Лекция 2

Криптанализ на класически криптосистеми

2.1 Общи бележки

В тази глава разглеждаме някои методи за криптанализ на класическите шифри. Ще направим едно общо допускане, известно като *принцип на Керхоф*, съгласно който използваната криптосистема е известна на криптаналиста във всичките ѝ детайли. технически това означава, че той разполага с точни Ще започнем с някои общи белеажки засягащи криптанализа. спецификации на използваните криптографски примитиви.

Най-общо атаките срещу една криптосистема могат да бъдат разделени на два вида:

- *пасивна атака* – това е атака, при която опонентът само наблюдава обменяните шифровани съобщения по комуникационния канал. Тук може да се допусне, че опонентът има достъп и до "черна кутия", осъществяваща шифриране, но не и дешифриране на съобщения. Пасивната атака е най-прост модел на атака и застрашава единствено конфиденциалността на съобщенията.

- *активна атака* – това е атака, при която опонентът се опитва да модифицира предаваните съобщения. Такава атака предствлява заплаха както за конфиденциалността на съобщенията, така и за целостта на съобщенията и за автентичността им.

В тази глава ще срещнем само пасивни атаки. В зависимост от наличната допълнителна информация те могат да бъдат разделени на няколко подвида.

(*a*) *Атака при известен криптотекст* (ciphertext-only attack): това е атака, при която опонентът се опитва да възстанови ключа или открития текст, наблюдавайки единствено криптотекста. Криптосистема, уязвима от такава атака се счита за изключително несигурна.

(*b*) *Атака при известен открит текст* (known plaintext attack): това е атака, при която опонентът разполага с няколко открити текста и съответните им криптотекстове. Всички криптотекстове са получени сизползване на един и същи ключ. На практика много често има достатъчно данни за започване на такава атака.

(*c*) *Атака при избран открит текст* (chosen plaintext attack): това е атака, при която опонентът може да избира откритите текстове, след което получава съответните криптотекстове. Това се получава, кагато, например, опонентът е получил временен достъп до криптооборудването.

(*d*) *Адаптивна атака при избран открит текст* (adaptive chosen plaintext attack): това е атака при избран открит текст, при която изборът на открит текст зависи от криптотекста, получен при предното запитване.

(*e*) *Атака при избран криптотекст* (chosen ciphertext attack): това е атака, при която опонентът избира криптотекст и след това получава съответния му открит текст. Един начин по който може да бъде започната такава атака е да получим достъп до обурудбването, използвано за дешифриране (но не и до ключа за дешифриране, който може да е сигурно вграден в оборудването. Целта е да сме в състояние в по-късен момент (без достъп до оборудването) да възстановим откриття текст при зададен криптотекст.

(*f*) *Адаптивна атака при избран криптотекст* (adaptive chosen ciphertext attack): това е атака при избран криптотекст, където изборът на всяка стъпка зависи от получените открити текстове при предните стъпки.

2.2 Криптанализ на субституционни шифри

Най-напред ще се спрем на криптанализа на някои субституционни шифри. При тях образът на всеки символ или рупа от символи след шифриране е еднозначно определен. Това позволява използването на статистическите свойства на езика, на който е написано съобщението. По-нататък приемаме, че откритият текст е на а нглийски език и е написан без пунктуация и интервали. Общата схема на атака срещу проста субституция по даден криптотекст следната:

(1) изследват се статистическите характеристики на криптотекста;

(2) сравняват се със съответните характеристики на “типичен” английски текст;

(3) тези характеристики трябва да са “близки”.

Приликата се очаква да пасте с увеличаване на дължината на наличния криптотекст, шифриран с един и същи ключ. Най-очевидната характеристика е честотата на буквите от английската азбука. По-долу даваме таблица на честотите на буквите в текст на английски език.

буква	честота	буква	честота	буква	честота
A	0.082	J	0.002	S	0.063
B	0.015	K	0.008	T	0.091
C	0.028	L	0.040	U	0.028
D	0.043	M	0.024	V	0.010
E	0.127	N	0.067	W	0.023
F	0.022	O	0.075	X	0.001
G	0.019	P	0.019	Y	0.020
H	0.061	Q	0.001	Z	0.001
I	0.070	R	0.060		

За удобство те се разделят на три групи според честотата на появяване в английски текст.

Висока честота		Средна честота		Ниска честота	
E	0.127	D	0.043	G	0.018
T	0.091	L	0.040	B	0.015
A	0.082	C	0.028	V	0.010
O	0.075	U	0.028	K	0.008
I	0.070	M	0.024	J	0.002
N	0.067	W	0.023	Q	0.001
S	0.063	F	0.022	X	0.001
H	0.061	Y	0.020	Z	0.001
R	0.060	P	0.020		

При криптанализ на текст, шифриран със субституционен шифър, се оказва полезно и знанието на най-често срещаните диграфи и триграфи. Тук представяме списък на най-често срещаните диграфи и триграфи, както и таблица с техните честоти, получени при преброяване на буквите във вестник с над 80000 символа.

Диграфи: TH HE IN ER AN RE ED ON ES ST
EN AT TO NT HA ND OU EA NG AS
OR TI IS ET IT AR TE SE HI OF

Триграфи: THE ING AND HER ERE ENT
THA NTH WAS ETH FOR DTH

TH	2161	ED	890	OF	731	THE	1771	TER	232
HE	2053	TE	872	IT	704	AND	483	RES	219
IN	1550	TI	865	AL	681	TIO	384	ERE	212
ER	1436	OR	861	AS	648	ATI	287	CON	206
RE	1280	ST	823	HA	646	FOR	284	TED	187
ON	1232	AR	764	NG	630	THA	255	COM	185
AN	1216	ND	761	CO	606				
ET	1029	TO	756	SE	595				
AT	1019	NT	743	ME	573				
ES	917	IS	741	DE	572				

Да разглдаме три криптотекста, получени от един и същи открит текст с дължина 165, получени чрез използване на

(1) проста субституция;

(2) шифър на Vigenére с ключ с дължина $m' = 3$;

(3) шифър на Vigenére с ключ с дължина $m'' = 6$;

Открит текст:	thepathoftherighteousmanisbesetonallsidesbytheinequitie
Криптотекст 1:	OINMLOIFUOINAPBIONFVHRLYPHSNHNOFYLKKHPGNHSXOINPYNTVPOPN
Криптотекст 2:	WVKSOZKCLWVKUWMKHKRAYPOTLGHHGKWCTDZRVWJHGHBHNHWTHEALHOH
Криптотекст 3:	VPTWEKJWUALVTQVOXVQCETEEKAQLWVVWCHPCUQSLWSABWLMEGYJPXZG

Открит текст:	softheselfishandthetyrannyofevilmanblessedishewhointhen
Криптотекст 1:	HFUOINHNKUPHILYGOINOXALYYXFUNCPKRLYSKNHHNGPHINJIFPYOINY
Криптотекст 2:	GUIHNHGKOTOVVGQRZKSZBFGQBERTKYWRPOTEZKVGKGWYKSCKCOQHNHB
Криптотекст 3:	ADMXYGATSJZUPPUHKJMIFVRPVNVJVXQATEEDTTZWVFQGOINJWXUXYGV

Открит текст:	ameofgoodwillshepherdstheweakthroughthevalleyofdarkness
Криптотекст 1:	LRNFUBFFGJPKKHINMINAGHOINJNLDOIAFVBIOINCLKKNXFUGLADYNHH
Криптотекст 2:	GPSUIUURRCLZRVVKSVKURYWVKZSGNHNUCAJVZKSBDZRHMUIRGUYTHGY
Криптотекст 3:	PTIFHODVHNKTAZLVRPTYHJVPTDIRMBWYSLIPIOIMCTALCFLHPYOEGAH

Честотите на буквите във всеки от трите криптотекста са представени в следната таблицата.

	1	2	3		1	2	3
A	5	3	9	N	25	5	3
B	3	5	2	O	14	7	5
C	2	6	5	P	11	3	10
D	2	2	4	Q	0	3	6
E	0	3	8	R	3	11	3
F	11	1	4	S	3	7	4
G	6	13	6	T	1	7	12
H	15	17	7	U	6	9	5
I	17	3	7	V	3	11	15
J	3	2	7	W	0	9	9
K	9	17	4	X	4	0	6
L	10	4	8	Y	10	6	6
M	2	2	5	Z	0	9	4

Да се спрем на първия криптотекст. В него високочестотните букви са

N	25	F	11
I	17	P	11
H	15	L	10
O	14	Y	10
		K	9

Освен това триграфът `OIN` se poyavyava 8 пъти, а диграфите `OI` и `IN` – съответно 9 и 10 пъти. Така с доста голяма сигурност можем да приемем, че

$$\texttt{OIN} \to \texttt{the}.$$

2.3 Криптанализ на полиалфабетни шифри

Полиалфабетните шифри от тип Vigenère или Beaufort се считат за абсолютно сигурни до средата на XIX в., когато пруският офицер F. W. Kasiski [1] предлага стартегия за атака срещу тези шифри. Ние ще опишем криптанализа на шифъра на Vigenère, използвайки и по–късни идеи на Kerckhoff и Friedman.

Нека е дадена случайна редица от букви, от която избираме по случаен начин две. Вероятността тези букви да съвпадат е $\sum_\alpha \left(\frac{1}{26}\right)^2 = \frac{1}{26} \approx 0.0385$. Нека сега изберем по случаен начин две букви от потенциално "безкраен" английски текст (по точно от безкрайна редица от букви, в която те се появяват с теоретичните си вероятности от таблица ?? от предния раздел). Вероятността тези букви да съвпаднат сега е $\sum_\alpha p^2(\alpha) \approx 0.065$. Можем да заключим, че при сравняване на два шифрирани текста очакваното количество съвпадения на букви е 7 на 100, ако те са шифрирани при изболзването на една и съща азбука и 4 на 100, ако са използвани различни азбуки. Това просто наблюдение стои в основата на метода на Kasiski за определяне на дължината на ключа.

Да разгледаме криптотекст $c_0c_1\dots c_{n-1}$ с дължина n, шифриран по метода на Vigenère's. Да означим с f_α броя на появяванията на буквата α в криптотекста. вероятността да изберем две еднакви букви е

$$I_C = \frac{\sum_\alpha f_\alpha(f_\alpha - 1)}{n(n-1)}. \tag{2.1}$$

Чилсото I_C наричаме индекс на съвпаденията. [2] Да допуснем, че използваната дължина на използвания ключ е m и да запишем криптотекста във вида

$$\begin{array}{llll} c_0 & c_m & c_{2m} & \cdots \\ c_1 & c_{m+1} & c_{2m+1} & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ c_{m-1} & c_{2m-1} & c_{3m-1} & \cdots \end{array} \tag{2.2}$$

[1] Friedrich W. Kasiski е роден на 29.10.1805 в Шлохау, Източна Прусия. През 1822 г. той постъпва на служба в източнопруския 33. пехотен полк Граф Роон. там той служи до 1852 г., когато се уволнява с чин майор. През 1863 г. реномираното берлинско издателство Mittler & Sohn публикува книгата му "Die Geheimschriften und die Dechiffrierkunst". Този кратък текст (95 стр.) води до революция в криптологията, но става известен едва след смъртта на автора си на 22.05.1881 г.

[2] То е въведено от William Friedman.

Символите, появяващи се в един и същи ред се шифрират при използването на една и съща азбука. Ще преброим по два начина очаквания брой на двойките позиции от криптотекста, съдържащи идентични символи. От една страна този брой е

$$\frac{1}{2}\sum_{\alpha} f_\alpha(f_\alpha - 1) = \frac{1}{2}n(n-1)I_C. \tag{2.3}$$

От друга страна нека първо изберем произволен символ от криптотекста (това може да се случи по n начина), а след това и втори символ. Ако вторият символ е в реда, съдържащ първия, вероятността за съвпадение е ≈ 0.065. Ако вторият символ е в друг ред вероятността е ≈ 0.038. Така ние очакваме $\approx \frac{1}{2}n(\frac{n}{m} - 1) \times 0.065$ двойки идентични символи, появяващи се в един и същи ред и $\approx \frac{1}{2}n(n - \frac{n}{m}) \times 0.038$ двойки от идентични символи, появяващи се в различни редове. Следователно,

$$\frac{1}{2}n(n-1)I_C \approx \frac{1}{2}n(\frac{n}{m} - 1) \times 0.065 + \frac{1}{2}n(n - \frac{n}{m}) \times 0.038, \tag{2.4}$$

откъдето

$$m \approx \frac{0.027n}{I_C(n-1) - 0.038n + 0.065}. \tag{2.5}$$

За съжаление тази формула не е много полезна, тъй като не дава точен резултат (особено за големи стоиу0ности на m). Това се вижа и от таблицата по-долу.

m	1	2	5	10	∞
I_C	0.065	0.052	0.043	0.041	0.038

Да развием по-подробно тази идея. Нека

$$\begin{aligned} M_1 &= (M_{1,0}, M_{1,1}, \ldots, M_{1,n-1}), \\ M_2 &= (M_{2,0}, M_{2,1}, \ldots, M_{2,n-1}) \end{aligned}$$

са две редици от независими еднакво резпределени случайни променливи над q-буквена азбуката $\mathbb{Z}_q$. Случайната величина $M_{i,j}$ приема стойност m с вероятност

$$P(M_{i,j} = m) = p(m), \quad 0 \le m < q, i = 1, 2, j = 0, 1, \ldots, n-1.$$

Дефинираме

$$k[M_1, M_2] = |\{j \mid 0 \le j < n : M_{1,j} = M_{2,j}\}|.$$

Ясно е, че

$$P_{plain}(M_{1,j} = M_{2,j}) = \sum_{m \in \mathbb{Z}_q} P_{plain}(M_{1,j} = M_{2,j} = m) = \sum_{m} p^2(m).$$

Нека $G \subset \S_q$ е подмножество от пермутации на елементите от азбуката $\mathbb{Z}_q$. Да разгледаме редиците от еднакво разпределени случайни величини,

$$\Pi^{(i)} = (\Pi_0^{(i)}, \ldots, \Pi_{n-1}^{(i)}), i = 1, 2,$$

вземащи стойности от G и мащи разпределение

$$P_{key}(\Pi_j^{(i)} = \pi) = q(\pi).$$

Ще шифрираме редиците M_1 и M_2, използвайки съответно $\Pi^{(1)}$ и $\Pi^{(2)}$. Означаваме образите на M_1 и M_2 чрез $C_1 = (C_{1,0}, C_{1,1}, \ldots, C_{1,n-1})$ и $C_2 = (C_{2,0}, C_{2,1}, \ldots, C_{2,n-1})$. за всяко $i = 1, 2$ и всяко $0 \leq j \leq n-1$ имаме

$$P_{cipher}(C_{i,j} = c) = \sum_{\pi \in G} q(\pi) p(\pi^{-1}(c)).$$

Налице са две възможни хипотези

H_0: M_1 и M_2 се шифрират при използването на идентични редици от субституции, т.е. $\Pi_1 = \Pi_2 = \Pi$;

H_1: M_1 и M_2 се шифрират при използването на две различни редици от субституции, т.е. $\Pi^{(1)} \neq \Pi^{(2)}$.

Очевидно имаме

$$\begin{aligned} P_{cipher}(C_{1,j} = C_{2,j} \mid H_0) &= \sum_c P(C_{1,j} = C_{2,j} = c \mid H_0) \\ &= \sum_c \sum_{\pi \in G} q(\pi) p^2(\pi^{-1}(c)) \\ &= \sum_{\pi \in G} q(\pi) \sum_{c \in \mathbb{Z}_q} p^2(\pi^{-1}(c)) \end{aligned}$$

Ако c пробягва $\mathbb{Z}_q$, то и $\pi^{-1}(c)$ пробягва $\mathbb{Z}_q$, откъдето

$$\begin{aligned} Pr_{cipher}\{C_{1,j} = C2, j \mid H_0\} &= \sum_{\pi \in G} q(\pi) \sum_m p^2(m) \\ &= \sum_m p^2(m). \end{aligned}$$

От друга страна имаме

$$\begin{aligned} Pr_{cipher}(C_{1,j} = C_{2,j} \mid H_1) &= \sum_c P_{cipher}(C_{1,j} = C_{2,j} = c \mid H_1) \\ &= \sum_c \sum_{\pi_1, \pi_2 \in G} q(\pi_1) q(\pi_2) p(\pi_1^{-1}(c)) p(\pi_2^{-1}(c)) \\ &= \sum_c \left(\sum_{\pi_1 \in G} q(\pi_1) p(\pi_1^{-1}(c)) \right) \left(\sum_{\pi_2 \in G} q(\pi_2) p(\pi_2^{-1}(c)) \right) \\ &= \sum_c \left(\sum_{\pi \in G} q(\pi) p(\pi^{-1}(c)) \right)^2 \\ &= \sum_c Pr^2_{cipher}\{C = c\}. \end{aligned}$$

Ако допуснем, че елементите на G са равновероятни, получаваме

$$\begin{array}{rcl} P_{cipher}(C_{1,j} = C_{2,j} \mid H_0) & = & 0.06875, \\ P_{cipher}(C_{1,j} = C_{2,j} \mid H_1) & = & 0.03846, \end{array}$$

т.е. очакваната стойност за $k[C_1, C_2]$ е $0.06875n$ при хипотезата H_0 и $0.03846n$ при хипотезата H_1.

Ще използваме направените наблюдения при крптанализа на криптотекст, шифриран с шифъра на Vigenère. Означаваме открития текст, ключа и криптотекста с

$$\begin{array}{rcl} m & = & (m_0, m_1, \ldots, m_{n-1}), \\ \pi & = & (\pi_0, \pi_1, \ldots, \pi_{r-1}), \\ c & = & (c_0, c_1, \ldots, c_{n-1}), \end{array}$$

където $c_i = \pi_{i \pmod r}(m_i)$. Да дефинираме редиците

$$\begin{array}{rcl} c^{(s)} & = & (c_0, c_1, \ldots, c_{n-s-1}), \\ {}^{(s)}c & = & (c_s, c_{s+1}, \ldots, c_{n-1}). \end{array} \tag{2.6}$$

Разсъжденията по-горе водят до следния резултат.

Теорема 2.1. Очакваната стойност на $k[{}^{(s)}c, c^{(s)}]$ е

$$E(k[{}^{(s)}c, c^{(s)}] = \begin{cases} (n-s)\sum_m p^2(m) & \text{ако } r \text{ дели } s; \\ (n-s)\sum_c P^2\{C = c\} & \text{ако } r \text{ не дели } s. \end{cases}$$

При горните ограничения за G и $q(\pi)$ можем да очакваме, че стойността на $k[{}^{(s)}c, c^{(s)}]/(n-s)$ е близо до $\sum_m p^2(m) \approx 0.065$ ако r дели s и близо до $\sum_c Pr\{C = c\} \approx 0.0385$ ако r не дели s. Така стойността на r може да бъде определена като намерим онези стойности за s, при които $k[{}^{(s)}c, c^{(s)}]/(n-s)$ е близо до 0.065. Със следващия пример илюстрираме тези идеи.

Много по-полезен при определяне на дължината на ключа е т.нар. *тест на Касиски*. Той почива на следното наблюдение. Ако два идентични сегмента открит текст се шифрират в идентични криптотекстове, то разстоянието между тях се дели на дължината на ключа. Обратно, ако наблюдаваме идентични учстъци от криптотекста с дължина поне 3, то много вероятно е те да се получават от идентични открити текстове.

Пример 2.2. По-долу са дадени два криптотешта шифриращи едно и също съобщение:

Открит текст	codebreakingisthemostimpo
Криптотекст 1:	FRGHEUHDNLQJLVWKHPRVWLPSR
Криптотекст 2:	OOBQBPQAIUNEUSRTEKASRUMNA
Открит текст	rtantformofsecretintellig
Криптотекст 1:	UWDQWIRUPRIVHFUHWLQWHOOLJ
Криптотекст 2:	RRMNRROPYODEEADERUNRQLJUG
Открит текст	enceintheworldtodayitprod
Криптотекст 1:	HQFHLQWKHZRUOGWRGDBLWSURG
Криптотекст 2:	CZCCUNRTEUARJPTMPAWUTNDOB
Открит текст	ucesmuchmoreandmuchmoretr
Криптотекст 1:	XFHVPXFKPRUHDQGPXFKPRUHWU
Криптотекст 2:	GCCEMSOHKARCMNBYUATMMDERD
Открит текст	ustworthyinformationthans
Криптотекст 1:	XVWZRUWKBLQIRUPDWLRQWKDQV
Криптотекст 2:	UQFWMDTFKILROPYARUOLFHYZS
Открит текст	piesandthisintelligenceex
Криптотекст 1:	SLHVDQGWKLVLQWHOOLJHQFHHA
Криптотекст 2:	NUEQMNBFHGEILFEJXIEQNAQEV
Открит текст	ertsgreatinfluenceuponthe
Криптотекст 1:	HUWVJUHDWLQIOXHQFHXSRQWKH
Криптотекст 2:	QRREGPQARUNDXUCZCCGPMZTFQ
Открит текст	policiesofgovernmentsyeti
Криптотекст 1:	SROLFLHVRIJRYHUQPHQWVBHWL
Криптотекст 2:	PMXIAUEQAFEAVCDNKQNREYCFI
Открит текст	thasneverhadachronicler
Криптотекст 1:	WKDVQHYHUKDGDFKURQLFOHU
Криптотекст 2:	RTAQZETQRFMDYOHPANGOLCD

Криптотекст 1 е получен при шифриране с моноалфабетен шифър, докато за криптотекст 2 е използван Vigenère с трибуквена ключова дума: MAY. Най-напред да преброим честотитте на буквите в двата криптотекста.

(TWO HISTOGRAMS)

на първата хистограма разпознаваме типичните черти на моноалфабетен шифър. Една буква се среща значително по-често от останалите (най-вероятно криптотекста за e) докато три букви не се появяват въобще (най-вероятно образите на три от буквите v,k,j,x,q,z). Втората хистограма е много по-плоска в смисъл, че никоя буква не доминира останалите, всички букви се появяват в криптотекста и има много-помалка разлика между най-често и най-рядко срещащите се.

Една възможна мярка за това колко плоска е една хистограма е вариацията

$$\sum_{\mathtt{A}}^{\mathtt{Z}}(p_\alpha - \frac{1}{26})^2.$$

Лесно получаваме

$$\sum_{\mathtt{A}}^{\mathtt{Z}}(p_\alpha - \frac{1}{26})^2 = \sum_{\mathtt{A}}^{\mathtt{Z}} p_\alpha^2 - \frac{1}{26} \approx \sum_{\mathtt{A}}^{\mathtt{Z}} p_\alpha^2 - 0.038.$$

Сега ще илюстрираме техниките за определяне на дължината на ключа при втория криптотекст. Нека запишем веднъж криптотекст 2 и веднага след това да го запишем втори път отмествайки го една позиция вдясно, т.е. първата буква под втората, втората под третатта и т.н. Да преброим броя на съвпаденията в двете криптограми. Повтаряме операцията, отмествайки криптотекста на две позиции, след това на три позиции и т.н. По този начин получаваме статистика, илюстрираща връзката между отместването и броя на съвпаденията. При отместване 1 адитивният шифър, използван в първия ред е различен от адитивния шифър, използван за втория. Ако разгледаме позициите, в които криптотекстовете се пресичат, не можем да очакваме извънредно голям брой съвпадения. Същото е в сила и за отместване на две позиции. При отместване на три позиции (което е и дължината на ключа) във всяка позиция, ǩоято криптотекстовете се пресичат, е използван един и същ адитивен шифър. Следователно тук трябва да очакваме значително по-голям брой съвпадения. Същото ще е в сила и за всички отмествания, които са кратни на 3. Разбира се при много голямо отместване двата криптотекста се пресичат в много по-малък брой букви и съвпаденията ще намаляват. В таблицата по-долу са представени отместванията и съвпаденията за криптограма 2.

D	C	%	D	C	%	D	C	%
1	9	4.054054	42	26	14.364461	83	5	3.571429
2	3	1.357466	43	6	3.333333	84	7	5.035971
3	14	6.363636	44	2	1.117318	85	5	3.623188
4	10	4.566210	45	7	3.932584	86	4	2.919708
5	13	5.963303	46	10	5.649718	87	9	6.617647
6	19	8.755760	47	9	5.113636	88	3	2.222222
7	5	2.314815	48	8	4.571429	89	9	6.716418
8	9	4.186047	49	7	4.022988	90	12	9.022556
9	11	5.140187	50	9	5.202312	91	14	10.606061
10	8	3.755869	51	3	1.744186	92	3	2.290076
11	9	4.245283	52	5	2.923977	93	4	3.076923
12	10	4.739336	53	4	2.352941	94	2	1.550388
13	7	3.333333	54	9	5.325444	95	7	5.468750
14	10	4.784689	55	7	4.166667	96	5	3.937008
15	13	6.250000	56	9	5.389222	97	7	5.555556
16	8	3.864734	57	16	9.638554	98	4	3.200000
17	8	3.883495	58	4	2.424242	99	4	3.225806
18	10	4.878049	59	6	3.658537	100	2	1.626016
19	4	1.960784	60	15	9.202454	101	5	4.098361
20	9	4.433498	61	7	4.320988	102	16	13.223140
21	11	5.445545	62	5	3.105590	103	4	3.333333
22	8	3.980099	63	6	3.750000	104	1	0.840336
23	4	2.000000	64	7	4.402516	105	9	7.627119
24	13	6.532663	65	7	4.430380	106	6	5.128205
25	10	5.050505	66	9	5.732484	107	7	6.034483
26	10	5.076142	67	6	3.846154	108	10	8.695652
27	13	6.632653	68	7	4.516129	109	3	2.631579
28	4	2.051282	69	4	2.597403	110	4	3.539823
29	8	4.123711	70	8	5.228758	111	12	10.714286
30	12	6.217617	71	4	2.631579	112	7	6.306306
31	12	6.250000	72	11	7.284768	113	3	2.727273
32	7	3.664921	73	7	4.666667	114	6	5.504587
33	19	10.000000	74	3	2.013423	115	7	6.481481
34	7	3.703704	75	6	4.054054	116	5	4.672897
35	8	4.255319	76	1	0.680272	117	11	10.377358
36	14	7.486631	77	6	4.109589	118	2	1.904762
37	9	4.838710	78	10	6.896552	119	6	5.769231
38	7	3.783784	79	6	4.166667	120	3	2.912621
39	8	4.347826	80	3	2.097902	121	4	3.921569
40	4	2.185792	81	13	9.154930	122	8	7.920792
41	7	3.846154	82	7	4.964539	123	8	8.000000

Тъй като ние всъщност знаем, че е използван ключ с дължина 3, очакваме да видим по-висока степен на съвпадане при отмествания, които са кратни на 3. Таблицата оправдава тези очаквания. Въпреки това можем да добием допълнителна увереност в дължината на ключа, ако извършим и някои допълнителни пресмятания по тази таблица. Ако означим дължината на клича с p, то при отместване, което не е кратно на p, очакваме приблизително 3.8% съвпадения, докато при отместване кратно на p, очакваме този процент да надхвърля 6.5%. В таблицата са представени отместванията, за които имаме голям процент на съвпаденията (в случая поне 8%).

съвпадения	отместване	разлагане
8.76%	6	2×3
10.00%	33	11×3
14.36%	42	$7 \times 2 \times 3$
9.64%	57	19×3
9.20%	60	$5 \times 2^2 \times 3$
9.15%	81	3^4
9.02%	90	$5 \times 2 \times 3^2$
10.60%	91	13×7
13.22%	102	$17 \times 2 \times 3$
8.70%	108	$2^2 \times 3^2$
10.71%	111	37×3
10.38%	117	13×3^2
8.00%	123	41×3

Като изключим отместване 91, всички отмествания имат множител 3. Всъщност 3 техният най-голям общ делител. Това потвърждава нашата увереност, че дължината на ключовата дума е 3.

Втората техника, която използваме е намирането на идентични редици от букви в криптотекста. Две идентични реици от символи в открития текст се шифрират различно в общия случай. Но ако позициите им в открития текст са такива, че първия символ на всяка от тях се шифрира с една и съща буква от ключа, те ще се шифрират в един и същи криптотекст. Така, ако разглеждайки криптограма открием идентични редици от символи, то е много вероятно разстоянието между тях да е кратно на дължината на ключа. Процесът на намиране на такива повтарящи се редици е известен като тест на Касиски. Да илюстрираме теста на Касиски с нашия криптотекст.

редица	разстояние	разлагане
`PQA`	150	$2 \times 5^2 \times 3$
`RTE`	42	$2 \times 7 \times 3$
`ROPY`	81	3^4
`DER`	57	19×3
`RUN`	117	13×3^2
`UNR`	12	$2^2 \times 3$
`CZCC`	114	$2 \times 19 \times 3$
`MNB`	42	$2 \times 7 \times 3$
`ARU`	42	$2 \times 7 \times 3$
`UEQ`	54	2×3^3

Отново най-големият общ делител на разстоянията е 3, навеждайки на мисълта, че дължината на ключа е наистина 3. Ако са необходими още доказателства, можем да запишем криптограмата в три реда, като в ред 1 са буквите в позиции 1,4,7,..., въвтория – тези в позиции 2,5,8,..., и в третия – тези в позиции 3,6,9,..., и да пресметнем индекса на съвпаденията за всеки от редовете. Получените резултати са

$$I_C(\text{ред } 1) = 0.0717117, I_C(\text{ред } 2) = 0.0636801, I_C(\text{ред } 1) = 0.0640504.$$

We assume that a Vigenere cipher has been used. Let two cihpertexts enciphered using a monoalphabetic cipher are given. Let their respective lengths be n and n'. The length of the combined cryptogram is $n + n'$. Further denote the frequencies of the letter λ in both cryptograms by f_λ and $f_{\lambda'}$. The incidence of coincidences for the combined cryptogram is

$$\frac{\sum_\lambda (f_\lambda + f'_\lambda)(f_\lambda + f'_\lambda - 1)}{(n+n')(n+n'-1)} = \frac{\sum_A^Z f_\lambda^2 + \sum_A^Z f'^2_\lambda + 2\sum_A^Z f_\lambda f'_\lambda - n - n'}{(n+n')(n+n'-1)}. \tag{2.7}$$

If we encipher the second message withanother monoalphabetic cipher the letter frequencies will change. Denote the new frequencies by g'_λ. In the above expression $\sum_A^Z f_\lambda f'_\lambda$ will be replaced by $\sum_A^Z f_\lambda g'_\lambda$. So, identical substitutions will result in a high value of I_C. We will have to look for such shifts which maximize $\sum f_\lambda g'_\lambda$. (Therefore, rows 1 and 2 are obtained by a shift of 14; rows 2 and 3 – by a shift of 24; rows 3 and 1 – by a shift of 12. Note that $14 + 24 \equiv 12 \pmod{26}$.)

2.4 Задачи

2.1 Направете криптанализ на четирите текста по-долу, имайки предвид, че първите три са шифрирани съответно чрез използване на афинен шифър, общ субституционен шифър и шифър на Vigeneère. Шифърът, използван при получаване на четвъртия криптотекст е неизвестен.

(a) афинен шифър

```
KQERE JEBCP PCJCR KIEAC UZBKR VPKRB CIBQC ARBJC VFCUP KRIOF
KPACU ZQEPB KRXPE IIEAB DKPBC PFCDC CAFIE ABDKP BCPFE QPKAZ
BKRHA IBKAP CCIBU RCCDK DCCJC IDFUI XPAFF ERBIC ZDFKA BICBB
ENEFC UPJCV KABPC YDCCD PKBCO CPERK IVKSC PICBR KIJPK ABI
```

(b) общ субституционен шифър

```
EMGLO SUDCG DNCUS WYSFH NSFCY KDPUM LWGYI COXYS IPJCK QPKUG
KMGOL ICGIN CGACK SNISA CYKZS CKXEC JCKSH YSXCG OIDPK ZCNKS
HICGI WYGKK GKGOL DSILK GOIUS IGLED SPWZU GFZCC NDGYY SFUSZ
CNXEO JNCGY EOWEU PXEZG ACGNF GLKNS ACIGO IYCKX CJUCI UZCFZ
CCNDG YYSFE UEKUZ CSOCF ZCCNC IACZE JNCSH FZEJZ EGMXC YHCJU
MGKUC Y
```

(c) шифър на Vigenère

```
KCCPK BGUFD PHQTY AVINR RTMVG RKDNB VFDET DGILT XRGUD DKOYF
MBPVG EGLTG CKQRA CQCWD NAWCR XIZAK FTLEW RPTYC QKYVX CHKFT
```

```
PONCQ QRHJV AJUWE TMCMS PKQDY HJVDA HCTRL SVSKC GCZQQ DZXGS
FRLSW CWSJT BHAFS IASPR JAHKJ RJUMV GKMIT ZHFPD ISPZL VLGWT
FPLKK EBDPG CEBSH CTJRW XBAFS PEZQN RWXCV YCGAO NWDDK ACKAW
BBIKF TIOVK CGGHJ VLNHI FFSQE SVYCL ACNVR WBBIR EPBBV FEXOS
CDYGZ WPFDT KFQIY CWHJV LNHIQ IBTKH JVNPI ST
```

(d) неизвестен шифър

```
BNVSN SIHQC EELSS KKYER IFGKX UMBGY KAMQL JTYAV FBKVT DVBPV
VRJYY LAOKY MPQSC GDLFS RLLPR OYGES EBUUA LRWXM MASAZ LGLED
FJBZA VVPXW ICGJX ASCBY EHOSN MULKC EAHTQ OKMFL EBKFX LRRFD
TZXCI WBJSI CBGAW DVYDH AVFJX ZIBKC GJIWE AHTTO EWTUH KRQVV
RGZBX YIREM MASCS PBLNH JMBLR FFJEL HWEYL WISTF VVYFJ CMHYU
YRUFS FMGES IGRLW ALSWM NUHSI MYYIT CCQPZ SICEH BCCMZ FEGVJ
YOCDE MMPGH VAAUM ELCMO EHVLT IPSUY ILVGF LMVWD VYDBT HFRAY
ISYSG KVSUU HYHGG CKTMB LRX
```

2.2 (a) Намерете броя на обратимите 2×2 матрици над $\mathbb{Z}_{26}$.

(b) Намерете броя на обратимите 2×2 матрици над $\mathbb{Z}_p$, където p е просто число.

(c) Намерете броя на обратимите матрици от ред $m \geq 2$ над $\mathbb{Z}_p$, където p е просто число.

(d) Намерете броя на обратимите 2×2 матрици K над $\mathbb{Z}_{26}$, за които $K = K^{-1}$. (Такива матрици се наричат инволюционни. Използването на шифъра на Хил с такива матрици е много удобно, тъй като в този случай шифрирането съвпада с дешифрирането.)

2.3 Известно е, че откритият текст `conversation` се шифрира в криптотекста `HIARRTNUYTUS` при използване на шифъра на Хил с неизвестно m. Определете шифриращата матрица.

2.4 Дешифрирайте следния криптотекст получен от шифъра “Автоключ” чрез пълно изчерпване на всички ключове.

```
MALVV MAFBH BUQPT SOXAL TGVWW RG
```