Sisteme de criptare polialfabetice

Prof. Dr. Adrian Atanasiu

February 13, 2011

Substituția unui caracter variază în text, în funcție de diverși parametri (poziție, context etc.).

Substituția unui caracter variază în text, în funcție de diverși parametri (poziție, context etc.).

Se pare că primul sistem de criptare poli-alfabetic a fost creat de Leon Battista în 1568.

Sistemul homofonic

Intermediar între sistemele mono și cele poli-alfabetice.
Principalul lui scop este de a evita atacul prin frecvența de apariție a caracterelor.

Sistemul homofonic

Intermediar între sistemele mono și cele poli-alfabetice.
Principalul lui scop este de a evita atacul prin frecvența de apariție a caracterelor.
Fiecărui caracter $a \in \mathcal{P}$ i se asociază o mulțime $H(a) \subset \mathcal{C}$ astfel încât:

1. $H(a) \cap H(b) = \emptyset \iff a \neq b$;

Sistemul homofonic

Intermediar între sistemele mono și cele poli-alfabetice.
Principalul lui scop este de a evita atacul prin frecvența de apariție a caracterelor.
Fiecărui caracter $a \in \mathcal{P}$ i se asociază o mulțime $H(a) \subset \mathcal{C}$ astfel încât:

1. $H(a) \cap H(b) = \emptyset \iff a \neq b$;
2. Dacă a apare mai frecvent în textele clare, atunci $card((H(a)) \geq card(H(b))$.

Sistemul homofonic

Intermediar între sistemele mono și cele poli-alfabetice.
Principalul lui scop este de a evita atacul prin frecvența de apariție a caracterelor.
Fiecărui caracter $a \in \mathcal{P}$ i se asociază o mulțime $H(a) \subset \mathcal{C}$ astfel încât:

1. $H(a) \cap H(b) = \emptyset \iff a \neq b$;
2. Dacă a apare mai frecvent în textele clare, atunci $card((H(a)) \geq card(H(b))$.

Criptarea unui caracter $a \in \mathcal{P}$ se face cu un element ales aleator din $H(a)$.
Pentru decriptarea lui $y \in \mathcal{C}$ se caută o mulțime $H(a)$ astfel ca $y \in H(a)$.

Exemplu

Să considerăm $\mathcal{P}=\{a,b\}$ *și*

$$H(a)=\{001,010\},\ H(b)=\{000,011,101,111\}.$$

Exemplu

Să considerăm $\mathcal{P}=\{a,b\}$ *și*

$$H(a)=\{001,010\},\ H(b)=\{000,011,101,111\}.$$

Pentru criptarea textului ab se poate folosi oricare din secvențele

$$001000, 001011, 001101, 001111,$$
$$010000, 010011, 010101, 010111$$

Exemplu

Să considerăm $\mathcal{P}=\{a,b\}$ *și*

$$H(a)=\{001,010\},\ H(b)=\{000,011,101,111\}.$$

Pentru criptarea textului ab se poate folosi oricare din secvențele

$$001000, 001011, 001101, 001111,$$
$$010000, 010011, 010101, 010111$$

Sistemul homofonic este mult mai rezistent la un atac bazat numai pe textul criptat, dar cedează ușor la un atac cu text clar ales.

Sistemul Playfair

Inventat în 1854 de Sir Charles Wheatstone. Promovat și susținut (pentru a fi adoptat ca cifru oficial al Marii Britanii) de către baronul Lyon Playfair de St. Andrews.

Sistemul Playfair

Inventat în 1854 de Sir Charles Wheatstone. Promovat și susținut (pentru a fi adoptat ca cifru oficial al Marii Britanii) de către baronul Lyon Playfair de St. Andrews.
Din cele 26 litere ale alfabetului se elimină una de frecvență minimă; alegem Q.
Restul literelor se aranjează arbitrar sub forma unui pătrat 5×5.

Sistemul Playfair

Inventat în 1854 de Sir Charles Wheatstone. Promovat și susținut (pentru a fi adoptat ca cifru oficial al Marii Britanii) de către baronul Lyon Playfair de St. Andrews.
Din cele 26 litere ale alfabetului se elimină una de frecvență minimă; alegem Q.
Restul literelor se aranjează arbitrar sub forma unui pătrat 5×5.
Construcția va fi exemplificată folosind tabloul:

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

Acest tabel va forma atât cheia de criptare cât și cea de decriptare.

Reguli de criptare/decriptare:

- Textul clar este separat în blocuri de câte două caractere (ignorând spațiile).

Reguli de criptare/decriptare:

- Textul clar este separat în blocuri de câte două caractere (ignorând spațiile).
  Nici un bloc să nu conțină aceiași literă, iar textul să fie de lungime pară.

Reguli de criptare/decriptare:

- Textul clar este separat în blocuri de câte două caractere (ignorând spațiile).
  Nici un bloc să nu conțină aceiași literă, iar textul să fie de lungime pară.
- Criptarea unui bloc: dacă cele două litere nu sunt plasate în tabel pe aceeași linie sau coloană, se cercetează colțurile dreptunghiului determinat de cele două litere.

Reguli de criptare/decriptare:

- Textul clar este separat în blocuri de câte două caractere (ignorând spațiile).
  Nici un bloc să nu conțină aceiași literă, iar textul să fie de lungime pară.
- Criptarea unui bloc: dacă cele două litere nu sunt plasate în tabel pe aceeași linie sau coloană, se cercetează colțurile dreptunghiului determinat de cele două litere.
  Perechea este criptată în celelalte două colțuri opuse.

Reguli de criptare/decriptare:

- Textul clar este separat în blocuri de câte două caractere (ignorând spațiile).
  Nici un bloc să nu conțină aceiași literă, iar textul să fie de lungime pară.
- Criptarea unui bloc: dacă cele două litere nu sunt plasate în tabel pe aceeași linie sau coloană, se cercetează colțurile dreptunghiului determinat de cele două litere.
  Perechea este criptată în celelalte două colțuri opuse.
  Dacă cele două litere se găsesc pe aceeași linie (coloană), se merge ciclic cu o poziție la dreapta (respectiv jos).

Reguli de criptare/decriptare:

- Textul clar este separat în blocuri de câte două caractere (ignorând spațiile).
  Nici un bloc să nu conțină aceiași literă, iar textul să fie de lungime pară.
- Criptarea unui bloc: dacă cele două litere nu sunt plasate în tabel pe aceeași linie sau coloană, se cercetează colțurile dreptunghiului determinat de cele două litere.
  Perechea este criptată în celelalte două colțuri opuse.
  Dacă cele două litere se găsesc pe aceeași linie (coloană), se merge ciclic cu o poziție la dreapta (respectiv jos).
  Deci CA se criptează în AX, WX în UG, CA în AX etc.

Reguli de criptare/decriptare:

- Textul clar este separat în blocuri de câte două caractere (ignorând spațiile).
  Nici un bloc să nu conțină aceiași literă, iar textul să fie de lungime pară.
- Criptarea unui bloc: dacă cele două litere nu sunt plasate în tabel pe aceeași linie sau coloană, se cercetează colțurile dreptunghiului determinat de cele două litere.
  Perechea este criptată în celelalte două colțuri opuse.
  Dacă cele două litere se găsesc pe aceeași linie (coloană), se merge ciclic cu o poziție la dreapta (respectiv jos).
  Deci *CA* se criptează în *AX*, *WX* în *UG*, *CA* în *AX* etc.

$$\text{AFARA PLOUA} \longrightarrow \text{XHHPPDPEPX.}$$

Proprietăți:

O permutare ciclică a liniilor și coloanelor nu modifică criptarea. De exemplu, pătratul

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

este echivalent cu cel inițial (ambele dau aceiași cheie de criptare).

Proprietăți:

O permutare ciclică a liniilor și coloanelor nu modifică criptarea.
De exemplu, pătratul

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

este echivalent cu cel inițial (ambele dau aceiași cheie de criptare).
Regulile de bază pot fi modificate sau completate după necesități.
Astfel, se poate adăuga din loc în loc câte o literă falsă (cu frecvență foarte redusă) care să modifice textul criptat.
Pătratul 5×5 poate fi înlocuit cu un dreptunghi 4×6 sau 3×8.

Securitate

Pentru a păstra cheia în siguranță, se recomandă memorarea acesteia.

Securitate

Pentru a păstra cheia în siguranță, se recomandă memorarea acesteia.

Cum o astfel de cheie este extrem de greu de memorat, se folosește un cuvânt cheie sau o propoziție cu toate literele distincte.

Acesta cuvânt este scris la începutul tabloului. Spațiile rămase

sunt completate cu restul literelor alfabetului, scrise în ordinea apariției lor.

În preajma primului război mondial, armata română folosea un dreptunghi 3×8 din care lipseau literele Q și K.
Cuvântul cheie era *ROMANESC*. Un astfel de tablou putea avea de exemplu forma

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

În preajma primului război mondial, armata română folosea un dreptunghi 3×8 din care lipseau literele Q și K.
Cuvântul cheie era *ROMANESC*. Un astfel de tablou putea avea de exemplu forma

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

Playfair rezistă la atacuri bazate pe frecvența apariției, dar nu și la cele prin text clar ales.

În preajma primului război mondial, armata română folosea un dreptunghi 3×8 din care lipseau literele Q și K.
Cuvântul cheie era *ROMANESC*. Un astfel de tablou putea avea de exemplu forma

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

Playfair rezistă la atacuri bazate pe frecvența apariției, dar nu și la cele prin text clar ales.
Implementări actuale folosesc o reprezentare binară a literelor; în plus, după ce s-a obținut o pereche criptată, aceasta se combină printr-un *XOR* cu perechea criptată anterior.

Sistemul Vigenere

Numele vine de la baronul francez Blaise de Vigenere (1523 – 1596) diplomat la curtea regelui Henry III (alte surse indică drept real inventator al sistemului pe Giovan Batista Belaso în 1553).

Sistemul Vigenere

Numele vine de la baronul francez Blaise de Vigenere (1523 – 1596) diplomat la curtea regelui Henry III (alte surse indică drept real inventator al sistemului pe Giovan Batista Belaso în 1553).

Reamintim codificarea:

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Sistemul Vigenere

Numele vine de la baronul francez Blaise de Vigenere (1523 – 1596) diplomat la curtea regelui Henry III (alte surse indică drept real inventator al sistemului pe Giovan Batista Belaso în 1553).

Reamintim codificarea:

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

Definim

$$\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26},\ \mathcal{K} = \mathbb{Z}_{26}^{+}$$

O cheie $K \in \mathcal{K}$ este un cuvânt având codificarea numerică

$$k_0 k_1 \ldots k_{p-1}$$

O cheie $K \in \mathcal{K}$ este un cuvânt având codificarea numerică

$$k_0 k_1 \dots k_{p-1}$$

Fie

$$\mathbf{a} = a_0 a_1 \dots a_n$$

codificarea textului clar care trebuie transmis.

O cheie $K \in \mathcal{K}$ este un cuvânt având codificarea numerică

$$k_0 k_1 \dots k_{p-1}$$

Fie

$$\mathbf{a} = a_0 a_1 \dots a_n$$

codificarea textului clar care trebuie transmis.
Textul criptat va fi

$$e_K(\mathbf{a}) = \mathbf{x} = x_0 x_1 \dots x_n$$

unde

$$x_i = a_i + k_{i\ (mod\ p)} \quad (mod\ 26)$$

Exemplu

Fie cuvântul cheie *FOCAR*; deci $p = 5$ și $K = 5\ 14\ 2\ 0\ 17$.

Exemplu

Fie cuvântul cheie *FOCAR*; deci $p = 5$ și $K = 5\ 14\ 2\ 0\ 17$.
Să criptăm cheie textul clar

NU POT VENI AZI

Codificarea sa este

$$\mathbf{a} = 13\ 20\ 15\ 14\ 19\ 21\ 4\ 13\ 8\ 0\ 25\ 8.$$

Exemplu

Fie cuvântul cheie *FOCAR*; deci $p = 5$ și $K = 5\ 14\ 2\ 0\ 17$.
Să criptăm cheie textul clar

NU POT VENI AZI

Codificarea sa este

$$\mathbf{a} = 13\ 20\ 15\ 14\ 19\ 21\ 4\ 13\ 8\ 0\ 25\ 8.$$

Sub fiecare număr din **a** se așează câte un număr din K; când cheia se termină, ea se reia ciclic, până se termină **a**.
Deci vom avea

13	20	15	14	19	21	4	13	8	0	25	8
5	14	2	0	17	5	14	2	0	17	5	14
18	8	17	14	10	0	18	15	8	17	4	22
S	*I*	*R*	*O*	*K*	*A*	*S*	*P*	*I*	*R*	*E*	*W*

Exemplu

Fie cuvântul cheie *FOCAR*; deci $p = 5$ și $K = 5\ 14\ 2\ 0\ 17$.
Să criptăm cheie textul clar

NU POT VENI AZI

Codificarea sa este

$$\mathbf{a} = 13\ 20\ 15\ 14\ 19\ 21\ 4\ 13\ 8\ 0\ 25\ 8.$$

Sub fiecare număr din **a** se așează câte un număr din K; când cheia se termină, ea se reia ciclic, până se termină **a**.
Deci vom avea

13	20	15	14	19	21	4	13	8	0	25	8
5	14	2	0	17	5	14	2	0	17	5	14
18	8	17	14	10	0	18	15	8	17	4	22
S	*I*	*R*	*O*	*K*	*A*	*S*	*P*	*I*	*R*	*E*	*W*

Decriptarea se realizează similar, scăzând (modulo 26) din codul caracterului criptat, codul caracterului corespunzător din cheie.

Sistemul Beaufort

Este o variantă a sistemului Vigenere. Relația de criptare este

$$x_i = k_{i\ (mod\ p)} - a_i \quad (mod\ 26), \qquad (i \geq 0)$$

Sistemul Beaufort

Este o variantă a sistemului Vigenere. Relația de criptare este

$$x_i = k_{i\ (mod\ p)} - a_i \quad (mod\ 26), \qquad (i \geq 0)$$

Avantajul sistemului Beaufort constă în faptul că ecuația de criptare se aplică și la decriptare

$$a_i = k_{i\ (mod\ p)} - x_i$$

Sistemul Beaufort este "**auto-dual**".

Sistemul Autoclave

Atribuit matematicianului Cardano.

Sistemul Autoclave

Atribuit matematicianului Cardano.
Aici cheia se folosește doar pe primele poziții, după care este utilizat drept cheie textul clar.

Exemplu

Cuvântul cheie COVOR și textul clar A VENIT TOAMNA

Text clar:	*A*	*V*	*E*	*N*	*I*	*T*	*T*	*O*	*A*	*M*	*N*	*A*
Cheie:	*C*	*O*	*V*	*O*	*R*	*A*	*V*	*E*	*N*	*I*	*T*	*T*
Text criptat	*C*	*J*	*Z*	*B*	*Z*	*T*	*O*	*S*	*N*	*U*	*G*	*T*

Securitate

Sistemul Vigenere a fost considerat unul din cele mai sigure sisteme de criptare.
În 1917 de exemplu, revista "Scientific American" îl considera imposibil de atacat.

Securitate

Sistemul Vigenere a fost considerat unul din cele mai sigure sisteme de criptare.
În 1917 de exemplu, revista "Scientific American" îl considera imposibil de atacat.
Numai că acest sistem a fost spart de Kasiski încă din 1863 (și independent de Babbage în 1854).

Criptanaliza sistemului Vigenere

Fie

$$\mathbf{x} = x_0 x_1 \dots x_{n-1}$$

textul criptat cu cheia

$$K = k_0 k_1 \dots k_{p-1}.$$

Criptanaliza sistemului Vigenere

Fie

$$\mathbf{x} = x_0 x_1 \dots x_{n-1}$$

textul criptat cu cheia

$$K = k_0 k_1 \dots k_{p-1}.$$

El se poate aranja ca o matrice cu p linii și $\lceil n/p \rceil$ coloane:

$$\begin{array}{cccc} x_0 & x_p & x_{2p} \dots & \\ x_1 & x_{p+1} & x_{2p+1} & \dots \\ & & \vdots & \\ x_{p-1} & x_{2p-1} & x_{3p-1} & \dots \end{array}$$

Elementele de pe prima linie au fost criptate după formula

$$x_{pr} = a_{pr} + k_0 \ (mod\ 26), \quad (r \geq 0)$$

adică folosind un sistem Cezar (k_0 fiind o valoare fixată din $\mathbb{Z}_{26}$). În mod similar și celelalte linii.

Elementele de pe prima linie au fost criptate după formula

$$x_{pr} = a_{pr} + k_0 \ (mod\ 26), \quad (r \geq 0)$$

adică folosind un sistem Cezar (k_0 fiind o valoare fixată din $\mathbb{Z}_{26}$).
În mod similar și celelalte linii.
Dacă se cunoaște lungimea p a cheii, problema se reduce la criptanaliza a p texte criptate cu Cezar – sistem de criptare monoalfabetic.

Elementele de pe prima linie au fost criptate după formula

$$x_{pr} = a_{pr} + k_0 \ (mod\ 26), \quad (r \geq 0)$$

adică folosind un sistem Cezar (k_0 fiind o valoare fixată din $\mathbb{Z}_{26}$).
În mod similar și celelalte linii.
Dacă se cunoaște lungimea p a cheii, problema se reduce la criptanaliza a p texte criptate cu Cezar – sistem de criptare monoalfabetic.
Sunt cunoscute două metode pentru aflarea lungimii cheii:

- *testul lui Kasiski*

Elementele de pe prima linie au fost criptate după formula

$$x_{pr} = a_{pr} + k_0 \ (mod\ 26), \quad (r \geq 0)$$

adică folosind un sistem Cezar (k_0 fiind o valoare fixată din $\mathbb{Z}_{26}$).
În mod similar și celelalte linii.
Dacă se cunoaște lungimea p a cheii, problema se reduce la criptanaliza a p texte criptate cu Cezar – sistem de criptare monoalfabetic.
Sunt cunoscute două metode pentru aflarea lungimii cheii:

- *testul lui Kasiski*
- *indexul de coincidențe* (definit în 1920 de criptanalistul american Wolfe Friedman).

Testul Kasiski

Constă în studiul textului criptat și aflarea de perechi de segmente de cel puțin 3 caractere identice.
Pentru fiecare astfel de pereche, se determină distanța dintre segmente.

Testul Kasiski

Constă în studiul textului criptat și aflarea de perechi de segmente de cel puțin 3 caractere identice.
Pentru fiecare astfel de pereche, se determină distanța dintre segmente.
După ce s-au găsit mai multe astfel de distanțe, valoarea lui p va fi cel mai mare divizor comun al lor (sau – eventual un divizor al acestuia).

Exemplu

Oscar interceptează următorul text criptat, despre care bănuie că s-a folosit Vigenere:

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

Oscar găsește secvența *QLNCJ* care apare de două ori, având distanța 27.

Oscar găsește secvența *QLNCJ* care apare de două ori, având distanța 27.
De asemenea, apar două cuvinte foarte asemănătoare: *AQLN* și *AOLN*, la distanța 57 unul de altul.

Oscar găsește secvența *QLNCJ* care apare de două ori, având distanța 27.
De asemenea, apar două cuvinte foarte asemănătoare: *AQLN* și *AOLN*, la distanța 57 unul de altul.
Deci putem bănui că avem de-a face cu un cuvânt cheie de lungime $cmmdc(27, 57) = 3$.

Rescriem textul pe coloane, fiecare coloană având trei elemente.

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

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

Numărăm frecvența apariției literelor pe fiecare linie:

	A	B	C	D	E	F	G	H	I	J	K	L	M
I	2	0	6	1	0	1	3	2	2	1	0	0	0
II	0	0	1	2	4	0	3	0	1	3	6	3	0
III	11	0	3	0	3	0	1	0	2	0	0	2	2

	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
I	0	3	1	3	2	7	0	0	1	8	0	0	0
II	0	0	0	0	4	0	2	0	6	2	0	1	3
III	3	1	1	0	3	2	3	4	0	0	0	0	1

În limba română, primele litere ca frecvență sunt $A - E - I$, aflate la distanță egală una de alta.

În limba română, primele litere ca frecvență sunt $A - E - I$, aflate la distanță egală una de alta.
Căutăm pe fiecare linie triplete de litere situate pe pozițiile

$$(k, k+4, k+8)$$

având o frecvență semnificativ de mare.

În limba română, primele litere ca frecvență sunt $A - E - I$, aflate la distanță egală una de alta.
Căutăm pe fiecare linie triplete de litere situate pe pozițiile

$$(k, k+4, k+8)$$

având o frecvență semnificativ de mare.

- Pentru linia 3 tripletul este chiar $A - E - I$ (16 apariții din 49 posibile); deplasare 0 în codul Cezar.

În limba română, primele litere ca frecvență sunt $A - E - I$, aflate la distanță egală una de alta.
Căutăm pe fiecare linie triplete de litere situate pe pozițiile

$$(k, k+4, k+8)$$

având o frecvență semnificativ de mare.

- Pentru linia 3 tripletul este chiar $A - E - I$ (16 apariții din 49 posibile); deplasare 0 în codul Cezar.
- Pentru prima linie, sunt două posibilități: $O - S - W$ (deplasare 14) sau $S - W - A$ (deplasare 18), ambele cu câte 18 apariții.

În limba română, primele litere ca frecvență sunt $A - E - I$, aflate la distanță egală una de alta.
Căutăm pe fiecare linie triplete de litere situate pe pozițiile

$$(k, k+4, k+8)$$

având o frecvență semnificativ de mare.

- Pentru linia 3 tripletul este chiar $A - E - I$ (16 apariții din 49 posibile); deplasare 0 în codul Cezar.
- Pentru prima linie, sunt două posibilități: $O - S - W$ (deplasare 14) sau $S - W - A$ (deplasare 18), ambele cu câte 18 apariții.
- Pentru linia a doua, tot două variante: $C - G - K$ (deplasare 2) cu 10 apariții, sau $R - V - Z$ (deplasare 14) cu 13 apariții.

Deplasările dau exact codificările cheii.

Deplasările dau exact codificările cheii.
Deci trebuie luate în considerare patru variante de cuvânt cheie:

$$OCA,\ ORA,\ SCA,\ SRA$$

Deplasările dau exact codificările cheii.
Deci trebuie luate în considerare patru variante de cuvânt cheie:

$$OCA,\ ORA,\ SCA,\ SRA$$

O simplă verificare reține drept cuvânt cheie *ORA*, care conduce la decriptarea corectă a textului (spațiile și semnele de punctuație se pun ulterior):

PELANGAPLOPIIFARASOTADESEAAMTRECUT
MACUNOSTEAUVECINIITOTITUNUMAICUNOSCUT
ACEASTAESTEPRIMASTROFAAUNEINPOEZIICELEBRE
DEMIHAIEMINESCU

Indexul de coincidențe

A doua metodă de aflare a lungimii cheii de criptare într-un sistem Vigenere se bazează pe un concept definit de Wolfe Friedman în 1920: *indexul de coincidențe*.

Indexul de coincidențe

A doua metodă de aflare a lungimii cheii de criptare într-un sistem Vigenere se bazează pe un concept definit de Wolfe Friedman în 1920: *indexul de coincidențe*.

Definiție

Dacă $\mathbf{x} = x_1 x_2 \dots x_n$ *este o secvență de n caractere alfabetice, se numește "index de coincidențe" al lui* $\mathbf{x}$ *probabilitatea ca două caractere din* $\mathbf{x}$*, alese aleator, să fie identice.*
Această valoare se notează $I_c(\mathbf{x})$.

Fie f_i frecvența de apariție în $\mathbf{x}$ a caracterului literal codificat i.

Fie f_i frecvența de apariție în $\mathbf{x}$ a caracterului literal codificat i.
Două litere din $\mathbf{x}$ pot fi alese în C_n^2 moduri.
Din acestea, sunt $C_{f_i}^2$ moduri ca ambele să aibă aceiași codificare i.
Deci

$$I_c(\mathbf{x}) = \frac{\sum_{i=0}^{25} C_{f_i}^2}{C_n^2} = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)}$$

Fie f_i frecvența de apariție în $\mathbf{x}$ a caracterului literal codificat i.
Două litere din $\mathbf{x}$ pot fi alese în C_n^2 moduri.
Din acestea, sunt $C_{f_i}^2$ moduri ca ambele să aibă aceiași codificare i.
Deci

$$I_c(\mathbf{x}) = \frac{\sum_{i=0}^{25} C_{f_i}^2}{C_n^2} = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)}$$

Să presupunem că $\mathbf{x}$ este un text în limba română.

Fie f_i frecvența de apariție în $\mathbf{x}$ a caracterului literal codificat i.
Două litere din $\mathbf{x}$ pot fi alese în C_n^2 moduri.
Din acestea, sunt $C_{f_i}^2$ moduri ca ambele să aibă aceiași codificare i.
Deci

$$I_c(\mathbf{x}) = \frac{\sum_{i=0}^{25} C_{f_i}^2}{C_n^2} = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)}$$

Să presupunem că $\mathbf{x}$ este un text în limba română.
Din tabelul frecvențelor de apariție ale literelor, notând p_i probabilitatea de apariție a caracterului codificat cu i $(0 \leq i \leq 25)$, valoarea pe care o putem estima pentru indexul de coincidențe este

$$I_c(\mathbf{x}) \approx \sum_{i=0}^{25} p_i^2 = 0,0788$$

Fie f_i frecvența de apariție în $\mathbf{x}$ a caracterului literal codificat i.
Două litere din $\mathbf{x}$ pot fi alese în C_n^2 moduri.
Din acestea, sunt $C_{f_i}^2$ moduri ca ambele să aibă aceiași codificare i.
Deci

$$I_c(\mathbf{x}) = \frac{\sum_{i=0}^{25} C_{f_i}^2}{C_n^2} = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)}$$

Să presupunem că $\mathbf{x}$ este un text în limba română.
Din tabelul frecvențelor de apariție ale literelor, notând p_i probabilitatea de apariție a caracterului codificat cu i $(0 \leq i \leq 25)$, valoarea pe care o putem estima pentru indexul de coincidențe este

$$I_c(\mathbf{x}) \approx \sum_{i=0}^{25} p_i^2 = 0,0788$$

Afirmație valabilă pentru orice criptare monoalfabetică.

Fie textul criptat $\mathbf{x} = x_0 x_1 \dots x_{n-1}$ aranjat într-o matrice $p \times \lceil n/p \rceil$ (p este un număr întreg pozitiv arbitrar), astfel

$$\begin{array}{rccccc} \mathbf{x_0} = & x_0 & x_p & x_{2p} \dots & \\ \mathbf{x_1} = & x_1 & x_{p+1} & x_{2p+1} & \dots \\ & & \vdots & & \\ \mathbf{x_{p-1}} = & x_{p-1} & x_{2p-1} & x_{3p-1} & \dots \end{array}$$

Fie textul criptat $\mathbf{x} = x_0 x_1 \dots x_{n-1}$ aranjat într-o matrice $p \times \lceil n/p \rceil$ (p este un număr întreg pozitiv arbitrar), astfel

$$\begin{array}{rcccc} \mathbf{x_0} = & x_0 & x_p & x_{2p} \dots & \\ \mathbf{x_1} = & x_1 & x_{p+1} & x_{2p+1} & \dots \\ & & \vdots & & \\ \mathbf{x_{p-1}} = & x_{p-1} & x_{2p-1} & x_{3p-1} & \dots \end{array}$$

Dacă p este lungimea cheii, atunci fiecare valoare $I_c(\mathbf{x_i})$ este apropiată de $0,0788$.

Fie textul criptat $\mathbf{x} = x_0 x_1 \dots x_{n-1}$ aranjat într-o matrice $p \times \lceil n/p \rceil$ (p este un număr întreg pozitiv arbitrar), astfel

$$\begin{array}{rcccc} \mathbf{x_0} = & x_0 & x_p & x_{2p} \dots & \\ \mathbf{x_1} = & x_1 & x_{p+1} & x_{2p+1} & \dots \\ & & \vdots & & \\ \mathbf{x_{p-1}} = & x_{p-1} & x_{2p-1} & x_{3p-1} & \dots \end{array}$$

Dacă p este lungimea cheii, atunci fiecare valoare $I_c(\mathbf{x_i})$ este apropiată de $0,0788$.
În caz contrar, șirul $\mathbf{x_i}$ va arăta mult mai aleator, fiind obținut prin amestecul unei secvențe de caractere criptate cu chei diferite.

Pentru o secvență complet aleatoare, valoarea indexului de coincidențe este

$$I_c \approx 26 \left(\frac{1}{26}\right)^2 = \frac{1}{26} = 0,0384$$

Valorile $0,0788$ și $0,0384$ vor constitui punctele de extrem pe care le poate lua I_c.

Pentru o secvență complet aleatoare, valoarea indexului de coincidențe este

$$I_c \approx 26 \left(\frac{1}{26}\right)^2 = \frac{1}{26} = 0,0384$$

Valorile $0,0788$ și $0,0384$ vor constitui punctele de extrem pe care le poate lua I_c.
Se iau diverse valori pentru p, până se găsește una care să se apropie cât mai mult de $0,788$ și nu de $0,384$.

Pentru o secvență complet aleatoare, valoarea indexului de coincidențe este

$$I_c \approx 26 \left(\frac{1}{26}\right)^2 = \frac{1}{26} = 0,0384$$

Valorile $0,0788$ și $0,0384$ vor constitui punctele de extrem pe care le poate lua I_c.
Se iau diverse valori pentru p, până se găsește una care să se apropie cât mai mult de $0,788$ și nu de $0,384$.
Aceea se poate considera – cu suficientă siguranță – lungimea cheii.

În etapa a doua, vom încerca să aflăm efectiv cheia $K = k_0 k_1 \dots k_{p-1}$.

În etapa a doua, vom încerca să aflăm efectiv cheia $K = k_0 k_1 \dots k_{p-1}$.
Dacă $n_1 = \lfloor n/p \rfloor$ este lungimea secvenței $\mathbf{x_i}$ $(0 \leq i < p)$, atunci distribuția de probabilitate ale celor 26 litere în $\mathbf{x_i}$ este

$$\frac{f_0}{n_1}, \frac{f_1}{n_1}, \dots, \frac{f_{25}}{n_1}$$

În etapa a doua, vom încerca să aflăm efectiv cheia $K = k_0 k_1 \dots k_{p-1}$.
Dacă $n_1 = \lfloor n/p \rfloor$ este lungimea secvenței $\mathbf{x_i}$ $(0 \leq i < p)$, atunci distribuția de probabilitate ale celor 26 litere în $\mathbf{x_i}$ este

$$\frac{f_0}{n_1}, \frac{f_1}{n_1}, \dots, \frac{f_{25}}{n_1}$$

Secvența $\mathbf{x_i}$ este obținută printr-o criptare Cezar cu o deplasare k_i.
Deci, situația ideală este când distribuția de probabilitate a deplasării

$$\frac{f_{k_i}}{n_1}, \frac{f_{k_i+1 \ (mod\ 26)}}{n_1}, \dots, \frac{f_{k_i+25 \ (mod\ 26)}}{n_1}$$

este cât mai apropiată de distribuția de probabilitate $p_0, p_1, \dots, p_{25}$ a limbii române.

Fie m $(0 \leq m \leq 25)$; definim expresia

$$F_m = \sum_{i=0}^{25} \frac{p_i \cdot f_{i+m}}{n_1}$$

Fie m $(0 \leq m \leq 25)$; definim expresia

$$F_m = \sum_{i=0}^{25} \frac{p_i \cdot f_{i+m}}{n_1}$$

Dacă $m = k_j$ $(0 \leq j \leq p-1)$, ne putem aștepta ca

$$F_m \approx \sum_{i=0}^{25} p_i^2 = 0,0788$$

Fie m $(0 \leq m \leq 25)$; definim expresia

$$F_m = \sum_{i=0}^{25} \frac{p_i \cdot f_{i+m}}{n_1}$$

Dacă $m = k_j$ $(0 \leq j \leq p-1)$, ne putem aștepta ca

$$F_m \approx \sum_{i=0}^{25} p_i^2 = 0,0788$$

Dacă $m \neq k_j$, atunci F_m va fi semnificativ mai mic decât această valoare.

Fie m $(0 \leq m \leq 25)$; definim expresia

$$F_m = \sum_{i=0}^{25} \frac{p_i \cdot f_{i+m}}{n_1}$$

Dacă $m = k_j$ $(0 \leq j \leq p-1)$, ne putem aștepta ca

$$F_m \approx \sum_{i=0}^{25} p_i^2 = 0,0788$$

Dacă $m \neq k_j$, atunci F_m va fi semnificativ mai mic decât această valoare.

Deci, după cel mult 25 încercări, se poate afla deplasarea k_j și deci a j-a literă din cheie.

Mulțumesc pentru atenție !