0	0.1	0.3	0.8	....
$Pgbest$	....	2	16	26	12	9	....
P	....	1	9	26	5	16	....
P'	....	1	16	26	5	9	....

Figure 1: Example for particle position update strategy as described in [17]

Thank You!