UNIVERSITATEA TITU MAIORESCU

Facultatea de INFORMATICĂ

Profesor univ. dr. ing.

RĂCUCIU CIPRIAN

Conf. univ. dr.ing.

ROGOBETE MARIUS

Conf. univ. dr.

NEN MADLENA

CRIPTOGRAFIE ȘI SECURITATEA INFORMAȚIEI

Curs pentru învățământ la distanță

BUCUREȘTI - 2023

Contents

CAPITOLUL 1

BAZELE MATEMATICE ALE SISTEMELOR DE SECRETIZARE

Cuvinte cheie: divizibilitate, număr prim, congruenţă, inele.

OBIECTIVE

Cunoaşterea şi înţelegerea următoarelor concepte din teoria numerelor:

- numere prime, numere prime Fermat şi Mersenne;
- congruenţe, congruenţa modulo n, congruenţa de gradul al II-lea, logaritmi discreţi în sisteme de secretizare cu chei publice;
- inele de polinoame, proprietăţile lor şi polinoame ciclotomice; funcţia lui Möbius.

1.1 Noţiuni de teoria numerelor

1.1.1 Numere prime

Având două numere naturale *m* şi *n*, spunem că *m* divide pe *n*, sau că *n* este multiplu al lui *m*, dacă există un număr natural *k* astfel încât: $n = m \cdot k$. În acest caz se scrie *m*|*n* sau *m:n*. Relaţia de divizibilitate pe *N* o vom nota cu '|' .

Pentru un număr $n \in N$ un număr $m \in N$ se numeşte *divizor* al lui *n* dacă *m*|*n*. Deoarece $n = 1 \cdot n$ şi $n = n \cdot 1$, avem *1*|*n* şi *n*|*n*, deci *1*şi *n* sunt divizori ai lui *n* pentru $\forall n \in N$. Numerele *1* şi *n* se numesc *divizori improprii* ai lui *n*, iar orice alt divizor al lui *n* se va numi *divizor propriu*. Orice număr natural *m* este divizor al lui 0, deci $0 = m \cdot 0$ şi 0=0|*m*.

Cu excepţia numărului 1 care are un singur divizor, orice număr natural $n > 1$ are cel puţin doi divizori distincţi, aceştia fiind 1 şi *n*. Un număr natural $n > 1$ care are numai doi divizori se numeşte număr *prim.*

Proprietăţi

1) Relaţia de divizibilitate este o relaţie de ordine pe *N*.

2) Pentru $\forall n \in N$ avem *n*|*n* deci relaţia este *reflexivă*.

3) Dacă *m* şi *n* sunt două numere naturale şi avem *m*|*n* şi *n*|*m*, putem scrie:

$$n = k_1 m,\ m = k_2 n \text{ unde } k_1, k_2 \in N$$

deci $n = k_1(k_2 n) = k_1 k_2 n = n(k_1 k_2)$, ceea ce arată că $n = 0$ sau $k_1 k_2 = 1$

Dacă $n=0$, atunci $m=k_2n=0=n$. Dacă $k_1k_2=1$, atunci se demonstrează că $k_1=k_2=1$, deci $m=n$. Putem spune că relaţia este *antisimetrică*.

4) Dacă *m*, *n* si *p* sunt numere naturale şi avem: *m*|*n* şi *n*|*p* atunci:

$$n=k_1m \text{ şi } p=k_2n=k_1k_2m$$

Deci *m*|*p* şi relaţia | este *tranzitivă*.

Definiţie. Un număr natural *p* >1 este număr prim dacă şi numai dacă pentru orice două numere naturale *m* şi *n* avem:

$$p=m\cdot n\Rightarrow m=1 \text{ sau } n=1.$$

Numerele prime sunt foarte importante în primul rând datorită faptului că *orice număr natural nenul se scrie în mod unic ca un produs de numere prime*. Acest rezultat, cunoscut sub numele de *teorema fundamentală a aritmeticii,* a devenit, prin generalizările care i s-au dat instrumentul de bază în multe capitole ale teoriei algebrice a numerelor şi ale algebrei abstracte. Numerele prime sunt de asemenea importante deoarece multe teoreme despre numere prime sunt uşor de formulat, dar foarte greu de demonstrat. Unele din aceste „teoreme" se dovedesc adevărate în toate cazurile accesibile calculului, prin mijloacele cunoscute până în prezent, dar aceste mijloace se dovedesc insuficiente pentru a se verifica valabilitatea generala a „teoremei" respective.

Una din primele probleme care s-a pus este dacă mulţimea numerelor prime este infinită sau nu. Răspunsul este dat de teorema lui Euclid.

Teorema: Mulţimea numerelor prime este infinită.

Demonstraţia se face foarte simplu dacă, prin reducere la absurd presupunem că mulţimea numerelor prime este finită. Fie $P=\{p_1,p_2,...,p_n\}$ această mulţime:

Considerăm numărul natural: $N=p_1\cdot p_2...p_n+1$. Deoarece *N>1* există un număr *p* astfel încât *p*|*N*. Deoarece $p\in P$, rezultă $p|p_1p_2...\ p_n$ şi deci $p|p_1p_2...\ p_n\ p|[N-(p_1p_2...\ p_n)]=1$. Aşadar *p*|*1*, ceea ce contrazice faptul că *p* este număr prim.

Legat de faptul că mulţimea numerelor prime este infinită, s-a pus problema distribuţiei acestor numere. Notând cu $\prod(x)$ numărul numerelor prime mai mici decât *x*, se pune problema găsirii unei formule de calcul pentru $\prod(x)$. Mai mulţi matematicieni au găsit experimental că: $\prod(x)\cong\frac{x}{\ln x}$, însă abia la sfârşitului secolului trecut *J. Hadammar* şi *Ch. J. de la Valle Paussin* au demonstrat că:

$$\lim_{x\to\infty}\frac{\prod(x)}{\frac{x}{\ln x}}=1$$

S-a pus şi problema dacă anumite mulţimi de numere prime, cu anumite proprietari sunt infinite sau nu. Astfel, numerele prime de forma $2^{2^n}+1$ se numesc *numere prime Fermat* iar numerele prime de forma 2^p-1, unde *p* este număr prim,

se numesc *numere prime Mersenne*. Nu se cunoaşte dacă mulţimea numerelor prime Fermat sau Mersenne este finită sau nu.

Teoremă. Dacă *a* şi *n* sunt numere naturale, $a \geq 1$ şi $n \geq 2$ astfel încât $a^n - 1$ este număr prim, atunci $a = 2$ şi *n* este număr prim.

Într-adevăr

$$a^n - 1 = (a-1)\left(a^{n-1} + a^{n-2} + \ldots + a + 1\right)$$

şi în ipoteza ca $a^n - 1$ este prim, rezultă că a – 1=1 deci $a = 2$.

Dacă *m*|*n*, deci $n = m \cdot k$ avem:

$$2^n - 1 = 2^{mk} - 1 = (2m)^k - 1 = \left(2^m - 1\right)\left[2^{m(k-1)} + \ldots + 2^m + 1\right].$$

Rezultă că: $2^{mk} - 1 = 2^m$ sau *mk* = *m*, deci *m* = *n* sau *k* = 1. Deci *n* este număr prim.

Se demonstrează că nu numai *n* este prim dacă şi numai dacă:

$\sum_{1 \leq m \leq n} \left(\left[\frac{n}{m} \right] - \left[\frac{n-1}{m} \right] \right) = 2$ unde $[x]$ = partea întreagă a lui *x*.

Se întâlnesc două situaţii:

a) *m*|*n*, deci *n* = *km* si atunci

$$\left[\frac{km}{m} \right] - \left[\frac{km-1}{m} \right] = k - k + 1 = 1;$$

b) *m* nu divide pe *n*, deci $n = km + s$; $1 \leq s \leq m$

$$\left[\frac{km+s}{m} \right] - \left[\frac{km+s-1}{m} \right] = k - k = 0.$$

Deci dacă *n* este un număr prim, atunci el are doar cei doi divizori improprii, iar daca *n* are $\tau(n)$ divizori atunci:

$$\sum_{1 \leq m \leq n} \left(\left[\frac{n}{m} \right] - \left[\frac{n-1}{m} \right] \right) = \tau(n)$$

1.2 CONGRUENŢE

1.2.1 Noţiunea de congruenţe

Definiţie. Fie *n* un număr întreg pozitiv. Pe inelul *Z* al numerelor întregi, definim o relaţie liniară, numită *congruenţa modulo n*. Dacă $a, b \in N$ spunem că *a* este congruent cu *b* modulo *n* şi scriem $a \equiv b \pmod{n}$, dacă *n*|(*a* – *b*).

Relaţia de congruenţă este o relaţie de echivalenţă. Într-adevăr:

1) Dacă $a \in N$ atunci *n*|*(a–a)* deci $a \equiv a \pmod{n}$. Deci relaţia este reflexivă.

2) Fie $a,b \in N$, atunci $n|(a–b)$ şi deci $n|-(a–b)$ sau $n|(b–a)$, adică $b \equiv a \pmod{n}$. Aşadar relaţia este *simetrică*.

3) Fie $a,b,c \in N$ astfel încât $a \equiv b \pmod{n}$ şi $b \equiv c \pmod{n}$. Atunci $n|(a–b)$ şi $n|(b–c)$ deci $n|((a–b) + (b–c))$ sau $n|(a–c)$ de unde rezultă $a \equiv c \pmod{n}$. Deci relaţia este *tranzitivă*.

Pentru $a \in N$ notăm cu: $d = \{b \in N \mid a \equiv b(\mathrm{mod}\, n)\}$ clasa de echivalenţă a lui a, numită clasa de resturi modulo n. Mulţimea claselor de resturi modulo n o vom nota Z_n :

$$Z_n = \left\{\hat{0},\hat{1},...,(n-1)\right\}$$

Pe mulţimea claselor de resturi modulo n se definesc operaţiile algebrice de adunare şi înmulţire a claselor de resturi modulo n în modul următor:

$$\begin{cases} a + \hat{b} = a + b \\ a \cdot \hat{b} = ab, \forall a, \hat{b} \in Z_n. \end{cases}$$

Mulţimea înzestrata cu operaţiile de adunare şi înmulţire a claselor de resturi formează un inel unitar şi comutativ. Aceasta se verifică simplu prin proprietăţile:

1. $\left(a + \hat{b}\right) + \hat{c} = a + \left(\hat{b} + \hat{c}\right)$;
2. $a + \hat{b} = \hat{b} + a$;
3. $a + \hat{0} = \hat{0} + a$;
4. $a + \left(-a\right) = \left(-a\right) + a = \hat{0}$;
5. $\left(a \cdot \hat{b}\right) \cdot \hat{c} = a \cdot \left(\hat{b} \cdot \hat{c}\right)$;
6. $a \cdot \left(\hat{b} + \hat{c}\right) = a \cdot \hat{b} + a \cdot \hat{c}$;
   $\left(\hat{b} + \hat{c}\right) a = \hat{b} \cdot a + \hat{c} \cdot a$;
7. $\hat{a} \cdot \hat{1} = \hat{1} \cdot \hat{a} = \hat{a}, \forall \hat{a}, \hat{b}, \hat{c} \in Z_n$.

O clasă de resturi este ireversibilă în inelul Z_n dacă şi numai dacă $(a, n) = 1$ (cel mai mare divizor comun).

Ne propunem să descriem subnivelele şi idealele inelului Z_n. Orice ideal şi orice subinel al lui Z_n este în particular un subgrup al grupului aditiv $(Z_n,+)$ al nivelului. Grupul $(Z_n,+)$ este ciclic, fiind generat de exemplu de elementul $\hat{1}$. Atunci orice subgrup al sau este de asemenea ciclic şi deci este de forma:

$$\left\langle \hat{d} \right\rangle = \left\{\hat{d}a | a \in Z\right\} \text{ unde } \hat{d} \in Z_n .$$

Un astfel de subgrup este în acelaşi timp şi subnivel şi ideal al inelului Z_n. Deci subinelele şi idealele lui Z_n coincid cu subgrupurile grupului $(Z_n,+)$ al inelului.

De exemplu să determinam idealele inelului Z_6:

$$Z_6 = \{\hat{0},\hat{1},\hat{2},\hat{3},\hat{4},\hat{5}\}.$$

Cum doar $\hat{1}$ şi $\hat{5}$ sunt inversabile în Z_6 rezultă că $\langle\hat{1}\rangle = \langle\hat{5}\rangle$. Considerând celelalte elemente din Z_6 obţinem:

$$\langle\hat{0}\rangle = \{\hat{0}\};$$
$$\langle 2\rangle = \langle 4\rangle = \{\hat{0},2,4\};$$
$$\langle\hat{3}\rangle = \{\hat{0},\hat{3}\}.$$

Deci Z_6 are patru ideale care sunt în acelaşi timp şi subinelele sale:

$$\{\hat{0}\},\{\hat{0},\hat{3}\},\{\hat{0},\hat{2},\hat{4}\}Z_6.$$

Mulţimea Z_n are n elemente. În general numerele întregi a_1, a_2,...,a_n formează un sistem complet de resturi modulo n dacă:

$$Z_n = \{\hat{a}_1,\hat{a}_2,\hat{a}_3,...,\hat{a}_1\}.$$

De exemplu mulţimea: {0,1,2,3,4,5} formează un sistem complet de resturi pentru Z_6, dar şi {– 6,7,14,9,16,–1} formează tot un sistem de resturi modulo 6.

Mulţimea claselor de resturi care sunt relativ prime cu modulul formează sistemul redus de resturi împreuna cu numărul 1. În cazul Z_6, {1,5} este sistem redus de resturi.

Mulţimea Z_n, înzestrată cu operaţiile de adunare şi înmulţire definite prin:

$$a + \hat{b} = a + b;$$
$$a \cdot \hat{b} = a \cdot b,$$

este un inel unitar şi comutativ, iar aplicaţia $p: Z \to Z_n$ definită prin $p(a) = \hat{a}, a \in Z$ este un morfism de inele surjectitve.

Intr-adevăr :

$$p(a+b) = a+b = a+\hat{b} = p(a)+p(b);$$
$$p(a\cdot b) = a\cdot b = a\cdot\hat{b} = p(a)\cdot p(b).$$

1.2.2 Congruenţa ax≡b (mod m*)*

Teoremă. Congruenţa $ax \equiv b \pmod m$ are soluţii dacă şi numai dacă d, cel mai mare divizor comun *al numerelor* a şi m ($d = (a, m)$), divide pe b.

Demonstraţie. Fie $\hat{x}_0 \in Z_m, x_0 \in Z$ o soluţie a ecuaţiei date. Atunci $ax_0 \equiv b(\mathrm{mod}\, m)$, deci $m/(ax_0 - b)$, adică există $y_0 \in Z$ astfel încât $ax_0 - b = my_0$. Dar există *d*Â*a*; *d*|*m*, rezultă $d/(ax_0 - my_0)$, deci *d*|*b*. Reciproc, să presupunem că *d*|*b*. Se ştie că, există x_0' şi y_0' numere întregi, astfel încât $ax_0' + my_0' = d$. Luând $b' = \frac{b}{d} \in Z$, avem $b = db' = a(x_0'b') + m(y_0'b')$, deci $a(x_0'b') \equiv b(\mathrm{mod}\, m)$, ceea ce arată că $x'_0\, b' \in Z_m$ este o soluţie a congruenţei date. Fie $d = (a, m)$ şi să presupunem că *d*|*b*. Atunci dacă $m' = \frac{m}{d} \in Z$ iar x_0 este o soluţie a congruenţei $ax \equiv b(\mathrm{mod}\, m)$, rezultă că $x_0, x_0 + m', x_0 + 2m', ..., x_0 + (d-1)m'$ sunt toate soluţii congruenţei date.

Evident că dacă (*a, m*)=1 atunci congruenţa $ax \equiv b$ (mod *m*) are soluţie unică. Aceasta înseamnă că dacă *m* şi *a* sunt relativ prime, atunci există totdeauna a^{-1} *(*mod *m).*

Definiţie: Un element $a \in Z_m$ este unitate în Z_m dacă există $b \in Z_m$ astfel încât *ab*=1 (mod *m*).

Se poate astfel descrie grupul $U(Z_m)$ care conţine unităţile inelului Z_m:

$$U(Z_m) = \{a_{i_1}, a_{i_2}, ..., a_{i_n}\}.$$

În acest caz numărul *n* se numeşte indicatorul Euler al lui *m* şi este egal cu numărul elementelor relativ prime cu *m* plus 1. Aceasta se notează $\varphi(m)$.

De exemplu $U(Z_{12})=\{1,5,7,11\}$, deci $\varphi(12) = 4$.

Teorema lui Euler. Pentru orice număr întreg *a* relativ prim cu *m* avem $a^{\varphi(m)} \equiv 1(\mathrm{mod}\, m)$.

Demonstraţie: Fie $\{a_1, a_2, ..., a_{\varphi(m)}\}$ un sistem redus de resturi modulo *m*. Atunci şi $aa_1, aa_2, ..., aa_{\varphi(m)}$ este de asemenea un sistem redus de resturi. Produsul $a_1a_2...a_{\varphi(m)} \equiv aa_1 \cdot aa_2 ... aa_{\varphi(m)}$ este comutativ, deci $a_1a_2...a_{\varphi(m)} \equiv a^{\varphi(m)} \cdot a_1 \cdot a_2 ... a_{\varphi(m)}$. De aici rezultă $a^{\varphi(m)} \equiv 1(\text{modulo } m)$.

Daca *m* = *p*, *p* fiind un număr prim, atunci orice număr *a* cu *p*|*a* avem (*a*, *p*) = 1 si prin urmare congruenţa $ax \equiv b(\mathrm{mod}\, m)$ are soluţie unică pentru orice $b \in Z$. În acest caz $U(Z_p) = Z_p - \{0\}$. este *corp*.

Este evident că în cazul *p* este număr prim, $\varphi(p) = p - 1$.

O consecinţă imediată a teoremei lui Euler o constituie *teorema lui Fermat*.

Pentru orice *p* număr prim şi $a \in Z$ astfel încât *p*|*a*, avem $a^{p-1} \equiv 1(\mathrm{mod}\, p)$.

Indicatorul Euler al unui număr întreg *m* este o funcţie numerică multiplicativă. Deci dacă $m = m_1 \cdot m_2$, atunci $\varphi(m) = \varphi(m_1) \cdot \varphi(m_2)$.

Dacă $m = p^{\alpha}$, p fiind număr prim, atunci în şirul $1,2,...,p,...,p^{\alpha}$ sunt $p^{\alpha} - p^{\alpha-1}$ termeni relativ primi cu p^{α}. Deci $\varphi\left(p^{\alpha}\right) = p^{\alpha} - p^{\alpha-1}$.

Pe această bază se poate calcula indicatorul Euler când $m = p_1^{\alpha_1} \cdot p_2^{\alpha_2} ... p_k^{\alpha_k}$.

$$\varphi(m) = m\left(1-\frac{1}{p_1}\right)\cdot\left(1-\frac{1}{p_2}\right)...\left(1-\frac{1}{p_k}\right).$$

De exemplu, fie $m = 120 = 2^3 \cdot 3 \cdot 5$

$$\varphi(120) = 120\left(1-\frac{1}{2}\right)\cdot\left(1-\frac{1}{3}\right)\cdot\left(1-\frac{1}{5}\right) = 120\cdot\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{4}{5} = 32.$$

Se demonstrează că $\sum_{d/n}\varphi(d) = n$.

De exemplu, pentru n = 12, avem

$$\sum_{d/n}\varphi(d) = \varphi(1)+\varphi(2)+\varphi(3)+\varphi(4)+\varphi(6)+\varphi(12) =$$
$$= 1+1+2+2+2+4 = 12.$$

Dacă notăm cu $\tau(n)$ numărul divizorilor lui n (inclusiv divizorii improprii) şi n este dat de relaţia: $n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} ... p_k^{\alpha_k}$, atunci $\tau(n)$ se determină cu ajutorul relaţiei $\tau(n) = (\alpha_1 + 1)\cdot(\alpha_2 + 1)...(\alpha_k + 1)$ şi este egal cu numărul termenilor produsului.

$$P = \left(1 + p_1 + p_1^2 + ... + p_1^{\alpha_1}\right)\left(1 + p_2 + p_2^2 + ... + p_2^{\alpha_2}\right)...\left(1 + p_k + p_k^2 + ... + p_k^{\alpha_k}\right).$$

Teorema chineză a resturilor

Fie $m_1, m_2, ..., m_s$ numere întregi cu $(m_i, m_j) = 1$ pentru orice $i, j \in \{1,2,...,s\}$, $i \neq j$ şi $b_1, b_2, ..., b_s$ numere întregi oarecare. Atunci există un număr întreg x, soluţie a sistemului de congruenţe:

$$x \equiv b_1 \pmod{m_1}$$
$$x \equiv b_2 \pmod{m_2}$$
$$.....$$
$$x \equiv b_s \pmod{m_s}$$

si, pentru orice altă soluţie y a aceluiaşi sistem avem $x \equiv y \pmod{m}$, unde $m = m_1 \cdot m_2 ... m_s$.

Demonstratie: Pentru fiecare $i \in \{1,2,...,s\}$ notăm $n_i = \frac{m}{m_i}$, cu $(n_i, m_i) = 1$.

Există numerele întregi u_i, ν_i astfel încât $u_i m_i + \nu_i n_i = 1$. Luăm $e_i = \nu_i n_i$ şi atunci este evident că $e_i = 1 \pmod{m_i}$ şi $e_i \equiv 0 \pmod{m_j}, i \neq j$. Luând $x \equiv \sum_{i=1}^{s} b_i e_i$ avem $x \equiv b_i e_i \pmod{m_j}$, deci $x \equiv b_i \pmod{m_i}$ ştiind că $e_i \equiv 1 \pmod{m_i}$.

Exemplu:

$$x \equiv 1 \pmod{7};$$
$$x \equiv 4 \pmod{9};$$
$$x \equiv 3 \pmod{5};$$

Avem:

$$m_1 = 7; b_1 = 1;$$
$$m_2 = 9; b_2 = 4;$$
$$m_3 = 5; b_3 = 3.$$

Deci

$$m = m_1 \cdot m_2 \cdot m_3 = 315;$$
$$n_1 = \frac{m}{m_1} = 45;$$
$$n_2 = \frac{m}{m_2} = 35;$$
$$n_3 = \frac{m}{m_3} = 63.$$

Soluţia generală a sistemului de congruenţe va fi $x \equiv \sum_{i=1}^{3} b_i \nu_i n_i \pmod{m}$.

Se calculează $\nu_i, (i = 1,2,3)$, astfel:

$$n_i \cdot \nu_i \equiv 1 \pmod{m_i}.$$

Deci

$$n_1\nu_1 \equiv 1(\mathrm{mod}\, m_1);$$
$$3\nu_1 \equiv 1(\mathrm{mod}\, 7);\ 45(\mathrm{mod}\, 7) = 3(\mathrm{mod}\, 7);$$
$$\nu_1 \equiv 3^{\varphi(7)-1};\ \varphi(7) = 6;$$
$$\nu_1 \equiv 3^5(\mathrm{mod}\, 7) = 5(\mathrm{mod}\, 7);$$
$$n_2\nu_2 \equiv 1(\mathrm{mod}\, m_2);\ 35(\mathrm{mod}\, 9) = 8(\mathrm{mod}\, 9);$$
$$8\nu_2 \equiv 1(\mathrm{mod}\, 9);\ \varphi(9) = 6;$$
$$\nu_2 \equiv 8^5(\mathrm{mod}\, 9) = 8\cdot 8\cdot 8\cdot 8(\mathrm{mod}\, 9);$$
$$\nu_2 \equiv 8(\mathrm{mod}\, 9);$$
$$n_3\nu_3 \equiv 1(\mathrm{mod}\, 5);\ 63(\mathrm{mod}\, 5) = 3(\mathrm{mod}\, 5);$$
$$3\nu_3 \equiv 1(\mathrm{mod}\, 5);$$
$$\nu_3 \equiv 3^3(\mathrm{mod}\, 5) = 2(\mathrm{mod}\, 5);$$
$$x = 1\cdot 5\cdot 45 + 4\cdot 8\cdot 35 + 3\cdot 2\cdot 63 = 148(\mathrm{mod}\, 315).$$

1.2.3 Congruenţa de gradul al II-lea

Fie congruenţa: $ay^2 + by + c \equiv 0(\mathrm{mod}\, s)$, $a \neq 0$.

Prin prelucrări succesive, această congruenţă poate fi adusă la forma $x^2 \equiv n(\mathrm{mod}\, m)$:

$$ay^2 + by + c \equiv 0(\mathrm{mod}\, s)/4a$$
$$4a^2y^2 + 4aby + 4ac \equiv 0(\mathrm{mod}\, 4as)$$
$$(2ay + b)^2 + 4ac - b^2 \equiv 0(\mathrm{mod}\, 4as).$$

Notăm $\begin{cases} 2ay + b \equiv x \\ b^2 - 4ac \equiv n \\ 4as \equiv m \end{cases}$ şi ecuaţia devine $x^2 \equiv n(\mathrm{mod}\, m)$.

Numărul relativ prim *m* se numeşte *rest pătratic* modulo *m*, dacă congruenţa $x^2 \equiv n(\mathrm{mod}\, m)$ are soluţii, şi *nerest pătratic* în caz contrar.

În cazul *m* = 2, congruenţa $x^2 \equiv n(\mathrm{mod}\, 2)$ are totdeauna soluţii, şi anume:

$$x = 1, \text{ pentru } n = 1;$$
$$x = 0, \text{ pentru } n = 0.$$

Pentru cazul *m* = *p* număr prim, $m \geq 3$, într-un sistem de resturi modulo *p* există $\frac{p-1}{p}$ resturi pătratice şi $\frac{p-1}{p}$ neresturi pătratice.

Stabilirea faptului că n este sau nu rest pătratic modulo p se poate face cu ajutorul simbolului lui Legendre $\left(\frac{n}{p}\right)$ care se defineşte astfel:

$$\left(\frac{n}{p}\right)=\begin{cases}1 & n \text{ rest pătratic mod } p\\ -1 & n \text{ nerest pătratic mod } p\end{cases}$$

O serie de proprietari ale simbolurilor lui Legendre permit cu uşurinţă calculul acestora:

a) $\left(\frac{n_1 n_2 \ldots n_s}{p}\right)=\left(\frac{n_1}{p}\right)\cdot\left(\frac{n_2}{p}\right)\ldots\left(\frac{n_s}{p}\right)$;

b) $\left(\frac{n_1 n_2^2}{p}\right)=\left(\frac{n_1}{p}\right)$;

c) $\left(\frac{-1}{p}\right)=(-1)^{\frac{p-1}{2}}$;

d) $\left(\frac{2}{p}\right)=(-1)^{\frac{p^2-1}{8}}$;

e) $\left(\frac{p}{q}\right)=(-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}\left(\frac{q}{p}\right)$.

Simbolul lui Legendre se utilizează în cazul în care p este număr prim. O generalizare a acesteia o reprezintă simbolul lui Jacobi care este valabil şi pentru $m = p_1 \cdot p_2 \ldots p_r$, $p_i =$ numere prime, $p_i > 2$

$$\left(\frac{a}{m}\right)=\left(\frac{a}{p_1}\right)\cdot\left(\frac{a}{p_2}\right)\ldots\left(\frac{a}{p_r}\right).$$

Pentru orice $a \equiv a_1 \pmod m$ avem $\left(\frac{a}{m}\right)=\left(\frac{a_1}{m}\right)$.

Simbolul lui Legendre se poate calcula şi cu ajutorul relaţiei lui Euler:

$$\left(\frac{n}{p}\right)\equiv n^{\frac{p-1}{2}} \pmod p.$$

Pentru această relaţie există şi o generalizare pentru resturi de ordin n şi anume: numărul a este rest de ordin n pentru modulul p atunci şi numai atunci când:

$$a^{\frac{p-1}{\mathrm{d}}} \equiv 1 \pmod p, \text{ unde } \mathrm{d}=(n, p-1)$$

Există în acest caz $\frac{p-1}{\mathrm{d}}$ resturi de ordin n faţă de modulul p.

Numărul soluţiilor congruenţei: $x^2 \equiv n(\mathrm{mod}\, m)$ îl notăm cu $N(n,m)$ şi se calculează astfel:

a) cazul $m = p^{\alpha}, \alpha \geq 1$, p număr prim:

$$N\left(n, p^{\alpha}\right) = \begin{cases} 0, \left(\frac{n}{p}\right) = -1 \\ 2, \left(\frac{n}{p}\right) = 1 \end{cases}$$

b) cazul $m = p^{\alpha}; n = p^{\beta} \cdot n_1; \alpha > \beta$:

$$N\left(p^{\beta} \cdot n_1, p^{\alpha}\right) = 0 \text{ dacă } \beta \text{ este impar;}$$

$$N\left(p^{\beta} \cdot n_1, p^{\alpha}\right) = p^{\frac{\beta}{2}} \cdot N\left(n_1, p^{\alpha-\beta}\right) \text{ pentru } \beta \text{ număr par.}$$

1.2.3 Logaritmi discreţi

Fie un sistem de secretizare cu chei publice de tip exponenţial, R.S.A., în care modulul *m* este făcut public, iar exponentul de criptare e este păstrat secret. Presupunem că un criptanalist a interceptat cel puţin o pereche $\left(M, M^e\right)$ şi încearcă să spargă sistemul, adică să găsească exponentul de decriptare *d*. Deci criptanalistul se găseşte în faţa problemei de a găsi $\log_M M^e (\mathrm{mod}\, m)$. Acesta este un caz special de calculare a logaritmilor discreţi. Tăria metodei constă tocmai în dificultatea calculării acestui logaritm. Dacă consideram ecuaţia $a^x = y$ pentru numere reale pozitive, problema găsirii lui *x* cunoscând pe *a* şi *y* este aproape aceeaşi cu *a* găsi pe *y* ştiindu-i pe *a* şi *x*. Găsirea soluţiei presupune efectuarea unor calcule mai mult sau mai puţin complicate şi căutarea în tabele de logaritmi. Problema este complet diferită dacă se lucrează cu logaritmi discreţi. Noţiunea generală de *logaritm discret* poate fi formulată după cum urmează. Fie *g* un element al unui grup finit *G* şi fie *y* un alt element din *G*. Atunci orice număr întreg pozitiv *x*, astfel încât $g^x = y$ se numeşte *logaritm_discret* al lui *y* în baza *g*. Evident, orice element *y* al lui *G* are un logaritm discret în baza *g* dacă şi numai dacă *G* este ciclic cu generatorul *g*. De exemplu, în grupul multiplicativ al numerelor întregi modulat, numai 1,2,4 au un logaritm discret în baza doi dar toate numerele au un logaritm discret în baza trei:

y	1	2	3	4	5	6	$\left(3^x = y\right)$
x	6	2	1	4	5	3	

Dacă $x \in GF\left(3^2\right)$, atunci cu ajutorul ecuaţiei $x^2 + 2x + 2 = 0$ se pot genera toate elementele câmpului $GF\left(3^2\right)$ ridicând la putere un element *x* numit element generator:

i	1	2	3	4	5	6	7	8
α^i	α	$\alpha+1$	$2\alpha+1$	2	2α	$2\alpha+2$	$\alpha+2$	1

şi atunci:

y	1	2	α	$\alpha+1$	$\alpha+2$	2α	$2\alpha+1$	$2\alpha+2$
$\log_\alpha y$	8	4	1	2	7	5	3	6

Grupurile cu cardinalitate mică nu prezintă dificultăţi de calcul. Există şi algoritmi eficienţi de calcul al logaritmilor discreţi pentru unele cazuri speciale. În general însă, calculul logaritmilor discreţi sunt consideraţi de aceeaşi dificultate ca şi algoritmii pentru factorizarea lui *m*.

Algoritmul propus de Silver, Pohlig şi Hellman, probabil cel mai bun algoritm general de acest tip, este exemplificat după cum urmează şi el lucrează atunci când factorii primi ai lui *m* sunt mici.

Fie $GF(q)$, $q = p^n$ un câmp finit. Consideram logaritmi discreţi în baza *g*, unde *g* este generator pentru $GF(q)$. Pentru fiecare divizor *d* prim al lui *q* – 1, calculăm numerele:

$$a(i,\mathrm{d}) = g^{i\left(\frac{q-1}{\mathrm{d}}\right)} (\mathrm{mod}\, q), 0 \le i < p\,.$$

Dacă fiecare *d* care îl divide pe *q* – 1 este mai mic atunci mărimea tabelei precalculate conţinând numerele auxiliare *a* (*i*, *d*) este acceptabilă, destul de uşor de realizat.

i \ α	2	3	5
0	1	1	1
1	180	48	59
2		132	42
3			125
4			135

De exemplu, *GF* (181) şi $q = 2$ (2 este într-adevăr un generator). Acum $180 = 2^2 \cdot 3^2 \cdot 5$ şi tabela numerelor *a* (*i*, *d*) arată ca mai sus.

Să calculăm acum logaritmul discret al lui 62 în baza 2. În general dacă $q - 1 = \pi \mathrm{d}^{\alpha}$, atunci pentru a găsi logaritmul discret *x* al lui *y* în baza *g* este suficient să găsim $\left(x, \mathrm{mod}\, \mathrm{d}^{\alpha}\right)$ pentru fiecare *d* din factorizarea lui *q* – 1. Folosind teorema chineză a resturilor *x* este uşor de calculat din valorile $\left(x, \mathrm{mod}\, \mathrm{d}^{\alpha}\right)$. Pentru a calcula $\left(x, \mathrm{mod}\, \mathrm{d}^{\alpha}\right)$, considerăm reprezentarea acestui număr în baza *d*:

$$\left(x, \mathrm{mod}\, \mathrm{d}^{\alpha}\right) = x_0 + x_1 \mathrm{d} + \ldots + x_{\alpha+1} \mathrm{d}^{\alpha-1}; 0 \le x_i \le \mathrm{d} - 1.$$

În exemplul dat considerăm factorul $\mathrm{d}^{\alpha} = 3^2$ şi scriem $\left(z, \mathrm{mod} 181\right) = x_0 + 3x_1$.

Pentru a afla x_0 calculăm $\left(y^{\frac{q-1}{\mathrm{d}}}, \mathrm{mod}\, q \right)$, care este egal cu *a*(*i*, *d*) pentru un număr *i*. Alegem $x_0 = i$. În exemplul dat, $\left(62^{60}, \mathrm{mod} 181\right) = 48$, deci $x_0 = 1$. Acest lucru este valabil, în general, deoarece $\left(g^{q-1}, \mathrm{mod}\, q\right) = 1$ deci.

$$y^{\frac{q-1}{\mathrm{d}}} \equiv g^{\frac{\alpha(q-1)}{\mathrm{d}}} = g^{\frac{x_0(q-1)}{\mathrm{d}}} \equiv a\left(x_0, \mathrm{d}\right)\left(\mathrm{mod}\, q\right).$$

Pentru a-l obţine pe x_1, calculăm mai întâi inversul g^{-x_0} al lui $g^{x_0}\left(\mathrm{mod}\, q\right)$ şi considerăm $y_1 = y g^{-x_0}$. Dacă acum $\left(y_1^{\frac{q-1}{\mathrm{d}^2}}, \mathrm{mod}\, q \right) = a\left(i, \mathrm{d}\right)$, atunci $x_1 = i$. Pentru a-l obţine pe x_2, considerăm numărul $y_2 = y g^{-x_0 - x_1 \mathrm{d}}$ şi calculăm $\left(y_2^{\frac{q-1}{\mathrm{d}^2}}, \mathrm{mod}\, q \right)$.

Procedura se repetă până când este găsit $\left(x, \mathrm{mod}\, \mathrm{d}^{\alpha}\right)$.

În exemplul dat găsim $y_1 = 31$. Aceasta implică faptul că $\left(y_1^{\frac{180}{3^2}}, \mathrm{mod} 181 \right) = \left(y_1^{20}, \mathrm{mod} 181\right) = 1$, deci $x_1 = 0$ şi $z \equiv 1\left(\mathrm{mod} 9\right)$.

Să considerăm acum factorul $\mathrm{d}^{\alpha} = 2^2$. Trebuie să determinăm $x_0 + 2x_1$. Deoarece $\left(62^{90}, \mathrm{mod} 181\right) = 1$, rezultă că $x_0 = 0$. Acum $y_1 = y = 62$ şi $\left(62^{45}, \mathrm{mod} 181\right) = 1$, de unde $x_1 = 0$ şi $z \equiv 0\left(\mathrm{mod} 4\right)$.

Acum considerăm $d^{\alpha}=5^{1}$. Trebuie determinat doar x_2. Deoarece $\left(62^{36},\mathrm{mod}181\right)=1$ rezultă $x_0=0$ şi $z\equiv 0\left(\mathrm{mod}5\right)$.

Deci avem de rezolvat congruenţele:

$$\begin{aligned} &z\equiv 0\left(\mathrm{mod}4\right)\\ &z\equiv 0\left(\mathrm{mod}5\right) \;\Rightarrow\; z\equiv 100\left(\mathrm{mod}180\right)\\ &z\equiv 0\left(\mathrm{mod}9\right) \qquad \log 62=100 \end{aligned}$$

1.3 Inele de polinoame

1.3.1 Proprietăţi generale

Fie R un inel comutativ şi unitar şi fie $R^{(N)}$ mulţimea şirurilor $f=\left(a_0,a_1,...,a_n,...\right), a_i\in R$ care au numai un număr finit de termeni a_i nenuli. Există deci un număr natural *m* astfel încât $a_i=0$, pentru orice *i* > *m*.

1) Şirurile $f=\left(a_0,a_1,...,a_n,...\right)$ şi $g=\left(b_0,b_1,...,b_n,...\right)$ sunt egale dacă şi numai dacă: $a_i=b_i$ pentru orice *i*. Pe mulţimea $R^{(N)}$ se definesc două operaţii algebrice, adunarea şi înmulţirea, în raport cu care $R^{(N)}$ devine un inel comutativ şi unitar.

2) Dacă $f,g\in R^{(N)}$, atunci $k>\max\left\{m,n\right\}$ $f+g=\left(a_0+b_0,a_1+b_1,...,a_n+b_n,...\right)$ aparţine de asemenea mulţimii $R^{(N)}$. Într-adevăr, dacă *m* şi *n* sunt două numere naturale astfel încât $a_i=0$ pentru orice *i* > *m* şi $b_j=0$ pentru orice *j* > *n* atunci $a_k+b_k=0$ pentru orice $k>\max\left\{m,n\right\}$.

3) Se verifică uşor că $\left(R^N,+\right)$ formează un grup abelian, elementul nul fiind (0,0,0,...).

4) Pentru orice element $f=\left(a_0,a_1,...,a_n,...\right)$ opusul său va fi $-f=\left(-a_0,-a_1,...,-a_n,...\right)$.

5) Înmulţirea pe $R^{(N)}$ se defineşte astfel:
$f\cdot g=\left(c_0,c_1,...,c_k,...\right)$, unde $c_k=\sum_{i+j=k}a_ib_j$ pentru orice *k* = 0,1,2,...

6) Să arătăm că $f\cdot g\in R^{(N)}$. Într-adevăr $c_k=0$ pentru orice $k>m+n$.

7) Înmulţirea pe $R^{(N)}$ este asociativă, comutativă şi are element unitatea, (1,0,0,...), proprietăţi care se verifică cu uşurinţă.

8) În plus operaţia de înmulţire este distributivă faţă de adunare, adică:

$$f(g+h) = fg + fh$$
şi
$$(f+g)h = fh + gh$$

În concluzie, s-a demonstrat că $\left(R^{(N)}, +\right)$ formează un inel unitar comutativ. Elementele acestui inel se numesc polinoame cu coeficienţi în *R*.

Gradul unui polinom se notează cu grad (*f*) şi este dat de coeficientul dominant a_n al polinomului *f*, deci: $\text{grad}(f) = n$ dacă $a_i = 0$ pentru orice $i > n$ şi $a_n \neq 0$.

Notăm prin *X* polinomul (0,1,0,...) care se numeşte nedeterminata *x*. Înmulţirea polinoamelor ne dă $X^n = \left(\underbrace{0,0,...,1,0,...}_{\text{nori}}\right)$ cu 1 aflat pe poziţia (*n* + 1).

Fie *f* un polinom de grad *n* ai cărui coeficienţi sunt $a_0, a_1, ..., a_n$, adică

$$f = (a_0, a_1, ..., a_n).$$

Folosind adunarea şi înmulţirea definite pe $R^{(N)}$ se obţine:

$$f = (a_0,0,0,...,0) + (0,a_1,0,...,0) + ... + (0,0,...,a_n,0) =$$

$$= (a_0,0,0,...) + (a_1,0,0,...)\cdot(0,1,0,...) + ... + (a_n,0,0,...)\cdot\left(\underbrace{0,0,...,1,0,...}_{\text{nori}}\right),$$

deci $f = a_0 + a_1x + a_2x^2 + ... + a_nx^n = \sum_{i=0}^{n} a_i x^i$.

Mulţimea $R^{(N)}$ se mai poate nota $R[x]$ şi se numeşte inelul polinoamelor în nedeterminata *X* cu coeficienţi în inelul *R*.

Fie un inel şi *f*, *g* două polinoame din $R[x]$, atunci:

1. $\text{grad}(f+g) \leq \max\left[\text{grad}(f), \text{grad}(g)\right]$;
2. $\text{grad}(fg) \leq \text{grad}(f) + \text{grad}(g)$.

Avem egalitate dacă *f* şi *g* sunt nenule iar cel puţin unul dintre coeficienţii dominanţi ai lui *f* şi *g* nu este divizor al lui zero.

Fie *R* un inel comutativ şi unitar şi inelul polinoamelor $R[x]$. Atunci au loc afirmaţiile:

- un element $a \in R$ este inversabil în *R* dacă şi numai dacă *a* este inversabil în $R[x]$;

- dacă *R* este domeniu de integritate, atunci $R[x]$ este de asemenea domeniu de integritate şi $U(R) = U\left(R[x]\right)$, unde $u: R \to R[x]$ este un morfism al lui *R* pe $R[x]$ şi $u(a) = (a,0,0,...)$.

Teoremă. Fie R un inel comutativ şi unitar şi $R[x]$ inelul polinoamelor de o nedeterminată cu coeficienţi în R şi $u: R \to R[x]$ morfismul canonic. Atunci oricare ar fi inelul comutativ unitar S, morfismul de inele $\nu: R \to S$ şi $x \in S$, există un unic morfism de inele $\varphi: R[x] \to S$ astfel încât $u(x) = x$ şi diagrama să fie comutativă.

Demonstraţie. Să definim mai întâi morfismul φ. Dacă $f \in R[x]$, $f = \sum_{i=0}^{m} a_i x^i$, atunci $\varphi(f) = \sum_{i=0}^{m} \nu(a_i) x^i$.

Arătam că φ are proprietăţile din enunţ. Fie $g = \sum_{i=0}^{n} b_i x^i$ un alt polinom din $R[x]$ şi să presupunem că $n \geq m$. Completând eventual polinomul f cu termeni ai căror coeficienţi sunt zero, putem scrie:

$$f = \sum_{i=0}^{n} a_i x^i,$$

unde $a_{m+1} = a_{m+2} = ... = a_n = 0$.

Atunci

$$\varphi(f+g) = \varphi\left(\sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n} b_i x^i\right) = \sum_{i=0}^{n} \nu(a_i + b_i) x^i =$$

$$= \sum_{i=0}^{n} (\nu(a_i) + \nu(b_i)) x^i = \sum_{i=0}^{n} \nu(a_i) x^i + \sum_{i=0}^{n} \nu(b_i) x^i = \varphi(f) + \varphi(g).$$

Dacă notăm cu c_k, coeficienţii produsului fg, avem $c_k = \sum_{i+j=k} a_i b_j$, şi cum ν este un morfism de inele obţinem $\nu(c_i) = \sum_{i+j=k} \nu(a_i) \cdot \nu(b_i)$.

Ţinând seama de acest lucru se verifică imediat că:

$$\nu(fg) = \nu(f) \cdot \nu(g).$$

Comutativitatea diagramei se verifică uşor. Într-adevăr dacă $a \in$ □, avem:

$$(\varphi \cdot u)(a) = \varphi(u(a)) = \varphi(a) = \varphi(ax^0) = \nu(a) x^0 = \nu(a).$$

Deci $\varphi \cdot u = \nu$.

1.3.2 Factorialitatea inelelor de polinoame

Un inel R, se numeşte inel factorial sau cu descompunere unică în factori primi (ireductibili) dacă orice element neinversabil şi nenul din R se descompune într-un

produs finit de elemente prime. Orice inel de polinoame de o nedeterminată este un inel factorial.

Teoremă. Fie R un inel factorial. Atunci inelul de polinoame $R[x]$ este factorial. Pentru demonstraţie avem nevoie de o serie de rezultate preliminarii.

Lema 1. Fie $a \in R$ şi $f = a_0 + a_1 x + ... + a_n x^n \in R[x]$. Dacă $a|f$, atunci $a|a_i$ oricare ar fi $i = 0,1,...,n$.

Demonstraţie. Cum a / f rezultă că:

$$f = a\left(b_0 + b_1 x + ... + b_n x^n\right) = ab_0 + ab_1 x + ...ab_n x^n ,$$

deci a / a_i.

Lema 2. Fie R un domeniu de integritate. Daca $p \in R$ este un element prim din R atunci p este un element prim în $R[x]$.

Demonstraţie. Fie $f, g \in R[x]$ astfel încât p / fg. Presupunem că $f = a_0 + a_1 x + ... + a_n x^n$ şi $g = b_0 + b_1 x + ... + b_m x^m$ şi că p / f şi p / g. Conform lemei anterioare, rezultă că există un a_k astfel încât p / a_k. Analog, din p / g rezultă că există b_l astfel încât p / b_l. Deci $p / a_0, p / a_1, ..., p / a_{k-1}$ şi $p / b_0, p / b_1, ..., p / b_{l-1}$.

Coeficientul lui x^{k+l} este

$$c_{k+l} = \sum_{i+j=k+1} a_i b_j = a_0 b_{k+1} + a_1 b_k + ... + a_{k-1} b_{l+1} + a_k b_l + a_{k+1} b_{l-1} + ... + a_{k+l} b_0 .$$

Deoarece $p / a_i b_i$ cu $i \neq k$ şi $j \neq l$, iar $p / a_k b_l$, rezultă că p / c_{k+l}, deci p / fg, deci contradicţie, deci trebuie ca p / f sau p / g.

Presupunem acum că inelul R este factorial şi fie $f = a_0 + a_1 x + ... + a_n x^n$ un polinom din $R[x]$. Vom nota cu $c(f)$ cel mai mare divizor comun al elementelor $a_0, a_1, ..., a_n \cdot c(f)$ se numeşte conţinutul polinomului f. Dacă $c(f) = 1$, atunci polinomul f se numeşte primitiv. Se observă că $f = c(f) \cdot f'$, unde f' este un polinom primitiv.

Lema lui Gauss (3). Dacă R este un inel factorial şi $f, g \in R[x]$ atunci $c(fg) = c(f) \cdot c(g)$.

Demonstraţie. Cum $f = c(f) \cdot f'$ şi $g = c(g) \cdot g'$, rezultă $fg = c(f) \cdot f' \cdot c(g) \cdot g' = c(fg) f'g'$.Deci $c(fg) = c(f) \cdot c(g) \cdot c(f'g')$.

Trebuie arătat că $c(f'g') = 1$. Presupunem că $c(f'g') \neq 1$, deci există $p \in$ □ element prim, astfel încât $p / c(f'g')$. Deci $p / f'g'$, deci p / f' sau p / g' rezultă că f' sau g' nu sunt primitive, rezultă o *contradicţie*.

Lema 4. Fie R un inel factorial şi $f, g \in$ □ $[x]$, unde g este un polinom primitiv. Dacă $a \in$ □ $, a \neq 0$, şi g / af, atunci g / f

Demonstraţie. Avem $af = gh, h \in R[x] ac(f) = c(g) \cdot c(h) = c(h)$ deoarece $c(g) = 1$.Cum $h = c(h) \cdot h'$, rezultă că $af = g \cdot c(h) \cdot h'$ sau $af = g \cdot a \cdot c(f) \cdot h' / : a \Rightarrow f = g \cdot c(f) \cdot h' = c(f) \cdot g \cdot h'$, deci g / f.

Lema 5. Fie *R* un inel factorial cu corpul de fracţii *K*, iar $f, g \in R[x]$ două polinoame primitive. Atunci *f* şi *g* sunt asociate în *R*[*x*] dacă şi numai dacă sunt asociate în inelul *K*[*x*].

Demonstraţie. Evident că dacă *f* şi *g* sunt asociate în *R*[*x*] sunt asociate şi în *K*[*x*]. Invers, presupunem că *f* şi *g* sunt asociate în *K*[*x*]. Deci există $u \in K[x]$, element inversabil, astfel încât $g = fu$. Cum $u \in K$, atunci putem scrie $u = \frac{a}{b}$ cu $a \in K$ şi $b \in K$, $a \neq 0, b \neq 0$. Deci *bg* = *af* şi *f*|*g* şi *g*|*f*, adică *f* şi *g* sunt asociate în inelul *R*[*x*].

Lema 6. Fie *R* un inel factorial cu *K* corpul său de fracţii şi $f \in R[x]$ un polinom primitiv cu grad $(f) \geq 1$. Atunci *f* este ireductibil în *R*[*x*] dacă şi numai dacă *f* este ireductibil în *K*[*x*].

Demonstraţie: Presupunem că *f* este ireductibil în *R*[*x*] şi fie *f* = *gh*, cu $g \in K[x]$ şi $h \in K[x]$ şi $\text{grad}(g) \geq 1$, $\text{grad}(h) \geq 1$. Evident putem scrie $g = \frac{a}{b} g_1$, $a, b \in K$, $(a,b) = 1$, şi $g_1 \in R[x]$. Analog $h = \frac{c}{d} h_1$, $c, d \in K$, $(c,d) = 1$, $h_1 \in R[x]$. În plus, $\text{grad}(g) = \text{grad}(g_1)$ şi $\text{grad}(h) = \text{grad}(h_1)$.

Deci $f = \frac{a}{b} g_1 \cdot \frac{c}{d} h_1 = \frac{ac}{bd} g_1 h_1$.

$g_1 = c(g_1) \cdot g_1'$;
$h_1 = c(h_1) \cdot h_1'$; (g_1' şi h_1' sunt polinoame primitive).

Obţinem $f = u \cdot g_1' \cdot h_1'$; $u = \frac{ac}{bd} c(g_1) \cdot c(h_1)$ - element inversabil din *R*. Deci *f* şi $g_1' \cdot h_1'$ sunt asociate în *K*[*x*]. Deci *f* şi $g_1' \cdot h_1'$ sunt asociate şi în *R*[*x*]. Deci $f = \nu \cdot g_1' \cdot h_1', \nu \in U[R]$. Cum grad $g_1 \geq 1$ şi grad $h_1 \geq 1$, aceasta implică că *f* nu este ireductibil în *R*[*x*] rezultă deci o *contradicţie*.

Invers, presupunem că *f* este ireductibil în *R*[*x*] şi că *f* = *gh*, $g, h \in R[x]$. Cum *f* este ireductibil în *K*[*x*] rezultă că *g* este inversabil sau *h* este inversabil în *K*[*x*]. Dacă *g* este inversabil în *K*[*x*], rezultă că $g \in K$, adică $g = a \in K$. Deci $f = ah$. Cum *f* este primitiv, rezultă că *a* este inversabil în *R*. Deci *f* este ireductibil în *R*[*x*].

Teoremă: Fie *R* un inel factorial. Atunci inelul de polinoame *R*[*x*] este factorial.

Demonstraţie: Fie $f \in R[x]$. Putem scrie $f = c(f) f_0$, unde f_0 este un polinom primitiv. Cum $f_0 \in K[x]$, iar $K[x]$ este un inel factorial, rezultă că $f_0 = f_1 f_2 ... f_n$, unde $f_1, f_2, ..., f_n \in K[x]$ şi sunt polinoame ireductibile. Putem scrie pentru f_i, $f_i = \frac{a_i}{b_i} g_i$,unde $a_i, b_i \in R$ şi $g_i \in R[x]$ este un polinom primitiv, deci g_i este ireductibil în $R[x]$. Rezultă că f_0 se scrie sub forma $f_0 = \frac{a}{b} g_1 g_2 ... g_n$, unde $a, b \in R$. Cum f_0 este primitiv şi produsul $g_1 g_2 ... g_n$ este primitiv. Din lema 6 rezultă $f_0 = u g_1 g_2 ... g_n, u \in U[R]$. Cum $c(f)$ este un produs finit de elemente prime în R, care sunt elemente prime şi în $R[x]$, rezultă conform lemei 3 că f este un produs de elemente ireductibile în $R[x]$. Scrierea lui în produs de elemente ireductibile în $R[x]$ este unică. Să presupunem că $f = f_1 f_2 ... f_n = g_1 g_2 ... g_m$, unde $f_i, g_i \in R[x]$ sunt elemente ireductibile în $R[x]$. Dacă grad $f_i \geq 1$, atunci evident $c(f_i) = 1$. Deci putem scrie:

$$f = f_1 f_2 ... f_s f_{s+1} ... f_n = g_1 g_2 ... g_r g_{r+1} ... g_m,$$

unde $f_{s+1}, ..., f_n, g_{r+1}, g_m$ au gradul ≥ 1. Din lema lui Gauss rezultă că $f_1, f_2, ..., f_s$ şi $g_1, g_2, ..., g_r$ sunt asociate în divizibilitate în R. Cum R este factorial, rezultă că $r = s$ şi abstracţie făcând de o renumerotare avem $f_i \sim g_i$**)** oricare ar fi $1 \leq i \leq s$. Rezultă: $f_{s+1} ... f_n = g_{r+1} ... g_m$.

Din lema 7 rezultă $m = n$, şi $g_k \sim f_n$ în $K[x]$ oricare ar fi $s+1 \leq k \leq n$, deci $g_k \sim f_n$ în $R[x]$. Deci descompunerea este unică.

1.3.3 Criterii de ireductibilitate

1) ***Criteriul lui Eisenstein***. Fie R un inel factorial şi K corpul funcţiilor sale, iar $f = a_0 + a_1 x + ... + a_n x^n$ din $R[x]$. Presupunem că există un element prim $g_k \sim f_n$ $p \in R$ cu proprietăţile:

i) $p / a_0, p / a_1, ..., p / a_{n-1}$;

ii) p / a_n;

iii) p^2 / a_0.

Atunci polinomul este ireductibil în $K[x]$.

Demonstraţie. Putem presupune că polinomul f este primitiv. prin reducere la absurd vom presupune că f este reductibil în $K[x]$. Deoarece f este primitiv, rezultă că el este reductibil şi în $R[x]$. Fie atunci $f = gh$ cu $g, h \in R[x]$, unde.

$$g = b_0 + b_1 x + ... + b_m x^m, \quad b_m \neq 0;$$

$$h = c_0 + c_1 x + ... + c_r x^r, \quad c_r \neq 0.$$

) Relaţia „~” înseamnă asociat în divizibilitate. Dacă $f \sim g$, rezultă $f | g$ şi $g | f$.

Prin identificare rezultă $a_0 = b_0 c_0$. Din p / a_0 rezultă p / b_0 sau p / c_0 şi cum p^2 / a_0 avem că p nu divide în acelaşi timp pe b_0 şi c_0. Să presupunem că p / b_0 şi p / c_0. Deoarece p / a_n este clar că polinomul g are coeficienţi care nu se divid cu p. Fie i minim astfel încât p / b_i şi să consideram $a_i = \sum_{k+l=i} b_k c_l = b_i c_0 + \sum_{j=1}^{i-1} b_j c_{i-j}$ cu $i \le m < n$.

Deoarece $p / b_0, p / b_i, ..., p / b_{i+1}$, rezultă că $p \left| \sum_{j=1}^{i=1} b_j c_{i-j} \right.$, $i \le m < n$, şi cum $p / b_i c_0$, avem evident că p / a_i, contradicţie.

Aplicaţie. Fie p un număr prim. Polinomul $f = x^{p-1} + ... + x + 1$ cu coeficienţi întregi este ireductibil în $Q[x]$. Într-adevăr este suficient să demonstram că polinomul $f(x + 1)$ este ireductibil în $Q[x]$. Avem:

$$f(x+1) = \frac{(x+1)^{p-1} + (x+1)^{p-2} + ... + (x+1) + 1}{1} = \frac{(x+1)^p - 1}{x+1-1} = \frac{(x+1)^p - 1}{x} =$$

$$= \frac{x^p + c_p^1 x^{p-1} + ... + c_p^{p-1} x + 1 - 1}{x} = x^{p-1} + C_p^1 x^{p-2} + ... + C_p^{p-1}$$

Observăm că $p / C_p^k, 1 \le k \le p-1$, $p / 1$ şi p^2 / C_p^{p-1}. Conform criteriului lui Eisenstein, $f(x+1)$ este ireductibil în $Q[x]$ şi deci $f(x)$ este de asemenea ireductibil în Q.

2) ***Criteriul de reducţie***: Fie R şi S două inele factoriale, $\varphi : R \to S$ un morfism (unitar) de inele, $\overline{\varphi} : R[x] \to S[x]$ morfismul de inele care extinde pe φ, adică dacă $f = a_0 + a_1 x + ... + a_n x^n \in R[x]$, atunci $\overline{\varphi}(f) = \varphi(a_0) + ... + \varphi(a_n) x^n$. Presupunem că f este un polinom primitiv în $R[x]$ astfel încât $\text{grad}(f) = \text{grad}(\overline{\varphi}(f))$ şi $\overline{\varphi}(f)$ este ireductibil în $S[x]$. Atunci polinomul f este ireductibil.

Demonstraţie. Prin absurd presupunem că f este reductibil în $R[x]$, adică $f = gh$ cu $g, h \in R[x]$, iar g şi h nu sunt inversabile în $R[x]$. Cum f este primitiv, atunci trebuie ca $\text{grad}(g) \ge 1$ şi $\text{grad}(h) \ge 1$. Dar din egalitatea $f = gh$ rezultă că $\overline{\varphi}(f) = \overline{\varphi}(g) \cdot \overline{\varphi}(h)$. Cum $\text{grad}(\overline{\varphi}(g)) \le \text{grad}(g)$ şi $\text{grad}(\overline{\varphi}(h)) \le \text{grad}(h)$ iar $\text{grad}(\overline{\varphi}(f)) = \text{grad}(f)$ rezultă că:

$$\text{grad}(\overline{\varphi}(g)) = \text{grad}(g) \ge 1;$$
$$\text{grad}(\overline{\varphi}(h)) = \text{grad}(h) \ge 1.$$

Deci $\overline{\varphi}(f)$ nu este ireductibil în $S[x]$, contradicţie.

Caz particular: Fie p un număr prim, $f = a_0 + a_1 x + ... + a_n x^n$ un polinom primitiv din $Z[x]$ şi $\overline{f} = \overline{a}_0 + \overline{a}_1 x + ... + \overline{a}_n x^n$ polinomul din $Z_p[x]$ astfel încât $\overline{a}_i$ este clasa de resturi modulo p a lui a_i, pentru orice $i = 0,1,...,n$. Dacă $\text{grad}(f) = \text{grad}(\overline{f})$ şi $\overline{f}$ este ireductibil atunci şi f este ireductibil.

Aplicaţie: Polinomul $f = x^5 - 5x^2 + 1$ are coeficienţi întregi şi este ireductibil în $Z[x]$.

Într-adevăr, aplicăm criteriul de reducţie pentru $p = 2$. Avem $\overline{f} = x^5 + x^2 + 1 \in Z_2[x]$, $\text{grad}(f) = \text{grad}(\overline{f})$. Demonstrăm că f este ireductibil în $Z_2[x]$. Deoarece $\overline{f}(\overline{0}) = \overline{1}$ şi $\overline{f}(\overline{1}) = \overline{1} \neq \overline{0}$, rezultă că f nu are divizori de gradul întâi.

Deci $x^5 + x^2 + 1 = (ax^2 + bx + c)(\alpha x^3 + \beta x^2 + \gamma x + \delta), a,b,c,\alpha,\beta,\gamma,\delta \in \square_2$.

Prin identificare rezultă:

$$\begin{cases} a\alpha = 1 \\ \alpha b + a\beta = 0 \\ a\gamma + b\beta + c\alpha = 0 \\ a\delta + b\gamma + c\beta = 1 \\ b\delta + c\gamma = 0 \\ c\delta = 1 \end{cases}$$

Acest sistem nu are soluţie, deci $x^5 + x^2 + 1$ este ireductibil.

1.3.4 Polinoame ciclotomice

Pentru orice număr natural $n \geq 1$, rădăcinile complexe ale polinomului $f_n(x) = x^n - 1$ se numesc rădăcini de ordin n ale unităţii. Vom nota cu U_n mulţimea acestora, adică

$$U_n = \{x \in C / x^n = 1\}.$$

Mulţimea U_n conţine exact n elemente şi anume:

$$x_k = \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n}; k = 0,1,2,...,n-1.$$

Mulţimea U_n înzestrata cu operaţia de înmulţire este un „grup ciclic”, adică un grup în care toate elementele sunt puteri ale unui anumit element. Dacă $\alpha = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}$, iar atunci se demonstrează că dacă $\alpha^p = 1$, atunci p este multiplu de n.

Un element $x_k = \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n}$ se numeşte generator al grupului U_n (dacă prin ridicări succesive la putere se generează toate elementele mulţimii U_n) dacă şi numai dacă $(k, n) = 1$.

Grupul ciclic U_n are $\varphi(n)$ generatori care sunt rădăcinile primitive de ordinul n ale unităţii.

Pentru n numai prin mulţimea rădăcinilor primitive de ordin n este

$$P_n = \left\{\alpha, \alpha^2, ..., \alpha^{n-1}\right\},$$

unde:

$$\alpha = \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n}, 0 \le k \le n-1.$$

Pentru orice $n \in N^*$, polinomul

$$f_n = \prod_{\alpha \in P_n} (x - \alpha)$$

se numeşte al n-lea *polinom ciclotomic* (al n-lea polinom de diviziune circulară).

Relaţia lui Dedekind: Pentru orice $n \ge 1$ are loc egalitatea

$$x^n - 1 = \prod_{d/n} f_d ,$$

unde produsul se face după divizorii naturali d ai numărului n. Această relaţie, a cărei demonstraţie este relativ simplă, poate fi utilizată ca o relaţie de recurenţă cu ajutorul căreia se pot determina din aproape în aproape polinoamele ciclotonice.

De exemplu, ne propunem să calculăm polinoamele ciclotonice f_1, f_2, f_3, f_4, f_5 şi f_6.

Evident $f_1 = x - 1$ şi apoi $x^2 - 1 = f_1 \cdot f_2$. Deci $f_2 = x + 1$ $x^3 - 1 = f_1 \cdot f_3$, rezultă $f_3 = x^2 + x + 1$;

$$x^4 - 1 = f_1 \cdot f_2 \cdot f_4, \text{ deci } f_4 = \frac{x^4 - 1}{(x-1)(x+1)} = x^2 + 1;$$

$$x^5 - 1 = f_1 \cdot f_5, \text{ deci } f_5 = \frac{x^5 - 1}{x - 1} = x^4 + x^3 + x^2 + 1;$$

$$x^6 - 1 = f_1 \cdot f_2 \cdot f_3 \cdot f_6, \text{ deci } f_6 = \frac{x^6 - 1}{(x-1)(x+1)(x^2 - x + 1)} = x^2 - x + 1.$$

Pentru $n = p$ număr prim:

$$f_p = \sum_{i=0}^{p-1} x^i .$$

Polinomul f_p este în acest caz ireductibil. Mai mult, se demonstrează (teorema lui Dedekind) că orice polinom $f_n, n \geq 1, n \in N$ este ireductibil în inelul $Z[x]$ şi în inelul $Q[x]$.

Polinoamele ciclotonice pot fi exprimate cu ajutorul funcţiei lui Möbius care este definită astfel:

$$\mu(n) = \begin{cases} 1, \text{dacă } n = 1; \\ (-1)^k, n = P_1 P_2 ... P_k, n \text{ liber de pătrate}; \\ 0, \text{dacă există } p^2 / n, p \text{ număr prim}. \end{cases}$$

Pentru orice $n \geq 1, n \in N$ avem egalitatea

$$f_n = \prod_{d/n} \left(x^d - 1 \right)^{\mu\left(\frac{n}{d}\right)}.$$

De exemplu

$$f_6 = (x-1)^{\mu(6)} \cdot (x-1)^{\mu(3)} \cdot \left(x^3 - 1\right)^{\mu(2)} \cdot \left(x^6 - 1\right)^{\mu(1)} =$$

$$= \frac{(x-1)\left(x^6 - 1\right)}{\left(x^2 - 1\right)\left(x^3 - 1\right)} = x^2 - x + 1,$$

deoarece

$$\mu(1) = 1;$$
$$\mu(2) = -1;$$
$$\mu(3) = -1;$$
$$\mu(6) = 1.$$

Pentru demonstrarea acestei relaţii se utilizează teorema de inversiune a lui Möbius.

Vom prezenta formulele de inversiune ale lui Möbius fără demonstraţie:

a) formula aditivă

Dacă $f(n)=\sum_{\substack{d\\ d/n}} g(d)$, atunci $g(n)=\sum_{\substack{d\\ d/n}} \mu(d) f\left(\frac{n}{d}\right)$, unde

$$\mu(d)=\begin{cases}1; & d=1\\ (-1)^k; & d=P_1P_2...P_k\\ 0; & d=P_1P_2...P_j^{\alpha},\quad \alpha>1\end{cases}$$

b) formula multiplicativă

Dacă $f(n)=\sum_{\substack{d\\ d/n}} g(d)$, atunci $g(m)=\sum_{\substack{n\\ n/m}} f(n)^{\mu\left(\frac{m}{n}\right)}$.

Test de autoevaluare

1. Care este corolarul definiţiei numerelor prime, referitor la modul de scriere al acestora.
2. Dacă *a* şi *n* sunt numere naturale, $a\geq 1$ şi $n\geq 2$ astfel încât a^n-1 este număr prim, ce valori pot avea *a* si *n*? Demonstraţi.

Tema de autoinstruire

1. Propuneţi o schemă logică de implementare a algoritmului Silver, Pohlig şi Hellman (independentă de limbajul de programare).

Lucrare de verificare

1. Enunţaţi teorema chineză a resturilor.
2. Ce este un număr prim?
3. Definiţi noţiunea de congruenţă.
4. Definiţi criteriul de ireductibilitate al lui Eisenstein.
5. Arătaţi că polinomul $f=x^5+5x^2+1$, care are coeficienţi întregi, este ireductibil în Z[x].
6. Enunţaţi teorema lui Euler.
7. Definiţi noţiunea generală de logaritm discret.

CAPITOLUL 2

BAZELE TEORETICE ALE SISTEMELOR SECRETE

Cuvinte cheie: cifrator, receptor, interceptor, mesaj, criptogramă, chei, algoritm cifrare, sistem de cifrare secret.

OBIECTIVE

- Înţelegerea modelului matematic al unui sistem de cifrare
- Definiţia sistemului de cifrare dată de Shanon
- Estimarea numărului de mesaje false, a numărului de decriptări cu cheie falsă şi a numărului de mesaje false.
- Sisteme de criptare compuse, proprietăţile cifrurilor obţinute şi aplicare lor în practică

2.1 Introducere

Primele informaţii referitoare la criptografie datează de acum circa 4000 de ani şi provin din Egiptul antic, dar abordarea teoretică a acesteia şi descrierea modelului matematic al unui sistem secret sunt de dată mult mai apropiată de zilele noastre.

O contribuţie deosebită a avut-o, pe această linie C. E. Shannon care a introdus o serie de concepte şi idei originale cum ar fi noţiunile de secret într-un sistem de cifrare şi redundanţa unui anumit limbaj.

2.2 Modelul matematic al sistemului secret

Figura 2.1 ilustrează un sistem de cifrare sau, cum este deseori numit, un sistem de secretizare.

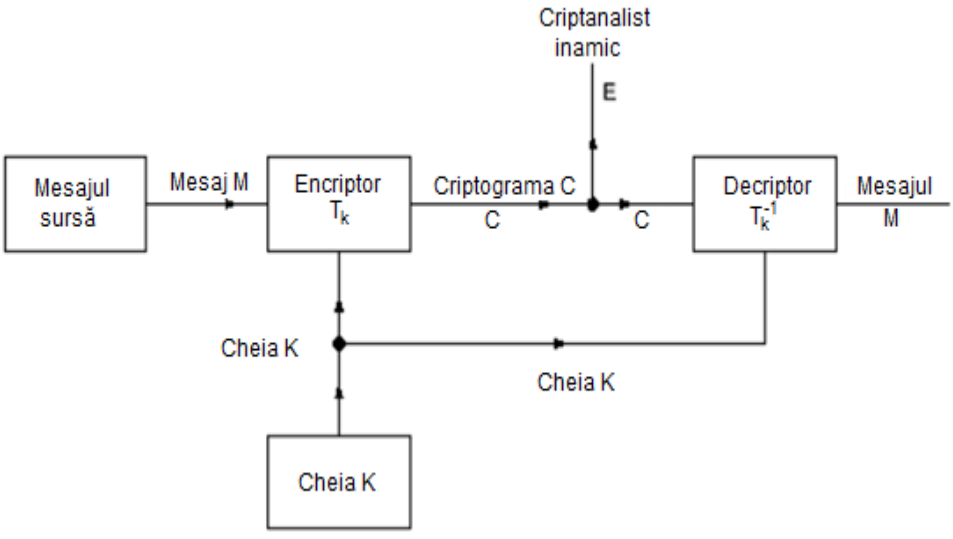


Fig. 2.1 Un sistem de cifrare

Interceptorul este introdus în schemă pentru a arăta unde este cel mai probabil să aibă loc interceptarea.

Înainte ca un mesaj să fie transmis cele doua parţi implicate numite cifrator şi receptor îşi stabilesc o cheie particulară din mulţimea disponibilă, cheie care este păstrată secretă.

Folosind mesajul ce urmează a fi transmis şi cheia aleasă, cifratorul îl cifrează cu ajutorul algoritmului utilizat înaintea transmiterii.

Mulţimea tuturor mesajelor posibile pe care cifratorul le poate transmite este numită „spaţiul mesajelor" şi o vom nota cu *M*. Mulţimea tuturor criptogramelor posibile este numită „spaţiul criptogramelor" şi o vom nota cu *C*, iar mulţimea tuturor cheilor o vom nota cu *K*. Mulţimile *M, C* şi *K* sunt mulţimi finite dar, de obicei prea mari pentru a le enumera explicit elementele. Procedura de cifrare poate fi redefinită după cum urmează:

- cifratorul şi receptorul îşi aleg o cheie;
- cifratorul îşi alege şi apoi foloseşte algoritmul de cifrare *f* ca să determine în mod unic $c \in C$.

Fiecare cheie $k_i \in K$ împreună cu algoritmul de cifrare determină o transformare $t_{k_i} : M \to C$.

Deci putem privi sistemul de cifrare ca un triplet (*M*,*T*,*C*) unde *T* cuprinde totalitatea transformărilor posibile. Este foarte posibil ca utilizând două chei diferite să obţinem aceeaşi transformare. Pot fi uneori mai multe chei decât transformări. Acest lucru poate determina ca la probabilităţi egale ale cheilor să obţinem probabilităţi inegale ale transformărilor. Pe de alta parte o anumită cheie determină în mod unic o transformare. Pentru simplificare vom considera că transformările şi cheile sunt aceleaşi.

O cerinţă de bază a unui sistem de cifrare este următoarea: cunoaşterea criptogramei, cheii şi a algoritmului trebuie să facă receptorul să determine în mod unic mesajul.

Deci dacă $C=t(m)$, atunci $m=t^{-1}(c)$.

Putem da acum definiţia lui Shannon pentru un sistem de cifrare: „**un sistem de cifrare este o familie de transformări ireversibile dintr-o mulţime de mesaje *M* într-o mulţime de criptograme. Cele trei mulţimi *T, M* şi *C* se consideră, de obicei finite**".

Diferenţa între cunoştinţele receptorului şi interceptorului constă în aceea că receptorul cunoaşte care *k* a fost folosită în timp ce interceptorul poate cunoaşte doar probabilităţile „apriori" ale diferitelor transformări. Dacă toate cheile sunt egal probabile atunci interceptorul are numai lista tuturor transformărilor posibile.

Definiţia unui sistem de cifrare permite ca transformările şi cheile să aibă probabilităţi diferite şi este important de observat că acest lucru se poate întâmpla şi chiar se întâmpla.

2.3 Reprezentarea sistemelor secrete

Sistemele secrete pot fi reprezentate în diferite moduri. Unul dintre acestea foloseşte schemele liniare arătate în figura următoare.

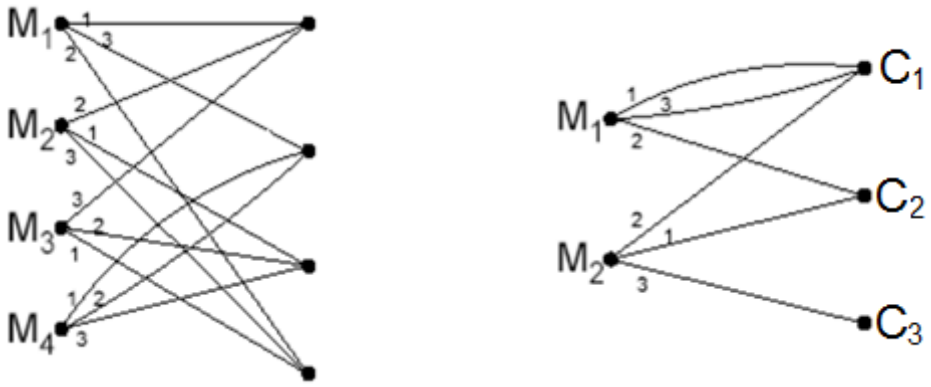


Fig. 2.2

Mesajele posibile reprezintă punctele din stânga iar criptogramele posibile puncte din dreapta. Dacă o anumită cheie (de exemplu 3) transformă mesajul M_2 în criptograma C_4, punctele M_2 şi C_4 se unesc cu o linie pe care se trece cifra 3. Pentru fiecare cheie din fiecare mesaj trebuie să rezulte numai o linie. Dacă acest text este adevărat şi pentru fiecare criptogramă, sistemul secrete se numeşte sistem închis. O metodă mai generală de descriere a sistemului secret constă în prezentarea operaţiei cu ajutorul căreia, operând asupra mesajului cu o cheie oarecare se poate obţine criptograma.

De asemenea se poate determina probabilitatea diferitelor chei fie indicând metoda de alegere a cheilor, fie descriind modul cum, de regulă, adversarul alege cheile. Probabilităţile mesajelor se determină din datele apriori care se cunosc despre conţinutul probabil al mesajului sau din orice informaţie specială care se referă la criptogramă.

Dacă fiecare mesaj sau criptogramă conţine *N* caractere dintr-un alfabet finit de *L* simboluri şi dacă se notează cu $R_c = \log_2 L$ viteza absolută de generare a limbajului, caracteristicile de bază ale sistemului arată ca în figura de mai jos:

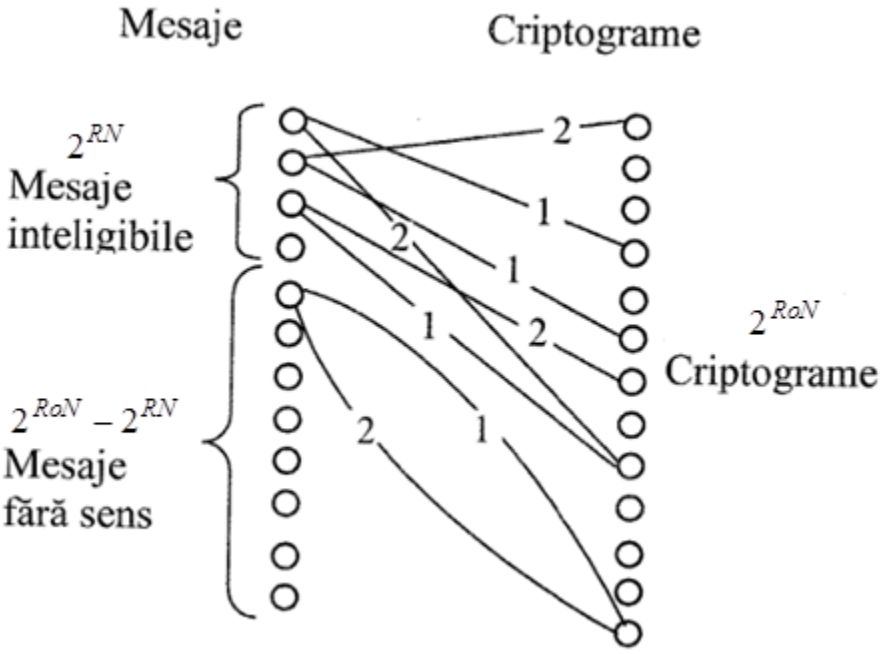


Fig. 2.3

Unde:

$L^n = 2^{RoN}$ este numărul de mesaje posibile şi egal cu numărul de criptograme:

2^{RN} - mulţimea mesajelor cu sens, cu probabilitatea apriori $a2^{-RN}$ (R viteza de generare a limbajului);

$2^{RoN} - 2^{RN}$ - mulţimea mesajelor fără sens în limbaj, atribuindu-li-se probabilitatea apriorică egală cu 0;

$2^{H(k)}$ - mulţimea cheilor, toate egal prealabile şi apriori independente de mesaj: $H(R)$ entropia cheii.

Se observă că o linie cu numărul i indică cifrarea mesajului din stânga cu criptograma din dreapta, când s-a folosit cheia i.

În figură sunt prezentate 12 mesaje şi două chei, deci $H(k)=1$.

Se observă că aceeaşi criptogramă poate rezulta prin cifrarea a două sau mai multe mesaje diferite dacă se folosesc chei diferite.

Mesajele obţinute prin descrierea unei criptograme folosindu-se o altă cheie decât cea utilizată la cifrare se numesc mesaje false. Numărul mesajelor false n_m este o variabilă aleatoare fiind determinată de cifru şi de criptogramă. Dacă n_m ia valori mari cu probabilităţi apropiate de 1 sistemul este sigur chiar dacă criptanalistul are posibilităţi nelimitate de calcul.

Poate exista şi situaţia când o criptogramă este descifrată în acelaşi mesaj cu mai multe chei. Astfel criptanalistul ştie care mesaj a fost transmis dar nu ştie care din chei a fost folosită. Această situaţie se numeşte descriptare cu cheie falsă. Dacă notăm cu n_k numărul de decriptări cu cheie falsă, atunci în cazul unui cifru cu o bună folosire a cheii trebuie ca $n_k = n_m$.

În timp ce n_m prezintă interes mai mare, este mai simplu de calculat n_k. Astfel dacă $l(c)$ reprezintă numărul de linii care se termină în criptograma *C*, atunci $n_k(c)$, numărul descifrărilor corecte cu cheie falsă când este recepţionată criptograma *C*, este:

$$n_k(c) = \max\{[l(c)-1],0\}$$

2.4 Compunerea sistemelor secrete

Fiind date două sisteme secrete *T* şi *R* ele pot fi combinate în diferite feluri pentru obţinerea unui nou sistem secret *S*.

Dacă *R* şi *T* au acelaşi spaţiu al mesajelor se poate forma o sumă ponderată:

$$S = pT + qR; \quad q + p = 1$$

Dacă *T* constă din prezentările: $T_1, T_2, ..., T_m$ cu probabilităţile $p_1, p_2, ..., p_m$ iar *R* din reprezentările: $R_1, R_2, ..., R_k$ cu probabilităţile $q_1, q_2 ... q_k$ sistemul rezultat, *S*, are reprezentările:

$T_1, T_2, ..., T_m$, $R_1, R_2, ..., R_k$ cu probabilităţile;

$PP_1, P\,P_2, ..., PP_m, q\,q_1, q\,q_2, ..., q\,q_k$

Generalizând se poate forma suma mai multor sisteme:

$$S = P_1T + P_2R + ... + P_mU; \quad \sum_{i=1}^{m} P_i = 1$$

O altă metodă de combinare a două sisteme secrete constă în formarea produsului lor, ca în figura de mai jos.

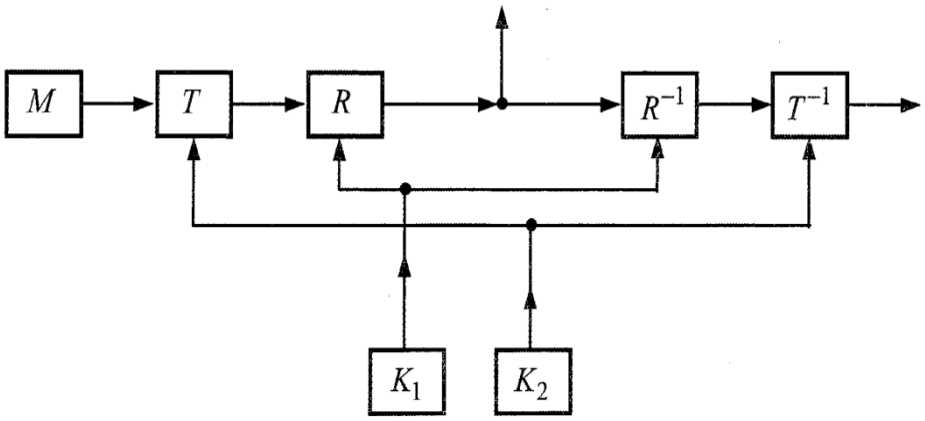


Fig. 2.4

Să presupunem că *T* şi *R* sunt două sisteme astfel încât spaţiul mesajelor sistemului *R* poate fi identificat cu spaţiul criptogramelor sistemului *T*. În acest caz se poate folosi sistemul *T* la mesajele iniţiale iar apoi sistemul *R*, la rezultatul primei operaţii ceea ce dă o operaţie rezultată *S* care se poate scrie sub forma:

$$S = RT$$

Cheia sistemul *S* constă atât din cheia sistemului *T* cât şi din cea a sistemului *R*, aceste chei alegându-se independent şi cu probabilităţile lor iniţiale. Dacă sistemul *T* are *m* chei cu probabilităţile $P_1, P_2 ... P_m$ iar sistemul *R* are *n* chei cu probabilităţile $P_j \cdot q_j, \quad i = 1, m$ şi $j = 1, n$.

Cifrurile de tip produs se folosesc în mod frecvent, dar trebuie remarcat că produsul, în general, nu este comutativ, adică nu totdeauna $RT = TR$. În schimb produsul este asociativ:

$$R(ST) = (RS)T = RST$$

Se verifică de asemene:

- legea asociativă ponderată pentru adunare;

$$p(p'T + q'R) + qS = pp'T + Pq'$$

- legea distributivă la stânga şi la dreapta:

$$T(pR + qS) = pTR + qTS.$$
$$(pR + qS)T = pRT + qST.$$

Sistemele la care spaţiile *M* şi *C* coincid (caz frecvent când succesiunea la litere se transformă tot în succesiune de litere) se numesc endomorfe.

Sistemul endomorf *T* poate fi ridicat la putere: T^n există deci. Dacă $T \cdot T = T$ sistemul se numeşte independent. Din aceste combinaţii rezultă metode pentru construirea unor noi tipuri de sisteme secrete din sisteme secrete date.

Proprietăţile prezentate mai sus se pot folosi pentru descrierea situaţiilor pe care le poate întâlni criptanalistul când încearcă să descifreze o anumită criptogramă. În mod real adversarul descifrează sistemul secret de tipul:

$$T = p_1 A + p_2 B + ... + p_n S + p'X; \sum_j p_j = 1$$

unde $A, B, ..., S$ sunt cifruri cunoscute cu probabilităţile lor apriori p_i iar $p'x$ reprezintă probabilitatea folosirii unui cifru nou necunoscut.

Sistemul la care pentru fiecare din transformările T_i, T_j, T_k există transformarea T_S astfel că, dacă cheile sunt echiprobabile are loc relaţia: $T_i T_j - 1_{Tk} = T_s$ se numeşte sistem omogen.

În caz contrar sistemul se numeşte neomogen.

Sistemele omogene au o serie de proprietăţi:

- în cifrul omogen operaţia care transformă spaţiul mesajelor în sine însuşi, $T_i^{-1} T_j$, formează un grup de ordinul *m*, unde *m* este numărul de chei diferite;
- produsul a două cifruri omogene comutative este un cifru omogen;
- mesajele se pot împărţi în clase de resturi $R_1' R_2' ... R_S'$.

Aceste clase de resturi sunt disjuncte, conţin toate mesajele posibile, dacă se cifrează un mesaj din clasa R_i se obţine o criptogramă din clasa R_i, fiecare mesaj din clasa de resturi R_i poate fi cifrat în fiecare criptogramă din R_i' cu ajutorul a exact K/φ_i chei diferite (*K* reprezintă numărul de chei iar φ_i numărul de mesaje din clasa R_i). Se spune despre două sisteme că sunt asemenea dacă există o transformare *A* care are o transformare $A^{-1}m$, astfel încât:

$$R = AS$$

Aceasta înseamnă că cifrarea cu *R* dă acelaşi rezultat ca şi cifrarea cu *S* dacă aceasta din urmă este aplicată ulterior transformării. Dacă notăm $R \approx S$ (*R* asemenea cu *S*) atunci se poate scrie:

$$R \approx R\left(\text{reflexivitate}\right)$$

$$\begin{cases} R \approx S \Rightarrow R \approx T\left(\text{tranzitivitatea}\right) \\ S \approx T \end{cases}$$

Dacă $R \approx S \Rightarrow S \approx R$ (simetrie)

Deci dacă $R \approx S$ atunci cele două sisteme sunt echivalente din punct de vedere al descifrării. Într-adevăr dacă criptanalistul advers interceptează o criptogramă din sistemul *S*, el o trece într-o criptogramă din sistemul *R* aplicând o transformare *A*. Invers criptograma deci sistemul *R* poate fi trecută într-o criptogramă din sistemul *S* prin transformarea A^{-1}. Din punct de vedere teoretic se pun o serie de probleme în ceea ce priveşte asigurarea secretului. Se pune întrebarea cât de rezistent este un sistem secret, dacă criptanalistul advers nu este limitat de timp şi are la dispoziţie specialişti şi mijloace tehnice pentru analizarea criptogramelor interceptate. De asemenea se pune întrebarea dacă o criptogramă oarecare are una sau mai multe soluţii şi ce volum de text cifrat într-un anumit sistem trebuie să fie interceptat pentru a se obţine o rezolvare unică. Trebuie dat răspunsul la întrebarea dacă există sisteme secrete la care nu se poate găsi rezolvarea unică indiferent de volumul textului cifrat interceptat. Pentru a putea defini un sistem secret perfect să presupunem că există un număr finit de mesaje $M_1, M_2 ... M_n$ cu probabilităţile lor apriorice $p_1, p_2 ... p_n$. Aceste mesaje sunt transformate în criptogramele $C_1, C_2 ... C_n$ astfel încât să se păstreze regula de cifrare $C = T_i m$. După interceptarea unei criptograme se pot calcula probabilităţile aposteriori ale diferitelor mesaje, $P\left(M/C\right)$. Din rezolvarea acestor calcule rezultă că *sistemul secret perfect* se defineşte ca un sistem la care pentru toate criptogramele *C*, probabilităţile sunt egale cu probabilităţile apriori ale mesajelor, fără a depinde de acestea. În caz contrar, când probabilităţile aposteriori se deosebesc de cele apriori pentru anumite mesaje de chei, adversarul poate să-şi corecteze acţiunile sale privind alegerea cheilor. Cu alte cuvinte, dacă un sistem secret este perfect, adversarul nu primeşte nici o informaţie suplimentară la recepţionarea unei noi criptograme. Dacă notăm *P*(*M*) = probabilitatea apriori a mesajului *M*; $P\left(C/M\right)$ = probabilitatea condiţionată a criptogramei *C* cu condiţia că s-a ales mesajul *M*.

$P(C)$ = probabilitatea că se obţine criptograma C;

$P(M/C)$ = probabilitatea aposterioară a mesajului M, în condiţia că s-a recepţionat criptograma C, atunci se poate scrie condiţia necesară şi suficientă pentru ca un sistem secret să fie perfect;

$$P(M/C) = \frac{P(M) \cdot P(C/M)}{P(C)}$$

Probabilitatea $P(C/M)$ nu trebuie să depindă de mesaj, deci $P(C/M) = P(C)$ şi $P(M/C) = P(M)$.

Astfel spus probabilitatea totală a tuturor cheilor care transformă mesajul M_i în criptograma C este egală cu probabilitatea totală a cheilor care transformă mesajul M_j în aceeaşi criptogramă C pentru orice M_i, M_j şi C.

În cazul sistemului secret perfect trebuie ca numărul mesajelor M să fie egal cu numărul criptogramelor C, deoarece, pentru orice i dat, transformarea T_i dă o corespondenţă univocă între toate mesajele M şi unele criptograme din C. Mai mult, există cel puţin o cheie care transformă pe M dat în orice criptogramă C.

Toate cheile care transformă pe M dat în diferite criptograme C trebuie să fie diferite, deci numărul de chei nu trebuie să fie mai mic decât numărul mesajelor M.

Un sistem secret perfect este reprezentat în figura de mai jos:

După cum se observă fiecare mesaj se leagă cu fiecare criptogramă cu câte o linie, toate cheile fiind echiprobabile. Cantitatea de informaţie obţinută la alegerea mesajului sau cheii este:

$$H(M) = -\sum_{i=1}^{m} P(M_i) \log P(M_i)$$

$$H(C) = -\sum_{j=1}^{m} P(C_j) \log P(C_j)$$

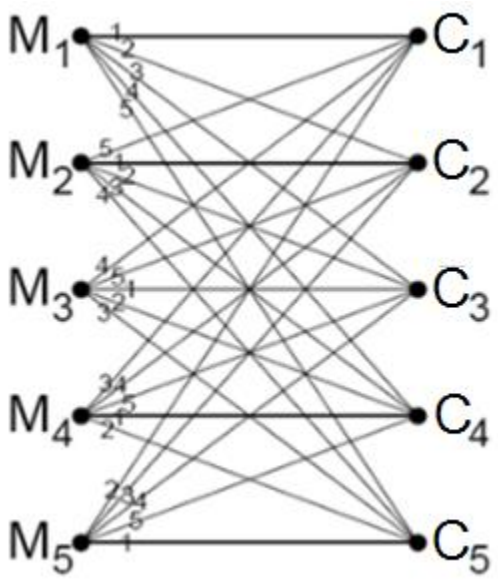


Fig. 2.5

Aceasta este maximă când mesajele sunt echiprobabile deci $P(M_i)=\frac{1}{n}$ şi rezultă:

$$H(M)_{\max} \le \log n$$

Această informaţie este complet ascunsă când nedeterminarea cheii este maximă, adică:

$$H(C) = \log n$$

Deci cantitatea de nedeterminare care poate fi introdusă într-un sistem secret nu poate fi mai mare decât nedeterminarea cheii, deci decât log *n*.

În practică sistemele perfecte se folosesc pentru transmiterea unor documente de importanţă deosebită sau când numărul de mesaje este mic. Neajunsul principal îl constituie faptul că aceste sisteme necesită chei cu volum mare.

Pentru depistarea unui text este necesar să se intercepteze un număr suficient de litere *N*. Dacă *N* este suficient de mare este posibilă rezolvarea univocă a deriptării. Probabilităţile apriori ale literelor pentru diferite mesaje şi cifruri se pot calcula din timp iar cele aposteori se pot determina după interceptarea criptogramelor. Pe măsură ce *N* creşte probabilitatea ca literele respective să reprezinte anumite mesaje creşte, iar pentru alte mesaje scade. Procesul continuă până când rămâne un singur mesaj cu probabilitatea tinzând către 1. Această situaţie este similară cu cea din teoria informaţiei unde semnalul transmis este deformat de perturbaţii, ceea ce a determinat introducerea noţiunii de entropie condiţionată ca măsură a nedeterminării semnalului transmis când se cunoaşte semnalul recepţionat.

Deci este normal să se folosească ca măsură a secretului entropia condiţionată a cheii şi a mesajului.

$$H(K/C) = -\sum_{K,C} P(K,C)\log P(K/C)$$

$$H(M/C) = -\sum_{C,M} P(C,M)\log P(M/C)$$

unde $P(K, C)$ este probabilitatea cheii K şi criptogramei C, $P(K/C)$ = probabilitatea aposterioară a cheii K dacă s-a interceptat criptograma C, iar $P(C, M)$, $P(M/C)$ probabilităţile analoge pentru mesaje.

Din aceste relaţii se observă că rezistenţa sistemului secret este egal cu zero când cheia are probabilitatea egală cu 1, iar la toate celelalte probabilităţi este 0.

Entropia depinde de numărul de N interceptate, ea scăzând când N creşte.

Rezistenţa sistemului secret are câteva proprietăţi interesante dintre care menţionăm:

- rezistenţa cheii $H(K/C)$ este o funcţie necrescătoare cu N. Dacă s-au recepţionat N litere rezistenţa primelor N litere ale mesajului este mai mică sau egală cu rezistenţa cheii, deci:

$$H(K/C,S) \quad H(K/C,N), \quad S \; N$$
$$H(M/C,S) \quad H(M/C,N)$$
$$H(M/C,N) \quad H(K/C,N)$$

- rezistenţa secretului în cazul unui sistem produs $S = T\,R$ nu este mai mică decât rezistenţa unui sistem R. Dacă M_1, C_1 şi C_2 sunt mesajul şi respectiv criptogramele corespunzătoare celor două cazuri atunci:

$$H(M/C_2) \geq H(M/C_1)$$

- pentru o sumă ponderată de sisteme rezistenţa H este limitată de inegalitatea:

$$\sum_i P_i H_i \geq H > \sum_i P_i H_i - \sum_i P_i \log P_i$$

unde H_i putând fi rezistenţa cheii sau rezistenţa mesajului.

Un sistem secret este ideal dacă $H(K/C)$ şi $H(M/C)$ nu tind spre zero când N tinde spre infinit. Sistemul este strict ideal dacă:

$$H(X/C) = H(K)$$

Un exemplu de sistem secret ideal este simpla substituire folosită la un limbaj artificial în care toate literele sunt egal probabile.

În cazul limbajelor naturale sistemele secrete pot în general să se apropie de caracteristica ideală dar acesta duce la creşterea rapidă a complexităţii sistemului.

Sistemele secrete reale au o serie de neajunsuri:

- sistemul trebuie să se găsească în strictă concordanţă cu limbajul deci cel care realizează cifrul trebuie să cunoască profund structura limbii;

- structura limbajelor naturale fiind complexă, înlăturarea reductanţei cere transformări complicate iar dispozitivul ce ar realiza această transformare ar trebui să fie foarte complicat;
- transformările folosite pot duce la multiplicarea erorilor. Eroarea produsă în transmiterea unei litere duce la erori într-un sector mare, definit de lungimea legăturilor statistice în limbajul iniţial.

Orice sistem de cifrare cu un număr finit de mesaje poate fi ilustrat în felul următor: în stânga se scrie o coloană de puncte reprezentând mulţimea mesajelor şi în dreapta o mulţime de puncte reprezentând mulţimea criptogramelor. Dacă un mesaj m_i se transformă cu o cheie K_j într-o criptograma C_k atunci o săgeată etichetată K_j va uni m_i cu C_k.

O dată definit sistemul secret se pot analiza cele cinci principii care stau la baza sistemelor secrete, principii elaborate de Shannon în 1940. Analiza va trebui să ţină seama de evoluţia tehnologiei în perioada care a trecut de atunci mai ales în ceea ce priveşte tehnica de calcul. Cele cinci principii propuse de Shannon erau:

1. Cantitatea de secretizare oferită;
2. Mărimea „cheii”;
3. Simplitatea operaţiunilor de cifrare şi descifrare;
4. Propagarea erorii;
5. Extensia mesajului.

Importanţa primului principiu este evidentă, complexitatea şi tăria sistemelor secrete fiind direct proporţională cu importanţa mesajelor transmise. În ceea ce priveşte „cheia” aceasta trebuie ţinută secretă, periodic ea trebuie schimbată şi transmisă corespondenţilor de către cel care gestionează sistemul de „chei”. Deci cheia ar trebui să fie cât mai scurtă posibil. Pe de altă parte „lungimea cheii” poate determina numărul de chei, deci mulţimea transformărilor făcute, care trebuie să fie cât mai mare. De aici rezultă că mărimea „cheii” trebuie să satisfacă două cerinţe contradictorii.

Uneori se poate împarţi sistemul de chei în clase cum ar fi:

- chei de sistem;
- chei de mesaj.

Operaţiile de cifrare şi descifrare pot fi făcute manual sau automat. Dacă se lucrează manual complexitatea poate duce la erori sau la creşterea timpului necesar efectuării operaţiilor respective. Dacă se lucrează automat complexitatea poate duce la necesitatea unor maşini sofisticate şi scumpe.

În anumite sisteme secrete apariţia unor erori de transmitere se poate propaga afectând porţiuni întregi din mesaj sau chiar întregul mesaj. De aceea este bine ca erorile de propagare să fie minimizate.

În multe cazuri în urma operaţiunii de cifrare lungimea textului creste. O astfel de extensie a mesajului este de nedorit pentru majoritatea sistemelor de comunicaţie.

Din cele arătate mai sus rezulta că există o anumită incompatibilitate între cerinţele celor cinci principii şi este practic imposibil să fie satisfăcute.

Problema realizării unor sisteme de secretizare „sigure” a preocupat dintotdeauna proiectanţii unor asemenea sisteme. Pentru a putea realiza sisteme cat

mai rezistente la atacuri criptanalitice, trebuie cunoscute ameninţările la adresa sistemelor secrete.

Trebuie presupus întotdeauna că un criptanalist are cunoştinţe despre sistemul de secretizare, dispune de mijloacele şi forţele necesare pentru a încerca să „spargă" sistemul secret analizat.

Proiectanţii trebuie să fie capabili să evalueze „tăria" sistemului realizat sau timpul de „acoperire" (timpul necesar unui criptanalist pentru a-l „sparge").

În scopul evaluării securităţii unui sistem se fac următoarele presupuneri care se consideră a fi condiţii fundamentale:

- C1. Criptanalistul are cunoştinţe complete despre sistemul de cifrare;
- C2. Criptanalistul a acumulat un volum considerabil de text cifrat;
- C3. Criptanalistul are la dispoziţie un anumit volum de text clar şi echivalentul său cifrat.

Condiţia C1 implică faptul că nu există siguranţă în sistemul de cifrare însuşi şi securitatea trebuie să fie dată de cheia utilizată. Firește, sarcina criptanalistului este mult mai grea dacă nu cunoaşte metoda folosită. Dar este posibil ca algoritmul de cifrare sa fie in posesia criptanalistului (uneori acesta este public).

În ceea ce priveşte celelalte două condiţii este posibil ca prin acumularea unui volum mare de text cifrat şi prin faptul că toate comunicaţiile între două surse încep cu un anumit antet, criptanalistul să aibă la dispoziţie text clar şi echivalentul său cifrat. O întrebare pe care trebuie să şi-o pună criptograful (realizatorul unui sistem de cifrare) este dificil să determini parţial sau în totalitate mesajul cunoscând criptograma plus o mică parte din textul clar echivalent. Manual, răspunsul depinde de anumiţi factori printre care şi lungimea textului cunoscut.

Desigur cele trei condiţii sunt destul de pesimiste din punct de vedere al criptografului, dar dacă un sistem de secretizare nu ţine seama de ele, acel sistem poate fi considerat necorespunzator.

Dezvoltarea sistemelor de cifrare a fost mult influenţată de demonstraţia lui Shannon ca sistemul cu şir cu unică folosinţă este practic imbatabil. Sistemul constă în aceea că sistemul de chei este un şir aleator de lungime mult mai mare decât mesajul şi care se utilizează o singura dată. De exemplu dacă mesajul este un şir de biţi:

$$m = 01001011101...,$$

iar cheia, un alt şir de biţi:

$$k = 101100010111000....,$$

atunci criptograma C se poate obţine „însumând" cele două şiruri (anticoincidenţă):

$$m = 01001011101...$$
$$k = 10110001011...$$
$$C = 11111010110.$$

Mulţi criptografi şi-au dat seama că dacă realizează un sistem mai bun decât acesta, sistemul realizat ar asigura un înalt grad de securitate. Un astfel de cifru este schematizat în figura 2.7.

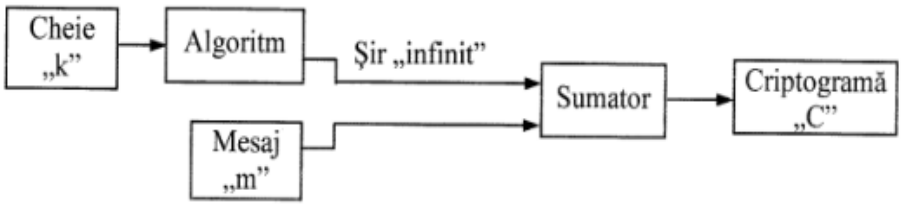


Fig. 2.6

Pe baza unui sistem de chei şi a fig. 2.6 unui algoritm, se obţine un şir „infinit” aleator.

Shannon a sugerat criptografilor utilizarea a două tehnici de cifrare numite de el „difuzie” si „confuzie”. Ideea de „difuzie” se referă la împrăştierea caracteristicilor statistice ale spaţiului mesajelor într-o structură statistică ce implica lungi combinaţii de litere din criptogramă. De exemplu se urmăreşte modificarea frecvenţei de apariţie a caracterelor din spaţiul mesajelor în sensul micşorării dispersiilor.

Efectul este ca un criptanalist trebuie să intercepteze o criptogramă mult mai lungă pentru a putea încerca o descifrare statistică.

Prin tehnica „confuziei” se urmăreşte corelarea cât mai complexă între criptogramă şi cheie aceasta implică dependenţa fiecărui caracter cifrat de întreaga structura de „cheie”. Sistemul secret cu acoperire unică prezentat mai sus se bazează pe tehnica „confuziei”. În plus acest sistem nu prezintă fenomenul de propagare a erorii necesitând în schimb o sincronizare perfectă între transmiţător şi receptor.

Spargerea acestui sistem implică cunoaşterea şirului aleator în întregime, cunoaşterea unor secvenţe mai scurte sau mai lungi nu compromit întregul sistem. În general s-ar putea asocia „tăria” unui sistem secret cu entropia cheii care este cu atât mai mare cu cât numărul lor este mai mare şi cu cât au probabilităţi mai apropiate.

$$H(k) = -\sum_{i=1}^{n} P_i \log_2 P_i$$

Dacă $P_i = \frac{1}{n}$ atunci $H(k) = \log_2 n$.

Se poate considera că sunt două concepte diferite în ce priveşte securitatea unui sistem: „securitate teoretică” şi „securitate practică”.

Din punctul de vedere al criptografului mai important este conceptul de „securitate teoretică”, el fiind interesat de şansele obţinerii unei soluţii unice prin încercarea fiecărei chei asupra criptogramei interceptate precum şi din determinarea cantităţii de criptograma necesară pentru obţinerea soluţiei unice.

În sistemul secret se poate recurge la alegeri statistice: acelea ale mesajului clar şi ale cheii. Entropia mesajului $H(M)$ este dată de relaţia:

$$H(M) = -\sum_{m \in M} P(m) \cdot \log_2 P(m)$$

şi reflectă siguranţa cu care putem aprecia ca un mesaj particular va fi transmis. Criptanaliştii încearcă să determine ce mesaj a fost transmis şi ce cheie a fost utilizată. Acest lucru ne conduce la ideea de entropie condiţionată, numită de Shannon, *echivocitate*. Dacă pentru mulţimea criptogramelor *C* şi întregul pozitiv *n* notăm cu C_n mulţimea criptogramelor de lungime *n*, putem defini echivocităţile:

$$H_C(M,n) = -\sum_{\substack{m \in n \\ c \in C_n}} P(c,m) \cdot \log P_c(m)$$

$$H_C(K,n) = -\sum_{\substack{k \in K \\ c \in C_n}} P(c,k) \cdot \log P_c(k)$$

$P(c, m)$ reprezintă probabilitatea cu care se recepţionează criptograma *C* fiind transmis mesajul *m*, iar $P_c(m)$ este probabilitatea aposteori cu care mesajul *m* a fost transmis ştiind că s-a recepţionat criptograma *C*. Probabilitatea de a fi transmis oricare mesaj este egală cu probabilitatea de a fi aleasă cheia corespunzătoare şi deci:

$$H_C(m,n) = H_C(k,n)$$

Dacă se consideră numărul literelor alfabetului egal cu 26 (deci mulţimea mesajelor are 26 de elemente) atunci, folosind logaritmul zecimal obţinem:

$$H_C(M,1) = 1,26$$
$$H_C(M,2) = 0,948$$
$$H_C(M,3) = 0,394$$
$$H_C(M,4) = 0,206$$

Echivocitatea dă o oarecare măsura cantitativă incertitudinii pe care o avem în prezicerea mesajului transmis (sau altfel spus cheia utilizată). Creşterea lui *n* nu garantează uşurarea decriptării.

Dacă alfabetului sursei de mesaje are *L* caractere iar sistemul foloseşte mesaje de lungime *n* atunci vor exista L^n mesaje posibile.

Dacă notăm $r_0 = \log_2 L$ atunci numărul de mesaje de lungime *n* posibile va fi: $2^{r_0 \cdot n}$. Cele $2^{r_0 \cdot n}$ mesaje posibile pot fi împărţite în două grupe:

- mesaje cu semnificaţie care au probabilităţi aproximativ egale $2^{r_n \cdot n}$ unde $r_n = \dfrac{H_c(m,n)}{n}$;
- mesaje fără semnificaţie care au probabilităţi neglijabile: $2^{r_0 n} - 2^{r_m n}$.

Dacă sistemul are *h* chei egal probabile, atunci $h = 2^{H(k)}$.

Probabilitatea ca un mesaj ales la întâmplare să aibă semnificaţie este: $2^{-r_0 n + r_n n} = 2^{-n(r_0 - r_n)}$. Astfel dacă un mesaj *m* se codifică într-o criptogramă *C* atunci probabilitatea de a găsi un mesaj cu semnificaţie prin descifrarea lui *C* este: $2^{-n(r_0 - r_n)}$. Dacă $r_0 = r_m$ atunci această probabilitate este egală cu 1. Dacă însă $r_n << r_0$ atunci această probabilitate este mică.

Diferenţa $d_n = r_0 - r_n$ se numeşte *redundanţă* a sistemului secret.

Deci dacă avem redundanţă d_n, probabilitatea de a obţine o decodificare semnificativă este 2^{-nd_n}. Dacă încercăm toate cheile posibile $\left(2^{H(k)}\right)$ atunci numărul decodificărilor cu semnificaţie va fi: $2^{H(k)} \cdot 2^{-nd_n} = 2^{H(k) - nd_n}$. Dacă $H(k)$ este mult mai mare decât nd_n atunci probabilitatea de a obţine o decodificare semnificativă este mai mare, deci probabilitatea de a determina mesajul corect este mică.

În practica mesajele nu au totdeauna aceeaşi lungime. Dacă *n* devine oricât de mare s-a constatat că redundanţa tinde asimptotic spre o anumită valoare *d*.

Dacă interceptăm primele *q* litere ale unei criptograme, numărul aşteptat de decodificări semnificative pentru aceste *q* poziţii este:

Aceasta sugerează să determinăm cel mai mic întreg n_0 care dă decodificarea unică aşteptată cu relaţia:

$$n_0 = \frac{H(k)}{d}$$

Vom încerca să determinăm n_0 pentru un cod monoalfabetic. Având 26! chei egal probabile, atunci $H(k) = \log_2 26! \cong 88,4$. Alfabetul are 26 de litere, deci $L = 26$ şi deci $r_0 = \log_2 26 \cong 4,7$. Pentru calculul lui *d* trebuie determinat debitul real al limbii r_n. Nu există nici o cale de a furniza o evaluare precisă a probabilităţii oricărui cuvânt al limbii, dar diferite studii statistice oferă anumite aproximări ale probabilităţilor diferitelor cuvinte. De exemplu pentru limba engleză Hellmann estimează $r_n \cong 1,5$, iar Deavours, $r_n \approx 1$. Deci $d = 4,7 - 1,5 = 3,2$.(redundanţa mai arată cât la sută din caracterele unui text sunt redundante. De exemplu dacă $d = 3,2$ iar $r_0 = 4,7$ atunci $\frac{3,2}{4,7} \approx 68\%$ din caractere sunt redundante.).

Este de aşteptat că pentru orice criptogramă de lungime mai mare ca 28 să avem soluţie unică. În practică însă e nevoie de mai mult de 28 de litere pentru a determină soluţia unică. Dar cu cât se interceptează mai multe criptograme, cu atât redundanţa limbajului creşte şi deci unicitatea soluţiei se obţine mai uşor. De asemenea numărul de caractere interceptate trebuie să facă astfel încât probabilitatea de utilizare a tuturor literelor să fie suficient de mare.

Una din cerinţele unui sistem secret aleator este ca pentru orice criptogramă dată *C*, decodificarea folosind toate cheile conduce la o selecţie aleatoare a tuturor mesajelor, în particular la o selecţie aleatoare atât a mesajelor semnificative cât şi nesemnificative. Aceasta implică în mod clar că numărul decodificărilor este mai mic decât putem garanta, dacă am ştiut că decodificarea unei criptograme, care a provenit de la un mesaj semnificativ trebuie să dea un mesaj semnificativ. Aceasta ultimă condiţie impune ca şi spaţiul criptogramelor să fie împărţit în două grupe:

- criptograme care provin de la codificarea mesajelor semnificative;
- criptograme care nu provin din codificarea mesajelor semnificative.

Dacă este posibilă o astfel de împărţire a spaţiului criptogramelor atunci se poate reduce spaţiul mesajelor astfel ca toate mesajele să fie semnificative.

Dar îndepărtarea mesajelor nesemnificative este echivalentă cu îndepărtarea redundanţei sistemului şi în acest caz orice sistem de cod ar satisface din punct de vedere al securităţii. Acest lucru este practic imposibil. Structura tuturor limbajelor naturale este mult mai complexă. Un sistem secret se poate considera ideal dacă creşterea volumului de text interceptat nu este în mod necesar de ajutor, deci dacă $H_c(k,n)$ nu tinde spre zero când *n* creste oricât de mult. Un exemplu de sistem ideal ar fi un cifru monoalfabetic pe un limbaj artificial în care toate literele sunt echiprobabile, iar literele succesive sunt alese independent.

Test de autoevaluare

1. Ce este redundanţa unui sistem secret şi care este utilizarea sa practică.
2. Care sunt tipurile de tehnici de cifrare propuse de Shannon.

Tema de autoinstruire

1. Propuneţi o schemă logică de combinare a mai multor sisteme secrete prin formarea produsului lor, pornind de la figura 2.4.

Lucrare de verificare

1. Estimarea numărului de mesaje false, a numărului de decriptări cu cheie falsă şi a numărului de mesaje false.
2. Sisteme de criptare compuse, proprietăţile cifrurilor obţinute şi aplicare lor în practică
3. Enunţaţi definiţia lui Shannon pentru un sistem de cifrare.
4. Enunţaţi cele cinci principii care stau la baza evaluării sistemelor secrete (principii elaborate de Shannon în 1940).
5. Definiţi procedura de cifrare.
6. Explicitaţi modul de construire a unui sistem secret S din două sisteme secrete R şi T.
7. Enunţaţi proprietăţile care se referă la rezistenţa sistemului secret.
8. Care sunt neajunsurile sistemelor secrete reale.

CAPITOLUL 3

SUCCESIUNI PSEUDOALEATOARE ÎN SECRETIZAREA INFORMAŢIEI

Cuvinte cheie: numere aleatoare, conceptul de aleator, generarea numerelor aleatoare.

OBIECTIVE

Înţelegerea şi cunoaşterea următoarelor noţiuni:

- numere aleatoare, secvenţe de numere aleatoare;
- aleatorismul şi teste de aleatorism: frecvenţă, serial, packer, autocorelare şi spectral;
- generarea succesiunilor pseudoaleatoare; scheme şi aplicaţii.

3.1 SUCCESIUNI DE NUMERE ALEATOARE

O succesiune finită de numere sau evenimente de orice fel este aleatoare dacă ea a fost obţinută într-un mod care să nu permită prevederea apariţiei diferitelor elemente ale succesiunii.

Noţiunea de *aleatoriu* se referă la condiţiile apriorice pentru formarea succesiunii şi nu la stabilirea aposteriorică a caracterului şi proprietăţilor succesiunii.

De cele mai multe ori succesiunile aleatoare se obţin în mod determinist dintr-o secvenţă scurtă cu ajutorul unor generatoare de succesiuni cum ar fi registrul de deplasare cu reacţii liniare.

Cunoaşterea mai multor secvenţe din şir nu permite stabilirea legii de generare şi de aceea se urmăreşte generarea unor succesiuni cu perioade lungi de repetiţie care sunt apoi testate asupra proprietăţilor lor aleatoare.

Un şir de numere reale

$$\{u_n\} = u_0, u_1, ..., u_n, ..., 0 \le u_i \le 1$$

se numeşte succesiune de numere aleatoare dacă ele sunt alese la întâmplare. Aceste succesiuni se dovedesc utile în multe tipuri de aplicaţii:

- simularea fenomenelor naturale;
- selectarea unui eşantion aleator pentru obţinerea de informaţii despre ceea ce poate constitui comportare tipică;
- analiza numerică;
- luarea deciziilor în criptografie.

Şirul $\{x_n\} = x_0, x_1, x_2, ...$ se numeşte şir b-nar dacă oricare termen al şirului este unul din numerele întregi 0,1,2,...,(b-1). Un şir binar (2-nar) este format din 0 şi 1.

La început cei ce voiau să obţină numere aleatoare în diferite lucrări ştiinţifice le realizau prin simularea unor evenimente aleatoare (aruncarea unei monede sau extragerea unei bile dintr-o urnă etc.). După introducerea calculatoarelor electronice obţinerea numerelor aleatoare se face prin intermediul unor programe. Asupra acestor metode s-a ridicat obiecţia cu privire la caracterul aleator al şirului generat deoarece fiecare număr obţinut este complet determinat de predecesorii săi. Şirurile de acest tip, generate în mod determinist sunt numite *succesiuni pseudoaleatoare*. Generarea succesiunilor aleatoare lungi s-a dovedit o operaţie dificilă. Pericolul constă în acea că şirul degenerează şi tinde să se stabilizeze la anumite cicluri de elemente. De aceea s-au elaborat metode adecvate care să garanteze obţinerea unor şiruri lungi de numere cu proprietăţi aleatoare.

O clasă de metode de generare o constituie metodele matematice. Una dintre aceste metode este şi metoda *congruenţial-liniară*.

În conformitate cu această metodă şirul $\{x_n\}$ se obţine pe baza unei relaţii de recurenţă,

$$x_{n+1} = ax_n + c \ (\text{modulo } m)$$

în care: - m este modulul şi $m > 0$;

- a se numeşte multiplicator, $0 < a < m$;

- c se numeşte increment, $0 \le c < m <$;

- x_0 este termenul iniţial, $0 \le x_0 < m$.

Aceşti parametri, care se mai numesc şi numere magice trebuie să ia astfel de valori încât să permită obţinerea unui şir cât mai mare.

Trebuie ca şirul de numere să aibă o distribuţie uniformă pe o mulţime finită (deci toate numerele să aibă aceeaşi probabilitate de obţinere).

Dacă se ia $m = 10$ şi $x_0 = a = c = 7$, atunci şirul obţinut va fi 6,9,0,7 după care se repetă. Şirul obţinut nu este aleator. În general şirurile care se obţin prin relaţii de recurenţă $x_{n+1} = f(x_n)$ au această proprietate de a se închide într-un ciclu (buclă) care apoi se repetă. Ciclul care se repetă se numeşte *perioad*ă. În exemplul de mai sus, perioada este 4, cu valorile alese, dar *perioada maxima* este 10.

Pentru rezultate bune trebuie ca parametrii m, a, c, x_0 să îndeplinească anumite condiţii:

- m să fie cât mai mare, având în vedere că perioada maximă este egală cu m;
- $(c, m) = 1$; c şi m să fie relativ prime între ele;
- a să fie de forma $a = p + 1$, unde p este un divizor prim al lui m.

Dacă $c = 0$ generatorul se numeşte congruenţial multiplicativ, relaţiile devenind:

$$x_{n+1} = ax_n \left(\text{modulo } m\right);$$
$$x_0 \neq 0 \; şi \left(x_0, m\right) = 1.$$

Numărul *a* trebuie să fie un element primitiv modulo *m*. Deci, dacă (*a*, *m*) = 1 iar λ este cel mai mic număr întreg pentru care $a^{\lambda} = 1\left(\text{mod}\, m\right)$, numărul λ este numit *ordinul lui a* modulo *m*. Orice număr *a* pentru care λ este maxim posibil, (în cazul nostru *m* – 1) se numeşte *element primitiv* modulo *m*.

Von Neumann a lansat ideea folosirii procedeelor aritmetice pentru generarea algoritmică de numere cu proprietăţi aleatoare. Astfel acesta propune metoda aşa numită a, părţii de la mijlocul pătratului, metoda care constă în următoarele:

- să presupunem că avem o reprezentare în baza *b* a numărului cu care lucrăm (*b* = 2 sau *b* = 10);
- dacă numărul are 2*a* cifre atunci pătratul său va avea 4*a* cifre (dacă numărul nu are 2*a* cifre se adaugă un 0 în faţă);
- se extrage din pătratul numărului un număr format din 2*a* cifre aflate la mijloc. Deci:

$$x_{n+1} = \left[\frac{x_n^2}{b^a}\right] - \left[\frac{x_n^2}{b^{3a}}\right] \cdot b^{2a},$$

unde prin [*x*] se înţelege partea întreagă a lui *x*. Dar, de exemplu, dacă

$$x_0 = 3792$$
$$a = 2$$
$$b = 10$$
$$x_0^2 = 14379264,$$

rezultă $x_1 = x_0$.

Notând $\left[\frac{x^2}{b^{3a}}\right] = k$ şi $\left[\frac{x^2}{b^a}\right] = \frac{x^2 - r}{b^a}$ atunci găsirea numărului care se repetă revine la studierea ecuaţiei diofantice:

$$x^2 - b^a x = r + b^{3a} k,$$

unde $x \in \left[0, b^{2a}\right] r, k \in \left[0, b^a\right]$.

A fost găsit ulterior un algoritm care permite determinarea numerelor repetabile în funcţie de *a*, *b* şi *r*.

Altă metodă este metoda *congruenţial aditivă* adică:

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \ldots + a_k x_{n-k} \quad \left(\text{modulo } m\right).$$

Dacă $m = p$ număr prim, atunci din teoria corpurilor finite se cunoaşte că şirul definit de $\{x_n\}$ cu relaţia de mai sus are *perioada maximă de lungime* $p^k - 1$. Dacă polinomul $f(x) = x^k - a_1 x^{k-1} - ... - a_k$ este primitiv modulo *p*, adică are o rădăcină care este element primitiv în corpul cu p^k elemente, constantele $a_1, a_2, ..., a_k$ se pot determina astfel încât perioada maximă să fie $\left(p^k - 1\right)$.

Pe lângă metodele matematice de generare a succesiunilor aleatoare pot fi utilizate unele procedee care se bazează pe fenomene fizice, de exemplu zgomotul electronic sau radioactiv.

De asemenea s-au realizat tabele cu numere întâmplătoare. Aceste tabele conţin de obicei numere întregi conform distribuite pe un interval. Cea mai mare tabelă de numere aleatoare este „A Million Random Digits and 100.000 Normal Deviates" publicată în 1955 de Rand Corporation.

3.2 Teste de aleatorism

3.2.1 Conceptul de aleatorism

Se pune problema de a decide dacă o secvenţă periodică are sau nu proprietăţi care să o apropie de o secvenţă aleatoare. Evident nici o secvenţă periodică nu este aleatoare în întregime. În criptografie ceea ce se doreşte de la o secvenţă periodică este să fie nedeterminată, astfel încât un criptanalist să nu poată prevedea ce urmează în continuare din secvenţa interceptată. În primul rând trebuie ca textul cifrat să fie mult mai scurt decât secvenţa periodică.

Aşa cum s-a văzut secvenţele utilizate sunt pseudoaleatoare generându-se cu ajutorul unor registre de deplasare cu reacţie. Dar textul cifrat prin utilizarea unei astfel de secvenţe trebuie să capete proprietăţile specifice unui text aleator.

Pentru a putea da o definiţie aleatorismului trebuie definit o serie de termeni specifici şirurilor aleatoare binare. Vom utiliza astfel termenul de *serie* ca fiind un număr de biţi identici. De exemplu în secvenţa 1011000101100 sun 8 serii:

- 3 serii de lungime 2;

- 1 serie de lungime 3.

Daca o anumită secvenţă de lungime *m* are perioada *p*, atunci simbolurile de pe poziţiile *m* şi $m + p$ satisfac relaţia:

$$S_{m+p} = S_m .$$

Pentru fiecare τ fixat putem compara biţii secvenţei date cu biţii secvenţelor translatate cu τ biţi.

Dacă notăm cu *A* numărul de coincidenţe şi cu *D* numărul (*p* – *A*), atunci se poate defini funcţia de autocorelaţie $C(\tau)$:

$$C(\tau)=\frac{A-D}{p}$$

Evident $C(p+\tau)=C(\tau)$ pentru orice τ, astfel că putem considera că τ verifică relaţia $0\le\tau<p$. Dacă $\tau = 0$ atunci avem autocorelaţie în fază, caz în care $A = p$ şi $D = 0$, iar $C(0) = 1$. Pentru $\tau\neq 0$ vom avea autocorelaţie în afara fazei.

Golomb a propus trei criterii de aleatorism pentru secvenţele binare:

C1. Dacă p este par atunci secvenţa de lungime p va avea $p/2$ biţi de valoare 0 şi $p/2$ biţi de valoarea 1. Dacă p este impar, atunci numărul de zerouri si numărul de cifre unu diferă printr-o unitate;

C2. Într-o secvenţă de lungime p, jumătate din numărul seriilor va avea lungimea 1, 1/4 va avea lungimea $2,...,1/2^k$ din numărul seriilor va avea lungimea k;

C3. Autocorelaţia defazată este constantă.

Vom considera în continuare aceste trei criterii ca o necesitate acceptată pentru a putea spune că o secvenţă este „bună”.

3.2.2 Teste de aleatorism

Pe lângă criteriile de aleatorism propuse de Gaulomb care sunt necesare dar nu şi suficiente se folosesc o serie de teste statistice pentru a determina proprietăţile aleatoare ale unei secvenţe. Vom prezenta în continuare cinci asemenea teste statistice care pot fi realizate pentru a demonstra o măsură cantitativă a aleatorismului.

Este de asemenea să stabilim anumite niveluri de încredere pentru fiecare test astfel încât să putem decide dacă o secvenţă a trecut sau nu testul respectiv. Vom nota în continuare cu n_0 numărul de zerouri şi cu n_1 numărul de „1”.

Testul 1: *Testul de frecvenţă*

Acesta este testul cel mai evident şi se aplică pentru a ne asigura ca $n_0 \cong n_1$. Pentru aceasta se calculează

$$\chi^2=\frac{(n_0-n_1)^2}{n},$$

unde $n=n_0+n_1$.

Evident că dacă $n_0=n_1$, rezultă $\chi^2=0\cdot\chi^2$ este cu atât mai mare cu cât este mai mare discrepanţa dintre frecvenţele observate şi cele aşteptate. Pentru a decide dacă valoarea obţinută este destul de bună astfel că secvenţa analizată să treacă testul, trebuie doar să comparăm valoarea obţinută cu un tabel al distribuţiei χ^2 (tabele care există şi în care găsim valoarea lui χ^2 pentru anumite niveluri de semnificaţie). Dacă valoarea obţinută este mai mică decât cea aflată în tabel, atunci secvenţa trece testul, în caz contrar ea este respinsă. De asemenea, secvenţa este respinsă dacă $\chi^2=0$, deoarece fiind prea bună poate fi suspectă.

Testul 2: *Testul serial*

Testul serial este folosit pentru a asigura că probabilităţile de tranziţii sunt egale sau foarte apropiate, ceea ce ar demonstra că fiecare bit este independent de predecesorul său. Să presupunem că perechea 01 apare de n_{01} ori, perechea 10 de n_{10} ori, 00 de n_{00} ori şi 11 de n_{11} ori. Evident că:

$$\begin{cases} n_{00} + n_{01} = n_0 \text{ sau } n_0 - 1 \\ n_{10} + n_{11} = n_1 \text{ sau } n_1 - 1 \end{cases}$$

(Apare $n_0 - 1$ sau $n_1 - 1$ deoarece într-o secvenţă de lungime *n* sunt doar n – 1 tranziţii).

Ideal ar fi ca $n_{00} = n_{01} = n_{10} = n_{11} = \frac{n-1}{4}$.

Pentru valorile $n_{00}, n_{01}, n_{10}, n_{11}, n_1$ şi n_0 determinate, se calculează valoarea:

$$\frac{4}{n-1}\sum_{i=0}^{l}\sum_{j=0}^{l}\left(n_{ij}\right)^2 - \frac{2}{n}\sum_{i=0}^{l}\left(n_i\right)^2 + 1,$$

unde *l* este numărul de caractere distincte.

Această valoare are o distribuţie χ^2 cu două grade de libertate. Există tabele corespunzătoare cu valorile lui χ^2. Se compară valoarea obţinută cu cea din tabele şi se decide dacă secvenţa trece testul sau nu.

Testul 3: *Testul packer*

Pentru orice număr întreg *m* sunt 2^m posibilităţi distincte pentru o secţiune de lungime *m*. În acest test împărţim secvenţa noastră în blocuri de lungime *m* şi apoi numărăm fiecare tip de secţiune de lungime *m* din secvenţa analizată.

Dacă frecvenţele sunt $f_0, f_1, ..., f_{2^m-1}$, atunci $\sum_{i=0}^{2^m-1} f_i = F = \left[\frac{n}{m}\right]$, unde cu $\left[\frac{n}{m}\right]$ am notat cel mai mare întreg care nu este mai mare decât $\frac{n}{m}$ (partea întreagă), *n* – lungimea totală a secvenţei. Atunci evaluăm:

$$\chi^2 = \frac{2^m}{F}\sum_{i=0}^{2^m-1}\left(f_i\right)^2 - F\ .$$

Se compara în final valoarea obţinută cu cea găsită în tabele ale variabilei χ^2 cu $2^m - 1$ grade de libertate şi se decide dacă secvenţa trece sau nu testul.

O variantă de aplicare a acestui test ar fi următoarea:

- se împarte secvenţa în blocuri de *m* biţi (de exemplu *m* = 4);
- se transformă în baza 10 numărul scris cu patru biţi obţinându-se numerele de la 0 la 15;
- şirul de numere astfel obţinut se împarte la rândul său în grupe de câte 5 a căror structură poate fi:

- cinci numere identice;
- patru plus unu;
- trei plus două;
- două plus două, plus unu;
- trei plus unu, plus unu;
- două plus unu, plus unu, plus unu;
- cinci numere diferite.

- se determină frecvenţele de apariţie ale acestor grupuri şi se compară cu valorile calculate pentru un şir de numere aleatoare;
- pe baza unor diferenţe admise se decide dacă secvenţa analizată trece sau nu testul.

Testul 4: *Testul de autocorelare*

Dacă secvenţa de *n* biţi pe care o testăm este $a_1, a_2, ..., a_n$ atunci se calculează:

$$A(d) = \sum_{i=1}^{n-d} a_i a_{i+d}; \quad 0 \le d \le n-1.$$

Dacă $d = 0$ atunci $A(0) = \sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} a_i = n_1$.

Dacă secvenţa are n_0 cifre 0 şi n_1 cifre 1 care sunt distribuite aleator atunci valoarea estimată a lui $A(d)$, $d \neq 0$ este

$$\mu = \frac{n_1^2 (n-d)}{n^2}$$

Pentru o secvenţă de *n* biţi se poate determina corelaţia care există între secvenţa dată şi orice deplasare circulară a acesteia cu *q* biţi ($0 < q < n$). De exemplu pentru $q = 1$ se calculează coeficientul de corelaţie serială *c*:

$$c = \frac{n(a_1a_2 + a_2a_3 + a_3a_4 + ... + a_{n-1}a_n) - (a_1 + a_2 + ... + a_n)^2}{n(a_1^2 + a_2^2 + ... + a_n^2) - (a_1 + a_2 + ... + a_n)^2}.$$

Acest coeficient apare frecvent în statistică si are valori cuprinse între -1 şi 1. Dacă *c* este 0, sau foarte apropiat de 0 indică o slabă corelaţie între biţii secvenţei, iar dacă *c* tinde spre 1 indică o dependenţă liniară totală.

Testul 5: *Testul spectral*

Se datorează lui COVEYOU şi Mc. PHERSON şi s-a impus datorită faptului că s-a dovedit că toate secvenţele care au trecut toate celelalte teste de aleatorism sunt acceptate şi de acest test, iar cele care au fost respinse de unul dintre teste sunt respinse şi de acesta.

Testul se bazează pe tehnica transformărilor Fourier aplicate funcţiilor de variabile întregi, transformări care permit scoaterea în evidenţă a caracterului aleator al unei secvenţe de numere întregi.

3.3 Scheme liniare şi neliniare pentru generarea succesiunilor pseudoaleatoare

Pentru a produce secvenţe aleatoare lungi dintr-un cod scurt, lucru dorit în criptologie, se utilizează în mod frecvent un generator format din registre de deplasare cu reacţie. Caracteristica principală este reprezantată de adăugarea unui circuit logic combinaţional cu funcţia de reacţie la registrul de deplasare. Prin efectul de reacţie se înţelege introducerea unei variabile care se aplică intrării seriale a registrului care va influenţa următoarea stare obţinută.

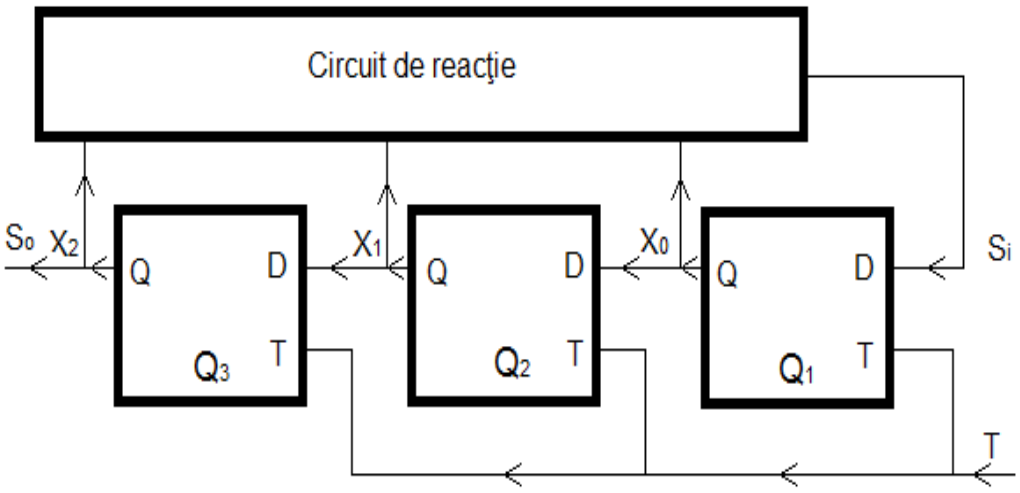


Fig.3.1

Un registru de deplasare cu trei celule de memorie tip CBB tip D şi un circuit de reacţie este dat in figura 3.1.

Un exemplu de registru de deplasare cu trei celule de memorie și un numărător modulo 2 este dat în figura de mai jos:

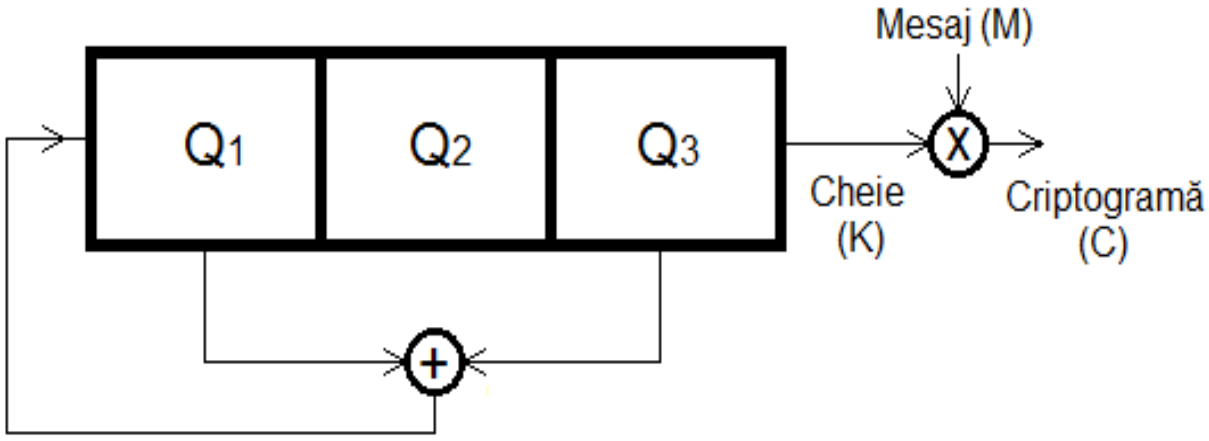


Fig.3.2

Dacă iniţial conţinutul registrului este 111, prin deplasări succesive combinaţiile formate din stările ieşirilor celor trei celule Q_1, Q_2, Q_3 vor fi:

mod 10	**Q_3,Q_2,Q_1**	**Si**
7	111	1
3	011	1
5	101	0
2	010	1
1	001	0
4	100	0
6	110	1
7	111	1

La ieşire se obţine şirul 1110100 care se care se repetă şi care se adună cu mesajul obţinându-se criptograma *C*. Perioada secvenţei depinde de numărul de celule care compun registrul precum şi de alegerea conexiunilor la sumatorul modulo 2.

Pentru tratarea teoretică a succesiunilor şi a schemelor generatoare, acestora li se ataşează polinoame ale căror coeficienţi (a_i) sunt elemente din câmpul *GF* (2).

Succesiunea obţinută la ieşire se poate de asemenea exprima polinomial, astfel:

$$a(x) = a_0 + a_1 x + ... + a_n x^n,$$

unde $a_0, a_1, ..., a_n$ constituie şirul de la ieşirea generatorului.

În exemplul de mai sus, $a(x) = 1 + x + x^2 + x^4$.

Polinoamele care sunt ataşate schemelor semnifică, prin coeficienţii acestora, conexiunile sumatoarelor de la diferite celule ale registrului.

De exemplu, în schema de mai sus, $g(x) = 1 + x^2$.

Conţinutul informaţional al registrelor se poate exprima de asemenea printr-un polinom de grad egal cu numărul de celule ale registrului de deplasare, cu puterile mici spre rangurile inferioare de la intrarea registrului şi cu puterile mari spre rangurile de la ieşire.

În figura de mai jos se prezintă un registru de deplasare cu conexiune inversă care calculează relaţiile de recurenţă liniară:

$$a_{i+k} = -\sum_{j=0}^{k-1} h_j \cdot a_{i+j}.$$

Se poate determina a_k şi următoarele valori, pe baza celor *k* valori anterioare ale succesiunii.

Structura schemei este determinată de polinomul:

$$h(x) = h_0 + h_1 x + ... + h_k x^k,$$

unde $h_0 \neq 0$ şi $h_k = 1$.

Conţinutul iniţial al registrului este secvenţa $a_0, a_1, ..., a_k$.

Prin deplasări succesive, la ieşirea schemei se obţine la început conţinutul registrului, apoi elementele formate pe baza combinaţiilor liniare, până când se va găsi un moment *n* după care succesiunea se repetă.

Perioada de repetiţie a succesiunii este dată de cel mai mic număr natural *n* pentru care polinomul $x^n - 1$ se împarte exact la $h(x)$:

$$\frac{x^n - 1}{h(x)} = g(x).$$

Polinomul $h(x)$ se numeşte *polinom caracteristic* al succesiunii $\{a_n\}$ şi al registrului de deplasare care o generează. Dacă gradul polinomului $h(x)$ este *k*, gradul polinomului $g(x)$ este $(n-k)$.

O succesiune generată de un registru de deplasare cu reacţie cu *k* ranguri (celule) are lungimea maximă dacă perioada sa este $2^k - 1$.

Dacă secvenţa $\{a_n\}$ are lungimea maximă atunci polinomul *h*(*x*) este ireductibil peste câmpul coeficienţilor. Reciproca acestei afirmaţii nu este adevărată, deci pot exista polinoame caracteristice ireductibile care nu dau succesiuni de lungime maximă.

De exemplu, polinomul: $h(x) = 1 + x + x^2 + x^3 + x^4$ este ireductibil peste *GF*(2) dar nu este primitiv. El divide polinomul $x^5 - 1$ iar schema respectivă generează o succesiune de lungime 5 care nu este de perioadă maximă $n = 2^4 - 1 = 15$.

Dacă *h*(*x*) este reductibil, *h*(*x*) = *s*(*x*) *t*(*x*), atunci exponentul lui *h*(*x*) este cel mai mic multiplu comun al exponenţilor lui *s*(*x*) şi *t*(*x*).

Când polinomul *h*(*x*) este ireductibil şi primitiv succesiunea obţinută la ieşirea registrului de deplasare nu depinde de condiţiile iniţiale (cu excepţia conţinutului zero peste tot care trebuie evitată).

Dacă $2^k - 1 = p$ este un număr prim, atunci fiecare polinom primitiv de grad *p* corespunde unei succesiuni de lungime maximă.

În cazul în care polinomul este un trinom de forma: $x^p + x^q + 1$, se demonstrează că rezultate mai bune se obţin dacă gradul *p* este astfel ales încât însuşi $2^p - 1$ să fie un număr prim (astfel de numere prime se numesc numere prime Mersenne).

Funcţionarea registrului poate fi descrisă matriceal. Fiecare stare a unui registru de deplasare cu reacţie având *k* ranguri poate fi considerată ca un vector *k* - dimensional. În acest caz, registrul de deplasare este un operator liniar care schimbă

fiecare stare în starea următoare. Matricea unui registru de deplasare cu reacţie este o matrice pătrată $k \times k$ având forma:

$$M = \begin{bmatrix} h_1 & 1 & 0 & \cdots & 0 \\ h_2 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ h_{k-1} & 0 & 0 & \dots & 1 \\ h_k & 0 & 0 & \dots & 0 \end{bmatrix},$$

unde h_i sunt coeficienţii polinomului caracteristic.

Dacă notăm cu $S(i)$ vectorul coloană având k elemente din $GF(2)$ care dau starea celulelor registrului la momentul i:

$$S(i) = \begin{pmatrix} a_0^i \\ a_1^i \\ \vdots \\ a_{k-1}^i \end{pmatrix},$$

unde cu a_j^i am notat starea celulei j la momentul i.

Se poate uşor observa că:

$$\begin{aligned} a_0^{i+1} &= h_1 a_0^i + h_2 a_1^i + \ldots + h_k a_{k-1}^i \\ a_1^{i+1} &= a_0^i \\ &\vdots \\ a_{k-1}^{i+1} &= a_{k-2}^i \end{aligned}$$

Deci starea registrului la momentul $i + 1$ va fi:

$$S(i+1) = M^T \cdot S(i)$$

şi

$$S(i+j) = \left(M^T\right)^j \cdot S(i).$$

Dacă se cunosc $2k$ biţi ai succesiunii de ieşire se poate stabili atât conţinutul iniţial al registrului cat si conexiunile la sumatoare prin rezolvarea unui sistem de $2k$ ecuaţii liniare, sau matriceal.

Dacă notăm cu $S(1)$ un vector coloană format din primii k biţi ai şirului de ieşire, $S(2)$ un alt vector coloană format din k biţi începând cu al doilea etc., se pot forma două matrice pătrate de dimensiune $k \times k$:

$$X(1)=\left[S(1),S(2),...,S(k)\right];$$
$$X(2)=\left[S(2),S(3),...,S(2k)\right].$$

Între cele două matrice există relaţia:

$$X(2)=T\cdot X(1).$$

Matricea *T* este matrice de trecere şi are forma:

$$T=\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ h_1 & h_2 & h_3 & \dots & h_k \end{pmatrix}.$$

Se observă cu uşurinţă legătura dintre matricele *M* şi *T*.
Deci *T* se poate calcula cu relaţia:

$$T=X(2)\cdot\left(X(1)\right)^{-1}.$$

Inversarea matricei $X(1)$ necesită cel mult operaţii de ordinul k^3 şi se pot uşor efectua cu ajutorul calculatorului pentru orice *k*.

Pentru funcţionarea mai sigură a schemelor secvenţiale registrul de deplasare se completează cu o logică neliniară. În figura de mai jos se prezintă un registru de deplasare cu reacţie pentru generarea unei succesiuni, care apoi este filtrată neliniar cu o funcţie *f* pentru a realiza o distribuţie uniformă între biţii de 0 si cei de 1, secvenţa care se poate utiliza drept cheie pentru cifrarea mesajelor:

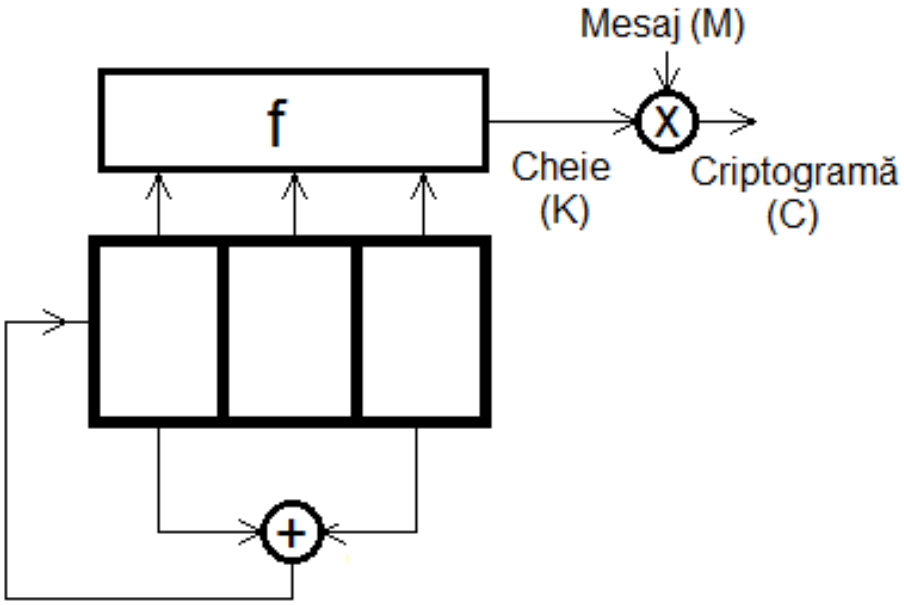


Fig.3.3

Există o clasă de generatori neliniari realizată pe baza de polinoame caracteristice primitive şi cu logică neliniară la două ranguri pentru a asigura ieşirea. Schema unui astfel de generator neliniar este dată mai jos:

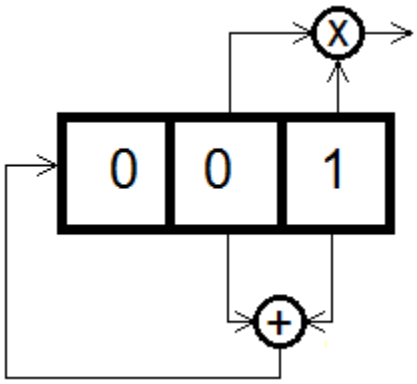


Fig.3.4

Succesiunea la ieşire poate fi privită ca o soluţie generală a ecuaţiei caracteristice a unui generator liniar echivalent, determinat de polinomul $X^6 + X^5 + X^4 + X^3 + X^2 + X + 1$:

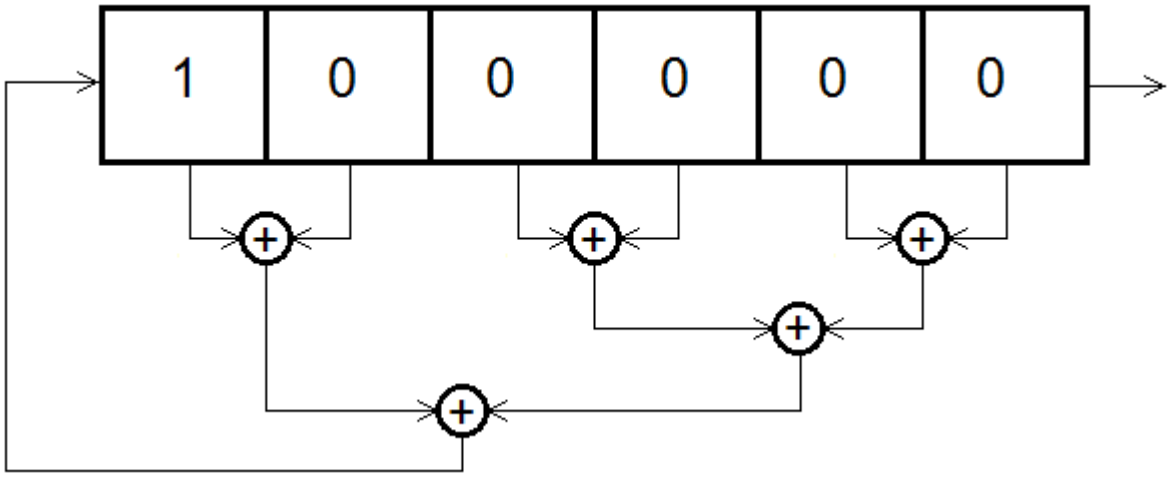
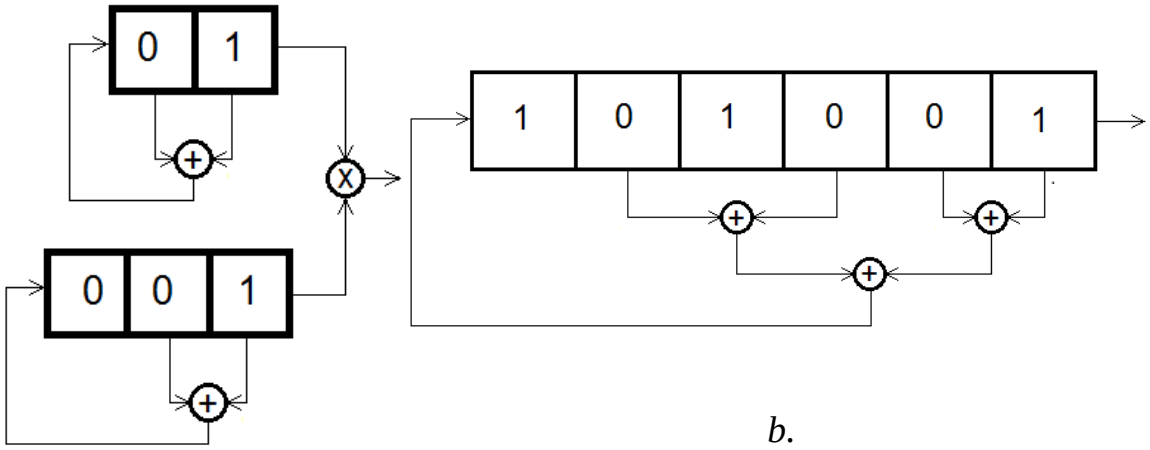


Fig.3.5

Analiza se poate extinde la generatoare având mai mult de două ranguri combinate neliniar, aşa cum se arata în figurile de mai jos:



a. *b.*

Fig.3.6

Schema din dreapta reprezintă echivalentul liniar al schemei din stânga.

Procesul neliniar analizat poate fi extins la sinteza unor generatoare mai complexe. Un astfel de generator este reprezentat mai jos:

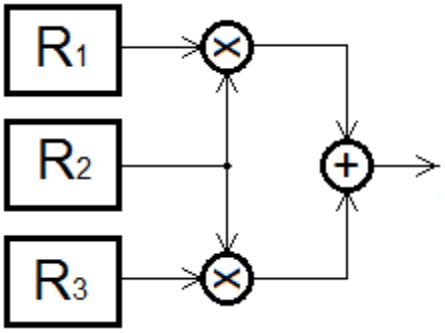


Fig.3.7

Dacă cele trei registre R_1, R_2 şi R_3 au polinoame caracteristice primitive de grade respectiv *r*, *s* şi *t* atunci se demonstrează că generatorul rezultat va avea complexitatea echivalenta cu: *rs* + (*s* + 1) *t*. Perioada de repetiţie va fi egală cu cel mai mic multiplu comun al numerelor: $2^r - 1; 2^s - 1; 2^t - 1$.

Succesiunea obţinută poate fi combinată cu alte succesiuni generate în mod analog într-un dispozitiv asemănător celui de mai jos:

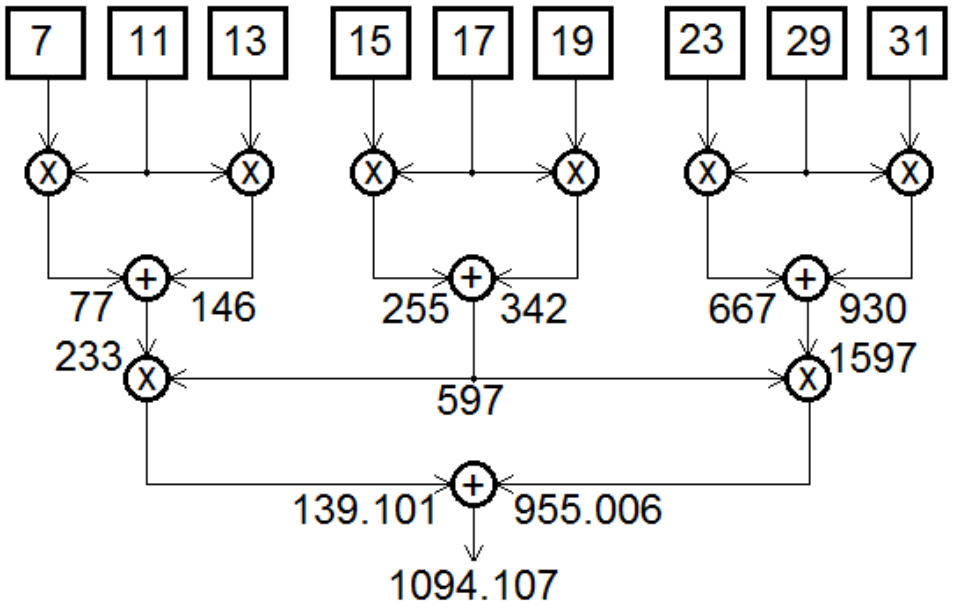


Fig.3.8

La aceste scheme se poate încerca maximizarea complexităţii succesiunii în funcţie de numărul total de ranguri conţinute în registre. Aşadar combinând registrele de deplasare cu reacţie liniară cu un număr mic de ranguri cu ajutorul operaţiilor neliniare se pot obţine generatoare foarte complexe cu perioade mari de repetiţie care pot fi folosite cu succes în criptografie.

Aplicaţie

Să analizam în continuare generatorul din figura de mai jos:

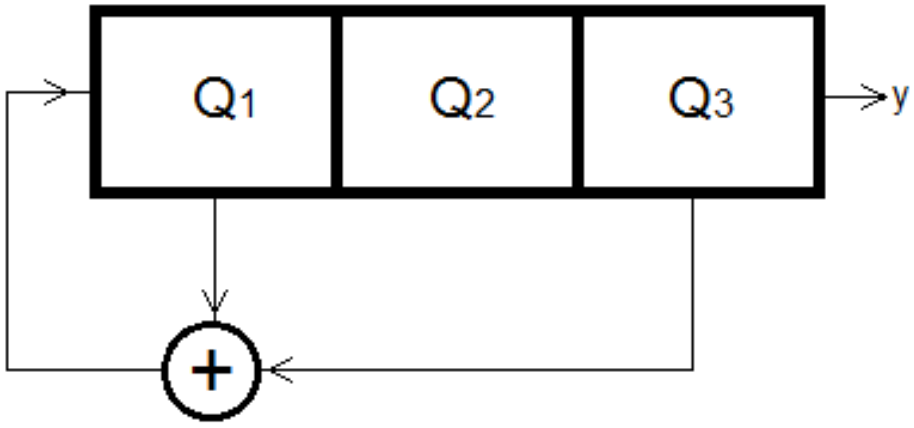


Fig.3.9

Dacă se încarcă iniţial registrul cu starea: 1,0,0 atunci tabela de adevăr care conţine funcţionarea acestuia este:

Q_1	Q_2	Q_3	y
1	0	0	0
1	1	0	0
1	1	1	1
0	1	1	1
1	0	1	1
0	1	0	0
0	0	1	1

Se defineşte funcţia de autocorelaţie $R(i)$ astfel:

$$R(i)=\sum_{k=1}^{2^n-1} Q(k)\cdot Q(k+i)=\begin{cases}2^n-1 \text{ pentru } i=0\\ 0 \text{ în rest}\end{cases}$$

$R(i)$ este o funcţie periodică cu perioada 2^n-1, deci $R(i)=R\left(i+2^n-1\right)$. $Q(k + i)$ reprezintă decalarea ciclică cu i biţi a secvenţei obţinută la ieşirea registrului.

Dacă îi asociem lui 0 valoarea numerică – 1, iar lui 1 valoarea numerică 1, atunci pentru ieşirea: $Y = 0011101$ se obţine:

$Q(k)$: -1 -1 1 1 1 -1 1
$Q(k + 0)$: -1 -1 1 1 1 -1 1
$Q(k + 1)$: 1 -1 -1 1 1 1 -1
$Q(k + 2)$: -1 1 -1 -1 1 1 1
$Q(k + 3)$: 1 -1 1 -1 -1 1 1
$Q(k + 4)$: 1 1 -1 1 -1 -1 1
$Q(k + 5)$: 1 1 1 -1 1 -1 -1
$Q(k + 6)$: -1 1 1 1 -1 1 -1

$R(0) = (-1)(-1) + (-1)(-1) + 1 + 1 + 1 + (-1)(-1) + 1 = 7$

$R(1) = (-1) + (-1)(-1) + (-1) + 1 + 1 + (-1) + (-1) = -1$

În mod analog: $R(2) = -1$

$R(3) = -1$

$R(4) = -1$

$R(5) = -1$

$R(6) = -1$

Existenţa maximului pronunţat în graficul funcţiei de autocorelaţie face posibilă utilizarea acestei secvenţe pseudoaleatoare ca secvenţa de sincronizare în transmisiunile de date. Dacă se intercalează în mod periodic la transmitere o astfel de succesiune şi se face corelaţia dintre secvenţa recepţionată cu o secvenţă generată local atunci la ieşirea corelatorului va apare un maxim ori de câte ori va apare secvenţa respectivă. Avantajul utilizării acestei secvenţe într-o astfel de aplicaţie constă în faptul că sincronizarea se poate face chiar dacă, „e" biţi ai acestuia au fost eronaţi, cu condiţia ca $e < \frac{2^n - 1}{2}$ (deci $e < 3$).

În registrele din schemele de generare a succesiunilor trebuie evitată situaţia în care toate celulele sunt în starea 0 deoarece nu se mai poate ieşi din această stare. Un exemplu de evitare a acestei stări este dat în schema de mai jos:

Presupunând că se cunosc primii 6 biţi ai succesiunii să determinăm acum structura şi starea iniţială a registrului. Pe baza celor 6 biţi cunoscuţi: 001110 se construiesc matricele $X(1)$ şi $X(2)$.

$$X(1) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}; \quad X(2) = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Din relaţia: $X(2) = T \cdot X(1)$ se obţine $T = X(2) \cdot X(1)^{-1}$.

Dar

$$X(1)^{-1} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

deci:

$$T = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Rezultă că $h_1 = 1; h_2 = 0; h_3 = 1$.

Primii trei biţi ai şirului ne arată starea iniţială a celor trei celule:

$$Q_3 = 0; Q_2 = 0 \text{ şi } Q_1 = 1.$$

Test de autoevaluare

1. Ce reprezintă numerele aleatoare si succesiunile pseudoaleatoare.
2. De ce sunt folosite testele de aleatorism şi care sunt tipurile de teste.
3. Care este utilizarea numerelor aleatoare.

Tema de autoinstruire

1. Propuneţi o schemă logică pentru aplicaţia prezentată la generarea succesiunilor pseudoaleatoare.

Lucrare de verificare

2. Definiţi metoda congruenţial-liniară de generare a succesiunilor pseudoaleatoare.
3. Explicitaţi metoda Von Neumann de generare algoritmică a numerelor cu proprietăţi aleatoare.
4. Definiţi metoda congruenţială aditivă de generare a succesiunilor pseudoaleatoare.
5. Care sunt criteriile de aleatorism ale lui Golomb pentru secvenţele binare.
6. Explicaţi metoda de testare a unui şir pseudoaleator prin aplicarea testului de frecvenţă.
7. Explicaţi metoda de testare a unui şir pseudoaleator prin aplicarea testului serial.
8. Explicaţi metoda de testare a unui şir pseudoaleator prin aplicarea testului poker.
9. Utilizând polinomul polinomul $f = x^5 + x^2 + 1$ cu coeficienţi în GF2 să se genereze o secvenţă pseudoaleatoare plecând din starea 00001.

CAPITOLUL 4

METODE DE CIFRARE

Cuvinte cheie: metode de cifrare, polinoame primitive, polinoame de permutare.

OBIECTIVE

Înţelegerea şi cunoaşterea următoarelor noţiuni:
- câmpuri Galois, funcţii de permutare;
- metode de cifrare, exemple de cifruri;
- simularea şi implementarea unor exemple practice de cifrare.

4.1 Corpuri Galois. Calcule

4.1.1 Câmpuri Galois

Fie câmpul *GF*(2) în care $GF(q)$ în care $q = p^n$ şi care se notează cu *F*. Elementele $x \in F$ pot fi reprezentate printr-un set de *q* polinoame:

$$x = C_0 t^{n-1} + C_1 t^{n-2} + ... + C_{n-2} t + C_{n-1}$$

unde $0 \le c_i < p$, iar *t* este un element primitiv al câmpului *F*, adică orice $x \neq 0$ poate fi exprimat prin $x = t^k, \quad 1 \le k \le q-1$.

În plus $t^{q-1} = 1$ unde $q - 1$ este cel mai mic întreg pozitiv cu această proprietate. Deci se poate scrie: $F = \left\{0, t, t^2, ..., t^{q-1}\right\}$.

Elementul *t* satisface o ecuaţie ireductibilă de forma:

$$f(t) = t^n + k_1 t^{n-1} + k_2 t^{n-2} + ... + k_{n-1} t + k_n = 0$$

unde $0 \le k_i < p$. Polinomul $f(t)$ se numeşte polinom primitiv peste câmpul *F* iar ecuaţia $f(t) = 0$ se poate folosi pentru a converti orice formă de puteri $\left(x = t^k\right)$ într-o formă polinomială $\left(x = \sum_{i=0}^{n-1} c_i t^{n-1-i}\right)$ exprimând $t^n = -\sum_{i=1}^{n} k_i t^{n-i}$.

De asemenea, fiecărui element $x \in F$ *i* se poate atribui o valoare numerică cu ajutorul relaţiei:

$$f = \sum_{i=0}^{n-1} C_i p^{n-1-i}, \quad 0 \le f \le q-1$$

Dacă $x=(C_0,C_1,...,C_{n-1})$ iar $y=(d_0,d_1,...,d_{n-1})$, atunci $x+y=(C_0+d_0,C_1+d_1,...,C_{n-1}+d_{n-1})(\text{modulo}p)$.

Dacă $x=t^k$ iar $q=t^m$ atunci $xy=t^{k+m}$ unde $k + m$ se calculează modulo (q – 1).

Aşadar orice element al câmpului F poate fi exprimat în două moduri:
- prin componentele sale $(C_0,C_1,...,C_{n-1})$;
- sub forma de puteri (t^k), în afara elementului $x = 0$.

Vom ilustra remarcile de mai sus considerând câmpul $GF(3^3)$ căruia îi asociem polinomul primitiv $f(t)=t^3+2t+1$. Elementele câmpului vor fi: $F=(0,t,t^2,...,t^{26})$.

$$0=(0,0,0) \quad cu \quad f_0=0$$
$$t=(0,1,0) \qquad f_1=3$$
$$t^2=(1,0,0) \qquad f_2=9$$
$$t^3=(0,1,2)$$

se obţine cu ajutorul polinomului primitiv $f(t)$, $t^3+2t+1=0$.

Calculul acesta se bazează pe faptul că valorile termenilor vor fi totdeauna pozitive și vor reprezenta $|x|mod(n)$ adică pentru -2 => 1 iar pentru -1 => 2, unde $t=3$, rezultând $t^3 = -2t-1 = t+2 = 3+2 = 5$, adică $f_3=5$.

Valorile următoare se calculează pornind de la această relație.

$$t^4 = (1,2,0) \quad f_4 = 15$$
$$t^5 = (2,1,2) \quad f_5 = 23$$
$$t^6 = (1,1,1) \quad f_6 = 13$$
$$t^7 = (1,2,2) \quad f_7 = 17$$
$$t^8 = (2,0,2) \quad f_8 = 20$$
$$t^9 = (0,1,1) \quad f_9 = 4$$
$$t^{10} = (1,1,0) \quad f_{10} = 12$$
$$t^{11} = (1,1,2) \quad f_{11} = 14$$
$$t^{12} = (1,0,2) \quad f_{12} = 11$$
$$t^{13} = (0,0,2) \quad f_{13} = 2$$
$$t^{14} = (0,2,0) \quad f_{14} = 6$$
$$t^{15} = (2,0,0) \quad f_{15} = 18$$
$$t^{16} = (0,2,1) \quad f_{16} = 7$$
$$t^{17} = (2,1,0) \quad f_{17} = 21$$
$$t^{18} = (1,2,1) \quad f_{18} = 16$$
$$t^{19} = (2,2,2) \quad f_{19} = 26$$
$$t^{20} = (2,1,1) \quad f_{20} = 22$$
$$t^{21} = (1,0,1) \quad f_{21} = 10$$
$$t^{22} = (0,2,2) \quad f_{22} = 8$$
$$t^{23} = (2,2,0) \quad f_{23} = 24$$
$$t^{24} = (2,2,1) \quad f_{24} = 25$$
$$t^{25} = (2,0,1) \quad f_{25} = 19$$
$$t^{26} = (0,0,1) \quad f_{26} = 1$$

Se observă că fiecare element al câmpului are două reprezentări şi *i* se asociază o valoare numerică. Pentru efectuarea operaţiilor de înmulţire şi împărţire se foloseşte reprezentarea exponenţială, iar pentru operaţiile de adunare şi scădere reprezentarea polinomială.

De exemplu:

$$t^5 \cdot t^{25} = t^{30} = t^4 = (1,2,0)$$
$$t^5 + t^{25} = (2,1,2) + (2,0,1) = (1,1,0) = t^{10}$$
$$t^3 : t^{14} = t^{29} : t^{14} = t^{15} = (2,0,0,0)$$

4.2 Funcţii polinomiale

Un polinom $f(x) = a_0x^n + a_1x^{n-1} + ... + a_{n-1}x + a_n$ cu coeficienţi $a_i \in GF(q)$ este numit polinom de permutare dacă pentru cele q elemente ale câmpului $(x_1, x_2, ..., x_q)$ imaginile $(f(x_1), f(x_2), ..., f(x_q))$ reprezintă o permutare a elementelor câmpului.

Fie de exemplu elementele câmpului *GF* (7) şi polinomul *f* (*x*) = 2*x* + 3. Acest polinom realizează următoarea permutare:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 0 & 2 & 4 & 6 & 1 \end{pmatrix}$$

Aceste funcţii de permutare pot fi utilizate în criptografie dacă literelor unui alfabet *L* li se asociază valori numerice corespunzătoare, iar cifrarea are loc după următoarea schemă:

$$\alpha \to x \to f(x) \to \alpha'$$

Deci unei litere α *i* se asociază valoarea numerică *x*. Se calculează *f* (*x*) care corespunde literei α'.

Această procedură realizează o substituţie simplă. În scopul de a da o complexitate mai mare se foloseşte o permutare polinomială cu mai mulţi parametrii care să fie modificaţi după o anumită regulă.

De exemplu polinomul: $f(x) = ax + b,\ a \neq 0\ \ a,b \in GF(q)$ asigură o substituţie simplă *a* şi *b* putând fi considerate ca elemente ale cheii de cifrare.

În general dacă permutarea este reprezentată prin: $y_i = f(x_i, a_1, a_2, ..., a_r)$, $a_i \in GF(q)$ unde $a_1, a_2, ..., a_r$ pot fi supuşi unor anumite restricţii impuse de funcţia utilizată, atunci descifrarea va impune utilizarea funcţiei de permutare inverse $f^{-1}(x)$.

Deci $x_i = f^{-1}(y_i, a_1, a_2, ..., a_r)$.

4.3 Metode de cifrare bazate pe funcţii de permutare

Vom prezenta mai multe metode bazate pe funcţiile de permutare care pot fi aplicate în raport cu numărul de litere al alfabetului *N* şi de câmpul asociat.

4.3.1 Metoda utilizată în cazul N=P^n

Vom ilustra această metodă printr-un exemplu.

Fie câmpul $GF\left(3^3\right)$, deci $N=3^3=27$ litere (cele 26 de litere ale alfabetului plus cuvântul - spaţiu), si fie polinomul de permutare: $y_i=a_ix_i+b_i$ cu $a_i\neq 0$. Între parametrii a_i şi b_i vom stabili o anumită regulă de recurenţă, de exemplu:

$$a_{i+2}=a_{i+1}+a_i;\quad i=1,2,3,...$$
$$b_{i+2}=b_{i+1}\cdot b_i;\quad i=1,2,3,...$$

şi valorile iniţiale:

$$a_1=(0,2,1)=t^{16}$$
$$a_2=(1,1,1)=t^{6}$$
$$b_1=(0,0,2)=t^{13}$$
$$b_2=(1,1,0)=t^{10}$$

Cifrarea cuvântului SECRET poate fi realizată astfel:

(a)	S	E	C	R	E	T
(b)	19	5	3	18	5	20
(c)	(2,0,1)	(0,1,2)	(0,1,0)	(2,0,0)	(0,1,2)	(2,0,2)
(d)	25	3	1	15	25	8
(e)	(0,2,1)	(1,1,1)	(1,0,2)	(2,1,0)	(0,1,2)	(2,2,2)
(f)	16	6	12	17	3	19
(g)	(0,0,2)	(1,1,0)	(2,2,0)	(1,2,2)	(1,2,0)	(1,1,2)
(h)	20	16	26	18	24	17
(i)	T	P	Z	R	X	Q

(a) textul în clar (x_i);
(b) valorile numerice asociate literelor;
(c) reprezentare polinomială;
(d) reprezentare exponenţială;
(e) a_i reprezentate polinomial;
(f) a_i reprezentate ca puteri;
(g) b_i reprezentate polinomial;
(h) valori numerice obţinute pentru reprezentările cifrate;
(i) criptograma (y_i).

Pentru descifrare se foloseşte funcţia inversă: $x_i=a_i^{-1}(y_i-b_i)$.

4.3.2 Metoda utilizată în cazul $M=p^n+1$

Carmichael a prezentat o metodă de cifrare bazată pe utilizarea funcţiilor de permutare introducând simbolul ∞ la elementele câmpului $GF\left(p^n\right)$, metoda care

poate fi utilizată atunci când numărul literelor din alfabet poate fi scris sub forma: $M = p^n + 1$.

Vom ilustra această metodă utilizând ca funcţie de permutare o funcţie raţională $R(x)$:

$$R(x) = y_i = \frac{a_i x_i + b_i}{c_i x_i + d_i},\ a_i b_i c_i \in GF\left(p^n\right) \text{ şi } a_i d_i - b_i c_i \neq 0.$$

Simbolul ∞ este reprezentat sub forma $\frac{x}{0}$ $(x \neq 0)$ unde $x \in GF\left(p^n\right)$. Astfel $R\left(-\frac{d_i}{c_i}\right) = \infty$ $c_i \neq 0$ şi $R(\infty) = \frac{a_i}{c_i}$, iar dacă $c_i = 0$ atunci $R(\infty) = \infty$.

Vom utiliza elementele câmpului $GF\left(5^2\right)$ cu polinomul primitiv $f(t) = t^2 + 4t + 2$. Alfabetul este cel al limbii engleze cu $26 = 5^2 + 1$ litere, iar corespondenta între litere şi elementele câmpului $GF\left(5^2\right)$ este dată mai jos:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
(a)	0	1	2	3	4	5	6	7	8	9	10	11	12	13
(b)	(0,0)	(1,0)	(1,3)	(4,3)	(2,2)	(4,1)	(0,2)	(2,0)	(2,1)	(3,1)	(4,4)	(3,2)	(0,4)	(4,0)
(c)		(1,2)	(2,3)											
(d)		(0,1)	(3,3)											
(e)		(3,4)	(1,2)											
(f)		(2,0)	(1,0)											

	O	P	Q	R	S	T	U	V	W	X	Y	Z
(a)	14	15	16	17	18	19	20	21	22	23	24	
(b)	(4,2)	(1,2)	(3,3)	(1,4)	(0,3)	(3,0)	(3,4)	(2,4)	(1,1)	(2,3)	(0,1)	
(c)	...											
(d)	...											
(e)	...											
(f)	...											

Linia (*a*) cuprinde reprezentarea exponenţială iar linia (*b*) reprezentarea polinomială a elementelor de câmp, iar (c), (d), (e) și (f) reprezintă parametrii a_i, b_i, c_i şi respectiv d_i.

Linia *(a)* folosește valorile obținute din:

$t^2 = -4t-2 = t+3$

$t^3 = t*t^2 = t*(t+3) = t^2+3t = t+3+3t = 4t+3$

$t^4 = t*t^3 = t*(4t+3) = 4t^2 + 3t = 7t+12 = 2t+2$

.....

Vom folosi ca regulă de variaţie a parametrilor a_i, b_i, c_i şi d_i selectând două matrice pătrate de ordinul 2 cu elemente ale câmpului $GF\left(5^2\right)$ şi stabilind:

$$M_{i+2} = M_i \times M_{i+1}, \quad \text{deci} \quad M_i \neq 0 \quad \text{şi} \quad M_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}$$

Fie de exemplu $M_1 = \begin{bmatrix} (1,2) & (0,1) \\ (3,4) & (2,0) \end{bmatrix}$; $M_2 = \begin{bmatrix} (2,3) & (3,3) \\ (1,2) & (1,0) \end{bmatrix}$ şi folosind regula stabilită calculăm M_3:

$$M_3 = \begin{bmatrix} (0,4) & (3,0) \\ (4,1) & (2,0) \end{bmatrix}$$

$$a_3 = (1,2)\cdot(2,3) + (0,1)\cdot(1,2) = t^{15}\cdot t^{23} + t^{24}\cdot t^{15} =$$
$$= t^{14} + t^{15} = (4,2) + (1,2) = (0,4)$$
$$b_3 = (1,2)\cdot(3,3) + (0,1)\cdot(1,0) = t^{15}\cdot t^{16} + t^{24}\cdot t =$$
$$= t^7 + t = (2,0) + (1,0) = (3,0)$$
$$c_3 = (3,4)\cdot(2,3) + (2,0)\cdot(1,2) = t^{20}\cdot t^{23} + t^7\cdot t^{15} =$$
$$= t^{19} + t^{22} = (3,0) + (1,1) = (4,1)$$
$$d_3 = (3,4)\cdot(3,3) + (2,0)\cdot(1,0) = t^{20}\cdot t^{16} + t^7\cdot t =$$
$$= t^{12} + t^8 = (0,4) + (2,1) = (2,0)$$

Pentru fiecare matrice calculată trebuie varificată condiția de ne-nulitate.
Se observă că:

$$a_1 d_1 - b_1 c_1 = (3,2) \neq 0$$
$$a_2 d_2 - b_2 c_2 = (3,1) \neq 0$$
$$a_3 d_3 - b_3 c_3 = (3,4) \neq 0$$

Cifrarea cuvântului COD se va face astfel:

	C	O	D
(a)	2	14	3
(b)	(1,3)	(4,2)	(4,3)
(c)	(3,0)	(2,3)	(3,1)
(d)	T	X	J

$$Y_1 = \frac{a_1 x_1 + b_1}{c_1 x_1 + d_1} = \frac{t^2 \cdot t^{15} + (0,1)}{t^{20} \cdot t^{15} + (2,0)} = \frac{(1,4) + (0,1)}{(3,2) + (2,0)} = \frac{(1,0)}{(0,2)} =$$

$$= \frac{t}{t^6} = t^{19} = (3,0)$$

$$Y_2 = \frac{a_2 x_2 + b_2}{c_2 x_2 + d_2} = \frac{t^{23} \cdot t^{14} + (3,3)}{t^{15} \cdot t^{14} + (1,0)} = \frac{(4,0) + (3,3)}{(4,1) + (1,0)} = \frac{(2,3)}{(0,1)} =$$

$$= t^{23} = (2,3)$$

$$Y_3 = \frac{a_3 x_3 + b_3}{c_3 x_3 + d_3} = \frac{t^{12} \cdot t^3 + (3,0)}{t^5 \cdot t^3 + (2,0)} = \frac{(1,2) + (3,0)}{(2,1) + (2,0)} = \frac{(4,2)}{(4,1)} =$$

$$= \frac{t^{14}}{t^5} = t^9 = (3,1)$$

Descifrarea se face folosind transformarea inversă:

$$x_i = -\frac{d_i y_i - b_i}{c_i y_i - a_i}$$

4.3.3 Metoda de cifrare a mai multor câmpuri şi funcţii de permutare

Metoda care va fi prezentată în continuare este echivalentă cu cifrarea independentă şi simultană a mai multor mesaje, fiecare cifrat printr-o metodă diferită dintre tipurile prezentate anterior.

Fie un alfabet *L* având *l* litere. Dacă se vor cifra câte *n* litere simultan, cifrare poligrafică, se poate spune că avem de-a face cu un alt alfabet *A* care are $N = l^n$ caractere (*n*-grame) de forma $\lambda = \overline{l_1 l_2 ... l_n}, l_i \in L$.

La cifrare textul clar compus din literele alfabetului *L* se împarte în blocuri de lungime *n*. Numărul *N* se poate scrie ca o suma de *k* numere astfel:

$$N = N_1 + N_2 + ... + N_k$$

în care $M_i = P_i^{m_i}$ sau $N_i = P_i^{m_i} + 1$.

Alfabetul *A* este astfel partiţionat în *K* subalfabeturi astfel încât $A = \bigcup_{i=1}^{k} A_i$ subalfabetele fiind disjuncte două câte două. Fiecărui subalfabet îi este destinat un câmp $GF\left(p_i^{m_i}\right) = F_i$ şi o funcţie de permutare $f_i(x)$ care poate fi polinomială dacă $N_i = p_i^{m_i}$ sau raţională dacă $M_i = p_i^{m_i} + 1$.

Modul în care cele *N* *n*-grame sunt împărţite în cele *k* subalfabete este următorul:

- se repartizează fiecărei litere ale alfabetului iniţial L o valoare numerică din secvenţă: 0,1,2,...,(l–1). Această repartiţie poate fi diferită pentru mesaje diferite.

- fiecare *n*-gramă se scrie sub formă numerică $\lambda = (v_1, v_2, ..., v_n)$ unde v_i este valoarea numerică a literei l_i din *n*-gramă;

- fiecărei *n*-grame i se ataşează un număr $\sigma = v_1 l^{n-1} + v_2 l^{n-2} + ... + v_{n-1} l + v_n$ care reprezintă valoarea numerică asociată acesteia;

- în subalfabetul A_1 vor fi incluse toate *n*-gramele pentru care $0 \le \sigma \le N_1 - 1$;

- în subalfabetul A_2 vor fi incluse toate *n*-gramele pentru care $N_1 \le \sigma \le M_1 + M_2 - 1$;

- în subalfabetul A_k vor fi incluse toate *n*-gramele pentru care $N_1 + N_2 + ... + N_{k-1} \le \sigma < N - 1$;

Se defineşte în continuare pentru fiecare *n*-gramă numărul ρ astfel $\rho = \sigma - (N_1 + N_2 + ... + N_{i-1})$ astfel încât $0 \le \rho \le N_i - 1$, *n*-gramă fiind inclusă în subalfabetul A_i.

Se scrie ρ sub formă polinomială:

$$\rho = C_o \rho_i^{m_i - 1} + C_1 \rho_i^{m_i - 2} + ... + C_{m_i - 2} \rho_i + C_{m_i - 1}$$

$$\rho = (C_o, C_1, C_2, ..., C_{m_i - 1})$$

Procedura de mai sus defineşte o bijecţie între *n*-gramele fiecărui subalfabet A_i şi elementele câmpului F_i. Cu ajutorul funcţiilor $f_i(x)$ se face apoi cifrarea după metodele descrise anterior.

Vom ilustra această procedură printr-un exemplu. Să consideram cazul $n = 2$ şi $l = 26$ iar repartizarea valorilor numerice pentru literele alfabetului normală:

A = 0	*H* = 7	0 = 14	*V* = 21
B = 1	*I* = 8	*P* = 15	*W* = 22
C = 2	*J* = 9	Q = 16	*X* = 23
D = 3	*K* = 10	*R* = 17	*Y* = 24
E = 4	*L* = 11	*S* = 18	*Z* = 25
F = 5	*M* = 12	*T* = 19	
G=6	N=13	U=20	

Numărul de bi-grame este $N = 26^2 = 676$.

Acesta se poate scrie astfel:

$$26^2 = 7^3 + 13^2 + (3^4 + 1) + (3^4 + 1)$$

Deci:

$$N_1 = 7^3 = 343 \qquad N_1 + N_2 = 512$$
$$N_2 = 13^2 = 169 \qquad N_1 + N_2 + N_3 = 594$$
$$N_3 = 3^4 + 1 = 82$$
$$N_4 = 3^4 + 1 = 82$$

Vom avea deci pentru subalfabete A_1, A_2, A_3 şi A_4.

O bigramă λ va aparţine: - subalfabetului A_1 dacă $0 \le \sigma \le 342$;

- subalfabetului A_2 dacă $343 \le \sigma \le 511$;
- subalfabetului A_3 dacă $512 \le \sigma \le 593$;
- subalfabetului A_4 dacă $594 \le \sigma \le 675$.

Fie diagrama: λ = *RM* = (17,12).

Deci: σ = 17 $\sigma = 17 \times 26 + 12 = 454$ şi de aici rezultă că această diagramă aparţine subalfabetului A_2 $(RM \in A_2)$. Se calculează ρ:

$$\rho = \sigma - N_1 = 454 - 343 = 111$$

Subalfabetului A_2 îi este asociat câmpul $GF\left(13^2\right)$ şi valorii 111 îi corespunde elementul (8,7) din $GF\left(13^2\right)$ deoarece $(8 \times 13 + 7 = 111)$.

Câmpului $GF\left(7^3\right)$ corespunzător primului subalfabet îi ataşăm polinomul primitiv $f(t) = t^3 - t - 5$. Celui de-al doilea subalfabet îi asociem câmpul $GF\left(13^2\right)$ şi polinomul primitiv $f(t) = t^2 + t + 2$, celui de-al treilea subalfabet îi asociem câmpul $GF\left(13^2\right)$ şi polinomul primitiv $f(t) = t^2 + t + 2$, la fel şi celui de-al patrulea subalfabet (A_4).

Cele patru funcţii de permutare utilizate vor fi:

- pentru A_1 $\quad y_i = x_i^5 + a_i$;
- pentru A_2 $\quad y_i = x_i^5 + b_i$;
- pentru A_3 $\quad y_i = c_i + \dfrac{1}{x_i^3 + 1}$;
- pentru A_4 $\quad y_i = d_i + \dfrac{1}{x_1^3 + 1}$.

Modul de variaţie a parametrilor este următorul:

$$a_{i+1} = (1,0,2)a_i + (2,3,6); \quad a_1 = (2,4,5)$$
$$b_{i+1} = (7,8)b_i; \qquad b_1 = (7,10)$$
$$c_{i+1} = (1,0,0,0)c_i; \qquad c_1 = (1,1,0,1)$$
$$d_{i+1} = (1,0,0,0)d_i; \qquad d_1 = (1,2,2,0)$$

În cazul nostru: $x_1 = (8,7)$.

Se calculează $y_1 = x_1^5 + b_1 = (8,7)^5 + (7,10)$.

Pentru descifrare se vor aplica funcţiile inverse care pentru A_2 este:

$$x_i^5 = y_i - b_i$$
$$x_i = \sqrt[5]{y_i - b_i}$$

Dacă n = 3 atunci vom avea $N = 26^3 = 17576$ care se poate scrie: $N = 26^3 = 7^5 + 3^6 + 3^3 + 13$ şi se vor putea cifra trigrame.

4.3.4 Metoda de cifrare prin folosirea unui singur câmp

Vom prezenta o metodă de cifrare în care numărul de litere ale alfabetului N este mai mic decât elementele câmpului asociat.

Fie $A = (\alpha_1, \alpha_2, ..., \alpha_N)$ alfabetul utilizat şi $F = GF(p^n)$ câmpul asociat, $(8^n > N)$.

Se face partiţionarea elementelor câmpului în două seturi:

$$S = (x_1, x_2, ..., x_N) \text{şi}$$
$$S' = F - S = (x_1', x_2', ..., x_{q-1}'); \quad q = p^n$$

Se stabileşte o bijecţie între cele *N* litere ale alfabetului *A* şi cele *N* elemente ale câmpului *S*.

Folosind o funcţie polinomială de permutare de forma $y_i = f(x_i, a_i)$ să cifrăm textul clar: $P_1P_2P_3...$. Se stabileşte bijecţia între literele textului clar şi valorile numerice $\lambda_1, \lambda_2, \lambda_3...$ unde $\lambda_i \in S$ şi se evaluează apoi $f(\lambda_i a_i) = g_i$.

Acum dacă $g_i \in S$ şi g_i corespunde literei C_i atunci P_i este cifrat în C_i. Dacă $g_i \in S'$ se repetă procedura obţinându-se şi aşa mai departe. Deoarece S' este finit, şi dacă prin alegerea făcută se evită situaţia:

$$f(x_i, a_i) = x_i$$

atunci după cel mult ($q - N$) paşi se va obţine un element din *S*.

Pentru un parametru α_i fixat dacă $x_j \neq x_k$ se va obţine $y_i \neq y_k$ însă prin variaţia parametrilor α_i este posibil ca pentru $x_i \neq x_k$ să se obţină $y_i = y_k$.

Vom ilustra această procedură prin două exemple:

Exemplul 1: Fie câmpul $GF(u_1)$ şi polinomul de permutare $y_i = a_i x_i + b_i$ unde: $a_i = 6^i \pmod{41}$ iar $b_{i+2} = b_{i+1} + b_i \pmod{41}$ şi $b_1 = 1$ iar $b_2 = 5$.

În setul *S* vor intra elementele câmpului care au valori între 0 şi 25, deci:

$$S = (0,1,2,...,25)$$

iar *S* = (0,1,2,...,25) iar *S* ' = (26,27,..,40)

Să cifrăm textul PERMUTARE.

	P	E	R	M	U	T	A	R	E
(a)	15	4	17	12	20	19	0	17	4
(b)	6	36	11	25	27	39	29	10	19
(c)	1	5	6	11	17	28	4	32	36
(d)	9	15*	14*	24	24	1*	4	2*	13*
(e)	J	P	O	Y	Y	H	E	C	N

(a) valorile numerice asociate textului clar

(b) $a_i = 6^i$ i = 1,2,3,4,5,6,7,8,9

(c) b_i $i = \overline{1,9}$

(d) valorile numerice ale literelor criptogramei

(e) criptograma.

Valorile marcate cu " * " s-au obţinut după mai multe iteraţii.

Pentru descifrare se foloseşte funcţia inversă:

$$v_i = 6^{40-i}(y_i - b_i)$$

aplicându-se aceleaşi iteraţii în cazul obţinerii unor elemente din *S*'

Exemplul 2: Fie un alfabet cu *N* = 1000 caractere utilizate, format din 26 de litere, $26 \times 26 = 676$ diagrame şi 298 dintre cele mai frecvente trigrame: $1000 = 26 + 26^2 \times 26 + 298$. Cifrarea se face în câmpul $GF(11^3)$ care conţine 1331 elemente şi care pot fi generate cu ajutorul polinomului primitiv: $f(t) = t^3 + 10t + 4$. Se utilizează funcţia de generare polinomială:

$$y_i = a_i x_i^3 + b_i$$

Elementele câmpului $GF(11^3)$ sunt partiţionate în două seturi: setul *S* de 1000 de elemente şi setul S' cu 331 elemente:

S = {0,1,2,...,999} şi S' = {1000,1001,...,1331}

Se face corespondenţa între cele 1000 de caractere ale alfabetului şi elementele setului S, sub forma:

A = 000
TE = 789
STR = 672 etc.

Pentru cifrare se împarte textul în clar în secvenţe de lungime 1,2 sau 3 elemente conţinute în *A*; fie acestea $\lambda_1\lambda_2\lambda_3$....

Se înlocuieşte secvenţa λ_i prin elemente de câmp x_i cu valoarea stabilită prin corespondenţă. Parametrii a_i şi b_i pot fi variaţi în modul următor:

$$a_i = t^{3i+1}$$

$$b_i = t^{n^2+n+2}$$

iar reprezentările cifrate se obţin cu relaţia: $y_i = a_i x_i^3 + b_i$ şi dacă $y_i \in S$ atunci el corespunde caracterului λ_k. Deci λ_i s-a cifrat prin λ_k.

Dacă însă $y_i \in S'$ atunci se repetă procedeul cu aceeaşi parametrii până când $y \in S$.

Ca text cifrat poate fi folosit şi cel obţinut din înşiruirea valorilor numerice ale elementelor $y_i \in S$ obţinute în urma procesului iterativ de cifrare.

Pentru descifrare se utilizează funcţia inversă: $x_i = \left[\left(y_i - b_i\right)/a_i\right]^{\frac{1}{3}}$ cu specificaţia că $t^{\frac{1}{3}} = t^{887}$.

Test de autoevaluare

1. Ce reprezintă o metodă de cifrare şi de căte tipuri sunt.
2. Care sunt diferenţele dintre metodele de cifrare bayate pe funcţii de permutare şi .
3. Care este utilizarea numerelor aleatoare.

Tema de autoinstruire

1. Încercaţi crearea unei scheme logice care să algoritmeze metoda Carmichael, în speţă exemplul ataşat acesteia.

Lucrare de verificare

1. Fie câmpul GF(26) căruia îi asociem polinomul de permutare *f(x)=6x+7*.

Să se cifreze următoarea expresie:

CODE

2. Fie câmpul GF(3^3) căruia îi asociem polinomul de permutare $y_i=a_i*x_i+b_i$

Între parametrii a_i şi b_i există următoarele relaţii de recurenţă:

$$a_1=t^1$$

$a_{i+2}=a_{i+1}+a_i$ $a_2=t^3$

$b_{i+2}=b_{i+1}*b_i$ cu valorile iniţiale $b_1=t^2$

$i=1,2,3,...$ $b_2=t^{15}$

Să se cifreze cuvântul înscris în tabelul următor cu completarea acestuia în mod corespunzător.

(a)	**C**	**O**	**D**	**E**
(b)				
(c)				
(d)				
(e)				
(f)				
(g)				
(h)				
(i)				

Unde:

(a) textul în clar (x_i)

(b) valorile numerice asociate literelor;

(c) reprezentare polinomială;

(d) reprezentare exponenţială;

(e) a_i reprezentate polinomial;

(f) a_i reprezentate ca puteri;

(g) b_i reprezentate polinomial;

(h) valori numerice obţinute pentru reprezentările cifrate;

(i) criptograma y_i

Nota 1: Valorile numerice asociate literelor alfabetului sunt prezentate în tabelul următor:

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

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 0

3. Fie câmpul GF(3^3) căruia îi asociem polinomul de permutare $y_i=a_i*x_i+b_i$
Între parametrii a_i şi b_i există următoarele relaţii de recurenţă:

$a_{i+2}=a_{i+1}+a_i$

$b_{i+2}=b_{i+1}*b_i$ cu valorile iniţiale

$i=1,2,3,...$

$a_1=t^1$

$a_2=t^3$

$b_1=t^2$

$b_2=t^{15}$

Să se cifreze cuvântul înscris în tabelul următor cu completarea acestuia în mod corespunzător.

(a)	**S**	**A**	**L**	**T**
(b)				
(c)				
(d)				
(e)				
(f)				
(g)				
(h)				
(i)				

Unde:

(a) textul în clar (x_i)

(b) valorile numerice asociate literelor;

(c) reprezentare polinomială;

(d) reprezentare exponenţială;

(e) a_i reprezentate polinomial;

(f) a_i reprezentate ca puteri;

(g) b_i reprezentate polinomial;

(h) valori numerice obţinute pentru reprezentările cifrate;

(i) criptograma y_i

Nota 1: Valorile numerice asociate literelor alfabetului sunt prezentate în tabelul următor:

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

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 0

4. Fie câmpul GF(26) căruia îi asociem polinomul de permutare $f(x)=6x+7$. Să se cifreze următoarea expresie:

GEST

CAPITOLUL 5

DISPOZITIVE ŞI MAŞINI CRIPTOGRAFICE

Cuvinte cheie: maşini de criptare, maşina Enigma, maşini poligramice şi tomogramice.

OBIECTIVE

Înţelegerea şi cunoaşterea următoarelor noţiuni:
- care sunt şi cum funcţionează dispozitivele de criptare;

Explicarea şi exemplificarea funcţionării maşinilor criptografice.

Criptografia a jucat un rol foarte important în istorie, iar producătorii de cifruri au dorit realizarea unor sisteme de cifrare cât mai rezistente. Dacă la început şi-au pus mai puţin problema operativităţii şi a productivităţii, odată cu creşterea volumului de corespondenţă ce trebuia cifrată, aceste probleme au devenit foarte importante. De asemenea, crescând complexitatea metodelor de cifrare a crescut şi riscul de a greşi. Toate acestea au impus necesitatea realizării unor dispozitive care să facă mai sigură şi mai rapidă activitatea de cifrare. Au apărut mai întâi simple rigle, discuri, abace, apoi acestea au fost mereu perfecţionate ajungându-se la adevărate maşini la început mecanice, apoi electromecanice ajungându-se astăzi la realizarea unor sisteme de cifrare bazate pe utilizarea calculatoarelor şi a microprocesoarelor specializate.

5.1 Dispozitive şi maşini criptografice

5.1.1 Evoluţia dispozitivelor criptografice

Dispozitivele şi maşinile criptografice pot fi grupate în şase generaţii care urmăresc evoluţia dezvoltării tehnicii şi tehnologiilor de realizare precum şi a metodelor şi principiilor de cifrare.

În prima generaţie pot fi incluse acele dispozitive simple bazate pe principiul riglei şi care au apărut începând se pare cu anul 475 î.e.n.

Dintre acestea cel mai vechi dispozitiv de cifrat a fost SKYTALA spartanilor care va fi descrisă ulterior.

A doua generaţie este generaţia maşinilor mecanice bazate pe cilindri, tamburi, roţi dinţate şi pârghii. Acestea au apărut la sfârşitul secolului al XIX-lea, s-au dezvoltat apoi rapid atingând apogeul în timpul primului război mondial. Aceste maşini utilizează, pentru asigurarea secretului, cheile în număr foarte mare şi practic nerepetabile. Deci, algoritmul de cifrare se complică dar se automatizează, iar cele mai perfecţionate maşini din această generaţie asigură şi tipărirea textelor clare şi a criptogramelor.

Acest lucru elimină greşelile de transcriere, măreşte siguranţa şi operativitatea cifrării.

A treia generaţie apare la mijlocul secolului al XIX-lea, fiind condiţionată de apariţia telegrafului. Ea se dezvoltă în paralel cu a doua generaţie împrumutându-i multe din principiile şi modalităţile practice de realizare. Această generaţie impune metode de cifrare şi criptanaliză, maşinile fiind în general electromecanice.

A patra generaţie este generaţia aparaturii electrice şi începe să se impună puţin înainte de izbucnirea primului război mondial. Primele realizări au utilizat sisteme cu relee sau se bazau pe transformări ale maşinilor de scris electrice şi pe utilizarea roţilor de cod.

Ele au fost considerate mult timp deosebit de rezistente chiar dacă în 1929 Friedman a reuşit să spargă un asemenea cod. Mai târziu prof. ABRAHAM SINKOV de la universitatea din ARIZONA a demonstrat eficacitatea utilităriii teoriei grupurilor la spargerea acestor coduri.

A cincea generaţie a debutat puţin înaintea izbucnirii celui de-al doilea război mondial, odată cu accentuarea dezvoltării electronicii şi apariţia calculatoarelor cu memorii magnetice.

Alături de telegraf, comunicaţiile prin radio influenţează tot mai mult criptologia impunând o dezvoltare rapidă a criptoanalizei.

A şasea generaţie apărută după deceniul al VI-lea al secolului XX se impune ca fiind generaţia microelectronicii, a informaticii, a procesoarelor specializate. Se bazează pe algoritmi complecşi, iterativi, pe utilizarea cheilor aleatoare şi unice, pe trecerea de la folosirea ca suporţi de informaţie a benzilor sau colilor de hârtie la dischete sau benzi magnetice.

Dar evoluţia acestor mijloace este departe de a fi încheiată ea continuând şi astăzi într-un ritm şi mai alert.

Înainte de a trece la prezentarea unor dispozitive şi maşini criptografice considerăm util să facem o prezentare a principalilor beneficiari ale acestora.

5.1.1 Utilizarea mijloacelor criptografice

Beneficiarii de mijloace criptoanalitice pot fi clasificaţi în cinci categorii mai importante:

- militari;

- diplomaţi;

- servicii de informaţii;

- oameni de afaceri;

- corespondenţi particulari.

Din aceste categorii militarii sunt de departe principalii utilizatori ai mijloacelor criptologice.

Un exemplu este edificator pe această linie. Astfel la foarte scurt timp după ce inginerul suedez Boris Hagelin a inventat maşina de codificat C-36, uzinele Smith-Carana au construit pentru armata SUA circa 14000 de exemplare ale acestei maşini cu denumirea „CONVERTER M-209”. Se disting două tipuri de maşini criptografice militare:

- modele „strategice” care funcţionează la eşaloanele superioare ale armatelor;

- modele „tactice” care servesc la transmisiuni de campanie.

Dintre acestea, maşinile „strategice” trebuie să asigure un grad superior de securitate şi de aceea sunt mai costisitoare. Mesajele prelucrate cu aceste maşini pot păstra valabilitatea şi importanţa, perioade mai îndelungate de timp de ordinul anilor sau chiar deceniilor.

Maşinile „tactice” nu trebuie să fie atât de complexe pentru că informaţiile de la acest nivel au o valoare limitată în timp. Astfel, poziţia unei subunităţi să zicem, nu mai prezintă nici un interes pentru inamic la un interval de câteva zile, chiar ore după transmiterea mesajului. Aceste maşini vor utiliza coduri de campanie care nu necesită în practică mult timp şi efort pentru a fi decodificate.

Ar fi greşit să se creadă că serviciile diplomatice nu reprezintă decât o piaţă restrânsă pentru dispozitivele de codificare. Să admitem că, de exemplu, 120 de state fac schimb de reprezentanţe, se obţine un total de peste 28000 de maşini criptografice dacă se consideră doar câte două maşini pentru ambasadă sau consulat şi alte două la Ministerul de Externe al fiecărui stat.

Serviciile de informaţii sunt la rândul lor mari „consumatoare” de maşini de cifrat dacă avem în vedere aria largă a activităţilor pe care le desfăşoară: contraspionaj, securitate naţională, şi în multe situaţii spionaj (economic, militar, etc.).

Dispozitivele şi maşinile criptografice folosite de industriaşi şi oameni de afaceri reprezintă mai puţin de 5% din totalul acestora. Mai trebuie luate în considerare aici organismele financiare (marile bănci, instituii de credit, companii de asigurări) care utilizează în mod frecvent maşinile criptografice.

Costul relativ mare al aparaturii de cifrare face ca particularii care le utilizează să fie extrem de puţini. În sfârşit organizaţiile criminale, în special cele ce se ocupă cu traficul de orice natură, recurg deseori la criptografie.

De exemplu, în perioada prohibiţiei navele care soseau în SUA cu alcool, primeau prin radio instrucţiuni codificate de la căpeteniile traficanţilor.

Dar numai între anii 1927–1929 o echipă condusă de o femeie Elisabeth Smith Friedman a descifrat peste 12000 din aceste mesaje. Iar poliţia dispunea la rândul ei de păzitoare de coastă dotate cu laboratoare de decodificare.

Article I. 5.2 Funcţionarea maşinilor criptografice

Facem în continuare o prezentare a dispozitivelor şi maşinilor criptografice respectând pe cât posibil ordinea cronologică a apariţiei lor, dar făcând în acelaşi timp distincţie între cele două categorii.

SKYTALA apărută în secolul V î.e.n. era un baston în jurul căruia se înfăşura spiră lângă spiră o panglica foarte îngustă de piele, papirus sau pergament pe care, pe generatoare se scriau literele mesajelor. După scrierea textului panglica era dublată, mesajul devenea indescifrabil, întrucât literele erau dezasamblate. Mesajul se putea descifra numai de persoana care dispunea de un baston de lungime şi grosime identice cu dimensiunile iniţiale pe care să fie înfăşurate din nou panglica primită. Ea realiza o transpoziţie, fiind o primă formă a acestei metode de cifrare.

Criptograful lui Alberti. Era alcătuit din două discuri concentrice cu diametre diferite, suprapuse. Fiecare disc era împărţit în 24 de sectoare pe care erau înscrise literele şi cifrele. Pe discul magnetic erau înscrise 20 de litere (fără H, J, K, U, W) şi cifrele de la 1 la 4, iar pe al doilea 23 de litere (fără H, K, Y) şi conjuncţia „ET”. Ordinea lor era arbitrară. Pentru cifrare se stabilea o cheie de exemplu *D* = *A*. Aceasta însemna că pentru cifrare litera *D* de pe discul mic se aşează în dreptul literei *A* de pe discul mare şi apoi începea cifrarea. Alberti recomanda pentru mărirea rezistenţei schimbarea cheii după un număr de cuvinte. Criptograful lui Alberti a fost perfecţionat de Silvester, Argenti şi alţii constituind un element de bază pentru criptografele de tip disc apărute ulterior. Silvester Porta a împărţit discurile în 26 de sectoare utilizând apoi toate cele 26 de litere, criptograful său permiţând o substituţie simplă dar complet literală.

Criptograful lui Wheatstone. A fost inventat în 1867, este realizat tot cu două discuri şi realizează o substituţie dublă cu cheie finită. cele două discuri erau fixe, pe unul din ele cele 26 de litere fiind dispuse în mod normal iar pe cel de-al doilea literele erau aşezate aleator conform unei chei. Peste cele două discuri se mişcau două ace într-un raport 27 la 26 de rotiri producându-se astfel un decalaj continuu şi ciclic între cele două alfabete.

Aparent se produce o cifrare cu autocheie, ceea ce i-a făcut pe mulţi specialişti să-l declare indescifrabil.

Mai târziu însă *Kerchoffer* a demonstrat că numărul alfabetelor este limitat şi că de fapt el realizează o substituire cu dublă cheie.

Cifrarea se face pornind de la o cheie, de exemplu *A* = *C*, adică acul mare se afla la *A* iar acul mic se afla la *C*. În continuare se duce acul mare la prima literă a textului clar şi se ia drept literă cifrată, litera pe care se opreşte acul mic. Operaţia se repetă pentru a doua literă ş.a.m.d, deplasarea făcându-se în sensul acelor de ceasornic.

La descifrare se procedează invers plecând din aceleaşi puncte determinate de cheie şi se deplasează acele până când acul mic ajunge la poziţia primei litere din criptogramă, prima literă a textului în clar fiind litera la care a ajuns acul mare.

Criptograful Bazaries. A fost inventat în 1891 fiind considerat o mare realizare din punct de vedere criptologic la vremea respectivă.

Era realizat dintr-un ax şi 20 de inele care aveau înscrise pe marginea exterioară un număr, o literă de identificare şi un alfabet de 25 de litere (fără W).

Cheia sistemului se formează din litere sau cifre şi reprezintă ordinea de montare a inelelor pe ax, ea fiind valabilă atât la cifrare cât şi la descifrare. Cifrarea se face împărţind textul clar în grupe de 20 litere şi apoi aranjând inelele, astfel încât, pe o generatoare să apară înscrise cele 20 de litere ale primei grupe. Textul cifrat poate fi considerat oricare alt grup de litere de pe o altă generatoare.

Descifrarea se face în mod identic.

Rezistenţa acestui criptograf constă în numărul mare al cheilor: 20!

S-a stabilit că cifrarea cu ajutorul cilindrului lui Bazaries a asigurat secretul mesajelor în funcţie de pregătirea criptanaliştilor, de la 6 ore până la 3 zile. De aceea, ca măsură de protecţie a secretului, cheia era schimbată la fiecare 6 ore.

Criptograful lui Soudart. A fost inventat în anul 1914 şi realizează un cifru bazat pe dubla transpoziţie. Principiul este asemănător celui al lui Bazaries având tot 20 de inele montate pe un ax. Marginea exterioară este împărţită în două, pe partea stângă erau trecute cele 25 de litere ale alfabetului (fără W), iar pe marginea din dreapta, în dreptul literelor erau trecute literele textului clar.

Cheia de cifrare era literală sau numerică putând fi formată din cel mult 20 de caractere (litere sau cifre de la 1 la 20) într-o ordine aleatoare. De exemplu cheia DISCRET va deveni în numeric 2461537 ceea ce determină montarea inelelor pe cilindru în ordinea dată de cheie. După montare se rotesc inelele până apare pe poziţia iniţială cuvântul cheie.

2	4	6	1	5	3	7
DM	IE	ST	CO	RD	EE	TC
ER	JI	TP	DT	SO	FG	UR
FA	KF	UI	EC	TE	GC	VL
GA	LS	VI	FC	UE	HP	XT

În locurile din dreapta rămase libere se înscriu literele textului dar pe linii de la stânga la dreapta şi de sus în jos. Fie textul clar: METODE CRIPTOGRAFICE CLASICE.

Punctul s-a folosit cu PT. Se scot inelele şi se montează în ordine numerică normală.

1	2	3	4	5	6	7
CO	DM	EE	IE	RD	ST	TC
DT	ER	FG	JI	SO	TP	UR
EC	FA	GC	KF	TE	UI	VL
FC	GA	HP	LS	UE	VI	XT

Se realizează a doua transpoziţie aliniind inelele după alfabetele înscrise la stânga. Criptograma va fi: **OTCCM/ RAAEG/ CPEIF/ SDOEE/ TPIIC/ RLT**.

Dintre criptografele de tip riglă exemplificăm:

Rigla Saint-Cyr. Aceasta realizează o substituţie polialfabetică. Se compune dintr-un suport fix pe care este înscris câte un alfabet într-o anumită ordine şi un cursor.

Cifrarea se execută simplu, plecând de la o cheie literală care realizează corespondenţa dintre litera de pe cursor şi litera respectivă de pe suport.

Litera cifrată se citeşte pe suport în dreptul literei clare de pe cursor. Cheia determină poziţia începutului alfabetului de pe cursor faţă de alfabetul de pe suport.

De exemplu, dacă prima litera a cheii este *D,* atunci cifrarea literei *C* va fi *F*.

Riglele au fost perfecţionate în permanenţă şi se folosesc şi azi. S-au realizat rigle la care alfabetul s-a completat cu cifre şi semne de punctuaţie. Cea mai reprezentativă este *rigla universală*. Aceasta este formată din 20 de riglete care se deplasează orizontal, astfel încât pe coloana orificiilor să apară cheia numeric convenientă. Această riglă oferă foarte multe posibilităţi cum ar fi: cifrarea numerică, duplicarea transpoziţiilor, utilizarea de chei multiple literale sau numerice.

Oricât de complicate ar fi, criptografele nu sunt dispozitive care ajută pe cel ce execută cifrarea să elimine greşeli de transcriere, să scurteze timpul de lucru. Spre deosebire de ele maşinile de cifrat automatizează procesul cifrării eliminând în întregime greşelile de transmisie, rămânând doar greşelile de operare dependente de instruirea şi antrenamentul cifratorului. Transformările realizate cu maşinile de cifrat nu sunt simple transpoziţii de elemente, existând posibilitatea realizării unor funcţii de transformare, bazate pe calcule complexe.

În ceea ce priveşte productivitatea, aceasta este mult superioară procedeelor de cifrare utilizate anterior. Perfecţionarea maşinilor de cifrat este indiscutabil legată de nivelul de dezvoltare al ştiinţei şi tehnicii din epoca respectivă.

Concepute pentru mecanizarea operaţiunilor de cifrare şi descifrare, maşinile clasice, tradiţionale constituie sisteme criptografice complexe, îndeplinind atât cerinţa transparenţei algoritmilor de cifrare-descifrare pentru utilizatori, cât şi mutarea ponderii secretului pe sistemul de chei utilizat. În linii mari o maşină de cifrat, respectiv descifrat, trebuie să cuprindă următoarele elemente:

- mulţimea (algoritmul) transformărilor;

- sistemul de chei;
- dispozitive auxiliare care să permită introducerea datelor şi cheilor, adaptarea acestora la canalul de legătura etc.

Din punct de vedere al metodei utilizate maşinile de cifrat polialfabetice pot fi împărţite în trei categorii:

- maşini polialfabetice simple care realizează substituţii diferite la fiecare literă a textului în clar în funcţie de numărul de alfabete utilizate. Nu pot exista mai mult de 26 de alfabete diferite şi acest impediment duce la apariţia repetărilor, factor deosebit de decisiv în bătălia descifrărilor;
- maşini polialfabetice bazate pe principiul autocifrării. Substituţiile realizate depind atât de numărul alfabetelor cât şi de caracteristicile textului cifrat. Aceasta duce le reducerea numărului de repetări, deci îngreunează lucrul criptanaliştilor dar complică în acelaşi timp descifrarea.
- maşini polialfabetice complexe la care există posibilitatea schimbării aleatoare a rangurilor diferitelor alfabete utilizate.

5.2.1 Maşina "Enigma"

Dintre maşinile din aceasta categorie vom prezenta maşina „Enigma".

Este o maşină polialfabetică complex. A fost fabricată în Germania şi larg utilizată în timpul celui de-al doilea război mondial. De altfel, enigma „Enigmei" încă nu a fost elucidată pe deplin. Specialiştii oscilează de la a-i atribui în întregime eşecul Germaniei, până la a o trata doar ca un factor conjunctural, complementar, în desfăşurarea celei mai mari conflagraţii mondiale din istoria omenirii şi deci fără contribuţie semnificativă la destinul beligeranţilor. Înainte de a încerca să tratăm şi noi acest subiect, să vedem ce a însemnat din punct de vedere criptografic, maşina Enigma.

Algoritmul criptografic al maşinii avea la bază o substituţie polialfabetică complexă, realizată cu ajutorul a trei discuri mobile, prevăzute fiecare cu câte 26 de contacte pe o faţă şi 26 de ace pe cealaltă.

Contactele şi acele sunt legate electric între ele; două câte două, dar nu direct, ci într-o anumită ordine.

Există şi un al patrulea disc, fix, prevăzut cu 26 de contacte, legate două câte două. Discurile sunt coaxiale şi montate în aşa fel încât acele unuia presează contactele celuilalt, realizându-se astfel 13 linii paralele, formate din câte 3 segmente cu variaţie independentă.

Circuitul electric se realizează de la sursă la tastă, care marchează prin apăsare litera clară (la cifrare) sau pe cea cifrată (la descifrare), prin contactul acesteia la o contraplacă cu 20 de contacte fixe, apoi prin discurile I, II şi III, după circuite

realizate din segmentele cu variaţie independentă, la discul al patrulea, fix, şi prin una din cele 13 perechi de contacte ale acestuia se înapoiază prin alt circuit, format din segmente de către cele trei discuri mobile, la un bec ce iluminează litera cifrată (la cifrare) sau pe cea clară (la descifrare).

Complexitatea maşinii constă în numărul de segmente, care realizează variaţia circuitelor, datorită rotirii celor trei discuri funcţie de tastele apăsate şi de caracteristicile componentei textului clar.

La fiecare apăsare de tastă, funcţie de cheia adoptată, cele trei discuri se deplasează unul faţă de celălalt, cu un anumit număr de poziţii, astfel încât se realizează de fiecare dată, un anumit circuit electric şi deci o anumită corespondenţă între litera clară apăsată şi litera cifrată iluminată.

Cele trei discuri, cu diferite scheme de legaturi între contactele şi acele acestora, se pot schimba între ele, realizând practic în acest fel o altă variantă a maşinii. De altfel, în afara celor trei discuri montate, fiecare maşină dispunea şi de un alt set de discuri cu legături complet diferite.

Dacă se consideră că fiecare dintre cele cinci discuri poate fi realizat într-un alt mod, se obţine o succesiune de patru alfabete diferite, din care trei fiind realizate de discurile mobile, sunt permutabile. Dar datorită legăturilor întreţesute se poate considera că la fiecare transformare de literă clară în litera cifrată contribuie şapte alfabete (sase realizate de traseele - dus şi întors - din discurile mobile şi unul în cel de-al patrulea disc).

Fără îndoială, aceste posibilităţi de schimbare a alfabetelor determină un număr foarte mare de transformări, greu de descifrat, chiar şi în cazul în care se posedă o variantă a maşinii de cifrat.

Un alt element caracteristic al maşinii îl constituie cheia; aceasta defineşte ordinea discurilor, poziţia iniţială şi deplasările relative ale discurilor în timpul cifrării unei anumite litere.

Deşi robustă şi comodă, datorită faptului că nu avea posibilitatea de a imprima textul cifrat sau clar (acestea se culegeau şi se notau sau se scriau la altă maşină) greşelile de transcriere erau destul de frecvente. La acestea se mai adăugau şi câteva necazuri tehnice produse de imperfecţiunea unor contacte, arderea becurilor şi scăderea capacităţii sursei de alimentare.

Pentru a vedea cum se lucra cu această maşină (din care armata noastră a achiziţionat şi exploatat circa 100 de bucăţi), vom reproduce câteva pasaje (cu eliminarea unor elemente caracteristice de limbaj) din „Instrucţiunile de întrebuinţare

a maşinii de cifrat Enigma", editate la 12 ianuarie 1937 de către Marele Stat Major, reactualizate cu unele precizări organizatorice la 13 ianuarie 1940.

„Extinderea întrebuinţării maşinii de cifrat Enigma la toate Secţiile Armatei se ordonă de către Comandamentul Suprem al Armatei".

Lungimea minimă a unei radiograme cifrate cu maşina de cifrat Enigma nu este limitată, dar cea maximă nu va depăşi 250 de litere.

Cheile se vor schimba zilnic la orele 00.00. Tabelele cu cheile zilnice se distribuie lunar.

Cheile zilnice pentru cifrare impun schimbarea:

- poziţiei discurilor (în cifre romane);

- poziţiei inelelor (în cifre arabe sau litere);

- legaturilor fişelor (în litere).

Corespondenţa dintre literele şi numerele cheii este normală: $A = 01$, $B = 02,...,Y = 25$ şi $Z = 26$.

Poziţia discurilor (II, I, III etc.) ne arată ordinea în care acestea sunt introduse în maşina de cifrat, de la stânga la dreapta.

Poziţia inelelor (13 08 11) ne arată aranjarea lor iniţială. Corespondenţa dintre poziţia discurilor şi ordinea numerelor ce reprezintă poziţia inelelor se realizează pentru fiecare caz în parte. Pentru exemplu dat: II--13, I--08 şi III--11.

Legaturile fişelor se realizează conform succesiunii indicate de perechile de litere din cheie.

Dacă cheia legaturilor fişelor este:

AO BI DV EH GZ KW LX MU RY QT.

atunci se realizează următoarele legaturi:

A cu O, B cu I, etc.

Cheia unei radiograme se va determina printr-o grupă cu cinci litere, din care primele două sunt fără semnificaţie (se vor schimba la fiecare radiogramă), iar următoarele trei indică cheia din „Tabelul cu chei".

La telegramele compuse din mai multe fragmente, fiecare fragment va avea litere diferite ca grupă de recunoaştere şi alte două litere fără semnificaţie. Grupa de recunoaştere se transmite prima şi nu se cifrează.

Cele trei litere care formează poziţia inelelor vor fi alese din combinaţii de la AAA la ZZZ, excluzându-se cele cu semnificaţie, prescurtări -RGT-, indicative de apel, comenzi de trafic, litere în ordine alfabetică normală sau inversă -ABC şi CBA etc.

Cifratorul aşează tamburii, literele în ferestruici şi cablurile de legătură conform cheii zilnice.

LA A DOUA RADIOGRAMA POZITIA INELELOR SE MODIFICĂ CU CHEIA DE RADIOGRAMĂ.

După aşezarea cheii de radiogramă, se bate textul în clar, iar literele luminate - cifrate- se culeg pe formularul de radiogramă, în grupe de câte cinci.

La descifrare, cifratorul va aranja maşina conform cheii zilnice. Va extrage apoi cheia de radiogramă (cele trei litere cifrate din preambulul radiogramei adăugate după cele trei litere de bază şi o va descifra, apăsând tastele corespunzătoare.

După descifrarea cheii de radiogramă se vor stabili poziţiile celor trei inele, în concordanţă cu poziţiile discurilor indicate în cheia zilnică.

În tabel, o cheie zilnică se prezintă astfel: 4;I II III;16 11 13; BN KE VZ CO DI FR HU JW LS TX; adq nuz opw vxz, unde:

-4 - era data (ziua);

-I - poziţia discurilor;

-16 - poziţia inelelor;

-BN - legăturile fişelor;

-nuz - grupele indicative.

Pentru exemplificare, o radiogramă va arăta astfel:

1755-139-wep hfi -

ulznu sgexu nfopr salmc

ydr jd.....

unde:

- 1755 - ora expedierii;

- 139 - numărul literelor;

- wep - poziţia de bază;

- hfi - cheia cifrată a telegramei sau radiogramei;

- ul - litere fără semnificaţie;

- znu - grupa de recunoaştere;

- sgexu - textul cifrat.

Pentru exerciţiu se vor folosi un anumit număr de combinaţii de chei, care însă nu se vor întrebuinţa şi pentru transmiterea radiogramelor.

După analiza acestor instrucţiuni, ne putem da seama că pentru a evita greşelile şi mai ales încurcăturile în fixarea cheilor, operatorii trebuiau pregătiţi şi antrenaţi.

De asemenea, o mare parte din posibilităţile maşinii erau puse în valoare sau estompate de sisteme de realizare a cheilor şi de frecvenţa lor de schimbare.

Greşelile de operare, de schimbare a cheilor, repetarea radiogramelor care aveau elemente iniţiale greşite şi lungimea excesivă a textelor erau elemente ce puteau compromite maşina, lucru care de altfel s-a şi întâmplat.

5.2.2 Maşini cu cifrare poligramică

Dacă maşinile polialfabetice realizează o cifrare caracter cu caracter, există şi tipuri de maşini care fac o cifrare a *n*-gramelor, prin *n*-grame înţelegând grupuri de *n* caractere.

Principiul cifrării poligramice a fost imaginat în deceniul trei al acestui secol de Lester S. Hill şi publicat în „The American Mathematical Moutley” în 1929 şi se bazează pe calcule matematice.

Se alocă literelor alfabetului valori numerice de la 0 la 25 în mod aleator. Se pot imagina diverse combinaţii între numerele asociate literelor, operaţii care se pot materializa practic cu ajutorul unei maşini de cifrat. Toate operaţiile se efectuează modulo 26. De exemplu, fie alfabetul:

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,5,10,21,6,25,14,24,18,0,11,17,
1,22,7,15,2,19,12,8,23,3,16,9,20,Y,Z,13,4.

Dorim să cifrăm trigrame ale textului clar:

C R I P T O G R A F I E

Aceste trigrame vor fi: CRI-PTO-GRA-FIE.

Alegând cheia de cifrare: **ENIGMATIC**, forma numerică fiind 25 7 0/ 24 22 5/ 23 0 21, pentru prima criptogramă se pot scrie ecuaţiile:

$$25\times C+7\times R+0\times I=25\times 21\times 7\times 12+0\times 0=11\rightarrow J$$
$$24\times C+22\times R+5\times I=24\times 21+22\times 12+5\times 0=14\rightarrow F$$
$$23\times C+0\times R+21\times I=23\times 21+0\times 12+21\times 0=15\rightarrow O$$

Deci trigramei *CRI* îi corespunde trigrama *JFO*.

Descifrarea se face în mod asemănător. Se pleacă de la trigrama cifrată şi folosind alţi coeficienţi se determină mesajul clar.

Practic se rezolvă sistemul:

$$\begin{cases} K_1x+K_2y+K_3z=C_1 \\ K_4x+K_5y+K_6z=C_2 \\ K_7x+K_8y+K_9z=C_1 \end{cases}$$

unde: *x*, *y*, *z* sunt valorile numerice ale literelor textului clar;

$K_1,K_2,...,K_9$ sunt valorile numerice ale literelor care constituie cheia;

C_1,C_2,C_3 sunt valorile numerice ale literelor criptogramei.

În cazul nostru vom avea:

$$\begin{cases} 25x+7y+0z=11 \\ 24x+22y+5z=14 \\ 23x+0y+21z=15 \end{cases} \pmod{26}$$

Soluţia sistemului este $x = 21$, $y = 12$, $z = 0$.

Această soluţie se obţine în mod practic din următoarele calcule:

$$\begin{cases} x=K_1'C_1+K_2'C_2+K_3'C_3 \\ y=K_4'C_1+K_5'C_2+K_6'C_3 \\ z=K_7'C_1+K_8'C_2+K_9'C_3 \end{cases}$$

În lucrarea „**SOME CRYPTOGRAPHIC APPLICATIONS OF PERMUTATION POLYNOMIALS**” Jack Levine şi J.V. Brawley prezintă o altă metodă de cifrare poligrafică prin utilizarea mai multor câmpuri şi permutarea polinomială.

Fie un alfabet *P* cu *L* litere. Cu cele *L* litere pot fi construite $N = L^n$ poligrame (fiecare având *n* litere) descrise de succesiunile:

$$\lambda = L_1 L_2 ... L_2;\ L_i \in P;\ i = \overline{1,n}$$

Considerăm acum cele *N* poligrame ca fiind litere ale unui alfabet *A*. La cifrare textul clar compus din literele alfabetului *P* se divide în grupuri de *n* litere obţinând astfel un număr de poligrame. Numărul *N* se poate descompune sub forma: $N = \sum_i N_i$, unde : $N_i = P_i^{m_i}$ sau $N_i = P_i^{m_i} + 1$, iar $P_i =$ numere prime.

Se poate realiza astfel o partiţionare a alfabetului *A* în *K* subalfabete, astfel că;

$$A = \bigcup_{i=1}^{k} A_i$$

Subalfabetului A_i, *i* se repartizează N_1 poligrame, un câmp Galois extins $GF\left(P_i^{m_i}\right)$ şi o permutare polinomială $P_i(x_i)$ dacă $N_i = P_i^{m_i}$ sau o funcţie raţională $R_i(x_i)$, dacă $N_i = P_i^{m_i} + 1$.

Modul în care se face împărţirea poligramelor este următoarea:

1. Se asociază fiecărei litere ale alfabetului *P* o valoare numerică din secvenţa 0,1,2,...,(*L* – 1).

2. Schimbăm orice poligramă $\lambda = L_{i_1} L_{i_2} \ldots L_{i_n}$ cu o formă numerică $\beta = V_{i_1} V_{i_2} \ldots V_{i_n}$, unde V_i este valoarea numerică a literei L_i.

3. Fie $\sigma = V_{i_1} L^{n-1} + V_{i_2} L^{n-2} + \ldots + V_{i_{n-1}} L + V_{i_n}$ pe care o definim ca fiind valoarea numerică repartizată poligramei $\lambda\left(0 \le \sigma \le L^n - 1\right)$.

4. În subalfabetul A_1 vor fi repartizate poligramele pentru care $0 \le \sigma \le N_1 - 1$, în subalfabetul A_2 acele poligrame pentru care $N_1 \le \sigma \le N_2 - 1$ ş.a.m.d.

5. Definim un număr ρ astfel:

$\rho = \sigma - \left(N_1 + N_2 + ... + N_{i-1}\right)$ asociat poligramelor subalfabetului A_i. Se observă că $0 \le \rho \le N_{i-1}$.

6. Se exprimă numărul ρ în baza P_i, astfel:

$$\rho = C_0 P_i^{m_i} + C_1 P_i^{m_i - 2} + \ldots + C_{m_{i-1}} = \left(C_0, C_1, ..., C_{m_{i-1}}\right),$$

unde $C_0, C_1, ..., C_{m_{i-1}}$ sunt elementele câmpului $GF\left(P_i^{m_i}\right)$.

7. Printr-o procedură oarecare stabilim o corespondenţă între poligramele fiecărui subalfabet A_i şi elementele câmpului $GF\left(P_i^{m_i}\right)$ care sunt cifrate cu folosirea unui polinom de permutare sau a unei funcţii de permutare, după cum $N_i = P_i^{m_i}$ sau $N_i = P_i^{m_i} + 1$.Vom ilustra diferitele etape ale procedurii printr-un exemplu. Se consideră cazul cifrării diagramelor, deci $n = 2$, $L = 26$ având repartizate literelor următoarele valori numerice:

$$A = 0, b = 1, C = 2, ..., Z = 25.$$

Vor fi deci $N = 26 \times 26 = 676$ pe care îl partiţionăm astfel:

$$676 = 7^3 + 13^2 + \left(3^4 + 1\right) + \left(3^4 + 1\right)$$

Deci $N_1 = 7^3 = 343$ cu câmpul $GF\left(7^3\right)$;

$N_2 = 13^2 = 169$ cu câmpul $GF\left(13^2\right)$;

$N_3 = N_4 = 3^4 + 1 = 82$ cu câmpul $GF\left(3^4\right)$.

Vom împărţi cele 676 diagrame în patru subalfabete, astfel:

subalfabetul A_1, dacă $0 \le \sigma \le 342$;

subalfabetul A_2, dacă $343 \le \sigma \le 511$;

subalfabetul A_3, dacă $512 \le \sigma \le 592$;

subalfabetul A_4, dacă $593 \le \sigma \le 675$.

Având de exemplu diagrama $RM = (17,12)$ calculam:

$$\sigma = 17 \times 26 + 12 = 454.$$

Deci diagrama *RM* face parte din subalfabetul A_2.

Calculăm $\rho = \sigma - N_1 = 454 - 343 = 111$, valoare pe care o exprimăm în baza 13.

$111=(8,7)$ (baza 13), adică $111=8\times13+7$.

Rezultă că *RM* = (8,7).

Dacă folosim polinomul de permutare de forma $y_i = x_i^5 + b_i$ având $b_1=(12,8)$, se obţine $y_i = 88$ deoarece $x_1 = 111$, iar $b_i = 164$. Toate aceste operaţii se fac modulo 169, câmpul utilizat fiind $GF\left(13^2\right)$.

$$\sigma = 88 + 343 = 431 = (16,15) = QP \text{ (adică } 16\times26+15).$$

Deci diagramei *RM* îi corespunde criptograma *QP*.

Această metodă necesitând o serie de calcule complicate impune utilizarea calculatorului pentru executarea operaţiei de cifrare şi descifrare.

5.2.3 Maşini tomogramice

O altă categorie de maşini de cifrat o constituie *maşinile de cifrat tomogramice.*

Fracţionarea literelor (tomogramia) în scopul cifrării, a fost şi continuă să fie, chiar şi cu mijloace moderne, o temă de studiu deosebit de atractivă. De la substituţii cu numere fracţionare, până la accesul pe bit al criptologiei moderne, iată gama procedeelor tomogramice utilizate în acest tip de cifrare.

Alfabetul lui Collon, trifidul lui Delastelle, alfabetul Morse şi codul Baudot iată numai câteva din cifrurile tomogramice.

Collon realiza un alfabet de substituţie cu grupe formate din două simboluri, bazate pe utilizarea combinatorie a 5 sau 6 simboluri diferite.

Trifidul lui Delastelle realiza substituţia prin grupe de câte trei cifre, rezultate din combinarea cifrelor1, 2 şi 3.

Alfabetul Morse, bazat pe combinaţii de puncte, linii şi spaţii şi mai ales codul Baudot sunt larg cunoscute şi utilizate şi în telegrafia modernă. Ele nu mai reprezintă de mult un secret pentru un mare număr de oameni şi, deci, din punct de vedere criptografic nu mai au demult nici o valoare.

Sunt considerate, acum, la fel ca şi alfabetele limbilor de circulaţie mondială şi utilizate ca elemente de reprezentare a textelor clare şi ca atare supuse unor transformări comparatorii, bazate pe diferite combinaţii.

Sistemul Vernam a fost inventat în decembrie 1917 şi apoi larg răspândit până la sfârşitul primului război mondial. El are marele merit că a realizat un dispozitiv

practic de mixare a textelor clare transpuse pe banda perforată (în cod Baudot) cu o succesiune de grupe, în acelaşi cod, transpuse pe altă bandă.

Din punct de vedere criptografic, procedeul de mixare reprezintă algoritmul sistemului, iar cheile - benzile suport ale grupelor utilizate pentru mixaj.

Se cunoaşte că în cod Baudot literele sunt reprezentate prin grupe de câte cinci simboluri (1 şi 0, perforaţii pentru 1 sau pentru 0 pe banda de hârtie, impulsuri electrice pozitive si respectiv negative, etc.).

De exemplu: $A = 11000$ şi $I = 01100$ etc.

Metoda Vernam se bazează pe o relaţie de tip modulo 2 între aceste simboluri, conform următoarei convenţii:

$$1 + 1 = 0$$

$$1 + 0 = 1$$

$$0 + 1 = 1$$

$$0 + 0 = 0$$

Dacă se consideră că textul clar este A şi cheia B se obţine ca text cifrat G, astfel:

$$A = 11000$$
$$B = 10011$$
$$G = 01011$$

Ca să mixeze impulsurile electrice, Vernam a inventat un dispozitiv realizat cu magneţi, relee şi bobine. Cum cifrarea şi descifrarea erau reciproce, acelaşi aparat era folosit la ambele operaţii.

Dacă cineva interceptează aceste mesaje, obţine doar secvenţe de 0 şi 1 sau şiruri de caractere fără semnificaţie. De asemenea, a crescut mult operativitatea cifrării şi s-au eliminat, aproape în întregime erorile de transcriere.

Prin sistemul Vernam s-a impus în lumea criptografiei schimbarea accentului preocupării principale de la algoritmul de criptare la strategia cheilor.

În primele zile, cheile pentru acest aparat se obţineau prin perforarea benzilor cu simboluri aleatoare.

Apoi s-au căutat şi s-au găsit şi alte metode de elaborare a cheilor.

Cu ajutorul simbolurilor utilizate în codul Baudot se poate realiza un tabel de 32 x 32, în care pe rând coloanele puteau să fie socotite chei.

Întrucât secretul sistemului Vernam constă tocmai în cheia folosită, s-a renunţat la utilizarea unor chei repetate în favoarea unora foarte lungi. Aceste chei foarte lungi prezentau însă inconveniente de manipulare. Dificultatea a fost rapid escaladată, datorită ideii inginerului Morshouse de a combina două chei scurte, ca şi cum una ar fi servit la cifrarea celeilalte.

Rezultatul obţinut era o cheie foarte lungă, denumită apoi cheie secundară. Cheia secundară se obţine din două chei primare scurte datorită diferenţei numerelor de perforaţii (semne) în cele două cazuri. Astfel dacă prima cheie primară avea 1000 de semne, iar a doua 999, datorită diferenţei de un semn, prin combinaţii succesive, se putea obţine o cheie cu 999000 de semne. Acest lucru însemnă că în loc de a utiliza o cheie de 4000 de m lungime se utilizau două chei de aproximativ 4 m.

Acest procedeu însă nu exclude complet repetiţiile, mai ales pe anumite sectoare de bandă şi impune cert marcarea sfârşitului cheii, neexcluzând în totalitate nici sensul, inconveniente prompt „speculate” de criptanalişti, ceea ce făcea ca rezistenţa sistemului Vernam să scadă pe măsura ce el era mai cunoscut şi mai des utilizat.

Saltul următor l-a făcut Mauborgne prin introducerea „cheii cu o singură utilizare”, rezultată din aplicarea pentru realizarea unei chei atât a principiului aleatorismului, cât şi al non-repetiţiei.

Sistemul Jammet realizat de Jammet Onde Electrique în august 1926, este analog sistemului Vernam bazându-se însă pe alfabetul Morse.

Problema cea mai grea pentru acest sistem a fost sincronizarea manipulatoarelor şi, bineînţeles, extragerea semnalului clar la recepţie din cel secretizat. Sistemul a reprezentat un interesant experiment tehnic, dar s-a dovedit fără valoare criptografică.

Sistemul Belin este un prim sistem de secretizare a transmiterilor imaginilor. Principiul de funcţionare se baza pe exploatarea optică a imaginii şi transformarea impulsurilor luminoase în impulsuri electrice. Succesul acestui tip de transmisii era asigurat de sincronizarea perfectă a discurilor transmiţătoare şi receptoare.

Pentru a evita recepţia unor astfel de imagini Belin a realizat un sistem de variaţie controlată a vitezelor celor două discuri, astfel ca orice interceptare şi încercare de reproducere produceau o imagine deformată, neinteligibilă.

Test de autoevaluare

3. Cum au evoluat maşinile criptografice şi care sunt principalii beneficiari;
4. Care sunt principalele clase de maşini de criptare clasificate după modul de funcţionare.

Tema de autoinstruire

2. Realizaţi schema logică a metodei Vernam, bazată pe exemplul din curs.

Lucrare de verificare

1. Descrieţi funcţionarea maşinii de cifrat ENIGMA.
2. Să se descrie funcţionarea următoarelor sisteme criptografice:
   a. Sistemul Vernam.

   b. Criptograful Bazaries.

3. Considerăm un cifru de permutare la care se fixează numerele naturale p, q. Textul clar se împarte în blocuri de câte p·q caractere. Fiecare astfel de bloc se scrie pe liniile unei matrici de p linii şi q coloane. Criptarea blocului se realizează scriind aceste matrici pe coloane. De exemplu, pentru p = 3, q = 4, textul clar MAINI CURATE se scrie:

M A I N

I C U R

A T E X

(textul s-a completat cu litera X). Textul criptat va fi MIAACTIUENRX.

Decriptaţi următorul text DJNOUDNAINPAPANONZ criptat într-un mod similar.

4. Folosind atacul prin forţă brută, decriptaţi mesajul WYPTBSJBYZ criptat cu un sistem Cezar.

5. Presupunem că Alice doreşte să transmită mesajul "TRIMIT AJUTOR" folosind un canal de comunicaţie nesigur. Atât ea cât şi Bob se presupune că au obţinut printr-o metodă necunoscută un set de chei generate absolut aleator şi că una dintre aceste chei este "AFVKGHOPERTS". Folosind sistemul de criptare Vernam cifraţi textul clar enunţat.

6. În condiţiile enunţate de exerciţiul de la punctul 5 descifraţi următorul text cifrat: T W D W O A O Y O K H J

7. Să se cripteze textul clar INAINTE SI LA DREAPTA folosind sistemul de criptare Hill cu matricea:

17 17 5 21 18 21 22 19 1

CAPITOLUL 6

ELEMENTE DE CRIPTANALIZĂ

Cuvinte cheie: limbaje naturale, decriptarea, spargerea sistemelor cripto.

OBIECTIVE

Înţelegerea şi cunoaşterea următoarelor noţiuni:
- limbaje naturale şi caracteristici;
- rolul *sistemului* şi al *cheii de criptare* în metodele de decriptare;
- elementele ce stau la baza spargerii sistemelor criptografice şi poligrafice.

6.1 Caracteristicile statistice ale limbajelor naturale

Din punct de vedere al teoriei informaţiei limba poate fi socotită ca o sursă de comunicări, care creează o oarecare cantitate de informaţie, măsurată în biţi. În orice limbă s-ar face exprimarea, comunicările se compun din combinaţii şi şiruri de litere şi cuvinte care nu sunt cu totul întâmplătoare cum ar părea la prima vedere. Şirurile de litere formează cuvinte şi fraze cu o structura specifică unei anumite limbi. Studierea unei astfel de structuri specifice fiecărei limbi permite să se obţină o economie de timp în transmiterea textelor cu ajutorul liniilor de telecomunicaţii, prin codificarea optimă a comunicărilor şi literelor.

În scopuri ilustrative, în studiul limbilor se folosesc uneori limbaje abstracte sau artificiale. În acest caz se consideră fie că toate literele au aceeaşi probabilitate şi se succed independent una de alta (*aproximaţie de ordin 0*), fie ca literele se succed independent dar cu probabilităţile limbilor reale (*aproximaţie de ordin 1*). Se poate impune de asemenea să se ţină seama de structura diagramelor, adică după fixarea unei anumite litere, litera următoare se alege în funcţie de frecventele cu care diferitele litere se succed după litera fixată (*aproximaţie de ordin 2*). Se poate tine seama de asemenea de structura trigramelor, adică literele se aleg în funcţie de probabilităţile de apariţie care depind de primele două litere ale trigramei (*aproximaţie de ordin 3*).

Aşadar teoretic se poate calcula entropia medie, H_0, care revine unei litere de text (în care literele au aceeaşi frecvenţă) dacă se cunoaşte numărul *n* de litere al alfabetului. De exemplu dacă alfabetul are 8 litere: (a, b, c, d, e, f, g, h) iar

probabilităţile lor de apariţie sunt $p = 0{,}125$ pentru fiecare literă, atunci entropia medie pe literă este dată de mărimea:

$$H_0 = \log_2 8 = 3 \text{ biţi/literă}$$
$$H_0 = \log_2 26 = 4{,}7 \text{ biţi/literă.}$$

În cazul alfabetului latin care conţine 26 de litere, entropia medie pe literă H_0 dă o indicaţie destul de vagă asupra entropiei unei limbi (de fapt H_0 reprezintă entropia alfabetului).

În realitate frecvenţele de apariţie a literelor nu sunt aceleaşi şi ţinând seama de probabilităţile de apariţie a diferitelor litere se obţine *entropia de ordin 1* cu relaţia:

$$H_1 = -\sum_{i=1}^{n} p_i \log_2 p_i < H_0$$

În tabelul de mai jos sunt prezentate entropiile de ordin 0 şi 1 ale câtorva limbi de circulaţie mondială, precum şi ale limbii romane:

Limba	H_0	H_1
Engleză	4,7	4,03
Rusă	5	4,34
Germană	4,75	4,08
Franceză	4,7	3,94
Română	4,7	4,47

Dar nici această mărime nu este precisă deoarece în calculul entropiei H_1 s-au neglijat dependenţele ce există între diferitele litere.

La limbile reale se observă că unele litere apar de cele mai multe ori împreuna. Pentru a ţine seama de aceasta se va lua în consideraţie şi probabilitatea de tranziţie $P(j/i)$: probabilitatea ca după litera α_i să urmeze litera α_j. O altă metodă echivalentă pentru redarea acestei structuri constă în stabilirea probabilităţilor diagramelor, $P(i, j)$:

$$P(i, j) = P(i) \cdot P(j / i)$$

Calculul entropiei medii pe litera H_2 va fi dată de relaţia:

$$H_2 = -\sum_{i,j} P(i, j) \cdot \log_2 P(j / i) \text{ biţi/literă}$$

Ilustrarea modului de calcul a probabilităţilor diagramelor poate fi făcută cu exemplu de mai jos:

Presupunem că alfabetul conţine trei litere A, B şi C cu următoarele probabilităţi:

$\sum P(i) = 1$ $\qquad\qquad$ $P(j/i)$

I	P (I)
A	0,33
B	0,60
C	0,07

	J		
I	A	B	C
A	0	0,8	0,2
B	0,5	0,5	0
C	0,5	0,4	0,1

$$P(A,A) = P(A) \cdot P(A/A) = 0,33 \cdot 0 = 0$$
$$P(A,B) = P(A) \cdot P(B/A) = 0,33 \cdot 0,8 = 0,264$$
$$P(A,C) = P(A) \cdot P(C/A) = 0,33 \cdot 0,2 = 0,066$$
$$P(B,A) = P(B) \cdot P(A/B) = 0,6 \cdot 0,5 = 0,3$$
$$P(B,B) = P(B) \cdot P(B/B) = 0,6 \cdot 0,5 = 0,3$$
$$P(B,C) = P(B) \cdot P(C/B) = 0,6 \cdot 0 = 0$$
$$P(C,A) = P(C) \cdot P(A/C) = 0,07 \cdot 0,5 = 0,035$$
$$P(C,B) = P(C) \cdot P(B/C) = 0,07 \cdot 0,4 = 0,028$$
$$P(C,C) = P(C) \cdot P(C/C) = 0,07 \cdot 0,1 = 0,007$$

$$H_2 = -0,264\log_2 0,8 - 0,066\log_2 0,2 - 2 \cdot 0,3\log_2 0,5 -$$
$$-0,035\log_2 0,5 - 0,021\log_2 0,4 - 0,007\log_2 0,1 = 3,3 \text{ biţi}$$

A fost calculată H_2 pentru diferite limbi, astfel pentru limba engleză s-a găsit $H_2 = 3,3$ biţi/literă iar pentru limba rusă $H_2 = 3,51$ biţi/literă.

Calculul frecvenţelor de apariţie a combinaţiilor de trei litere şi mai multe este necesară pentru descrierea statistică a unei limbi. Deşi foarte complexe, calculele referitoare la stabilirea frecvenţei trigramelor au fost efectuate pentru mai multe limbi europene, calculându-se entropia de ordinul 3, H_3. Aceasta este 2,98 biţi/literă pentru limba engleza, 3 biţi/literă pentru limba rusă.

Calculul frecvenţelor de apariţie a combinaţiilor de 4,5 sau 6 litere este foarte complicat şi până acum nu s-a efectuat analitic pentru nici o limbă. S-au făcut însă unele determinări experimentale care arată că pentru texte mai lungi care cuprind 100 de litere şi mai mult entropia scade în jurul valorii de
1 bit/liter.

Studierea limbii ca mesaj conduce la necesitatea calculării cantităţii de informaţie pe care el o conţine, iar măsura cantităţii de informaţie o constituie entropia limbii. O valoare mare a acesteia denotă o anumită claritate a limbii. Din considerentele de mai sus se vede că eficacitatea cu care limba (ca sursă) produce informaţie, depinde de distribuţia probabilităţilor de apariţie a literelor care compun comunicarea; cu cât distribuţia este mai uniformă cu atât cantitatea de informaţie adusă de literă este mai mare. Raportul dintre entropia medie pe litera şi entropia maximă, $H_{\max} = \log_2 n$ (n fiind numărul de litere utilizat), se numeşte entropie relativă şi exprimă gradul de compresie al comunicării în funcţie de structura ei statistică.

Cantitatea $R = 1 - \frac{H}{H_{\max}}$ se numeşte *redundanţă* şi reprezintă un indicator util în cazul masajelor lungi.

Până nu demult frecvenţa de apariţie a literelor se determină manual solicitând o muncă laborioasă în special pentru texte lungi.

Dacă toate combinaţiile de litere ar forma cuvinte atunci într-un alfabet de 30 de litere ar exista:

30 de cuvinte de o literă;

30^2 cuvinte de două litere;

.....

30^n cuvinte formate din n litere.

În realitate în fiecare limbă vorbită numărul de cuvinte este mai mic de 50000, deci numai o parte foarte mică dintre combinaţii formează cuvinte, restul fiind combinaţii fără sens.

Fie un mesaj M de lungime N şi să presupunem că numărul literelor distincte din M este egal cu $n\left(n \le N\right)$. Dacă L este numărul tuturor literelor din alfabetul limbii în care este redactat textul. Dacă la mesajul M ataşăm pe rând câte o literă din alfabet obţinem un număr de L texte de lungime (N + 1) din care: (L – n) vor avea (n + 1) litere distincte restul de n texte vor rămâne cu n litere distincte.

Deci, în medie, numărul literelor distincte pentru fiecare din cele L texte este:

$$\frac{(L-n)(n+1)+n^2}{L}=\frac{n(L-1)+L}{L}$$

Dacă notăm cu Δ_n variaţia numărului de litere distincte corespunzătoare unei variaţii ΔN a lungimii textului avem:

$$\Delta_n=\left[\frac{n(L-1)+L}{L}-n\right]\cdot\Delta N=\frac{L-n}{L}\cdot\Delta n$$

Această relaţie poate fi asimilată cu o ecuaţie diferenţială cu variabile separabile, adică:

$$\mathrm{d}n=\frac{L-n}{L}\mathrm{d}N$$

Prin integrare se obţine:

Constanta k se determină din condiţiile la limită: $n=0$ pentru $N=0$ rezultă $n=0$.

$$0=L-e^{-k};\quad k=-\ln L;$$

deci

$$n=L\left(1-e^{-\frac{N}{L}}\right)$$

Corelând această relaţie cu rezultatele obţinute printr-o cercetare statistică rezultă că relaţia finală este de forma:

$$n=L_0\left(1-e^{-\frac{N}{L}}\right)$$

În această relaţie L_0 reprezintă numărul literelor din alfabet cu o frecvenţă de apariţie mai mare decât 1%. În tabelul următor se dau valorile indicatorului „numărul literelor distincte” (n) pentru texte de diferite lungimi calculate cu relaţia de mai sus pentru diferite limbi:

Limba	L_0	$N=30$	$N=40$	$N=50$	$N=100$
Italiană	18	12,69	13,98	15,57	17,96
Română	18	12,69	13,98	15,57	17,96
Germană	21	14,8	16,31	18,16	20,95
Engleză	21	14,8	16,31	18,16	20,95
Franceză	19	13,39	14,76	16,43	18,98

Din analiza acestor date rezultă că pentru texte având mai mult de 100 de litere intervine o anumită stabilitate, în sensul că în astfel de texte vor apărea în medie, toate cele L_0 litere din alfabet şi deci nu mai are rost calculul numărului de litere distincte cu aceasta relaţie.

În tabelul de mai jos se prezintă frecvenţele relative de apariţie ale literelor în câteva limbi europene determinate la un text de 1000 de litere:

Engleză		Rusă		Română		
litera	frecvenţa	litera	frecvenţa	litera	text ziar	text literă
e	0,131		0,090	a	0,105	0,093
t	0,105		0,072	b	0,0092	0,0092
a	0,086		0,062	c	0,053	0,054
o	0,08		0,053	d	0,04	0,039
n	0,071		0,045	e	0,12	0,1
r	0,068		0,040	f	0,008	0,009
i	0,063		0,038	g	0,006	0,008
s	0,061		0,035	h	0,007	0,005

Engleză		Rusă		Română		
litera	frecvenţa	litera	frecvenţa	litera	text ziar	text literă
h	0,053		0,028	i	0,13	0,11
d	0,038		0,024	j	0,0014	0,002
1	0,034		0,025	k	–	–
f	0,029		0,023	l	0,046	0,046
c	0,028		0,021	m	0,039	0,04
m, u	0,025		0,018	n	0,056	0,064
p,y, q	0,020		0,016	o	0,025	0,04
w	0,015		0,014	p	0,026	0,031
b	0,014		0,013	q	–	–
v	0,009		0,012	r	0,062	0,07
k	0,004		0,010	s	0,024	0,046
x	0,002		0,009	t	0,065	0,053
J,g, z	0,001		0,007	u	0,074	0,06
			0,006	v	0,012	0,013
			0,006	w	–	–
			0,003	x	–	–
			0,0002	y	–	–
				z	0,008	0,008
				ţ	0,006	0,001
				ş	0,17	0,02

Un alt indicator util în criptanaliză este raportul dintre numărul vocalelor şi cel al consoanelor. S-au făcut cercetări statistice pe această temă şi în tabelul următor sunt sintetizate unele rezultate:

Limba %	Română	Fran-ceză	Itali-ană	Germa-nă	Engleză
vocale	49,4	43,36	47,74	38,86	37,4
consoane	50,6	56,64	52,26	61,14	62,6

Rangul pe care îl ocupă literele în cadrul cuvântului constituie un indicator preţios în criptanaliză. Cercetarea literelor iniţiale şi finale ale cuvintelor oferă o serie de rezultate specifice dat fiind că parte iniţială din cuvânt deţine un loc principal în privinţa cantităţii de informaţie conţinută în cuvântul respectiv. În tabelul de mai jos sunt date în ordine descrescătoare a frecvenţelor de apariţie cele mai des întâlnite litere iniţiale şi finale:

Limba	Rangul literei	Litere
română	iniţială	SCPAIDMNTUVL…
	finală	EAITRUMD…
germană	iniţială	DASIEWV…
	finală	NERSTHDI…
engleză	iniţială	TASIOWCFRDB…
	finală	ESTNDYOR
italiană	iniţială	ADOVRCUSP…
	finală	EAOILN
franceză	iniţială	LEDASCPV
	finală	?

Cuvintele scurte cele mai frecvente constituie de asemenea un indicator preţios pentru criptanalişti. În tabelul de mai jos sunt prezentate cele mai frecvente cuvinte scurte din câteva limbi europene:

Limba	Cuvinte scurte
Franceză	A, DE, DU, EST, ET, LA, LE, LES, PAR, QUE…
Germană	AN, DASS, DEN, DER, DIE, ES, IN, IST, UND…
Engleză	A, AND, AT, FOR, HAS, IN, IT, IS, OF, THE…
Italiană	A, CHE, CON, DI, E, HA, HO, IN, PER

Cercetările statistice efectuate asupra cuvintelor au condus la rezultate deosebit de interesante atunci când s-a adoptat drept criteriu numărul silabelor componente. S-a stabilit că indiferent de limba folosită, distribuţia de probabilitate a cuvintelor după numărul silabelor componente este o distribuţie de tip Poisson.

Astfel dacă $P(n)$ este probabilitatea de apariţie a unui cuvânt format din n silabe şi cu I lungimea medie a cuvintelor, în silabe, într-o anumită limbă avem:

$$I = 1 \cdot P(1) + 2 \cdot P + ... + n \cdot P(n)$$

$$P(n) = \frac{a^{n-1} \cdot e^{-a}}{(n-1)!}, \quad \text{unde} \quad a = I - 1$$

Luând în considerare apropierea limbă-cod (limba fiind considerată ca un cod limitat probabilistic) se poate calcula cantitatea de energie informaţională conţinută într-un mesaj oarecare.

Energia informaţională este o noţiune introdusă de matematicianul român Octav Onicescu şi se defineşte astfel:

- fie A un sistem oarecare şi $A_1, A_2, ..., A_n$ diferitele sale stări. Notând cu $P_1, P_2, ..., P_n$ probabilităţile stărilor $\left(\sum_{i=1}^{n} P_i \right)$, mărimea $I(A)$ calculată cu relaţia:

$$I(A) = P_1^2 + P_2^2 + ... + P_n^2 = \sum_{i=1}^{n} P_i^n$$

se numeşte energie informaţională. Acest indicator reflectă starea de organizare a sistemului. În cazul în care $P_1 = P_2 = ... = P_n = \frac{1}{n}$ energia informaţională este minimă: $I(A) = \frac{1}{n}$.

Deoarece probabilităţile P_i nu se cunosc, ele vor fi estimate cu ajutorul frecvenţelor relative, astfel că expresia energiei informaţionale devine:

$$I(A) = \sum_{i=1}^{n} \left(\frac{n_i}{N} \right)^2 = \frac{1}{N^2} \sum_{i=1}^{n} n_i^2$$

unde n_i reprezintă frecvenţa absolută a stării A_i, iar $N = \sum_{i=1}^{n} n_i$.

Cea mai bună aproximaţie pentru $I(A)$ se obţine folosind pentru probabilităţile P_i o estimare eficientă şi absolut corectă şi în acest caz expresia energiei informaţionale devine:

$$I(A) = \sum_{i=1}^{n} \frac{n_i (n_{i-1})}{N(n-1)}$$

Cercetările statistice au determinat probabilităţile de apariţie a cuvintelor în funcţie de lungimea lor. Pentru limba română aceste probabilităţi sunt date în tabelul de mai jos:

Lungimea	Probabilităţile de apariţie
1	0,0206
2	0,2723
3	0,1066
4	0,1396
5	0,1146

6	0,0966
7	0,0990
8	0,0603
9	0,0505
10	0,0173
11	0,0152
12	0,0063
13	0,0033
14	0,0006

Se observă că, având probabilitatea de apariţie de 0,2723 cuvintele de două litere sunt cele mai frecvente. Lungimea medie a cuvântului se poate calcula cu relaţia:

$$l_{med} = \sum_{i=1}^{n} P_i l_i$$

n = cea mai mare lungime a unui cuvânt din limba română. S-a obţinut lungimea medie $l_{med} = 4,6$ litere/cuvânt.

Dacă se ia *n* = 14 şi dacă toate cuvintele ar fi egal probabile atunci:

$$l_{med} = \frac{1}{14}\sum_{i=1}^{14} l_1 = 7,5 \text{ litere.}$$

Pentru un limbaj natural *L* şi un text *T* de lungime *n* al cărui vocabular conţine *N* cuvinte avem:

$$F_\gamma \cdot \gamma = k\left(L,n\right); \quad \gamma = 1,2,...,N,$$

lege cunoscută sub numele de legea ESTOUP-ZIPP şi în care F_γ reprezintă frecvenţa cuvântului γ, iar *k* este o constantă care diferă în funcţie de limbă şi de lungimea textului *n*. Există şi o extindere naturală a acestei legi:

$$F_\gamma \cdot \gamma^a 0 = ka_0,$$

unde a_0 este o constantă reală pozitivă caracteristică limbajului. Pentru calculul lui a_0 şi a constantei k_{a_0}, se poate folosi relaţia:

$$k_a(L) = \frac{1}{N}\sum_{\gamma=1}^{N} F_\gamma \cdot \gamma^a$$

Pentru limba română s-a calculat $a_0 = 0{,}8$ şi $k_{a_0} = 655$.

O altă lege utilă criptanaliştilor este legea rang-frecvenţă: într-o listă de frecvenţă produsul dintre rangul unui cuvânt şi frecvenţa corespunzătoare este constantă.

Prin *rang* se înţelege numărul de ordine al unui cuvânt în cadrul unei liste de frecvenţe. *Lista de frecvenţe* este un vocabular special în care cuvintele sunt aşezate în ordinea descrescătoare a frecvenţelor cu care apar în text. *Vocabularul* unui text cuprinde numărul total de cuvinte diferite din text.

Există şi legea MANDELBROT:

$$F_\gamma = P \cdot L \cdot (\gamma + a)^{-b},$$

în care: *L* este lungimea textului în cuvinte; *a*, *b* doi parametrii caracteristici ai limbii; *p* este o constantă care se calculează cu relaţia:

$$\frac{1}{P} = \frac{1}{(1-a)^b} + \frac{1}{(2-a)^b} + \ldots + \frac{1}{(N-a)^b}$$

Un indicator interesant pentru limbajele naturale legat de numărul şi lungimea şirurilor repetate (secvenţe) care pot să apară şi care este implicit legat de transformările la care sunt supuse textele prin aplicarea unor prelucrări criptografice a fost introdus de ELLIOT FISCHER. Fie *A* un alfabet finit şi fie *L*(*s*) lungimea unui şir *S* din A_1, unde A_1 reprezintă mulţimea tuturor şirurilor de lungime finită din *A* care se repetă.

Notăm cu *S*(*i*, *j*) un subşir al lui *S* care începe cu poziţia *i* şi se termină cu poziţia *j*. Un şir *R* obţinut prin concatenarea a două şiruri *S* şi *Q* îl notăm: *R* = *SQ*.

Spunem că *R* este reproductibil din *S* dacă *Q* este un subşir al lui *S*. Spunem că *R* este productibil din *S* dacă şirul RQ_1 format prin ştergerea ultimului simbol din *Q* este reproducibil din *S*. Deci reproductibilitatea presupune *copierea*, iar productibilitatea permite apariţia unui simbol nou la sfârşitul operaţiei de copiere.

Dându-se un şir *S* din *A* îl putem considera fie ca rezultat al unei concatenări fie ca rezultat al unui proces de producere în care fiecare nouă secvenţă este productibilă din secvenţa deja existentă.

6.2 Metode de decriptare

Prin decriptarea unui sistem de cifrare *S* se înţelege evidenţierea textelor clare pe baza analizelor criptogramelor realizate prin intermediul lui *S*, fără a avea cunoştinţe despre *sistemul utilizat* şi *cheia folosită*. Cifrarea are ca efect perturbarea unor caracteristici statistice ale limbajelor de redactare a mesajelor, perturbări mai accentuate sau mai puţin accentuate în funcţie de complexitatea sistemului respectiv.

După ce volumul textului interceptat a trecut de *punctul de unicitate* criptograma are o rezolvare unică. Textul cifrat poate fi decriptat prin încercarea succesivă a tuturor cheilor posibile. Această metodă numită *metoda sistemului complet* de încercări nu este însă practică. De exemplu dacă o cheie are 26! posibilităţi $\left(\cong 2\cdot 10^{26}\right)$ care apar în cazul unei simple substituţii şi care se consideră a avea un volum mic.

Dacă adversarul (criptanalistul) are un calculator pentru încercarea cheilor cu o viteza de 1 cheie/microsecundă şi dacă el va găsi cheia adevărată făcând jumătate din încercările posibile ar avea nevoie de $\frac{2\cdot 10^{26}}{2\cdot 60^2\cdot 24\cdot 365\cdot 10^6}\cong 3\cdot 10^2$ ani.

Deci metoda sistemului complet de încercări nu poate fi făcută în practică. În analiza criptografică se foloseşte metoda încercărilor de o factură specială, încercările făcându-se în ordinea descrescătoare a probabilităţilor de utilizare a cheilor, mai mult decât atât încercările se referă la o grupă cât mai mare de chei şi nu la o cheie anume.

În acest caz mulţimea cheilor se împarte în submulţimi care conţin aproximativ acelaşi număr de chei. Cu ajutorul unor încercări se determină submulţimea cheilor din care face parte cheia utilizată.

În cazul de mai sus numărul de încercări s-ar reduce la $26\times 5=130$ de încercări.

Deci în timp ce metoda sistemului complet de încercări face că numărul acestora să fie egal cu numărul cheilor, metoda împărţirii în subgrupe reduce numărul încercărilor şi-l face să fie egal cu volumul cheii exprimat în biţi.

Deci procesul de reducere al numărului de încercări constă în împărţirea mulţimii cheilor în submulţimi echiprobabile, iar numărul mediu de încercări va fi:

$$h = \frac{H(c)}{\log 2}.$$

Dacă fiecare încercare are *S* rezultate posibile, iar fiecare rezultat corespunde posibilităţii găsirii cheii într-una din *S* grupe echiprobabile, atunci numărul mediu de încercări va fi:

$$h = \frac{H(c)}{\log s}.$$

Dacă *S* = 2, fiecare încercare aduce 1 bit de informaţie faţă de $\frac{1}{1,25}$ biţi în cazul sistemului complet de încercări.

În analiza criptografică repetările din textul criptat care pot fi cauzale sau accidentale, au un rol însemnat.

Repetările cauzale reprezintă repetările cifrării textului clar care a trecut prin prelucrarea criptografică.

Repetările accidentale sunt cele care prin circumstanţe neprevăzute provin din cifrările diferitelor elemente ale textului clar. Găsirea repetărilor de diferite lungimi (4,5 sau mai multe caractere) duce de obicei la rezolvarea criptogramelor.

Alteori dacă repetările de diferite lungimi sunt insuficiente se pune problema dacă ele sunt cauzale sau accidentale, evaluându-se semnificaţia criptografică a elementelor cifrate repetate.

În analiza criptografică prin text aleator se înţelege textul în care elementele cifrate vor apare aproximativ cu aceeaşi probabilitate.

Pentru exemplificare presupunem că avem un text aleator de *N* elemente ale unui sistem secret cu *n* elemente diferite (de exemplu *N* = 50 litere ale unui alfabet care conţine *n* = 26 litere); probabilitatea de apariţie a unui anume element va fi $p_i = \frac{1}{n} = \frac{1}{26}$.

În mod similar dacă dispunem de un text cu *N* = 376 de diagrame din cele 676 posibile probabilitatea de apariţie a unei anumite diagrame este $p_d = \frac{1}{676}$.

Dar nu toate cele *n* elemente posibile vor apare în textul de *N* elemente. Numărul de elemente care nu apar este un indicator semnificativ în criptanaliză.

Pentru a putea fi utilizat este necesar să se cunoască distribuţia teoretică a elementelor care nu apar (blancuri). Probabilitatea că există un număr de exact *r* blancuri este:

$$P(r)=\frac{n!}{r!}\sum_{i=0}^{n-r}(-1)^i\frac{1}{n!(n-r-i)!}\left(1-\frac{r+i}{n}\right)^N .$$

S-au tabulat valorile lui $P(r)$ pentru $N = n = 10$ (deci cazul cifrelor).

r	$P(r)$
0	0,00036288
1	0,0163296
2	0,13608
3	0,3556224
4	0,34514424
5	0,1285956
6	0,01718892
7	0,00067176
8	0,000004549
9	0,000000001

Numărul probabil al blancurilor dintr-un text aleator cu *N* elemente, într-un sistem în care există *n* caractere diferite este dat de relaţia:

$$B_N = n\left(1-\frac{1}{n}\right)^N$$

S-a determinat numărul mediu de blancuri pe mulţimi de *n* elemente ale unui text aleator format din *N* elemente. Rezultatele sunt prezentate în tabelul de mai jos pentru cazul $N = 200$ şi $n = 10$.

r	200 $P(r)$	Frecvenţa (f)	$r \times f$
0	0,08	0	0
1	3,26	8	8
2	27,22	22	44
3	71,12	72	216
4	69,02	72	288
5	25,72	21	105
6	3,44	4	24
7	0,14	1	7
8	0	0	0
9	0	0	0
	200	200	692

Deci numărul mediu al blancurilor pe submulţimi de 10 simboluri distincte este:

$$\frac{692}{200} = 3,46$$

Pentru un text aleator de $N = 100$ diagrame din cele $n = 676$ posibile se poate calcula numărul estimat de blancuri:

$$B = 676\left(1-\frac{1}{676}\right)^{100} \cong 582$$

Deci textul va conţine 94 de diagrame distincte.

Prin *text nealeator* se înţelege un text în care elementele sunt bine stabilite în concordanţă cu prelucrarea lor criptografică. Deci din punct de vedere statistic textul clar şi textul nealeator sunt identice. Dacă *n* elemente posibile ale unui sistem au probabilităţile de apariţie $P_1, P_2, ..., P_n$ atunci media blancurilor într-un text compus din *N* elemente:

$$B_N = (1-P_1)^N + (1-P_2)^N + ... + (1-P_N)^N,$$

sau cu o bună aproximaţie:

$$B_N = \sum_{i=1}^{n} e^{-NP_i}$$

Pentru un text în limba engleză cifrat monoalfabetic numărul blancurilor are valoarea din tabelul de mai jos:

n	Numărul blancurilor	*n*	Numărul blancurilor
10	18,50	110	5,64
20	14,13	120	5,46
30	11,55	130	5,21
40	10,03	140	5,04
50	8,84	150	4,88

60	7,98	160	4,78
70	7,33	170	4,67
80	6,74	180	4,56
90	6,29	190	4,44
100	5,83	200	4,4

Pentru texte cu caracter nealeator, în analizele criptografice se utilizează aşa numitul test.

Să considerăm un text având *N* caractere dintr-un sistem cu *n* elemente posibile. Presupunem că din cele *n* elemente sunt posibile $f_1, f_2,..., f_N$ elemente astfel încât $f_1 + f_2 + ... + f_N = N$.

Deci elementul α_1 apare de f_1 ori, elementul α_2 apare de f_2 ori şi aşa mai departe.

Dacă definim mărimea Φ cu ajutorul relaţiei:

$$\Phi = f_1(f_1 - 1) + f_2(f_2 - 1) + ... + f_n(f_n - 1),$$

atunci se poate calcula valoarea medie a acestuia ca fiind $E(\Phi)$:

$$E(\Phi) = S_2 N(N - 1),$$

unde S_2 reprezintă suma pătratelor probabilităţilor de apariţie a fiecărui element din cele *n* elemente posibile din sistem. Pentru un sistem aleator:

$$S_2 = \sum_{i=1}^{n} \frac{1}{n^2} = \frac{1}{n}$$

Pentru un text aleator în care gruparea literelor a fost monografică (caracter cu caracter), diagrafică (grupuri de două caractere cifrate odată) şi trigrafice (grupuri de trei caractere cifrate odată) s-a calculat obţinându-se valorile următoare:

- grupare monografică, $S_2 = 0,038$;

- grupare digrafică $S_2 = 0,0015$;

- grupare trigrafică $S_2 = 0,000057$.

Pentru un text clar şi pentru mai multe limbi s-au obţinut următoarele valori:

Limba	$E(\Phi)$	
	text monografic	text digrafic
Engleză	0,0661 N (N – 1)	0,0069 N (N – 1)
Franceză	0,0778 N (N – 1)	0,0093 N (N – 1)
Rusă	0,0529 N (N – 1)	0,0058 N (N – 1)
Germană	0,0762 N (N – 1)	0,0112 N (N – 1)
Română	0,08595 N (N – 1)	

Având la dispoziţie aceste date statistice se poate stabili din ce limbă face parte următorul text şi ce fel de cifrare s-a folosit:

IBMQO PBIUO MBBGA JCZOF MUUQB

Textul conţinând 25 de caractere are următoarea distribuţie:

A B C D E F G H I J K L M N O P Q U Z R S T V W X Y

1 5 1 0 0 1 1 0 2 1 0 0 3 0 3 1 2 3 1 0 0 0 0 0 0 0

Valoarea testului Φ va fi:

$$\Phi = 1(1-1)+5(5-1)+1(1-1)+1(1-1)+1(1-1)+2(2-1)+1(1-1)+$$
$$+3(3-1)+3(3-1)+1(1-1)+2(2-1)+3(3-1)+1(1-1)=42$$

În limba engleză $E(\Phi)=0,0661\times 25\times 24=39,6$ pentru cifrare monografică. Pentru un text aleator:
$E(\Phi)=0,038\times 25\times 24=22,8$.

Deoarece 39,6 este mult mai apropiat de 42 decât 22,8 se poate spune că mai curând este un text în engleză decât un text întâmplător.

Dacă se iau în considerare pătratele numerelor care reprezintă frecvenţele de apariţie ale literelor:

$$\Psi = f_1^2 + f_2^2 + ... + f_n^2,$$

atunci $\Psi = S_2 N^2 + (1 - S_2) N$ care reprezintă testul Ψ pentru diagrame.

Multe tipuri de cifruri pot fi sparte cu ajutorul analizei statistice, mai ales când textul interceptat are mai mult de 200 de caractere.

O metodă care poate fi utilizată în descifrare este *metoda cuvintelor probabile*. Cuvintele probabile sunt acele cuvinte sau expresii care pot fi întâlnite mai ales într-un mesaj ca urmare a faptului că sunt caracteristice pentru sursa de mesaje respectivă.

Cuvintele probabile pot fi considerate comune sau silabele care se întâlnesc mai frecvent într-o limbă; de exemplu: are, cu, şi, ile, lor etc. pentru limba română. Metoda se poate utiliza astfel:

- presupunând că o anumită parte a criptogramei reprezintă un anumit cuvânt al mesajului se găseşte o parte a cheii. Această parte se foloseşte pentru descifrarea celorlalte părţi ale criptogramei. Metoda cuvintelor probabile s-a dovedit destul de eficace în special pentru mesaje destul de lungi. Sunt puţine cifruri clasice care având un volum mic al cheii rămân mult timp nedecriptate prin metoda cuvintelor probabile. De aceea această metodă se foloseşte pentru verificarea calităţii cifrurilor. În practică metoda se utilizează împreună cu alte metode.

6.3 Spargerea sistemelor criptografice

Orice sistem criptografic conţine două elemente esenţiale:

- procedeul general de cifrare;

- cheia specifică utilizată.

De exemplu cifrul lui Caesar este un procedeu general monoalfabetic care operează asupra unui alfabet standard şi în care cheia specifică este cheia 3. Deci relaţia de cifrare ar fi:

$$c_i = m_i + 3$$

unde c_i este valoarea numerică a literei criptogramei nu valoarea numerică a literei din mesajul clar m_i.

Presupunând că nu este cunoscută cheia:

$$c_i = m_i + a\left(\text{modulo } 26\right),$$

unde $m_i \in \{A, B, C, ..., Z\}$ iar a trebuie determinată. Pentru un criptanalist această problemă nu este dificilă, el neavând de efectuat decât 25 de încercări. Este posibil ca numai prin efectuarea deplasărilor succesive ale literelor unui singur cuvânt din textul cifrat să se afle cheia specifică. De exemplu fie cuvântul din criptograma: K I T K C T C T C Q

K	*I*	*T*	*K*	*C*	*T*	*C*	*T*	*C*	*Q*
J	*H*	*S*	*J*	*B*	*S*	*B*	*S*	*B*	*P*
I	*G*	*R*	*I*	*A*	*R*	*A*	*R*	*A*	*O*
H	*F*	*Q*	*H*	*Z*	*Q*	*Z*	*Q*	*Z*	*N*
G	*E*	*P*	*G*	*Y*	*P*	*Y*	*P*	*Y*	*M*
F	*D*	*O*	*F*	*X*	*O*	*X*	*O*	*X*	*L*
E	*C*	*N*	*E*	*W*	*N*	*W*	*N*	*W*	*K*
D	*B*	*M*	*D*	*V*	*M*	*V*	*M*	*V*	*J*
C	*A*	*L*	*C*	*U*	*L*	*U*	*L*	*U*	*I*

Deci după 8 deplasări ale literelor s-a obţinut un cuvânt cu sens în limba română. Aşadar cheia de cifrare este –8 iar relaţia de cifrare este $c_i = m_i + 8$ (modulo 26).

În acest caz s-a presupus că se cunoaşte procedeul general de cifrare şi s-a reuşit stabilirea cheii specifice. Mergând de la simplu la complex se presupune că există situaţia când trebuie determinat şi procedeul de cifrare prin metode criptanalitice. Determinarea procedeului general de cifrare se poate face cu ajutorul frecvenţei relative de apariţie a diferitelor litere din alfabet. În acest scop se consideră un eşantion de text clar şi se determină frecvenţa de apariţie a literelor, folosind în acest sens relaţia:

$$f_i = \frac{n_i}{N}$$

unde *N* este numărul de litere pe care îl conţine textul clar iar n_i reprezintă numărul de apariţii a literei α_i .Se determină apoi frecvenţele de apariţie ale literelor din textul criptat şi prin

comparaţie se pot stabili anumite corespondente între literele din textul criptat şi literele alfabetului. Pentru exemplificare să analizam o criptogramă obţinută prin transformări de tipul: $c = am + b$, unde perechea *a, b* reprezintă cheia specifică. Această transformare nu mai realizează o simplă translaţie a literelor din alfabet ci o substituţie oarecare, iar constanta *a* nu trebuie să fie divizor al lui 26. Fie criptograma:

JAHGZTGTH JFJMTRTOXG NGFAMXZGHWFNT TJMT X XATGHMFT QFWFNFOH NHGT JXOFNFMH RPOMH MTUHNFMHMT

Lungimea criptogramei este 80 iar frecvenţele de apariţie ale literelor sunt:

Litera	Frecvenţa	Litera	Frecvenţa
A	3	P	1
F	10	Q	1
G	7	R	2
H	10	T	12
J	5	U	1
M	9	W	2
N	6	X	5
O	4	Z	2

Comparând aceste frecvenţe cu frecvenţele literelor din limba română se pot face unele corespondente:

- $T_c \to E_m$ (T din criptogramă corespunde lui E);

- $F_c \to J_m$? (F din criptogramă corespunde lui I).

Pe baza acestor corespondente se poate scrie sistemul:

$$\begin{cases} 19 \equiv 4a + b \\ 5 \equiv 8a + b \end{cases} \quad \text{(modulo 26)}$$

Scăzând cele două ecuaţii între ele se obţine:

$$-4a \equiv 14 \quad \text{(modulo 26)}$$
$$22a \equiv 14 \quad \text{(modulo 26)}$$
$$11a_0 \equiv 7 \quad \text{(modulo 13)}$$
$$a_0 \equiv 7 \cdot 11^{11} \text{(modulo 13)}$$
$$a_0 = 3 \text{ şi } a_1 = 16$$

Pentru a_0 se obţine:

iar pentru: $a_1 = 16$ se obţine:

$$b = 5 - 11 = 5 + 15 = 20$$

Deci s-au obţinut două soluţii:

$$c = 3\text{m} + 7 \quad \text{(modulo 26)}$$
$$c = 16\text{m} + 20 \quad \text{(modulo 26)}$$

Aplicând pe rând cele două transformări numai prima soluţie duce la obţinerea unui text cu înţeles în limba română:

SPARGEREA SISTEMELOR CRIPTOGRAFICE ESTE O OPERAŢIE DIFICILĂ CARE SOLICITĂ MULTĂ TENACITATE

Pentru descifrare s-a folosit relaţia:

$$m = 3^{-1}(c-7) \quad \text{(modulo 26)}$$

Dar:

$$3^{-1}\left(\bmod 26\right) = 9 \quad \left(\text{modulo } 26\right)$$
$$-7\left(\bmod 26\right) = 19 \quad \left(\text{modulo } 26\right)$$

Deci $m = 9c + 15$ (modulo26)

A doua soluţie nu este posibilă deoarece (16,26) = 2 şi deci 16 nu are invers în mulţimea claselor de resturi modulo 26.

Să analizăm un alt exemplu în care s-au folosit substituţii simple cu reprezentări unice şi uniforme. Fie criptograma:

33. 15. 44. 35. 14. 15. 32 .15./ 13. 42. 24. 41. 44. 35. 22. 42. 11. 21. 24. 13. 15./ 44. 42. 15. 12. 45. 24. 15./ 43. 44. 45. 14. 24. 11. 44. 15./ 13. 11./ 45. 34./ 13. 11. 41. 24. 44. 35. 32./ 43. 41. 15. 13. 24. 11. 31./ 11. 32./ 43. 13. 42. 24. 43. 45. 32. 45. 24./ 32. 11./ 34. 35. 24./

Pentru început facem următoarele ipoteze:

- mesajul clar este redactat în limba română;

- criptograma respectă împărţirea pe cuvinte a mesajului, fiecare literă având ca reprezentare cifrantă un grup de două cifre;

- nu sunt folosite semne de punctuaţie.

Vom căuta acum să vedem în ce măsura ipotezele de mai sus se verifică. În primul rând lungimea medie a cuvintelor este:

$$l_{med} = \frac{2 \cdot 8 + 1 \cdot 13 + 3 \cdot 7 + 4 \cdot 2 + 1 \cdot 3 + 1 \cdot 9}{13} = 5{,}83 \text{ litere}$$

Valoarea de 5,83 este foarte apropiată de lungimea medie a cuvintelor din limba română.

Frecvenţele de apariţie ale reprezentărilor cifrante sunt prezentate în tabelul următor:

Reprezentare cifrantă	11	12	13	14	15	21	22	23	24	25
Frecvenţă	7	1	6	2	8	1	1	–	9	–
Reprezentare cifrantă	31	32	33	34	35	41	42	43	44	45
Frecvenţă	–	6	1	2	4	3	4	4	6	5

Criptograma conţine în total 70 de litere din care 17 distincte. Pentru un text în limba română, de lungime $N = 70$ se obţine:

$$l_0 = 26\left(1 - e^{-\frac{70}{26}}\right) = 16{,}80\,.$$

Deci ipoteza că avem de-a face cu o substituţie de litere ca reprezentări unice de lungime 2 devine plauzibilă. Reprezentarea cifrantă 24 este cea mai frecventă apărând de 9 ori şi s-ar putea face corespondenţa $24 \to A$ (*I* sau *E*).

Aşezând reprezentările cifrante cele mai frecvente în ordinea descrescătoare a numărului de apariţie s-ar putea face următoarele corespondente:

Reprezentare cifrantă	24	15	11	13	44	32
literă	A	E	I	R	T	N

Înainte de a înlocui aceste posibile echivalente în criptogramă facem următoarele observaţii:

- cea mai frecventă reprezentare din finalul cuvintelor este 15, de aici corespondenţa $15 \to E$;

- diagramele 11.32 şi 32.11 cele mai plauzibile ar fi: AL-LA sau NU-UN, dar în corelaţie cu frecvenţele de apariţie ale literelor decidem $11 \to A;\ 32 \to L$.

Deci în tabelul de mai sus operăm două modificări:

Reprezentare cifrantă	24	15	11	13	44	32
literă	A	E	I	R	T	N

- reprezentările cele mai frecvente ca literă iniţială sunt 43 şi 13, iar în limba română C şi S, deci putem avea $S \to 13\,(43)\ C \to 43\,(13)$.

Dar există un cuvânt care începe cu 43.13 care nu poate fi CS… ci SC... Aşadar $43 \to S$ şi $13 \to C$. Răspândind în criptogramă echivalentele probabile stabilite se obţine în final mesajul:

METODELE CRIPTOGRAFICE TREBUIE STUDIATE CA UN CAPITOL SPECIAL AL SCRISULUI LA NOI, iar tabelul de substituţie este:

V 0	1	2	3	4	5
1	A	B	C	D	E
2	F	G	H	I	J
3	K	L	M	N	O
4	P	R	S	T	U
5	V	W	X	Y	Z

Literele care nu apar în text s-au trecut în tabel pe baza observaţiei ca primă linie formată din litere care apar în criptogramă a fost scrisă în ordine alfabetică normală.

În exemplul de mai sus am folosit faptul că criptograma a respectat împărţirea reală pe cuvinte a mesajului. Acest lucru nu este însă esenţial şi o criptogramă poate fi descifrată şi în cazul în care nu sunt puse în evidenţă spaţiile dintre cuvinte.

Vom analiza în continuare tehnicile folosite pentru spargerea sistemelor criptografice polialfabetice, la care deci lungimea cheii este mai mare decât 1.

Decriptarea acestor sisteme comportă două etape:

- determinarea lungimii cheii de cifrare;

- determinarea efectivă a alfabetelor folosite şi apoi descoperirea textului clar.

În cifrarea polialfabetică cu cât numărul de alfabete este mai mare cu atât distribuţia reprezentărilor cifrante este mai uniformă.

Dacă mesajul este scurt sau daca numărul alfabetelor folosite la cifrare este mic atunci examinarea distribuţiei frecvenţelor poate fi total neconcludentă. Neuniformitatea distribuţiei variabilei aleatoare implicată de un mesaj se caracterizează prin coeficientul M_R (măsura neuniformităţii) care se defineşte astfel:

$$M_R = \sum_{i=1}^{N} \left(P_i - \frac{1}{N} \right)^2$$

unde P_i este probabilitatea de apariţie a literei α_i.

Se observă că dacă:

$$P_1 = P_2 = ... = P_N = \frac{1}{N},$$

atunci $M_R = 0$.

Având în vedere că $\sum_{i=1}^{N} P_i = 1$, atunci M_R se mai poate scrie:

$$M_R = \sum_{i=1}^{N} \left(P_1^2 - \frac{2P_i}{N} + \frac{1}{N} \right) = \sum_{i=1}^{N} P_i^2 - \frac{2}{N} N + N \frac{1}{N^2} =$$

$$= \sum_{i=1}^{N} P_i^2 - \frac{1}{N}$$

Iar dacă $N = 26$, atunci $M_R = \sum_{i=1}^{26} P_i^2 - \frac{1}{26}$.

Pentru text clar având 1000 de litere coeficientul M_R are valoarea determinată experimental de 0,047592. Aşadar M_R poate lua valori cuprinse între 0 şi 0,047492, şi permite stabilirea cu o anumită probabilitate dacă cifrarea a fost monoalfabetică sau polialfabetică.

Dificultatea constă în calcularea sumei pătratelor probabilităţilor de apariţie a literelor. Având în vedere că P_i^2 se poate interpreta că probabilitatea evenimentului ca doua litere selectate la întâmplare din textul considerat să fie identice, apare posibilitatea de a estima $\sum_{i=1}^{N} P_i^2$ fără a cunoaşte P_i cu probabilitatea ca două litere luate la întâmplare să fie identice ceea ce este echivalent cu calculul raportului dintre numărul dubletelor şi numărul total de diagrame din textul considerat.

Dacă notăm cu f_A frecvenţa absolută a literei *A* din textul considerat atunci numărul dubletelor AA va fi:

$$C_{f_A}^2 = \frac{1}{2} f_A (f_A - 1)$$

Dacă numărul total de litere din text este *N* atunci probabilitatea unui dublet este:

$$P_i^2 = \frac{f_i(f_i - 1)}{N(N-1)}$$

iar $\sum_{i=1}^{N} P_i^2 = \frac{\sum_{i=1}^{N} f_i(f_i - 1)}{N(N-1)}$ şi semnifică şansa ca două litere dintr-o distribuţie să fie identice, iar acest număr se numeşte indice de coincidenţă (I_c) putând lua valori cuprinse între 0,038 şi 0,086 la un alfabet format din 26 de litere. Limita inferioară corespunde unei distribuţii uniforme, iar cea superioară unui cifru monoalfabetic.

O măsură a modului în care numărul de alfabete folosite la cifrare influenţează indicele de coincidenţă se poate determina statistic şi I_c se poate exprima cu ajutorul relaţiei:

$$I_c = \frac{1}{m} \cdot \frac{N-m}{N-1} \cdot 0,086 + \frac{m-1}{m} \cdot \frac{N}{N-1} \cdot 0,038$$

în care: *N*-lungimea textului cifrat;

m - numărul alfabetelor folosite;

Dacă *N* = *m* atunci I_c tinde spre valoarea 0,038 iar dacă *m* = 1 atunci $I_c = 0,086$.

S-au determinat valorile lui I_c pentru texte mai lungi de 1000 de litere şi pentru diferite valori ale lui *m* (numărul de alfabete utilizat), rezultatele fiind prezentate în tabelul de mai jos:

m	I_c
1	0,086
2	0,068
5	0,047
7	0,044
10	0,042
> 10	0,038

În cazul utilizării indicelui de coincidenţă pentru determinarea numărului de alfabete utilizate la cifrare trebuie să se ţină seama de caracterul statistic al lui I_c; deci pentru mesaje scurte valoarea calculată poate fi foarte diferită faţă de valoarea aşteptată. De asemenea, trebuie avut în vedere că valoarea scontată a lui I_c presupune folosirea alfabetelor de cifrare de acelaşi număr de ori. Deci cuvântul cheie nu trebuie să conţină litere identice.

Un alt indicator pentru stabilirea numărului de alfabete utilizat se poate obţine datorită proprietăţilor algoritmului de cifrare polialfabetic şi anume:

- dacă se scrie textul clar sub cuvântul cheie, atunci literele textului clar de pe o coloană se cifrează monoalfabetic

prin corespondenţa stabilită de litera cheii din coloana respectivă.

Ca urmare, dacă se repetă un cuvânt sau un şir de litere în textul clar şi dacă se întâmplă ca literele identice să fie în aceeaşi coloană, atunci şi în textul cifrat apar secvenţe de litere identice. Observaţia este prezentată mai jos şi este esenţială în spargerea sistemelor.

Să presupunem că o cheie de cifrare de lungime *K* şi *d* este multiplu de *K*, deci $d = K \times m$ atunci este evident că porţiunile [*AB*] şi [*CD*] vor fi cifrate cu aceeaşi cheie şi dacă ele cuprind porţiuni identice atunci vor apare în criptogramă ca secvenţe identice.

Deci calculând distanţele dintre secvenţele identice din criptogramă şi luând factor comun cel mai frecvent al acestora vom putea găsi lungimea cheii.

Apoi se împarte textul cifrat în coloane (în număr de *K*) fiecare reprezentând cifrări monoalfabetice.

Fie criptograma:

WBHNS/ PGHRF/ SGGGH/ EFHVW/ CCFHK/ ASNSP/ GBWCH/ HRSFC/ ZFXZD/ VCDTS/ SGFVQ/ BDNSV/ KWQTS/ WGZH.

Se observă că secvenţele NSPG se repetă la o distanţă de 24 de litere, grupul NS mai apare odată la distanţa de 30 de litere, iar grupul TS mai părea odată la 15 caractere. Putem trage concluzia că lungimea cheii este 3. Deci vom avea trei criptograme cifrate monoalfabetic:

1) W N G F G E V C K N G C R C X V T G Q N K T G;

2) B S H S G F W F A S B H S Z Z C S F B S W S Z;

3) H P R G H H C H S P W H F F D D S V D V Q W H.

Aplicând acum observaţiile de la cifrurile monoalfabetice şi rezolvând unele congruenţe modulo 26 se găseşte pentru prima criptogramă cheia 2, pentru a doua cheia 14, iar pentru a treia cheia 3.

Cu acestea soluţia apare simplu:

UNELE METODE DE SECRETIZARE IMPLEMENTATE PE CACULATOARELE PERSONALE ŞI ÎN REŢELE.

Dacă sistemele de cifrare au folosit reprezentări multiple proporţionale, vom urmări reducerea acestora la sisteme bazate pe substituţii simple cu reprezentări unice uniforme.

În cazul reprezentărilor multiple pentru aceeaşi literă a mesajului clar rezultă că numărul elementelor cifrante este mai mare decât numărul de litere din alfabet şi trebuie găsite acele reprezentări cifrante care sunt echivalente în sensul că reprezintă aceeaşi literă.

Dacă de exemplu în criptogramă apar două secvenţe asemănătoare cum ar fi:

13 11 42 70 53

13 11 42 70 53

este aproape sigur că 42 şi 18 sunt reprezentări cifrante ale aceeaşi litere.

Toate echivalentele care se fac trebuie verificate căutând alte secvenţe care le cuprind.

Metoda dă rezultate însă necesită un volum mare de text interceptat.

6.4 Spargerea sistemelor poligrafice

Pentru o mai bună mascare a caracteristicilor statistice ale variabilei aleatoare implicate de textul cifrat se recurge adeseori la un algoritm de cifrare în care unitatea de prelucrare este un şir de litere (două sau mai multe).

Procedura de decriptare a unui astfel de cifru se exemplifică în cele ce urmează:

1	FFTVT	SPUCM	BAUKA	BVCAA	PMYIK	KWEXO
2	UKABY	ECUXJ	XXYCY	PETKE	HOYRX	WCMZY
3	XIOGIP	ZDAZI	FFZVP	JNNCH	GIOUL	WGAZC
4	IMSXR	ZPBTV	LPDAN	FFFZI	BLYLP	UDUKP
5	CMERG	IXGZV	VYPMD	FEGVC	SYDQZ	VVYPM
6	DFDPU	RPUPZ	EBTSO	UVQTV	TSKIX	XUSFL
7	XWAST	YKGNB	GKZRA	DYCLW	HKHW	CXXFL
8	IMLPJ	FLCYE	CUZJX	XOGFR	MORVK	XXZZ
9	JRFFA	BNANP	XXUUA	CQUNU	RMDFD	FQUJM
10	SOYLX	XUFRI	JHZRN	FRFIM	ZZXNF	ITAZI

11	VAJEF	IGYXF				
12	HQDFN	NGYFF	DHXJO	AGIWN	XOABW	EKSVU
13	WWDUL	HOUKG	NBRYL	PCHNF	GAKGO	UFBET
14	PUAAN	LRMWX	OUQUG	AKGOU	BANQD	FNN

Sarcina primordială a criptanalistului este de a obţine anumite date privind sistemul criptografic folosit la cifrare. Pentru aceasta se stabileşte distribuţia monografică a variabilei aleatoare implicate de textul cifrat şi se calculează indicele de coincidenţă:

$$I.C. = \frac{\sum_{i=A}^{Z} f_i\left(f_i - 1\right)}{N\left(N-1\right)}$$

În cazul criptogramei date, N = 418 caractere, rezultă I.C. = 0,036 ceea ce arată că cifrarea nu este monoalfabetică. Pentru a verifica dacă cifrarea este sau nu polialfabetică se examinează şirurile de litere care se repetă în criptogramă. În criptograma recepţionată astfel de şiruri de litere cu lungimea minimă egală cu 4 sunt următoarele:

Şiruri de litere care se repetă	Poziţia primei litere		l – lungimea intervalului	Divizorii primi ai lui l
ZVVYPMOF	129	145	16	2
HQDFNN	330	412	62	2; 31
GAKGOU	380	404	24	2; 3
TVTS	3	169	166	2; 83
KIXX	173	235	82	2; 41

Este foarte puţin probabil că repetiţiile formate din 8, respectiv din 6 litere să fie accidentale. Singurul factor prim comun al intervalelor de repetiţie fiind 2 se verifică dacă s-a executat sau nu cifrarea cu ajutorul a două alfabete. Totuşi 2 fiind factor pentru fiecare din intervalele de repetiţie se pune întrebarea dacă sistemul de cifrare este digrafic. Pentru a răspunde la această întrebare se întocmeşte tabelul de frecvenţe de apariţie a diagramelor din textul cifrat.

Într-un fel sistemul digrafic se poate considera ca un sistem monoalfabetic însă pe un alfabet de $26^2 = 676$ caractere. Fiecare dintre aceste caractere (diagrame) are o frecvenţă caracteristică în textul clar. În acest caz se poate aplica raţionamentul de la decriptarea monoalfabetică. Calculând pentru un eşantion de text clar (2500 de diagrame) suma pătratelor frecvenţelor caracteristice ale celor 676 de diagrame rezultă:

$$\sum_{i=AA}^{ZZ} P_i^2 = 0,0116,$$

ceea ce arată că gradul de neuniformitate a distribuţiei (generate) digrafice, ţinând seama că pentru o distribuţie uniformă acest coeficient este egal cu (1/676) = 0,0015.

Cu ajutorul frecvenţelor de apariţie a digramelor se determină indicele de coincidenţă:

$$I.C. = \frac{\sum_{i=AA}^{ZZ} f_i\left(f_i - 1\right)}{N\left(N-1\right)} = 0,00989 \approx 0,001$$

unde $N = 209$ este numărul de digrame din criptogramă.

Având în vedere că I.C este o aproximare liniară a sumei $\sum P_i^2 = 0,0116$, rezultă că sistemul folosit este digrafic; în cadrul acestui sistem ori de câte ori apare o digramă din textul clar, ea este întotdeauna înlocuită cu aceeaşi digramă de cifru. În continuare se pune problema stabilirii tipului sistemului digrafic de cifrare precum şi a cheii specifice de cifrare. Există mai multe tipuri de sisteme digrafice. De exemplu, transformarea liniară ilustrată de ecuaţia matricială:

$$\begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}$$

Pentru a verifica acest lucru se realizează o serie de texte speciale. Astfel dacă s-ar putea identifica două sau mai multe corespondente de digrame de forma $C_1C_2 \to M_1M_2$ atunci cu ajutorul echivalentelor numerice ale literelor C_1, C_2, M_1 şi M_2 se poate obţine un sistem de congruenţe de forma:

$$\begin{cases} M_1 = C_1b_{11} + C_2b_{12} \pmod{26} \\ M_2 = C_1b_{21} + C_2b_{22} \pmod{26} \end{cases}$$

Matricea sistemului fiind inversă matricei din transformarea liniară de mai sus rezultă:

$$\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}^{-1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Căutarea digramelor componente se face în mod similar cu căutarea literelor pentru soluţionarea unui sistem de cifrare monoalfabetică. Dacă s-ar putea stabili două corespondente de forma $C_1C_2 \to M_1M_2$ atunci se pot scrie patru congruenţe, suficiente pentru calcularea necunoscutelor $b_{ij}\left(i, j = \overline{1,2}\right)$. Daca apar ambiguitati pentru valorile coeficientilor b_{ij} atunci este necesar să se examineze un text cifrat suplimentar pentru obţinerea soluţiei corecte. Desigur se presupune că echivalentele numerice ale literelor sunt cunoscute.

În căutarea digramelor corespunzătoare se pleacă de la frecvenţele de apariţie a digramelor şi a şirurilor de litere din textul cifrat şi din textul în clar în eşantion.

Digrame în ordine descrescătoare (la un text clar):

TI	*TO*
AT	*IC*
DE	*OK*
IN	*RA*
ER	*EN*
TE	*MI*
UL	*NT*
RE	*PE*
AR	*SI*
TA	*UN*
ST	*EC*
CA	*IE*
	IT

Din exemplul:

OU	*LP*
XX	*QU*
DF	*TV*
FG	*ZI*
PU	*ZV*
AB	*FF*
GA	

Se încearcă 3 comparaţii:

$$OU_c \to TI_m$$
$$XX_c \to AT_m$$
$$DF_c \to DE_m$$

Acestea conduc la un sistem de congruenţe:

$$\begin{aligned}
&19 \equiv 14b_{11} + 20b_{12}\\
&8 \equiv 14b_{21} + 20b_{22}\\
&0 \equiv 23b_{11} + 23b_{12} \pmod{26}\\
&19 \equiv 23b_{21} + 23b_{22}\\
&3 \equiv 3b_{11} + 5b_{12}\\
&4 \equiv 3b_{21} + 4b_{22}
\end{aligned}$$

din rezolvare rezultă un sistem incompatibil.

$$OU \to IE$$
$$XX \to TI$$
$$DF \to TA$$

$$\begin{cases}
8 \equiv 14b_{11} + 20b_{12}\\
4 \equiv 14b_{22} + 20b_{22}\\
19 \equiv 23b_{11} + 23b_{12} \pmod{26}\\
8 \equiv 23b_{21} + 23b_{22}\\
19 \equiv 3b_{11} + 5b_{12}\\
0 \equiv 3b_{22} + 5b_{22}
\end{cases}$$

Se rezolvă primele 4 grupându-le astfel:

$$\begin{cases} 8 \equiv 14b_{11} + 20b_{12} \,|\cdot 23 \\ 19 \equiv 23b_{11} + 23b_{12} \,|\cdot 20 \end{cases} \quad \begin{cases} 4 \equiv 14b_{21} + 20b_{22} \\ 8 \equiv 23b_{21} + 23b_{22} \end{cases}$$

$$184 \equiv 322b_{11} + 460b_{12}$$

$$380 \equiv 460b_{11} + 460b_{22}$$

Soluţiile 1,4,5 şi 8 verifică şi congruenţele din sistemul dat mai sus (cu 6 ecuaţii).

Având în vedere că soluţia 1 duce la matricea neinversabilă:

$$P = \begin{bmatrix} 5 & 6 \\ 1 & 4 \end{bmatrix}$$

deci aceasta nu poate fi o soluţie a transformării liniare:

$$P_8 = \begin{bmatrix} 18 & 19 \\ 15 & 17 \end{bmatrix}$$

Test de autoevaluare

1. Enumeraţi pricipalele statistici ale limbajelor naturale ;
2. Ce reprezintă un text nealeator.
3. Ce formă are măsura neuniformităţii şi care este rolul ei în analiza cifrării monoalfabetice sau polialfabetice .

Tema de autoinstruire

1. Realizaţi schema bloc a spargerii unei criptograme realizată prin utilizarea metodei cifrului lui Caesar, bazată pe exemplul din curs.

Lucrare de verificare

1. Să considerăm că s-a interceptat următorul text, criptat cu un sistem monoalfabetic (nu se ştie exact ce sistem a fost utilizat). Prin metode de criptanaliză să se descifreze textul de mai jos.

lqakc sp gcxk aca pcmgqb kq kxc pkersmpqsb vk vsmgxkbc mkacpc tcacpbqlqs
vk cgele cmtxq ms nocxgsb mbxcsp vk exsgk oxcbqsbcbk texbslk spclbk gcxk
cmgqpvkcq bxkgcbexslk gqxbslk xktxknkpbcq tkpbxq mbxcsps qp cfkxbsmakpb mqtcxcbex vcx lsatkvk pq bxkrqscq mc zsk txkc gqxsems psgs mc mk cmbktbk mc czlk acxk lqgxq vk lc gkl gq gcxk fkpkcq sp gepbcgb