Criptanaliza criptosistemului Vigenère

În cele ce urmează, vom identifica literele $A, B, \ldots, Z$ cu numerele $0, 1, \ldots, 25$. Fie x un cuvânt peste $\{0, 1, \ldots, 25\}$ şi $k \in \{0, 1, \ldots, 25\}^m$ o cheie arbitrară, $m \geq 1$. Criptarea Vigenère a lui x cu cheia k va conduce la criptotextul y dat prin[1]

$$y_i = (x_i + k_{((i-1) \bmod m+1)}) \bmod 26,$$

pentru orice $1 \leq i \leq |x|$.

Textul obţinut din y extrăgând simboluri din n în n poziţii începând cu poziţia a j-a va fi notat cu $y_{n,j}$, pentru orice $n \geq 1$ şi $1 \leq j \leq n$. Este interesant de remarcat că, dacă y este obţinut din x prin criptare Vigenère cu cheia k, atunci are loc relaţia[2]

$$y_{m,j} = SHIFT(x_{m,j}, k_j),$$

pentru orice $1 \leq j \leq m$, unde $m = |k|$.

Pentru determinarea cheii k, având un criptotext y, se va proceda astfel:

1. Determinarea lungimii cheii (m) - folosind *testul indexului de coincidenţă*:

   Pentru un text $\alpha \in \{0, 1, \ldots, 25\}^*$, indexul său de coincidenţă, notat cu $IC(\alpha)$, reprezintă probabilitatea ca, extrăgând la întâmplare două simboluri din α, acestea să coincidă. Mai exact, $IC(\alpha)$ este dat prin

   $$IC(\alpha) = \sum_{i=0}^{25} \frac{f_i(\alpha)}{|\alpha|} \frac{f_i(\alpha) - 1}{|\alpha| - 1},$$

   unde $f_i(\alpha)$ reprezintă numărul de apariţii ale simbolului i în α.

   (a) Dacă α este un text normal în limba engleză sau un text obţinut dintr-un text normal în limba engleză extrăgând simboluri din m în m poziţii, $IC(\alpha)$ se poate aproxima prin $\sum_{i=0}^{25} p_i^2 \cong 0.065$, unde p_i reprezintă probabilitatea de apariţie a simbolului i într-un text normal de dimensiune rezonabilă din limba respectivă (engleză);

   (b) Este interesant de remarcat că criptarea folosind criptosistemul $Sh(26)$ (sau chiar $S(26)$) nu modifică acest indicator. Mai exact,

   $$IC(SHIFT(\alpha, s)) = IC(\alpha),$$

   pentru orice $0 \leq s \leq 25$.

[1] Dacă indexarea se face începând cu poziţia 0, atunci formula poate fi simplificată la $y_i = (x_i + k_{(i \bmod m)}) \bmod 26$, pentru orice $0 \leq i \leq |x| - 1$.

[2] În general, $SHIFT(\alpha, s)$ denotă şirul obţinut prin înlocuirea fiecărui simbol din α cu simbolul aflat cu s poziţii mai la dreapta în alfabet (circular). Mai exact, $(SHIFT(\alpha, s))_i = (\alpha_i + s) \bmod 26$, pentru orice $1 \leq i \leq |\alpha|$.

Astfel, $IC(y_{m,j}) = IC(x_{m,j}) \cong 0.065$, pentru $m = |k|$ şi orice $1 \leq j \leq m$. Următorul algoritm va conduce la găsirea lungimii cheii:

Determină_lungimea_cheii(y)
`input:` y, un criptotext;
`output:` m, lungimea cheii folosite;
`begin`
  $m := 0$;
  `repeat`
    $m := m + 1$;
  `until` $IC(y_{m,1}) \cong IC(y_{m,2}) \cong \cdots \cong IC(y_{m,m}) \cong 0.065$
`end`.

2. Determinarea efectivă a cheii $(k_1, \ldots, k_m)$ - folosind *testul indexului de coincidenţă mutuală*.
Pentru $\alpha, \beta \in \{0, 1, \ldots, 25\}^*$, indexul lor de coincidenţă mutuală, notat cu $MIC(\alpha, \beta)$, reprezintă probabilitatea ca, extrăgând la întâmplare câte un simbol din α şi câte un simbol din β, acestea să coincidă. Mai exact, $MIC(\alpha, \beta)$ este dat prin

$$MIC(\alpha, \beta) = \sum_{i=0}^{25} \frac{f_i(\alpha)}{|\alpha|} \frac{f_i(\beta)}{|\beta|}.$$

   (a) Dacă α şi β sunt texte normale în limba engleză (sau texte obţinute din texte normale în limba engleză extrăgând simboluri din m în m poziţii), $MIC(\alpha, \beta)$ se poate aproxima prin $\sum_{i=0}^{25} p_i^2 \cong 0.065$;

   (b) Dacă α este un text normal în limba engleză (sau un text obţinut dintr-un text normal în limba engleză extrăgând simboluri din m în m poziţii), $MIC(\alpha, \beta)$ se poate aproxima prin $\sum_{i=0}^{25} p_i \frac{f_i(\beta)}{|\beta|}$.

Deoarece $x_{m,j} = SHIFT(y_{m,j}, -k_j)$ şi $MIC(textnormal, x_{m,j}) \cong 0.065$, pentru orice $1 \leq j \leq m$, unde $m = |k|$, putem construi următorul algoritm pentru determinarea efectivă a cheii[3]:

Determină_cheia(y,m)
`input:` y, un criptotext şi m, lungimea cheii;
`output:` $k_1, \ldots, k_m$, componentele cheii folosite;
`begin`
  `for` j:=1 `to` m `do`
    `begin`
      $s := -1$;
      `repeat`
        $s := s + 1$;
      `until` $MIC(textnormal, SHIFT(y_{m,j}, s)) \cong 0.065$
      $k_j := (26 - s)$ `mod` 26;
    `end`
`end`.

[3]Pentru calcularea indexului de coincidenţă mutuală care intervine în cadrul condiţiei de la bucla `repeat-until` se va folosi aproximarea prezentată la punctul 2(b).