Супституцијске криптографске шифре

***Бернадин Ибрахимпашић*, Бихаћ**

Увод

Приватност и тајност одувек су били актуелна тема у свим периодима људске цивилизације. Људи су на разне начине покушавали заштитити своје поруке. Криптографија је научна дисциплина која се бави проучавањем метода за слање порука у таквом облику, да их само онај коме су намењене може разумети. Реч криптографија је састављена од грчких речи *криптос* (κρυπτος - сакривен, тајан) и *графеин* (γραφειν – писати) и значи тајнопис. Појам криптографије, као тајног писања, се проширује и на тајно преношење говора и слика. На тај начин се долази до општег појма тајног споразумевања или сакривања информација. Криптоанализа је научна дисциплина која се бави проучавањем поступака за читање скривених порука без познавања начина њиховог настанка, док је криптологија грана науке која обухвата криптографију и криптоанализу. Криптографију и криптоанализу неки називају близаначким или реципрочним дисциплинама. По својим функцијама, оне су доиста, као предмет и његова слика у огледалу. Што једна гради, друга разграђује. Међутим, њихове се природе фундаментално разликују. Криптографија је теоретска и апстрактна, а криптоанализа је емпиријска и конкретна.

Методе криптографије су математичке. Криптографске трансформације су чиста математика. Тако, на пример, у криптографији имамо такве математичке појаве као што су измена редоследа примарних елемената (слова алфабета), сабирање и одузимање унутар ограничених скупова, трансформације координата дијаграма или линеарне алгебарске трансформације. Једноставан пример такве трансформације, којој је сврха тајност, је: $y = ax + b$, једначина у којој је x слово поруке, y слово које га замењује, док су a и b константе које одређују трансформацију. Када се једном утврди прикладан алгебарски поступак, математичке операције са словима се лако обављају. Зато, операције и резултати криптографије важе једнако универзално и вечно као математичке операције и њихови резултати.

Основни задатак криптографије је омогућавање двема особама, од којих је једна пошиљалац, тј. особа која жели саопштити поруку, а друга прималац, тј. особа којој је порука послата, да комуницирају преко несигурног комуникацијског канала (телефонска линија, радио таласи, рачунарска мрежа, итд.) на начин да трећа особа (њихов противник, тзв. нападач) не може разумети њихове поруке. Поруку коју пошиљалац жели послати примаоцу зовемо *отворени текст* (енгл. plaintext). То може бити текст на њиховом матерњем или неком другом језику, нумерички подаци или било шта друго. Пошиљалац трансформише отворени текст, користећи унапред договорени *кључ*. Тај поступак се зове шифрирање (криптирање), а добијени резултат *шифрат* (шифрирана порука, криптограм, криптат, криптирана порука) (енгл. ciphertext). Након тога пошиљалац шаље шифрат путем неког комуникацијског канала. Противник прислушкујући дозна садржај шифрата, али не може одредити отворени текст и разумети поруку. За разлику од њега, прималац, који зна кључ којим је порука шифрирана, може дешифрирати (декриптирати) шифрат и одредити отворени текст.

На основу сведочанстава великих историчара и филозофа Херодота, Плутарха и других, можемо тврдити да старим Грцима дугујемо проналазак првог познатог система за шифрирање. То је шифра СКИТАЛЕ која се користила у доба Ликурга у IX веку п.н.е. Шифра

се састојала од штапа одређене дебљине, око којег се намотавала кожна трака на коју се порука исписивала вертикално. После исписивања, трака би се одмотала тако да на њој остане испремештан низ знакова који је могао прочитати само онај ко је имао штап једнаке дебљине. Шифра (криптографски алгоритам) је математичка функција која се користи за шифрирање и дешифрирање. Ту се ради о две функције од којих је једна за шифрирање, а друга за дешифрирање. Аргументи функције за шифрирање су кључ и отворени текст, а аргументи функције за дешифрирање су кључ и шифрат. Скуп свих могућих вредности кључева зове се простор кључева. Криптосистем се састоји од шифре и свих могућих отворених текстова, шифрата и кључева. Из реченог, добијамо следећу формалну дефиницију.

Дефиниција 1. Криптосистем је уређена петорка $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ за коју важи:
1. $\mathcal{P}$ је коначан скуп свих могућих отворених текстова,
2. $\mathcal{C}$ је коначан скуп свих могућих шифрата,
3. $\mathcal{K}$ је простор кључева, тј. коначан скуп свих могућих кључева,
4. за сваки кључ $K \in \mathcal{K}$ постоји алгоритам шифрирања $e_K \in \mathcal{E}$ и одговарајући алгоритам дешифрирања $d_K \in \mathcal{D}$, где су $e_K : \mathcal{P} \to \mathcal{C}$ и $d_K : \mathcal{C} \to \mathcal{P}$ функције са својством да је $d_K(e_K(x)) = x$ за сваки отворени текст $x \in \mathcal{P}$.

Очигледно је да функције e_K морају бити инјекције, јер, кад би било, $e_K(x_1) = e_K(x_2) = y$, $x_1, x_2 \in \mathcal{P}, y \in \mathcal{C}$, онда прималац не би могао одредити да ли y треба дешифрирати као x_1 или x_2, тј. $d_K(y)$ не би било дефинисано. С обзиром на тајност кључа, криптосистеми се деле на симетричне и асиметричне. Код симетричних криптосистема, тј. криптосистема с тајним кључем, пошиљалац и прималац би тајно изабрали кључ K помоћу којег би генерисали функције e_K за шифрирање и d_K за дешифрирање. У овом случају је d_K исти као и e_K или се из њега може једноставно израчунати. Из тог разлога, сигурност симетричних криптосистема лежи у тајности кључа, што и представља велики недостатак, јер пошиљалац и прималац пре шифрирања морају бити у могућности да размене тајни кључ преко неког сигурног комуникацијског канала, помоћу курира или се лично срести. То је некада тешко изводиво, нарочито ако су они на великој удаљености и ако су комуникацијски канали, који су им на располагању, поприлично несигурни. Поред тога, тајни кључ се мора често мењати, јер шифрирање више пута истим кључем смањује сигурност. У ову групу спадају супституцијске и транспозицијске шифре, као и њихове комбинације.

Идеју једног криптосистема, потпуно другачијег типа него симетрични криптосистем, изнели су 1976. године Whitfield Diffie и Martin Hellman. Назвали су га криптосистем јавног кључа. Идеја се састојала у томе да се користе функције за шифрирање e_K из којих је практично немогуће, у неком разумном времену, израчунати функцију за дешифрирање d_K. У том случају би функција за шифрирање e_K могла бити јавна.

1. Цезарова шифра

Познати римски војсковода и државник, Јулије Цезар, у комуникацији са својим пријатељима се користио шифром у којој су се слова отвореног текста замењивала словима која су се налазила три места даље од њих у алфабету. Цезарову шифру можемо записати на следећи начин:

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

На овај начин, реч TANGENTA би била шифрирана као WDQJHQWD.

Уколико се ради с отвореним текстом на српском језику, тада се слова Č и Ć мењају словом C, а Dž, Đ, Lj, Nj, Š и Ž мењају редом са DZ, DJ, LJ, NJ, S и Z. Данас се Цезаровом шифром називају све шифре истог облика и с помаком различитим од 3.

Пошто енглески (међународни) алфабет има 26 слова, шифру ћемо дефинисати над скупом $\mathbf{Z}_{26} = \{0, 1, 2, \ldots, 25\}$. Имамо следећу кореспонденцију:

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
0	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

Нека је $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbf{Z}_{26}$. Сада за $0 \leqslant K \leqslant 25$ дефинишемо

$$e_K(x) = (x + K) \bmod 26 \quad \text{и} \quad d_K(y) = (y - K) \bmod 26.$$

Очигледно је $d_K(e_K(x)) = x$, што се и захтева у дефиницији криптосистема. За $K = 3$ се добија оригинална Цезарова шифра.

Пример 1. Декриптирати шифрат PWFLZOJAFH добијен Цезаровом шифром.

Решење. Како је простор кључева јако мали (има их свега 26), задатак можемо решити методом грубе силе, тј. тако да испитамо све могуће кључеве, док не добијемо неко решење које има смисла. За $0 \leqslant K \leqslant 25$ добијамо отворени текст d_K:

d_0 :	P	W	F	L	Z	O	J	A	F	H
d_1 :	O	V	E	K	Y	N	I	Z	E	G
d_2 :	N	U	D	J	X	M	H	Y	D	F
d_3 :	M	T	C	I	W	L	G	X	C	E
d_4 :	L	S	B	H	V	K	F	W	B	D
d_5 :	K	R	A	G	U	J	E	V	A	C

Видимо да је кључ $K = 5$, а да је отворени текст KRAGUJEVAC.

2. Афина шифра

Нека је $\mathcal{P} = \mathcal{C} = \mathbf{Z}_{26}$, те нека је $\mathcal{K} = \{(a, b) \in \mathbf{Z}_{26} \times \mathbf{Z}_{26} : \mathrm{NZD}(a, 26) = 1\}$. За $K = (a, b) \in \mathcal{K}$ дефинишемо

$$e_K(x) = ax + b \bmod 26 \quad \text{и} \quad d_K(y) = a^{-1}(y - b) \bmod 26.$$

Ова се шифра зове афина јер су функције шифрирања афине. Како је

$$d_K(e_K(x)) = d_K(ax + b) = a^{-1}(ax + b - b) = a^{-1}ax = x,$$

то је и услов $d_K(e_K(x)) = x$ задовољен. Овде a^{-1} означава мултипликативни инверз броја a у прстену $\mathbf{Z}_{26}$. Пошто 26 није прост број, то $\mathbf{Z}_{26}$ није поље, па стога ни сви елементи из $\mathbf{Z}_{26}$ немају мултипликативни инверз, него их имају само они бројеви из $\mathbf{Z}_{26}$ који су релативно прости с 26, тј. за које важи да је $\mathrm{NZD}(a, 26) = 1$. То можемо представити табелом:

a	1	3	5	7	9	11	15	17	19	21	23	25
a^{-1}	1	9	21	15	3	19	7	23	11	5	17	25

Пример 2. Шифрирати отворени текст TANGENTA афином шифром где је $K = (11, 4)$.
Решење.

$$\begin{aligned} T: &\quad 19 \cdot 11 + 4 = 213 \equiv 5 (\mathrm{mod}\, 26) = \mathrm{F} \\ A: &\quad 0 \cdot 11 + 4 = 4 \equiv 4 (\mathrm{mod}\, 26) = \mathrm{E} \\ N: &\quad 13 \cdot 11 + 4 = 147 \equiv 17 (\mathrm{mod}\, 26) = \mathrm{R} \\ G: &\quad 6 \cdot 11 + 4 = 70 \equiv 18 (\mathrm{mod}\, 26) = \mathrm{S} \\ E: &\quad 4 \cdot 11 + 4 = 48 \equiv 22 (\mathrm{mod}\, 26) = \mathrm{W}. \end{aligned}$$

Добијамо да је шифрат FERSWRFE

Пример 3. Дешифрирати шифрат UFSHP SUFIP XFLBK ZLWFJ FUPJF UZRDZ CFXDS FCLUM BGCDL UMFJF UPJFU ZXFCF LCOZI P добијен афином шифром.

Решење. У овом случају је број могућих кључева једнак $12 \cdot 26 = 312$. Овај број кључева се сматра довољно малеким у дешифрирању, па би се и овде могла применити метода грубе силе, тј. могли би се испитати сви могући кључеви (наравно уз помоћ рачунара). Међутим, постоји и елегантнији начин уколико нам је познато којим је језиком писан отворени текст. Претпоставимо да нам је познато да је у овом примеру текст писан српским језиком. За решавање нам је потребна само чињеница да су у српском језику најфреквентнија слова A, E, I, O и N и то у датом редоследу. У нашем шифрату, најфреквентнија слова су F, које се јавља 13 пута, U које се јавља 8 пута, те P, L, Z и C који се јављају по 5 пута. Упркос чињеници да је наш шифрат прекратак за статистичку анализу, можемо претпоставити да је слово F шифрат слова A, а да је U шифрат слова E. Проверимо нашу претпоставку. Имамо да је:

$$e_K(\mathrm{A}) = a \cdot 0 + b = b \quad \text{и} \quad e_K(\mathrm{E}) = a \cdot 4 + b = 4a + b.$$

Уз нашу претпоставку да је $e_K(\mathrm{A}) = \mathrm{F}$ и $e_K(\mathrm{E}) = \mathrm{U}$ добијамо да је $b = 5$ и $4a + b \equiv 20 (\mathrm{mod}\, 26)$, међутим, не постоји $a \in \{1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25\}$ који задовољава дату конгруенцију, што значи да овај систем нема решења, па је наша претпоставка погрешна.

Претпоставимо поново да је F шифрат слова A, али да је сада L шифрат слова E. Тада имамо да је $b = 5$ и $4a + b \equiv 11 (\mathrm{mod}\, 26)$. Одавде добијамо да је $a = 21$ и $b = 5$, тј. кључ је $K = (21, 5)$, па су функције за шифрирање и дешифрирање:

$$e_K(x) = 21x + 5 \bmod 26 \quad \text{и} \quad d_K(y) = 21^{-1}(y - b) = 5(y - 5) \bmod 26.$$

Применимо ли сада добијену функцију d_K на наш шифрат, добијамо отворени текст који нема смисла, што значи да је наша претпоставка била поново погрешна.

Претпоставимо поново да је F шифрат слова A, али да је сада P шифрат слова E. Тада имамо да је $b = 5$ и $4a + b \equiv 15 (\mathrm{mod}\, 26)$. Одавде добијамо да је $a = 9$ и $b = 5$, тј. кључ је $K = (9, 5)$, па су функције за шифрирање и дешифрирање:

$$e_K(x) = 9x + 5 \bmod 26 \quad \text{и} \quad d_K(y) = 9^{-1}(y - b) = 3(y - 5) \bmod 26.$$

Применимо ли сада добијену функцију d_K на наш шифрат, добијамо отворени текст

TANGENTA JE ČASOPIS ZA MATEMATIKU I RAČUNARSTVO
DRUŠTVA MATEMATIČARA SRBIJE.

3. ДЕКРИПТИРАЊЕ СУПСТИТУЦИЈСКЕ ШИФРЕ

Цезарова и афина шифра су специјални случајеви супституцијске шифре, која је дефинисана са $\mathcal{P} = \mathcal{C} = \mathbf{Z}_{26}$ и $\mathcal{K}$ који се састоји од свих пермутација скупа $\{0, 1, 2, \ldots, 25\}$. За сваку пермутацију $\pi \in \mathcal{K}$ дефинишемо: $e_\pi(x) = \pi(x)$, и $d_\pi(y) = \pi^{-1}(y)$, где је π^{-1} инверзна пермутација од π. У овом случају имамо $26! \approx 4 \cdot 10^{26}$ могућих кључева, па је дешифрирање грубом силом, тј. испитивањем свих могућих кључева, практично немогуће, чак и уз помоћ рачунара. Међутим, шифрат добијен супституцијском шифром, могуће је лако дешифрирати користећи статистичка својства језика којим је писан отворени текст. Основна метода је анализа фреквенције слова, где се броји појављивање сваког слова у шифрату, те се упоређује с познатим подацима. Врло је вероватно да најфреквентнија слова шифрата одговарају најфреквентнијим словима језика. Та вероватноћа расте с дужином шифрата. Такође, корисни могу бити и подаци о најчешћим биграмима (паровима слова) и триграмима (низовима од три слова) у језику. Зачеци анализе фреквенција слова се могу наћи у XIV веку, у делу арапског аутора Ибу ад-Дураихима, а претпоставља се да су ту методу у исто време познавали и италијански криптографи.

Основни подаци о фреквенцији слова, израженој у промилима, у српском и енглеском језику, уз напомену да се претпоставља да у тексту нема интерпункцијских знакова нити размака између речи (тиме би дешифрирање било пуно лакше), те да се слова č, ć, dž, đ, lj, nj, š и ž замењују на пре описани начин, су:

српски језик:

A	E	I	O	N	S	T	R	U	J	D	M	V	L	K	P	C	Z	G	B	H	F
123	93	92	88	63	60	49	48	46	46	41	35	34	34	33	30	26	22	16	14	5	2

енглески језик:

E	T	A	O	I	N	S	R	H	L	D	C	U	M	F	G	P	W	Y	B	V	K	J	X	Q	Z
125	93	80	76	73	71	66	61	55	41	40	30	27	25	23	20	20	19	17	15	10	7	2	2	1	1

Најфреквентнији биграми у српском језику су: JE (20, 1), NA (16, 2), AN (15, 2), ST (15, 0), AS, RA, KO, EN, IS, IJ, NI, DA, OS, SE, PO, ED (сви преко 10, 0). Такође треба истакнути и најфреквентније реципрочне биграме чија фреквенција у оба случаја прелази 8 промила: NA - AN, AS - SA, DA - AD, NI - IN, SE - ES и EN - NE. Најфреквентнији биграм је JE, иако J није међу најфреквентнијим словима. Фреквенција овог биграма је приближно једнака половини јављања свих биграма које слово J образује с осталим словима. Истакнимо још да скоро 80% биграма има облик самогласник - сугласник или сугласник - самогласник. Најфреквентнији триграми, са фреквенцијом преко 3 промила су STO, IST, STA, OJE и ENA. У енглеском језику, најфреквентнији биграми су TH, HE, IN, ER, RE, ON, AN, EN и AT, а најфреквентнији триграми су THE, ING, AND, ION и TIO.

Пример 4. Дешифрирати шифрат:

FXBXM	CRGDW	BXODF	XCRMO
XONMO	N IX IX	DCRFB	XKLXB
XAXOR	AXON I	RQCRM	ODPGD
OLRY J	CRBXW	XJXKX	BDLXW
YNCXB	CRMPG	MONOP	ZNCM I
RMNEL	R		

добијен супституцијском шифром, ако је познато да је отворени текст на српском језику.

Решење. Анализирајмо фреквенције слова и биграма помоћу следеће табеле:

A:	X (2)	=	2
B:	C (1), D (1), X (5)	=	7
C:	M (1), R (6), X (1)	=	8
D:	C (1), F (1), L (1), O (1), P (1), W (1)	=	6
E:	L (1)	=	1
F:	B (1), X (2)	=	3
G:	D (2), M (1)	=	3
H:		=	0
I:	R (2), X (2)	=	4
J:	C (1), X (1)	=	2
K:	L (1), X (1)	=	2
L:	R (2), X (2)	=	4
M:	C (1), I (1), N (1), O (4), P (1)	=	8
N:	C (2), E (1), I (2), M (1), O (1)	=	7
O:	D (2), L (1), N (4), P (1), R (1), X (1)	=	10
P:	G (2), Z (1)	=	3
Q:	C (1)	=	1
R:	A (1), B (1), F (1), G (1), M (4), Q (1), Y (1), * (1)	=	11
S:		=	0
T:		=	0
U:		=	0
V:		=	0
W:	B (1), X (1), Y (1)	=	3
X:	A (1), B (4), C (1), D (1), I (1), J (1), K (2), M (1), O (4), W (2)	=	18
Y:	J (1), N (1)	=	2
Z:	N (1)	=	1.

Видимо да су у шифрату најфреквентнија слова: X(18), R(11), O(10), C(8), M(8), N(7), B(7) и D(6), а најфреквентнији биграми су: CR(6), BX(5), MO(4), ON(4), RM(4), XB(4), XO(4). За очекивати је да је $e(A) = X$. Уочавајући реципрочне биграме веће фреквенције BX и XB, за претпоставити је да је $e(N) = B$. На основу фреквенција од R и CR, можемо претпоставити да је $e(J) = C$ и $e(E) = R$. Ако покушамо открити шифрат биграма ST, онда су кандидати MO и ON. Када бисмо претпоставили да је то ON, то би значило да је $e(T) = N$, што би нас довело до тога да се биграм TJ (чији је тада шифрат NC) појављује 2 пута, што је мало вероватно. Стога је логичније претпоставити да је MO шифрат за ST, тј. да је $e(S) = M$ и $e(T) = O$. Од свих високофреквентних слова су нам још преостали I и O, па је за очекивати да је $e(I) = N$ и $e(O) = D$.

За сада имамо следеће претпоставке:

отворени текст:	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
шифрат:			X				R				N	C			B	D				M	O					

Искористимо ли ове претпоставке у шифрату, добијамо:

FXBXM	CRGDW	BXODF	XCRMO	XONMO
ANAS	J E O	NATO	AJEST	ATIST
N IXI X	DCRF B	XKL XB	XAXOR	AXON I
I A A	OJE N	A AN	A ATE	ATI

RQCRM	ODPGD	OLRYJ	CRBXW	XJXKX
E JES	TO O	T E	JENA	A A A
BDLXW	YNCXB	CRMPG	MONO P	ZNCMI
NO A	IJAN	JES	S T I T	IJS
RMNE L	R			
E S I	E			

Сада већ имамо преко 70% отвореног текста, што нам даје довољно елемената да можемо почети одгонетавати неке речи. Тако је очигледно да је прва реч DANAS, тј. да је $e(\mathrm{D}) = \mathrm{F}$. Такође се уочава реч STATISTIKA, тј. да је $e(\mathrm{K}) = \mathrm{I}$. Сада је већ могуће добити комплетан отворени текст који гласи: *Danas je poznato, da je statistika, kao jedna grana matematike, vješto upotrebljena za lagano razbijanje supstitucijske šifre.*

Ако погледамо алфабет шифрата, видимо да он изгледа овако:

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

Ово је супституцијска шифра која се назива Цезарова шифра с кључном речи. У њој кључ представља кључна реч (у нашем примеру је то FREKVENCIJA) и број између 0 и 25 (у нашем примеру је то 3) који означава место у алфабету на којем почињемо писати кључну реч и то без понављања слова. Видимо да је, упркос великом простору кључева, супституцијска шифра прилично једноставна за декриптирање. То је било познато још у XV веку када је у Италији почела употреба хомофона, тј. када су се најфреквентнија слова шифрирала с више различитих симбола. Међутим, иако то повећава сигурност шифре, добра анализа фреквенције слова, биграма и триграма, може лако довести до решења.

4. Вигнерова шифра

Видели смо неке шифре моноалфабетског система, тј. шифре где сваком слову отвореног текста одговара јединствено слово шифрата. За разлику од њих Вигнерова шифра спада у полиалфабетске криптосистеме. Код ове шифре се свако слово отвореног текста може пресликати у једно од m могућих слова, где m представља дужину кључа. Блез де Вигнер је 1586. године објавио књигу у којој је описао полиалфабетски криптосистем који се данас назива Вигерова шифра. Дефинисао ју је на следећи начин: за фиксан природан број m дефинишемо $\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbf{Z}_{26})^m$. За кључ $K = (k_1, k_2, \ldots, k_m)$ дефинишемо

$$e_K(x_1, x_2, \ldots, x_m) = (x_1 + k_1, x_2 + k_2, \ldots, x_m + k_m)$$

$$d_K(y_1, y_2, \ldots, y_m) = (y_1 - k_1, y_2 - k_2, \ldots, y_m - k_m)$$

где су све операције у $\mathbf{Z}_{26}$, тј. где је $17 +_{26} 20 = 11$, јер је 11 остатак при дељењу $17 + 20 = 37$ са 26.

Пример 5. Шифрирати отворени текст KRIPTOGRAFIJA JAVNOG KLJUČA помоћу Вигнерове шифре, где је $m = 5$, а кључна реч је KRUPA.

Решење. Нумерички еквивалент кључа је $K = (10, 17, 20, 15, 0)$. Нумерички еквивалент отвореног текста је

$$(10, 17, 8, 15, 19, 14, 6, 17, 0, 5, 8, 9, 0, 9, 0, 21, 13, 14, 6, 10, 11, 9, 20, 2, 0).$$

Сада одредимо нумерички еквивалент шифрата:

	10	17	8	15	19	14	6	17	0	5	8	9	0	9	0	21	13	14	6	10	11	9	20	2	0
	10	17	20	15	0	10	17	20	15	0	10	17	20	15	0	10	17	20	15	0	10	17	20	15	0
$+_{26}$	20	8	2	4	19	24	23	11	15	5	18	0	20	24	0	5	4	8	21	10	21	0	14	17	0

Видимо да је шифрат: UICET YXLPF SAUYA FEIVK VAORA. Ако ово прикажемо на следећи начин:

кључ:	K	R	U	P	A	K	R	U	P	A	K	R	U	P	A	K	R	U	P	A	K	R	U	P	A
отворени текст:	K	R	I	P	T	O	G	R	A	F	I	J	A	J	A	V	N	O	G	K	L	J	U	C	A
шифрат:	U	I	C	E	T	Y	X	L	P	F	S	A	U	Y	A	F	E	I	V	K	V	A	O	R	A

може се уочити, на пример, да се слово A пресликало у слова P, U и A. Такође се уочава да су се у слово U пресликала слова K и A. У овој оригиналној варијанти Вигнерове шифре се кључ понавља у недоглед. Сигурнија варијанта је с аутокључем. У тој варијанти, отворени текст генерише кључ, па због тога ова варијанта спада у проточне криптосистеме.

Пример 6. Шифрирати отворени текст KRIPTOGRAFIJA JAVNOG KLJUČA помоћу Вигнереове шифре с аутокључем, ако је кључна реч KRUPA.

Решење. Прикажимо то на следећи начин:

отворени текст:	K	R	I	P	T	O	G	R	A	F	I	J	A	J	A	V	N	O	G	K	L	J	U	C	A
кључ:	K	R	U	P	A	K	R	I	P	T	O	G	R	A	F	I	J	A	J	A	V	N	O	G	K

Користећи нумеричке еквиваленте које смо већ одредили у прошлом примеру, имамо:

	10	17	8	15	19	14	6	17	0	5	8	9	0	9	0	21	13	14	6	10	11	9	20	2	0
	10	17	20	15	0	10	17	8	15	19	14	6	17	0	5	8	9	0	9	0	21	13	14	6	10
$+_{26}$	20	8	2	4	19	24	23	25	15	24	22	15	17	9	5	3	22	14	15	10	6	22	8	8	10

Шифрат је: UICET YXZPY WPRJF DWOPK GWIIK. Прикажемо ли и ово на другачији начин, имамо:

кључ:	K	R	U	P	A	K	R	I	P	T	O	G	R	A	F	I	J	A	J	A	V	N	O	G	K
отворени текст:	K	R	I	P	T	O	G	R	A	F	I	J	A	J	A	V	N	O	G	K	L	J	U	C	A
шифрат:	U	I	C	E	T	Y	X	Z	P	Y	W	P	R	J	F	D	W	O	P	K	G	W	I	I	K

И овде уочавамо да се слово A пресликало у слова P, R, F и K, док су се, на пример, у слово I пресликала слова R, U и C. Уколико поседујемо само шифрат добијен помоћу Вигнерове шифре, први корак при дешифрирању је одређивање дужине кључне речи. Размотрићемо две методе. Прва метода, коју је увео Фридрих Касиски 1863. године, назива се Касискијев тест. Ова се метода заснива на чињеници да ће два идентична сегмента отвореног текста бити шифрирани на исти начин ако се њихове позиције разликују за неки чинилац од m, где је m дужина кључне речи. Обратно, ако уочимо два идентична сегмента у шифрату, дужине барем 3, тада је вероватно да они одговарају идентичним сегментима отвореног текста. Сходно реченом, у Касискијевом тесту се у шифрату траже идентични сегменти дужине барем 3, те се одреде удаљености између њихових почетних положаја. Ако на тај начин добијемо удаљености $d_1, d_2, d_3, \ldots$, онда је за очекивати да m дели већину вредности d_i. Друга метода за одређивање дужине кључне речи користи тзв. индекс коинциденције. Тај је појам увео 1920. године Вилиам Фридман у књизи „Индекс коинциденције и његове примене у криптографији“, која се сматра једном од најважнијих публикација у историји криптографије.

Дефиниција 2. Нека је $x = x_1 x_2 \ldots x_n$ низ од n слова. Индекс коинциденције од x, у ознаци

$I_c(x)$, је вероватноћа да су два случајна елемента из x једнаки.

Нека су $f_0, f_1, \ldots, f_{25}$ редом апсолутне фреквенције од A, B, C, ..., Z у x. Како два елемента из x можемо одредити на $\binom{n}{2}$ начина, а за сваки $i = 0, 1, 2, \ldots, 25$ постоји $\binom{f_i}{2}$ начина одабира два пута i-тог слова, то важи формула:

$$I_c(x) = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)}.$$

Претпоставимо ли да x представља неки текст на енглеском језику и означимо ли очекиване вероватноће појављивања слова A, B, ..., Z редом са $p_0, p_1, \ldots, p_{25}$, тада је за очекивати да важи

$$I_c(x) \approx \sum_{i=0}^{25} p_i^2 = \left(\frac{125}{1000}\right)^2 + \left(\frac{93}{1000}\right)^2 + \left(\frac{80}{1000}\right)^2 + \cdots + \left(\frac{1}{1000}\right)^2 \approx 0,066.$$

Исти закључак важи и ако је шифрат x добијен из отвореног текста на енглеском језику помоћу неке моноалфабетске шифре. Ту ће се поједине вредности испремештати, али ће величина $\sum p_i^2$ остати непромењена. Аналогно овоме, за потпуно случајан низ имамо

$$I_c(x) \approx 26 \cdot \left(\frac{1}{26}\right)^2 = \frac{1}{26} \approx 0,038.$$

Ставимо ли $\kappa_r = 0,038$ ($r = \text{random}$) и $\kappa_p = 0,066$ ($p = \text{plaintext}$), видимо да су оне довољно далеко једна од друге, па се на овај начин одређује тачна дужина кључне речи или се потврђује претпоставка добијена помоћу Касискијевог теста. Вредност κ_p за поједине језике је следећа: у српском језику је 0,065, француском 0,078, немачком 0,076, шпанском 0,078, италијанском 0,074, док је у руском 0,053. Ова метода се у пракси примењује тако да се шифрат $y = y_1 y_2 \ldots y_n$, добијен Вигнеровом шифром, растави на m поднизова $z_1, z_2, \ldots, z_m$, тако да y запишемо по колонама у матрицу димензија $m \times (n/m)$. Ако n није дељиво са m матрицу можемо допунити произвољним текстом или једноставно посматрати матрицу са непотпуном задњом врстом. Редови ове матрице су управо тражени поднизови $z_1, z_2, \ldots, z_m$. Ако је m једнак дужини кључне речи, онда би сви $I_c(z_i)$ требао да буду приближно једнаки вредности κ_p за језик којим је писан отворени текст. С друге стране, ако m није дужина кључне речи, онда ће сви $I_c(z_i)$ изгледати више-мање случајни, будући су добијени помацима помоћу различитих кључева. Због тога, за примену ове методе није неопходно познавање језика на којем је писан отворени текст. Довољно је да се за тај језик κ_p знатно разликује од $\kappa_r = 0,038$.

Пример 7. Дешифрирати шифрат добијен Вигнеровом шифром, ако је познато да је отворени текст на српском језику.

RCHCOKGXZQ	OPTHUVOOMR	WHPRGBIGSC	LSOQERCKIB
FOCTWVOAWX	QAHMDCJXKG	UHDOCXWSMU	VFXAVCYDXC
NXPVCLIGQU	KRTQXWYXUG	FCEIECYXWX	EOEWBPOYCL
GNPANCJCWI	CZDDECXPAV	TSQOCUHJRG	XIZUWUSHSN
CBYIBOWYII	CGPZMCDDAX	WBDKUEAYIR	TSSVLKASIP
JCSIUKFXAG	UHGIXCBYMI	QJXUVTOVWO	RIGCLGGAIX
C			

Решење. Одредимо прво дужину кључне речи. Применимо прво Касискијев тест. Можемо уочити неколико триграма који се појављују по два пута у шифрату. То су GUH с почетком на позицијама 60 и 210 ($210 - 60 = 150 = 2 \cdot 3 \cdot 5 \cdot 5$), FXA с почетком на позицијама 72 и 207 ($207 - 72 = 135 = 3 \cdot 3 \cdot 3 \cdot 5$), CLG с почетком на позицијама 119 и 234 ($234 - 119 = 115 = 5 \cdot 23$), CBY с почетком на позицијама 161 и 216 ($216 - 161 = 55 = 5 \cdot 11$) и IXC с почетком на позицијама 214 и 239 ($239 - 214 = 25 = 5 \cdot 5$). Примећујемо да је највероватнија дужина кључне речи $m = 5$. Погледајмо сада другу методу. За $m = 1$ је $I_c(z_1) = 0,045$, за $m = 2$ је $I_c(z_1) = 0,039$ и $I_c(z_2) = 0,050$. За $m = 3$, вредности I_c су 0,048, 0,046 и 0,035, за $m = 4$ су 0,040, 0,042, 0,035 и 0,054, док су за $m = 5$ те вредности 0,073, 0,051, 0,063, 0,068 и 0,057. Сада већ с прилично великом вероватноћом закључујемо да је дужина кључне речи једнака 5. Још нам је преостало да одредимо кључну реч. То се може учинити помоћу међусобног индекса коинциденције двају низова.

Дефиниција 3. Нека су $x = x_1x_2 \dots x_n$ и $y = y_1y_2 \dots y_{n'}$ два низа од n, односно n' слова. Међусобни индекс коинциденције од x и y, у ознаци $MI_c(x, y)$, је вероватноћа да је случајни елемент од x једнак случајном елементу од y. Ако фреквенције од A, B, ..., Z у x и y означимо с $f_0, f_1, \dots, f_{25}$, односно с $f'_0, f'_1, \dots, f'_{25}$ редом, онда је

$$MI_c = \sum_{i=0}^{25} \frac{f_i \cdot f'_i}{nn'}.$$

Уколико знамо на којем је језику писан отворени текст или то барем претпостављамо, онда је $p_i = \dfrac{f_i}{n}$. Нека је m дужина кључне речи $K = (k_1, k_2, \dots, k_m)$ и нека су низови z_1, $z_2, \dots, z_m$ добијени на раније описани начин. Са $f_0^j, f_1^j, \dots, f_{25}^j$ означимо редом фреквенције слова A, B, ..., Z у низу z_j који има n_j слова. Да бисмо одредили j-то слово k_j кључне речи K, за сваки $t = 0, 1, 2, \dots, 25$ одредимо вредности

$$M_t = \sum_{i=0}^{25} \frac{p_i f_{i-t}^j}{n_j},$$

где је операција одузимања по модулу 26, и нека је h такав да је $M_h = \max\{M_t : 0 \leqslant t \leqslant 25\}$. Сада је $k_j \equiv -h \pmod{26}$, па понављајући наведени поступак за све $j = 1, 2, \dots, m$ добијамо кључну реч, након чега је могуће извршити дешифровање. Да бисмо одредили кључну реч у нашем примеру, морамо шифрат преложити по колонама у матрицу која има $m = 5$ врста. Тако добијамо

RKOVWBLRFV	QCUXVCNLKW	FCEPGCCCTU	XUCOCCWETK	JKUCQTRGC
CGPOHISCOO	AJHWFYXIRY	CYOONJZXSH	ISBWGDBASA	CFHBJOIG
HXTOPGOKCA	HXDSXDPGTX	EXEYPCDPQJ	ZHYYPDDYSS	SXGYXVGA
CZHMRSQITW	MKOMAXVQQU	IWWCAWDAOR	USIIZAKIVI	IAIMUWCI
OQURGCEBWX	DGCUVCCUXG	EXBLNIEVCG	WNBIMXURLP	UGXIVOLX

Рачунајући наведене вредности за наш пример, добијамо да је:

$$\begin{aligned}
&j = 1: \quad \max M_t = M_{24} = 0,068 \quad \Rightarrow k_1 = -24 \bmod 26 = 26 - 24 = 2 \quad (\text{C})\\
&j = 2: \quad \max M_t = M_{12} = 0,063 \quad \Rightarrow k_2 = -12 \bmod 26 = 26 - 12 = 14 \quad (\text{O})\\
&j = 3: \quad \max M_t = M_{11} = 0,064 \quad \Rightarrow k_3 = -11 \bmod 26 = 26 - 11 = 15 \quad (\text{P})\\
&j = 4: \quad \max M_t = M_{18} = 0,069 \quad \Rightarrow k_4 = -18 \bmod 26 = 26 - 18 = 8 \quad (\text{I})\\
&j = 5: \quad \max M_t = M_{24} = 0,063 \quad \Rightarrow k_5 = -24 \bmod 26 = 26 - 24 = 2 \quad (\text{C})
\end{aligned}$$

Сада када имамо клучну реч COPIC, лагано добијамо да отворени текст, у који смо додали дијакритичке знакове („квачице“), знакове интерпункције и размаке, представља почетне стихове познате песме за децу „Јежева кућица“ коју је написао Бранко Ћопић:

PO ŠUMI, ŠIROM, BEZ STAZE, PUTA
JEŽURKA JEŽIĆ POVAZDAN LUTA.
LOVOM SE BAVI ČESTO GA VIDE,
S TRISTA KOPALJA NA JURIŠ IDE.
I VUK I MEDO, PA ČAK I OVCA,
POZNAJU JEŽA, SLAVNOGA LOVCA.
JASTREB GA ŠTUJE, VUK MU SE SKLANJA,
ZMIJA GA ŠARKA PO SVU NOĆ SANJA.
PRED NJIM DAN HODA, ŠIRI SE STRAVA,
NJEGOVIM TRAGOM, PUTUJE SLAVA.

Литература

[1] A. Dujella, M. Maretić, *Kriptografija*, Element, Zagreb, 2007.

[2] B. Ibrahimpašić, *Matematičke osnove kriptografije javnog ključa*, Magistarski rad, PMF Sarajevo, 2004.

[3] B. Ibrahimpašić, *RSA kriptosustav*, OML, Vol. 5 (2), 101112, 2005.

[4] D. Kahn, *The Codebreakers. The Story of Secret Writing, IIII*, Macmillan Co., New York, 1967. (prevod: Šifranti protiv špijuna, Centar za informacije i publicitet, Zagreb, 1979.)

[5] A. J. Menezes, P. C. Oorshot, S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, Boca Raton, 1996.

[6] D. R. Stinson, *Cryptography. Theory and Practice*, CRC Press, Boca Raton, 1996.

Статијата прв пат е публикувана во списанието ТАНГЕНТА на ДМ на Србија во 2007/08 година