METODOLOGIA SPARGERII SISTEMELOR MONOALFABETICE DE CRIPTARE

Afanas Dorin, dr., conf. Univ. UST
Gorceag Alexandrina, student, UST
Șestacov Andrei, lector universitar, AMFA

Adnotare. În prezenta lucrare sunt cercetate cifrurile de substituție monoalfabetice de criptare: Cezar, afin și posibilitățile de spargere a lor, fiind structurată în: introducere, spargerea sistemelor de criptare monoalfabetice prin forță brută, spargerea sistemelor de criptare monoalfabetice prin analiza frecvenţei apariţiei literelor într-o limbă, bibliografie.

Cuvinte cheie: criptare, decriptare, cheie de criptare, cifrul Cezar, cifrul afin, spargerea sistemelor monoalfabetice de criptare.

Annotation. In this paper investigates monoalphabetic substitution ciphers encryption: Caesar, affine and breaking their possibilities is structured into: introduction, breaking encryption systems monoalfabetice brute force, breaking encryption systems monoalfabetice by analyzing the incidence of letters in a language, bibliography.

Keywords: encryption, decryption, encryption key, Caesar cipher, affine cipher, monoalphabetic breaking encryption systems.

1. Introducere

Necesitatea de a cripta informaţiile a existat dintotdeauna. Astăzi, mai mult ca niciodată, informaţii confidenţiale se transmit zilnic între diferite instituţii guvernamentale şi/sau companii. Pentru ca astfel de date confidenţiale să nu ajungă în mâna celor care le caută există criptografia.

Trecând prin procesul de criptare, mesajele sau informațiile sânt codificate cu ajutorul unui algoritm rezultând mesaje care nu pot fi decriptate fără el.

Aceasta se realizează de obicei utilizând o cheie de criptare, care specifică modul în care mesajele vor fi codificate. Astfel, cei care nu au acces la cheia de criptate nu pot descifra mesajul. Cel căruia îi sunt destinate aceste mesaje trebuie să aibă algoritmul de criptare pentru a le putea citi.

În trecut, criptarea era folosită de către armată şi guverne pentru a transmite informații secrete. În prezent, majoritatea companiilor folosesc criptarea pentru a proteja datele care sunt transmise atât intern cât şi extern. Criptarea poate fi utilizată pentru a proteja atât datele de pe unitățile de stocare interne cât și externe, astfel, în cazul pierderii sau furtului de informații să poată avea acces doar persoana autorizată, adică persoana care deține cheia de criptare. Sistemele de criptare sunt prezente în comerțul electronic, telefoane mobile, device-uri wireless, device-uri Bluetooth, bancomate și automate bancare.

Atunci când navigăm pe internet realizăm multe schimburi de informații. De exemplu, atunci când efectuăm o comandă pe internet trebuie să creăm un cont online și să furnizăm o parte din informațiile noastre personale. Pentru a realiza unele tranzacții, informațiile de tip standard (numele, adresa de e-mail, adresa poştală și numărul de telefon) nu sunt îndeajuns, astfel trebuie să furnizăm şi CNP-ul, seria și numărul de buletin, o parolă etc. Aceste date sunt destul de importante și nu trebuie furnizate oricui, iar aici criptarea joacă un rol important.

Comerțul pe internet este într-o continuă creştere. Este foarte ușor să realizezi o tranzacție de pe orice device care este conectat la internet. Astfel, cea mai mare problemă a internetului rămâne securitatea datelor, în special a celor private sau personale.

Principalele tipuri de date pe care oamenii preferă să nu le partajeze sunt: informații despre cardurile de credit; CNP-ul sau numărul și seria actului de identitate; corespondență privată; detalii personale; informații private despre companii; informații despre conturi bancare.

Există mai multe metode de criptare. Noi vom cerceta metodele clasice care se bazau pe substituţii.

Prima cheie de criptare care a fost recunoscută și aplicată la nivelul Statelor Unite ale Americii se numește DES (Data Encryption Standard). Aceasta a fost aprobată și implementată în anul 1970 și avea dimensiunea de 56 biți. O astfel de cheie conține până la 70 de catralioane de posibile combinații. Însă un atac în forță poate decripta foarte ușor informațiile criptate cu o astfel de cheie. Tocmai de aceea DES a fost înlocuită cu AES (Advanced Encryption Standard), ce are o dimensiune cuprinsă între 128 și 256 biți. Majoritatea oamenilor cred că o astfel de cheie este îndeajuns de sigură pentru mult timp de acum înainte, numărul de combinații posibile fiind imens (oare în realitate fiind așa ?).

La etapa actuală, pentru a utiliza eficient criptarea informației, este necesar de cunoscut nu numai metodele de criptare, dar și neajunsurile lor.

În cazul când cheia de criptare este cunoscută, decriptarea informației este simplă. De aceea ne punem scopul să decriptăm informația interceptată fără a cunoaște cheia de criptare.

2. Spargerea sistemelor de criptare monoalfabetice prin forță brută

Să cercetăm poziția unui criptanalist *O*. Se admite că el întotdeauna are la dispoziție facilități de calcul excelente, adesea superioare celor de care dispun cei doi parteneri: expeditorul *X* și destinatarul *Y*. Mai mult, se poate considera că *O* cunoaște

sistemul de criptare. S-a constatat că acest lucru se întâmplă practic totdeauna. Ce nu cunoaşte însă criptanalistul este cheia. Cel mai simplu atac constă în parcurgerea tuturor cheilor posibile şi verificarea textului criptat, până se găseşte cheia corectă. Un astfel de atac se numeşte *atac prin forţă brută* şi el mereu reuşeşte, deoarece întotdeauna există o cheie în *K*, care a fost folosită la criptare. Deci, în cazul când numărul cheilor posibile este mic, această cheie se poate afla foarte uşor după un număr finit de încercări.

Dacă analizăm, în calitate de exemplu, cifrul lui Cezar şi cel afin, atunci devine imediat clar că ele sunt extrem de vulnerabile la atacul prin forţă brută, deoarece pentru cifrul lui Cezar în limba română există numai 31 chei posibile, iar pentru cifrul afin – 929 de chei.

Mai jos prezentăm câte un exemplu pentru cifrul lui Cezar şi cel afin cum ele pot fi sparte.

Exemplul 1 (cifrul lui Cezar). Decriptaţi textul interceptat: VDQDCIÎCQÎHXZUF luând în consideraţie şi spaţiile goale, fără a cunoaşte cheia, ştiind că mesajul a fost scris în limba română şi criptat după algoritmul lui Cezar (vezi [1]).

Rezolvare. În acest caz cunoaştem limba în care a fost scris mesajul şi cunoaştem că cifrul de criptare este cifrul lui Cezar. Deoarece numărul de chei posibile este mic (numai 31 de chei), vom aplica atacul prin forţă brută. Cu acest scop vom verifica pe rând toate cheile posibile, până când se va obţine un text cu sens. Notăm cu *sl* spaţiul liber şi-i punem în corespondenţă poziţia 31 din alfabet.

În funcţie de lungimea cheii, corespondenţa dintre literele textului clar şi cele ale textului cifrat devine:

x	0	1	2	3	4	5	...	30	31
textul clar	A	Ă	Â	B	C	D	...	Z	*sl*
k = 1	Ă	Â	B	C	D	E	...	*sl*	A
k = 2	Â	B	C	D	E	F	...	A	Ă
k = 3	B	C	D	E	F	G	...	Ă	Â
k = 4	C	D	E	F	G	H	...	Â	B
k = 5	D	E	F	G	H	I	...	B	C
...	...	...	...	...	...	...	...	...	...

Observăm că sistemul presupune substituţia fiecărei litere cu litera corespunzătoare în alfabetul rotit cu *k* poziţii.

Decriptând fiecare caracter în corespondentul său clar se obţine, consecutiv:

– pentru *k* = 1: UCPCBHIBPIGWYŢE;

– pentru *k* = 2: UBOBÂGHÂOHFVXTD;

– pentru *k* = 3: TÂNÂĂFGĂNGEUWŞC;

– pentru $k = 4$: ȘĂMĂAEFAMFDȚVSB;

– pentru $k = 5$: SALA DE LECTURĂ.

Deci textul criptat este ” SALA DE LECTURĂ”, iar cheia este $k = 5$.

Exemplul 2 (cifrul afin). Decriptați textul interceptat:

LSJSAQĂBZBZEȚJSZVQXĂYHSQSȘG

dacă el a fost scris în limba română utilizând cifrul afin (vezi [1]).

Rezolvare. Deoarece mesajul interceptat este foarte scurt, atunci vom aplica atacul prin forță brută. Numărul a trebuie să fie reciproc prim cu numărul 32, deci $a \in \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31\}$, iar b va lua 32 de valori, cu excepția $a = 1$ și $b = 0$, care nu duce la nici o criptare. Deci trebuie să verificăm $16 \cdot 32 - 1 = 511$ chei posibile.

Pentru $a = 1$ și $b \in \{1, 2, 3, 4, 5, ..., 31\}$ vom obține cifrul lui Cezar. În acest caz trebuie să alcătuim 31 de tabele corespunzătoare cheilor: (1, 1); (1, 2); (1, 3); ...; (1, 30); (1; 31).

Dacă textele obținute n-au nici un sens semantic, atunci continuăm mai departe alcătuirea tabelelor pentru $a = 3$ și $b \in \{0, 1, 2, 3, 4, 5, ..., 31\}$. Este necesar de alcătuit 32 de tabele. Continuăm așa până nu obținem un text cu sens. În cazul nostru se obține textul clar:

MISIUNEA A FOST ÎNDEPLINITĂ

cu cheia de criptare $k = (5, 3)$.

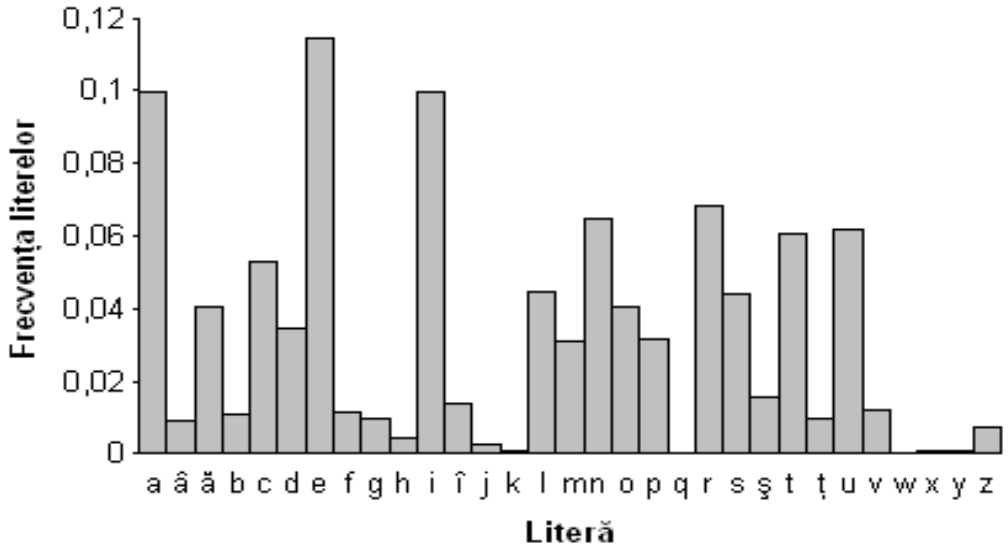
3. Spargerea sistemelor de criptare monoalfabetice prin analiza frecvenţei apariţiei literelor într-o limbă

Punctul slab al sistemelor de criptare monoalfabetice constă în frecvenţa de apariţie a caracterelor în text. Dacă un text criptat este suficient de lung şi se cunoaşte limba în care este scris textul clar, atunci sistemul poate fi spart printr-un atac bazat pe frecvenţa apariţiei literelor într-o limbă.

Sunt construite diverse structuri de ordine relativ la frecvenţa apariţiei literelor în fiecare limbă. De obicei, cu cât un text criptat este mai lung, cu atât frecvenţa literelor folosite se apropie de această ordonare generală. O comparare între cele două relaţii de ordine (cea a caracterelor din textul criptat şi cea a literelor din alfabetul limbii curente) conduce la realizarea câtorva corespondenţe (literă text clar – literă text criptat), ceea ce stabileşte în mod univoc cheia de criptare. Pentru sistemul Cezar este suficientă stabilirea unei singure perechi; pentru sistemul afin sunt necesare două perechi, etc.

Slăbiciunea cifrurilor monoalfabetice este dată de faptul că distribuţia lor de frecvenţă reflectă distribuţia alfabetului folosit vezi ([2, 3, 4]).

Pentru limba română frecvenţa literelor este prezentată în figura și tabelul de mai jos:



A	Ă	Â	B	C	D	E	F	G	H	I
9,95	4,06	0,91	1,07	5,28	3,45	11,47	1,18	0,99	0,47	9,96

Î	J	K	L	M	N	O	P	Q	R
1,40	0,24	0,11	4,48	3,10	6,47	4,07	3,18	0,00	6,82

S	Ș	T	Ț	U	V	W	X	Y	Z
4,40	1,55	6,04	1,00	6,20	1,23	0,03	0,11	0,07	0,71

Exemplul 3 (vezi [3]). S-a interceptat următorul text, criptat în limba română fără diacritice cu un sistem monoalfabetic (nu se cunoaște exact ce sistem a fost utilizat):

„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”.

Decriptați textul, stabiliți sistemul de criptare și determinați cheia.

Rezolvare. Vom aplica atacul bazat pe frecvenţa apariţiei literelor într-o limbă. Frecvența literelor în alfabetul român fără diacritice este prezentată în tabelul de mai jos:

A	B	C	D	E	F	G	H	I	J	K
14,92	1,07	5,28	3,45	11,47	1,18	0,99	0,47	11,36	0,24	0,11

L	M	N	O	P	Q	R	S	T

4,48	3,10	6,47	4,07	3,18	0,00	6,82	5,95	7,04

U	V	W	X	Y	Z
6,20	1,23	0,03	0,11	0,07	0,71

La prima etapă, vom stabili de câte ori apare în text fiecare caracter. Vom obține tabelul:

caracter	*c*	*k*	*x*	*b*	*s*	*q*	*g*	*p*	*m*
frecvență	39	38	27	25	23	20	19	18	18

caracter	*l*	*e*	*p*	*a*	*v*	*b*	*n*	*o*	*f*	*z*
frecvență	11	9	8	7	7	2	2	2	2	2

Din tabel observăm, că caracterele cele mai frecvente sunt *c* și *k*. Pe de altă parte, cele mai frecvente caractere din limba română fără diacritice sunt *a, i* și *e* (textul nu este destul de mare pentru a putea face o distincție netă). La sigur, $a \in \{c, k\}$. Sunt patru opțiuni posibile, din care trei se elimină rapid. Rămâne de abordat $a \to c$ și $e \to k$. Vom nota cu litere mari caracterele din textul clar. Prin înlocuirea lui *c* cu *A*, a lui *k* cu *E*, textul devine:

lqaEA sp gAxE aAa pAmgqb Eq ExA pEersmpqsb vE vsmgxEbA mEaApA tAaApbqlqs vE Agele Amtxq ms noAxgsb mbxAsp vE exsgE oxAbqsbAbE texbslE spAlbE gAxE AmgqpvEAq bxEgAbexslk gqxbslE xEtxEnEpbAq tEpbxq mbxAsps qp AfExbsmaEpb mqtAxAbex vAx lsatEvE pq bxErqsAq mA zsE txEA gqxsems psgs mA mE AmbEtbE mA AzlE aAxE lqgxq vE lA gEs gq gAxE fEpEAq sp gepbAgb.

Cuvântul ExA de pe primul rând are caracterul din mijloc (x) de frecvență ridicată (27 de apariții). Deci el trebuie să corespundă unei litere frecvente din limba română și în plus să aibă semnificație semantică.

Concluzie: acest cuvânt este ”ERA”. Deci $R \to x$. Facem substituția si vom obține textul:

lqaEA sp gARE aAa pAmgqb Eq ERA pEersmpqsb vE vsmgREbA mEaApA tAaApbqlqs vE Agele AmtRq ms noARgsb mbRAsp vE eRsgE oRAbqsbAbE teRbslE spAlbE gARE AmgqpvEAq bREgAbeRsleR gqRbslE REtREnEpbAq tEpbRq mbRAsps qp AfERbsmaEpb mqtARAbeR vAR lsatEvE pq bRErqsAq mA zsE tREA gqRsems psgs mA mE AmbEtbE mA AzlE aARE lqgRq vE lA gEs gq gARE fEpEAq sp gepbAgb.

În acest text, cuvântul REtREnEpbAq are corespondent în limba română numai pe REPREZENTA {I, M, U}. De aici se obțin decriptările $P \to t$, $Z \to n$, $N \to p$ și $T \to b$. Pentru ultimul caracter – *q*, nu facem deocamdată nici o opțune. Noul text va fi:

lqaEA sp gARE aAa NAmgqT Eq ERA NEersmNqsT vE vsmgRETA mEaANA PAaANTqlqs vE Agele AmPRq ms ZoARgsT mTRAsN vE eRsgE oRATqsTATE PeRTslE sNAlTE gARE AmgqNvEAq TREgATeRsleR gqRTslE REPREZENTAq PENTRq mTRAsNs qN AfERTsmaENT mqPARATeR vAR lsaPEvE Nq bRErqsAq mA zsE PREA gqRsems Nsgs mA mE AmTEPTE mA AzlE aARE lqgRq vE lA gEs gq gARE fENEAq sN geNTAgT.

Acum lucrurile încep să se simplifice:

cuvântul PENTRq este corect numai pentru $U \to q$, AmTEPTE pentru $S \to m$. Apoi din cuvântul NASgUT obținem $C \to g$, din SUPARATeR obținem $O \to e$, iar din fENEAU deduce $V \to f$. Efectuând aceste substituții, vom obține următorul text:

lUaEA sp CARE MAM NASCUT EU ERA NEOrsSNUsT DE vsSCRETA SEaANA PAaANTUlUs DE ACOlO ASPRU Ss ZoARCsT STRAsN vE ORsCE oRATUsTATE PORTslE sNAlTE CARE ASCUNvEAU TRECATORslOR CURTslE REPREZENTAU PENTRU STRAsNs UN AfERTsSaENT SUPARATOR vAR lsaPEvE NU bRErqsAU SA zsE PREA CURsOms NsCs SA SE ASTEPTE mA AzlE aARE lUCRU vE lA CEs CU CARE VENEAU sN CONTACT.

Ultimele caractere se deduc imediat: $L \to l$, $M \to a$, $B \to r$, $I \to s$ și $D \to v$. În fine obținem:

LUMEA IN CARE MAM NASCUT EU ERA NEOBISNUIT DE DISCRETA SEMANA PAMANTULUI DE ACOLO ASPRU SI ZGARCIT STRAIN DE ORICE GRATUITATE PORTILE INALTE CARE ASCUNDEAU TRECATORILOR CURTILE REPREZENTAU PENTRU STRAINI UN AVERTISMENT SUPARATOR DAR LIMPEDE NU TREBUIAU SA FIE PREA CURIOSI NICI SA SE ASTEPTE SA AFLE MARE LUCRU DE LA CEI CU CARE VENEAU IN CONTACT (textul provine din romănul ”Viața ca o coridă” de Octavian Paler). Evident, că dacă se cunoștea sistemul de criptare (afin, Cezar etc.) criptanaliza se simplifică mult.

Ne punem scopul în continuare să determinăm sistemul de criptare şi cheia. Pentru aceasta alcătuim tabelul corespondenței obținute:

0	1	2	3	4	5	6	7	8	9	10
A	B	C	D	E	F	G	H	I	J	K
C	R	G	V	K	Z	O		S		
2	17	6	21	10	25	14		18		

11	12	13	14	15	16	17	18	19
L	M	N	O	P	Q	R	S	T

L	A	P	E	T		X	M	B
11	0	15	4	19		23	12	1

20	21	22	23	24	25
U	V	W	X	Y	Z
Q	F				N
16	5				13

Din cele obținute mai sus tragem concluzia, că sistemul de criptare la sigur nu este sistemul cavalerilor de Malta. De asemenea acest sistem nu este sistemul lui Polybios, deoarece fiecărei litere i se pune în corespondență numai o literă, iar în sistemul lui Polybios fiecărei litere i se pune în corespondență o pereche de litere sau cifre. Probabil, sistemul aplicat este un sistem a lui Cezar sau un sistem afin. Deoarece sistemul lui Cezar este un sistem particular al sistemului afin, putem presupune, că sistemul aplicat la criptare este un sistem afin.

Cheia de criptare în sistemul afin are forma: $e_k(x) = ax + b \pmod{26}$. Substituind primele două valori ale lui x și $e_k(x)$ din tabelul de mai sus obținem:

$e_k(0) = a \cdot 0 + b = 2 \pmod{26}$, de unde $b = 2$ și $e_k(1) = a \cdot 1 + 2 = 17$, de unde $a = 15$. Deci cheia de criptare este $k = (15; 2)$.

Literele *H, J, K, Q, W, X* și *Y* n-au făcut parte din textul interceptat. Însă cunoscând funcția de criptare noi cu ușurință obținem corespondențele:

$$H \to D,\ J \to H,\ K \to W,\ Q \to I,\ W \to U,\ X \to J \text{ și } Y \to Y.$$

În fine, obținem următorul tabel complet al corespondenței textului criptat:

A	B	C	D	E	F	G	H	I	J	K
C	R	G	V	K	Z	O	D	S	H	W

L	M	N	O	P	Q	R	S	T
L	A	P	E	T	I	X	M	B

U	V	W	X	Y	Z
Q	F	U	J	Y	N

Următoarea dată, când se va intercepta alt text cu aceeași cheie $k = (15; 2)$, noi vom putea cu ușurință să decriptăm mesajul, chiar dacă eventual el va conține și unele din literele care n-au participat la criptarea primului text.

Observație. Funcția de criptare poate fi aflată utilizând oricare două corespondențe din tabelul de mai sus, pur și simplu calculele vor fi mai voluminoase. De exemplu, dacă luăm corespondențele $0 \to 2$ și $11 \to 11$, atunci obținem $b = 2$ și $11 = 11 \cdot a + 2 \pmod{26}$. Deci din această ecuație trebuie să găsim un a întreg în Z_{26}. Vom obține:

$11a = 9$ – nu există a întreg; $11a = 35$ – nu există a întreg;

$11a = 61$ – nu există a întreg; $11a = 87$ – nu există a întreg;

$11a = 113$ – nu există a întreg; $11a = 139$ – nu există a întreg;

$11a = 165$ – soluția ecuației este $a = 15$.

Bibliografie

1. Afanas D., Gorceag A. Cifruri de substituție monoalfabetice și posibilitățile de spargere a lor. Conferinţa ştiinţifică interuniversitară „Evoluţia ştiinţei militare în contextul noilor ameninţări la securitatea naţională şi regională”. Chișinău, 6 decembrie, 2018. 8 p.
2. Schneier B. Applied Cryptography. John Wiley and Sons, 1995.
3. Zgureanu A. Criptarea și securitatea informației. Note de curs. Chișinău, 2013. 148 p.
4. http://en.wikipedia.org/wiki/Caesar_cipher# History_and_usage.