CRIPTANALIZA. REZULTATE ŞI TEHNICI MATEMATICE

Ediţia I apărută la: Ed. Univ. Buc, 2004, ISBN 973575975-6.

Vasile PREDA, Emil SIMION şi Adrian POPESCU

Ediţia a doua 2011

Cuprins

Capitolul 6

CRIPTANALIZA CIFRURILOR CLASICE

Deciphering is, in my opinion, one of the most fascinanting of arts, and I fear I have wasted upon it more time than it deserves.
Charles Babbage, 1864

6.1. Substituţia simplă şi multiplă

6.1.1. Substituţia simplă

Operaţia de cifrare se bazează pe o *corespondenţă* biunivocă între *alfabetul clar* notat prin $\mathcal{A}$ şi *alfabetul cifrat* notat prin $\mathcal{C}$. Pentru exemplificarea ideilor vom prezupune că alfabetul clar este format din cele 26 de litere ale *limbii române* (fără diacritice) plus *delimitatorul de cuvânt* spaţiul. Alfabetul cifrat poate fi format din aceeleaşi caractere sau doar din cele 26 de litere ale limbii române caz în care spaţiul se va înlocui cu cea mai puţin frecventă literă (Q) sau se va ignora pur şi simplu.

Corespondenţa dintre cele două alfabete (presupunem că delimitatorul de cuvânt este înlocuit cu litera Q) poate fi:

-aleatoare;

-pseudoaleatoare: plecând de la o parolă se construieşte alfabetul cifrat.

Dacă în cazul corespondenţei aleatoare lucrurile sunt cât se poate de clare, vom prezenta pe scurt o metodă de construcţie a corespondenţei în cel de-al doilea caz. Pornind de la o parolă, alfabetul cifrat este construit după următorul algoritm:

-se scriu, o singură dată, în ordinea apariţiei, literele din parolă;

-se scriu literele alfabetului ce nu apar în parolă.

Corespondenţa între cele două alfabete se realizează după regula alfabet în alfabet după o permutare fixă σ (aceasta poate fi chiar permutarea identică iar la decriptare se aplică aceeaşi procedeu dar cu inversa permutării σ).

În funcţie de forma permutării substituţia se numeşte:

-*directă* (alfabetul cifrat are acelaşi sens lexicografic cu alfabetul clar, sunt în total 26 astfel de substituţii). Exemplu de substituţie directă:

A	B	C	D	E	F	G	H	I	J	K	L	M
G	H	I	J	K	L	M	N	O	P	Q	R	S

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
T	U	V	W	X	Y	Z	A	B	C	D	E	F

-*inversă* (alfabetul cifrat are sens invers lexicografic cu alfabetul clar, sunt în total 26 de astfel de substituţii). Exemplu de substituţie inversă:

A	B	C	D	E	F	G	H	I	J	K	L	M
U	T	S	R	Q	P	O	N	M	L	K	J	I

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
H	G	F	E	D	C	B	A	Z	Y	X	W	V

Reamintim aici trei exemple celebre (vechile coduri ebraice) de substituţii reciproce (dacă litera $\mathcal{X}$ se substituie cu litera $\mathcal{Y}$ atunci $\mathcal{Y}$ se va substitui cu $\mathcal{X}$) şi anume:

-*atbash* (prima jumătate a literelor alfabetului se mapează în cea de-a două jumătate în ordine invers lexicografică):

A	B	C	D	E	F	G	H	I	J	K	L	M
Z	Y	X	W	V	U	T	S	R	Q	P	O	N

-*albam* (prima jumătate a literelor alfabetului se mapează în cea de-a două jumătate în ordine lexicografică):

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

-*atbah*:

A	B	C	D	J	K	L	M	E	S	T	U	V
I	H	G	F	R	Q	P	O	N	Z	Y	X	W

Definiţia 6.1.1. Un *cifru de substituţie liniar* de la $\mathbf{Z}_m$ la $\mathbf{Z}_m$ (m fiind numărul de caractere al alfabetului sursă) poate fi descris prin funcţia $f : \mathbf{Z}_m \to \mathbf{Z}_m$ definită prin $f(x) = \alpha x + \beta$ cu $\gcd(\alpha, m) = 1$, funcţia de descifrare fiind $f^{-1}(x) = \alpha^{-1}(x - \beta)$. Cheia de cifrare sunt numerele α şi β.

Observaţia 6.1.1. Primul manual de analiză criptografică a cifrurilor de substituţie a fost scris de *Al-Kindi* [2] în anul 850 AD.

Observaţia 6.1.2. Cifrul de substituţie are *proprietatea de confuzie* (ascunderea legăturii dintre textul clar şi textul cifrat).

6.1.2. Substituţia multiplă

În cazul substituţiei multiple (*substituţie poligrafică*) M caractere din textul clar sunt substituite în N caractere de text cifrat. Evident pentru ca operaţia de cifrare să fie reversibilă trebuie ca $M \geq N$. Dacă $M = N = 2$ atunci cifrul se numeşte cifru de *substituţie digrafică*.

Definiţia 6.1.2. Un cifru de *substituţie poligrafică liniar* de la $\mathbf{Z}_m^k$ la $\mathbf{Z}_m^k$ (m fiind numărul de caractere al alfabetului sursă şi k dimensiunea $k-$gramei) poate fi descris prin funcţia $f : \mathbf{Z}_m^k \to \mathbf{Z}_m^k$ definită prin $f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \boldsymbol{\beta}$ cu matricea $\mathbf{A}$ inversabilă, funcţia de descifrare fiind $f^{-1}(\mathbf{y}) = \mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\beta})$. Cheia de cifrare este matricea $\mathbf{A}$ şi vectorul $\boldsymbol{\beta}$.

Identificarea cifrului de substituţie poligrafică

Un cifru de substituţie poligrafică se poate detecta foarte uşor după frecvenţa $N-$gramelor. Distribuţia acestora este departe de a fi uniformă. Evident, o condiţie necesară de identificare a acestui tip de cifru este aceea ca dimensiunea mesajului cifrat să fie suficient de mare. O procedură de detectare a lui N se bazează, de exemplu, pe minimul entropei H calculate din $R-$grame:

$$H(m; N) = \min_R H(m; R).$$

Alte proceduri de identificare corectă a parametrului n sunt prezentate în secţiunea *Proceduri de identificare a sistemului.* Pentru a se asigura biunivocitatea parametrul M trebuie să fie egal cu N. Construcţia tabelei de substituţie poligrafică se realizează cu ajutorul frecvenţei $N-$gramelor din textul cifrat.

Cifrul celor 4(2) tabele rectangulare

În continuare vom prezenta două sisteme de cifrare de tip substituţie digrafică. Literele alfabetelor (clar şi cifrat) sunt trecute într-un careu de 5×5 (litera I fiind asimilată literei J). Modul de aranjare al literelor albabetelor clar este fixat apriori (de exemplu ordine lexicografică) iar al alfabetelor cifrate în funcţie de parolă. Textul clar este preprocesat astfel încât acesta să fie compatibil cu matricea de cifrare (delimitatorul de cuvânt este ignorat sau este înlocuit cu cea mai puţin frecventă literă, litera I este asimilată cu litera J, şi în fine dacă este cazul mai adjuncţionăm o literă la text pentru a avea un număr par de digrame). Pentru cifrul celor 4 tabele regula de cifrare este dată de regula dreptunghiului: o digramă din textul clar se cifrează în digrama corespunzătoare diagonalei secundare a dreptunghiului determinat de cele două caractere ale digramei clare. Modul de aranjare al celor patru alfabete ($P1$ şi $P2$ sunt alfabetele clare iar $C1$ şi $C2$ sunt alfabetele cifrate în funcţie de parolă) este prezentat mai jos:

P1	C1
C2	P2

.

Algoritmul prezentat mai sus degenerează în algoritmul celor două tabele (verticale sau orizontale) după cum urmează: $P1 \equiv C1$ şi $P2 \equiv C2$ respectiv $P1 \equiv C2$ şi $P2 \equiv C1$. Dacă literele ce formează digrama se află pe aceeaşi coloană, respectiv linie, atunci reprezentarea cifrată a acestora sunt ele însele, respectiv acestea scrise în ordine inversă. Cele două tabele de cifrare sunt:

P1[C1]
C2[P2]

,

respectiv

P1[C2]	P2[C1]

.

Cifrul Playfair

Cifrul Playfair (numele acestui sistem de cifrare provine de la Lordul englez Playfair) este unul dintre cele mai cunoscute sisteme de cifrare digrafice (transformă un grup de 2 litere într-un grup de alte două litere). Acest sistem de cifrare este foarte simplu de folosit dar mult mai sigur decât sistemele de substituţie monoalfabetice. Descriem în continuare modul de utilizare al acestui sistem de cifrare. Literele alfabetului sunt trecute într-un careu de 5×5 (litera I fiind asimilată literei J). Textul clar este preprocesat astfel încât acesta să fie compatibil cu matricea de cifrare (delimitatorul de cuvânt este ignorat sau este înlocuit cu cea mai puţin frecventă literă,

litera I este asimilată cu litera J, şi în fine dacă este cazul mai adjuncţionăm o literă la text pentru a avea un număr par de digrame). Regula de cifrare este următoarea:

i) Dacă digrama ce se doreşte cifrată nu are literele pe aceeiaşi linie sau coloană, atunci regula de cifrare este *regula dreptunghiului*, traseul fiind pe verticală de la cea de-a doua literă a digramei către prima literă.

ii) Dacă digrama ce se doreşte cifrată are literele pe aceeiaşi linie, atunci se aplică regula: *cifrează la dreapta, descifrează la stânga.*

iii) Dacă digrama ce se doreşte cifrată are literele pe aceeiaşi coloană, atunci se aplică regula: *cifrează în jos, descifrează în sus.*

Observaţia 6.1.3. Dacă o digramă apare în textul clar în ordine inversă atunci acelaşi lucru se va întâmpla şi în textul cifrat.

Observaţia 6.1.4. Cifrul Playfair nu are regulă pentru cifrarea literelor duble: digramele ce conţin două litere identice sunt sparte prin introducerea artificială a unei alte litere.

Observaţia 6.1.5. Cifrul Playfair apare ca o extindere, în sensul reducerii numărului de tabele rectangulare folosite (de la două la unul), al cifrului cu 2 tabele.

Metoda cea mai freventă de atac a acestui tip de cifru constă în analiza frecvenţei digramelor de text clar combinată cu metoda comparaţiei patternurilor din textul cifrat cu patternuri din dicţionar. Spectaculos este faptul că în *Manualul Teroristului* [104], în capitolul 13 intitulat *Scrieri Secrete, Coduri şi Cifruri*, se prezintă tehnici steganografice, tehnici de codificare şi o serie de metode de cifrare clasice cum ar fi substituţia simplă şi substituţia multiplă.

6.2. Substituţia polialfabetică

6.2.1. Caracteristicile şi identificarea sistemelor de substituţie polialfabetică

Un sistem de cifrare de tip substituţie polialfabetică este generalizarea sistemului de cifrare de substituţie monoalfabetică. Fie un sistem de cifrare polialfabetic compus dintr-un număr N de alfabete. Fiecare alfabet reprezintă o permutare (stabilită în funcţie de parolă) a alfabetului de intrare. Algoritmul de cifrare constă în substituirea celei de a $i-$a litere m din textul clar cu litera corespunzătoare din cel de al $i \bmod N$ alfabet. Sistemele polialfabetice sunt uşor de identificat prin aplicarea tehnicii frecvenţelor secvenţelor decimate din textul cifrat sau a valorii maxime a funcţiei $Kappa$ (v. paragraful Proceduri de identificare a sistemului).

6.2.2. Atacul sistemelor polialfabetice

Atacul sistemelor polialfabetice este similar cu atacul a N sisteme de substituţie monoalfabetică. Deci, o procedură de *tip divide et impera* are o compexitate de $O(N)$. Procedura este descrisă în continuare:

Intrare: Textul cifrat de lungime M suficient de mare.

Ieşire: Textul clar corespunzător sistemului de cifrare polialfabetic.

Pas 1. Determină numărul de alfabete N.

Pas 2. Pentru $j = 0$ to 4 execută:

pentru $i = 1$ to $N - j$ execută:

aplică procedura de reconstrucţie parţială (pe baza frecvenţelor $(j+1)-$gramelor) a alfabetelor $i, \ldots, i+j$.

Pas 3. Conform celor N alfabete reconstruieşte textul clar.

Observaţia 6.2.1. Procedura descrisă mai sus are ca parametru implicit de analiză numărul maxim de legături 4 : astfel, $1-$gramele sunt caracterele, $2-$gramele sunt dubleţii, etc.

6.3. Soluţia unui cifru de substituţie

În criptanaliză, soluţia unui cifru de substituţie parcurge următoarele etape:

a) *analiza criptogramelor:*

a1) pregătirea unui tabel de frecvenţe;

a2) căutarea repetiţiilor;

a3) determinarea tipului de sistem utilizat;

a4) pregătirea unei foi de lucru;

a5) pregătirea unui alfabet individual;

a6) tabelarea repetiţiilor lungi.

b) *Clasificarea vocalelor şi consoanelor prin studierea:*

b1) frecvenţelor;

b2) spaţiilor;

b3) combinaţiilor de litere;

b4) repetiţiilor.

c) *Identificarea literelor:*

c1) partiţionarea literelor în clase de probabilitate;

c2) verificarea presupunerilor;

c3) înlocuirea valorilor corecte în criptogramă;

c4) descoperirea altor valori pentru a avea soluţia completă.

d) *Reconstrucţia sistemului:*

d1) reconstrucţia tabelei de cifrare;

d2) reconstrucţia cheilor folosite în operaţia de cifrare;

d3) reconstrucţia cheilor sau a cuvintelor cheie ce au fost folosite pentru construcţia şirurilor de alfabete.

6.4. Transpoziţia

Dacă substituţia schimbă valoarea unui caracter din textul cifrat, permutarea schimbă poziţia pe care acest caracter apare în textul cifrat şi nu valoarea sa. Operaţia de cifrare prin permutarea $\sigma \in S_N$ se realizează prin permutarea caracterelor din blocurile adiacente, de lungime N, ale textului clar. Un caz particular al permutării este aceela al transpoziţiei de lungime N. Textul este scris într-o tabelă (completă sau nu) cu N coloane, literele fiind scrise linie cu linie. Pornind de la o parolă literală se construieşte o parolă numerică (spre exemplu se asociează fiecărei litere din parolă poziţia sa în scrierea lexicografică a acesteia). Textul este apoi citit coloană cu coloană într-o ordine stabilită apriori (crescător, descrescător, etc.). Pentru descifrare se aplică acelaşi algoritm dar cu parola numerică σ^{-1}.

Observaţia 6.4.1. Cifrul de transpoziţie (mai general cifrul de permutare) are *proprietatea de difuzie* (disiparea redundanţei textului clar de-a lungul textului cifrat).

6.5. Sisteme mixte

Sistemele mixte de cifrare au la bază o cifrare succesivă a mesajului prin metoda substituţiei iar apoi prin metoda transpoziţiei sau viceversa. Tot ceea ce trebuie să facem acum este să atacăm sistemul de cifrare de la ultima sa componentă către prima. Remarcăm faptul că în cazul substituţiei simple aceasta este comutativă cu operaţia de transpoziţie deci se poate aborda mai întâi substituţia iar apoi transpoziţia. În cazul utilizării unui sistem polialfabetic, cu număr necunoscut de alfabete, recomandarea este ca după stabilirea, prin metode statistice, a numărului de alfabete, să se abordeze concomitent identificarea efectivă a alfabetelor şi al transpoziţiei utilizate. În cazul utilizării unui sistem poligrafic (tabele de cifrare) şi o transpoziţie este recomandabilă o tehnică de tip backtracking.

6.6. Proceduri de identificare a sistemului

Procedurile de criptanaliză prezentate în cadrul acestui paragraf sunt bazate pe calculul unor estimatori pentru o serie de funcţii de test (vezi *Friedman* [20], *Bauer* [6], *Preda* şi *Simion* [58]). Textul clar trebuie să fie omogen din punct de vedere

statistic. Dacă textul nu este omogen, atunci, cu ajutorul unei proceduri, acesta se poate diviza în părţi omogene. Procedurile ce urmează ne permit să identificăm modelul de cifrare, structura statistică a textului clar şi în cele din urmă soluţia problemei.

Să notăm prin $T = (t_1, \ldots, t_M)$ şi $T^{'} = (t_1^{'}, \ldots, t_M^{'})$ două şiruri de lungime M din acelaşi vocabular Z_N (de lungime N). Notăm prin m_i şi $m_i^{'}$ frecvenţele de apariţie ale celei de a $i-a$ litere din alfabetul sursă din şirul T respectiv $T^{'}$. Avem deci:

$$\sum_{i=1}^{N} m_i = M$$

şi

$$\sum_{i=1}^{N} m_i^{'} = M.$$

În cele ce urmează vom descrie funcţiile de decizie *Kappa, Chi, Psi* şi *Phi*.

6.6.1. Funcţia Kappa

Funcţia de test Kappa este definită prin:

$$Kappa(T, T^{'}) = \frac{\sum\limits_{i=1}^{M} \delta(t_i, t_i^{'})}{M},$$

unde δ este simbolul lui *Kronecker.*

Avem inegalităţile:

$$0 \leq Kappa(T, T^{'}) \leq 1,$$

cu egalitate în partea stângă pentru $t_i \neq t_i^{'}$ (pentru orice i) respectiv $T \equiv T^{'}$. Testul Kappa este similar cu testul de corelaţie. Câteodată $Kappa$ se va nota mai simplu prin K.

Avem următoarele teoreme de invarianţă (vezi *Bauer* [6]):

Teorema 6.6.1. *Pentru toate sistemele de cifrare de tip substituţie polialfabetică valoarea lui Kappa a două texte de lungimi egale, cifrate cu aceeaşi cheie, este invariantă.*

Teorema 6.6.2. *Pentru toate sistemele de cifrare de tip transpoziţie valoarea lui Kappa a două texte de lungimi egale, cifrate cu aceeaşi cheie, este invariantă.*

Valoarea medie a lui $Kappa(T, T^{'})$, pentru T şi $T^{'}$ definite de două surse Q respectiv $Q^{'}$ (cu probabilitatea simbolurilor p_i respectiv $p_i^{'}$ pentru $i = = 1, \ldots, N$) este:

$$E[Kappa(T, T^{'})] = \sum_{i=1}^{N} p_i^{'} p_i.$$

Dacă sursele Q şi $Q^{'}$ sunt identice (acest lucru se notează prin $Q \equiv Q^{'}$), adică $p_i = p_i^{'}$, atunci:

$$E[Kappa(T, T^{'})] = \sum_{i=1}^{N} p_i^2.$$

Teorema 6.6.3. *Pentru două surse identice Q şi $Q^{'}$ are loc inegalitatea:*

$$\frac{1}{N} \leq E[Kappa(T, T^{'})] \leq 1,$$

cu egalitate în partea stângă pentru distribuţia uniformă, iar în partea dreaptă pentru sursa deterministă (adică există un indice i pentru care $p_i = 1$).

Observaţia 6.6.1. Dispersia funcţiei Kappa se calculează după formula:

$$D^2(X) = E((E(X) - X)^2).$$

Observaţia 6.6.2. Funcţia Kappa poate fi folosită la atacul sistemelor polialfabetice: diferenţa dintre valorile maxime ale valorilor $Kappa(T^{(i)}, T)$ (am notat prin $T^{(i)}$ textul T deplasat ciclic cu i poziţii la dreapta) este divizibil cu numărul de alfabete utilizate. Pentru a găsi alfabetele utilizate se poate aplica o tehnică de tip divide et impera şi o procedură bazată pe maximul frecvenţei sau atac de tip stereotip.

6.6.2. Funcţia Chi

Funcţia Chi este definită prin formula:

$$Chi(T, T^{'}) = \frac{\sum_{i=1}^{N} m_i m_i^{'}}{M^2}.$$

Vom nota această funcţie şi prin litera grecească χ. Au loc următoarele teoreme de invarianţă (vezi *Bauer* [6]):

Teorema 6.6.4. *Pentru toate sistemele de cifrare de tip substituţie monoalfabetică valoarea lui Chi a două texte de lungimi egale, cifrate cu aceeaşi cheie, este invariantă.*

Teorema 6.6.5. *Pentru toate sistemele de cifrare de tip transpoziţie valoarea lui Chi a două texte de lungimi egale, cifrate cu aceeaşi cheie, este invariantă.*

Avem inegalitatea:

$$Chi(T, T^{'}) \leq 1,$$

cu egalitate pentru $T \equiv T^{'}$. Pentru un text cu distribuţie uniformă T ($m_i = \frac{M}{N}$) şi un text arbitrar $T^{'}$ avem:

$$Chi(T, T^{'}) = \frac{1}{N}.$$

Valoarea medie a lui Chi a două texte T şi $T^{'}$ de lungime egală M peste aceeaşi vocabular Z_N se calculează din probabilităţile p_i şi $p_i^{'}$ ale frecveţei de apariţie ale celui de al i-lea caracter în sursele stochastice Q şi $Q^{'}$ ale textelor:

$$E[Chi(T, T^{'})] = \sum_{i=1}^{N} p_i p_i^{'}.$$

Dacă sursele Q şi $Q^{'}$ sunt identice atunci $p_i = p_i^{'}$ pentru orice i şi

$$E[Chi(T, T^{'})] = \sum_{i=1}^{N} p_i^2.$$

Avem inegalitatea:

$$\frac{1}{N} \leq E[Chi(T, T^{'})] \leq 1,$$

cu egalitate în partea stângă pentru distribuţia uniformă iar în partea dreaptă pentru sursa deterministă.

6.6.3. Funcţia Psi

Funcţia Psi este legată de funcţia Chi prin formula:

$$Psi(T) = Chi(T, T).$$

Vom nota această funcţie prin litera grecească ψ. În mod similar ca pentru funcţia χ avem următoarele teoreme de invarianţă (vezi *Bauer* [6]).

Teorema 6.6.6. *Pentru toate sistemele de cifrare de tip substituţie monoalfabetică valoarea lui Psi este invariantă.*

Teorema 6.6.7. *Pentru toate sistemele de cifrare de tip transpoziţie valoarea lui Psi este invariantă.*

Avem inegalitatea:

$$\frac{1}{N} \leq Psi(T) \leq 1,$$

cu egalitate în partea stângă pentru un text cu distribuţie uniformă ($m_i = \frac{M}{N}$), şi în partea dreaptă dacă T este construit dintr-un singur caracter.

Valoarea medie a lui Psi al unui text de lungime M din alfabetul Z_N se calculează din probabilităţile p_i ale frecvenţei de apariţie a celui de al i-lea caracter din sursa stochastică Q ce produce textul:

$$E[Psi(T)] = \sum_{i=1}^{N} p_i^2.$$

Are loc inegalitatea:

$$\frac{1}{N} \leq E[Psi(T)] \leq 1,$$

cu egalitate în partea stângă pentru un text cu distribuţie uniformă şi în partea dreaptă pentru o sursă deterministă.

Avem teorema $Kappa - Chi$ (vezi *Bauer* [6]):

Teorema 6.6.8. *Pentru două texte T şi T' peste aceeaşi vocabular şi de aceeaşi lungime, valoarea $Chi(T, T')$ este media aritmetică a valorilor $Kappa(T^{(i)}, T')$ ($T^{(i)}$ este textul T rotit ciclic cu i poziţii la dreapta obţinut prin formula $t_j^* = t_{(j-i-1) \bmod M+1}$ pentru $j = 1, \ldots, M$):*

$$Chi(T, T') = \frac{1}{M} \sum_{i=0}^{M-1} Kappa(T^{(i)}, T').$$

Corolarul următor este cunoscut sub numele de teorema $Kappa - Psi$:

Corolar 6.6.1. Pentru un text T de lungime M peste vocabularul Z_N avem:

$$Psi(T) = \frac{1}{M} \sum_{i=0}^{M-1} Kappa(T^{(i)}, T).$$

6.6.4. Funcţia Phi

Vom defini acum o nouă funcţie denumită *Phi* care este dată prin formula:

$$Phi(T) = \frac{\sum_{i=1}^{N} m_i(m_i - 1)}{M(M-1)}.$$

Funcţia *Phi* va fi notată prin litera grecească φ. În cazul în care $T \equiv T'$ avem $Kappa(T^{(0)}, T) = 1$, în timp ce pentru $i \neq 0$ valorile lui $Kappa(T^{(i)}, T)$ sunt relativ mici. Deci cazul $i = 0$ este atipic pentru medie şi deci este mult mai natural să extindem media numai peste cele $M - 1$ cazuri:

$$\frac{1}{M-1} \sum_{i=1}^{M-1} Kappa(T^{(i)}, T).$$

Avem teorema $Kappa - Phi$:

Teorema 6.6.9. (*Bauer* [6]) *Pentru un text T de lungime M peste vocabularul Z_N valoarea lui $Phi(T)$ este media valorilor $Kappa(T^{(i)}, T)$ ($T^{(i)}$ este textul T rotit ciclic cu i poziţii la dreapta):*

$$Phi(T) = \frac{1}{M-1} \sum_{i=1}^{M-1} Kappa(T^{(i)}, T).$$

Avem de asemenea următoarele teoreme de invarianţă (vezi *Bauer* [6]) pentru funcţia *Phi*.

Teorema 6.6.10. *Pentru toate sistemele de cifrare de tip substituţie monoalfabetică valoarea lui Phi este invariantă.*

Teorema 6.6.11. *Pentru toate sistemele de cifrare de tip transpoziţie, valoarea lui Phi este invariantă.*

Teorema 6.6.12. *Dacă T este o secvenţă distribuită uniform atunci*

$$Phi(T) = \frac{M - N}{(M-1)N}$$

unde M este lungimea secvenţei iar N numărul de simboluri (caractere).

Teorema 6.6.13. *Dacă notăm prin $Phi_c(T)$ valoarea funcţie Phi calculată pe $c-$grame atunci:*

$$\lim_{c\to\infty} Phi_c(T) = 0.$$

Valoarea c pentru care $Phi_c(T) = \max_r Phi_r(T)$ este dimensiunea blocului pe care se realizează codificarea datelor după procesul de cifrare.

Observaţia 6.6.3. i) Valoarea medie a lui $Phi(T)$ este egală cu valoarea medie a lui $Psi(T)$ deci este egală cu $\sum_{i=1}^{n} p_i^2$.

ii) Funcţia Phi este utilă deoarece literele rare ($m_i = 1$) sunt eliminate.

iii) $Phi(T) = 0$ dacă şi numai dacă $0 \leq m_i \leq 1$.

iv) Funcţia Phi este utilă în identificarea limbii folosite în textul clar, cu condiţia ca sistemul de cifrare folosit să fie o tranpoziţie şi/sau o substituţie monoalfabetică.

v) Funcţia Phi a fost propusă iniţial de *Solomon Kullback* din considerente stochastice.

6.7. Funcţii simetrice de frecvenţă a caracterelor

Teoremele de invarianţă pentru Psi au loc pentru toate funcţiile de frecvenţă m_i a caracterelor care sunt simetrice. Cea mai simplă funcţie polinomială netrivială este $\sum_{i=1}^{N} m_i^2$. Aceasta aparţine familiei de funcţii:

$$Psi_a(T) = \begin{cases} (\sum_{i=1}^{N} (\frac{m_i}{M})^a)^{\frac{1}{a-1}} & \text{dacă } 1 < a < \infty \\ \exp(\sum_{i=1}^{N} (\frac{m_i}{M}) \ln(\frac{m_i}{M})) & \text{dacă } a = 1 \\ \max_{1,\ldots,N} \frac{m_i}{M} & \text{dacă } a = \infty \end{cases}$$

cu $\sum_{i=1}^{N} \frac{m_i}{M} = 1$. Observăm că Psi_2 este Psi. Cantitatea logaritmică $-\ln Psi_a(T)$ se numeşte entropia *Rényi de ordinul a* a lui T (Alfred Rényi 1921-1970, matematician ungur) . Familia are următoarea reprezentare:

$$-\ln Psi_a(T) = \begin{cases} -\frac{1}{a-1} \ln(\sum_{i=1}^{N} (\frac{m_i}{M})^a) & \text{if } 1 < a < \infty \\ -\sum_{i=1}^{N} (\frac{m_i}{M}) \ln(\frac{m_i}{M}) & \text{if } a = 1 \\ -\max_{1,\ldots,N} \ln(\frac{m_i}{M}) & \text{if } a = \infty. \end{cases}$$

Entropia Rényi de ordinul 1 se numeşte *entropie Shannon* (Claude Shannon 1916-2000, inginer American, părintele teoriei informaţiei) iar entropia Rényi de ordinul 2 este *entropia Kullback.* Toate proprietăţile lui Psi se generalizează la proprietăţi ale lui Psi_a.

6.8. Atac stereotip asupra cifrurilor de substituţie

Fie $t_1, \ldots, t_M$ mesajul cifrat de lungime M (suficient de mare). Notăm prin $c_1, \ldots, c_N$ literele care apar în mesajul cifrat şi prin $p_1, \ldots, p_N$ frecvenţa de apariţie a acestora. Presupunem că sistemul de cifrare folosit este o substituţie monoalfabetică. Notăm prin $c_1^*, \ldots, c_N^*$ alfabetul textului clar şi prin $p_1^*, \ldots, p_N^*$ probabilitatea apariţiei literelor alfabetului. O metodă de rezolvare a acestui sistem de cifrare este atacul brut care constă în generarea tuturor permutărilor alfabetului. Numărul total de permutări ce trebuiesc generate este de $N!$ care este imens şi deci trebuie găsită o altă cale de atac. Tehnica este de a diviza mulţimea literelor din textul cifrat şi din alfabetul clar în K clase astfel încât dispersia din fiecare clasă este mică. Notăm prin m_i numărul de litere din cea de a $i-a$ clasă corespunzătoare textului cifrat şi prin n_i numărul de litere din cea de a $i-a$ clasă corespunzătoare alfabetului clar. Notăm cu $k_i = \max(m_i; n_i)$ şi $l_i = \min(m_i; n_i)$. Astfel trebuie generate numai $\prod_{i=1}^{K} A_{k_i}^{l_i}$ (avem $\sum_{i=1}^{K} n_i = \sum_{i=1}^{K} m_i = N$) număr relativ mic în raport cu $N!$. Pentru a realiza această clasificare vom proceda de exemplu pentru literele $c_1, \ldots, c_N$ după cum urmează:

1) Sortăm literele $c_1, \ldots, c_N$ în ordinea crescătoare a probabilităţilor $p_1, \ldots, p_N$. Notăm prin $\bar{c}_1, \ldots, \bar{c}_N$ şi prin $\bar{p}_1, \ldots, \bar{p}_N$ literele sortate, respectiv probabilităţile sortate.

2) Calculăm $\delta_i = \bar{p}_i - \bar{p}_{i-1}$ $(i = 2, \ldots, n)$. Cele mai mari $K-1$ valori ale lui δ_i sunt puncte delimitatoare ale claselor.

Observaţia 6.8.1. Pentru texte de lungime mică, în loc de probabilitatea p_i se foloseşte f_i frecvenţa absolută a literei i.

6.9. Atac de tip frecvenţă maximă a cifrurilor de substituţie

Fie $t_1, \ldots, t_M$ text clar de lungime M cu $t_i \in \{0, \ldots, 2^L - 1\}$ (textul este codificat pe L-biţi). Notăm prin $\sigma \in S_{2^L}$ cheia (permutarea) şi prin $c_1, \ldots, c_M$ textul permutat (cifrat). Energia informaţională a textului este:

$$\sum_{i=0}^{2^L-1} \left(\frac{f_i}{M}\right)^2$$

unde f_i este frecvenţa de apariţie a caracterului i ($i = 0, \ldots, 2^L - 1$). Utilizând inegalitatea Cauchy-Buniakovsky-Schwartz obţinem:

$$\sum_{i=0}^{2^L-1} \frac{f_i}{M} p_{\sigma_i} \leq \sqrt{\sum_{i=0}^{2^L-1} \left(\frac{f_i}{M}\right)^2 \sum_{i=0}^{2^L-1} p_{\sigma_i}^2},$$

unde p_{σ_i} este probabilitatea de apariţie a caracterului i în textul T. Deoarece

$$\sum_{i=0}^{2^L-1} \left(\frac{f_i}{M}\right)^2 = \sum_{i=0}^{2^L-1} p_{\sigma_i}^2,$$

vom obţine

$$\sum_{i=0}^{2^L-1} \frac{f_i}{M} p_{\sigma_i} \leq \sum_{i=0}^{2^L-1} \left(\frac{f_i}{M}\right)^2.$$

Avem egalitate dacă şi numai dacă:

$$p_{\sigma_i} = \frac{f_i}{M} \text{ pentru orice } i.$$

Deci pentru a obţine permutarea optimă σ (cheia sistemului de cifrare) trebuie rezolvată problema următoare:

$$\begin{cases} \max\limits_{\sigma} \sum\limits_{i=0}^{2^L-1} \frac{f_i}{M} \hat{p}_{\sigma_i} \\ \sum\limits_{i=0}^{2^L-1} \frac{f_i}{M} = 1, \end{cases}$$

unde $\hat{p}_{\sigma_i}$ este un estimator al lui p_{σ_i}.

Observaţia 6.9.1. Această procedură se poate folosi şi ca un test de confirmare a rezultatului obţinut prin atacul de tip stereotip.

6.10. Concluzii

Funcţiile de test definite în cadrul acestui capitol sunt invariante în cazul folosirii anumitor sisteme de cifrare cum ar fi sistemele de cifrare de tip transpoziţie şi de substituţie mono sau polialfabetică. Aceste funcţii de test pot fi folosite în identificarea limbii folosite în textul clar, a sistemului de cifrare folosit, a cheii şi în cele din urmă a textului clar. Funcţia *Kappa* este folosită în identificarea sistemului de cifrare şi în identificarea limbii folosite în textul clar. Procedura de identificare a

limbii se bazează pe compararea valorilor lui Psi şi Phi ale textului cifrat cu valorile lui Psi şi Phi ale fiecărei limbi (aceste funcţii de test devin în acest moment teste de confirmare). Funcţiile Chi sunt folosite în atacuri de *tip isolog* (texte diferite cifrate cu aceeiaşi cheie). Testele $Kappa, Chi, Psi$ şi Phi se pot efctua pe digrame sau trigrame, etc.

Alte aplicaţii criptografice sunt cele de tip căutare inteligentă a parolelor sau de tip dicţionar modificat (conform unor reguli lexicale).

Cele mai bune rezultate sunt obţinute dacă se analizează textul format numai din litere mari sau mici. Textele pot fi cu delimitator de cuvânt (spaţiu) sau fără delimitator de cuvânt.

Descrierea completă a acestor tipurilor de cifruri prezentate în cadrul acestui capitol precum şi principalele moduri de atac se poate găsi de exemplu în *Bauer* [6].

6.11. Aplicaţii

Exerciţiul 6.11.1. Care este diferenţa dintre proprietatea de confuzie şi proprietatea de difuzie a unui algoritm de cifrare?

Exerciţiul 6.11.2. Să se construiască alfabetul cifrat cu ajutorul parolei de cifrare $TESTARESISTEM$ iar apoi să se cifreze mesajul "IN CRIPTOGRAFIE NICI O REGULA NU ESTE ABSOLUTA". Permutarea ce realizează corespondenţa este:

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

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

Răspuns. Corepondenţa dintre alfabetul clar şi alfabetul cifrat (înainte de realizarea permutării) este:

A	B	C	D	E	F	G	H	I	J	K	L	M
T	E	S	A	R	I	M	B	C	D	F	G	H

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
J	K	L	N	O	P	Q	U	V	W	X	Y	Z

Corepondenţa dintre alfabetul clar şi alfabetul cifrat după realizarea permutării este:

A	B	C	D	E	F	G	H	I	J	K	L	M
Z	Y	X	W	V	U	Q	P	O	N	L	K	J

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
H	G	F	D	C	B	M	I	R	A	S	E	T

Mesajul clar se procesează astfel încât spaţiul este înlocuit cu cea mai puţin frecventă literă:

$$INQCRIPTOGRAFIEQNICIQREGULAQNUQESTEQABSOLUTA.$$

Mesajul cifrat va fi:

$$OHDXCOFMGQCZUOVDHOXODCVQIKZDHIDVBMVDZYBGKIMZ.$$

Exerciţiul 6.11.3. Să se cifreze mesajul "SI IN CRIPTOGRAFIE TACEREA ESTE AUR" cu ajutorul metodei celor 4 tabele iniţializate cu parolele de cifrare $CRIPTOGRAFIE$ şi $TEST$.

Exerciţiul 6.11.4. Să se construiască matricea de cifrare Playfair cu ajutorul parolei $CRIPTOGRAFIE$ iar apoi să se cifreze mesajul "SI IN CRIPTOGRAFIE TACEREA ESTE AUR".

Răspuns. Matricea Playfair se obţine trecând literele din parolă o singură dată în careul de 5×5, iar apoi celelalte litere ale alfabetului în ordinea lexicografică:

C	R	I/J	P	T
O	G	A	F	E
B	D	H	K	L
M	N	Q	S	U
V	W	X	Y	Z

Mesajul este preprocesat prin introducerea literei Q ca delimitator de cuvânt, adjuncţionându-se la finalul mesajului (pentru ca acesta să aibă lungime pară) litera Q:

$$SIQINQCRIPTOGRAFIEQTACEREAQESTEQAURQ.$$

Respectând regulile de cifrare Playfair mesajul cifrat devine:

$$QPXAQSRIPTCEDGFETAUIOIGTOFUAUPAUEQIN.$$

Exerciţiul 6.11.5. Să se cifreze prin metoda transpoziţiei ($N = 12$), pornind de la parola $CRIPTOGRAFIE$ mesajul "SI IN CRIPTOGRAFIE TACEREA ESTE AUR".

Răspuns. Vom construi secvenţa numerică de cifrare asociind fiecărei litere din parolă indicele din ordinea lexicografică: astfel literele din parolă, scrise în ordine lexicografică sunt:

1	2	3	4	5	6	7	8	9	10	11	12
A	C	E	F	G	I	I	O	P	R	R	T

deci parola $CRIPTOGRAFIE$ produce permutarea: 2 10 6 9 12 8 5 11 1 4 7 3.

Textul clar este scris într-o tabelă cu 12 coloane:

2	10	6	9	12	8	5	11	1	4	7	3
S	I	Q	I	N	Q	C	R	I	P	T	O
G	R	A	F	I	E	Q	T	A	C	E	R
E	A	Q	E	S	T	E	Q	A	U	R	Q

Deoarece lungimea textului nu este divizibilă cu 12 vom completa ultimul rând cu o secvenţă cunoscută (în acest caz caracterul Q). Textul cifrat se obţine citind coloanele tabelei de cifrare în ordinea indicată de parola numerică:

$$IAASGEORRQPCUCQEQAQTERQETIFEIRARTQNIS.$$

Descifrarea se va realiza în mod similar folosind permutarea inversă σ^{-1}.

Dacă dimensiunea transpoziţiei N este mai mică decât lungimea parolei atunci se vor reţine N caractere din parolă.

Exerciţiul 6.11.6. Studiaţi comutativitatea operatorilor de cifrare substituţie mono/polialfabetică şi a transpoziţiei.

Exerciţiul 6.11.7. Implementaţi algoritmul de decriptare a unei transpoziţii.

Exerciţiul 6.11.8. Implementaţi algoritmul de decriptare a unei subtituţii simple.

Exerciţiul 6.11.9. Implementaţi algoritmul de decriptare a unui cifru obţinut prin compunerea unei transpoziţii şi a unei substituţii simple.

Exerciţiul 6.11.10. Implementaţi algoritmul de decriptare a unui cifru polialfabetic.