https://doi.org/10.53656/math2022-1-4-kri

Educational Technologies
Образователни технологии

КРИПТОГРАФИЯ И КРИПТОАНАЛИЗ С MS EXCEL

Гл. ас. д-р Деян Михайлов
Икономически университет – Варна

Резюме. В статията са дадени примери за реализация в средата на MS Excel на три добре познати криптографски алгоритъма – шифрите на Цезар, на Виженер и на Хил. Демонстрирано е успешно разбиване на тези шифри с атака с груба сила, честотен анализ и с известни двойки съобщение-криптограма. За целите на криптоанализа са определени относителните честоти на буквите и е изчислен индексът на съвпадение (K) за съвременния български език. Класическият метод за честотен криптоанализ е допълнен с оценка на взаимната корелация между честотата на буквите в естествения език и в криптирания текст. Показани са действия с матрици по модул в средата на MS Excel.

Ключови думи: криптоанализ; MS Excel

Увод

При изучаването на коя да е предметна област има две главни направления – придобиване на теоретични знания и усвояване на практически умения под формата на решаване на задачи или провеждане на експеримент. Характерно за областта на криптографията и криптоанализа е, че решаването на задача или провеждането на експеримент включват голям брой операции, много от които се повтарят многократно. Решаването им „на ръка, с молив и хартия“ изисква много време и е непривлекателно. Поради това е целесъобразно да се използват подходящи програмни средства.

В настоящата статия се предлага използване на MS Excel. В него има редица функции, които са подходящи за реализиране на криптографски преобразувания. Excel дава възможност за нагледно демонстриране на целия процес на криптиране или криптоанализ на информацията.

Представени са три вида атаки срещу криптографски системи – атака с груба сила (Brute-force Attack), честотен анализ (Frequency Analysis Attack) и атака с известни двойки съобщение – криптограма (Known-plaintext Attack).

Действието на атаките е показано върху три от т.нар. исторически шифри – на Цезар (Caesar), на Виженер (Vigenere) и на Хил (Hill). Те не се при-

лагат в чистия си вариант, но рационалните им идеи се използват в редица съвременни криптографски алгоритми. Цикличното изместване по реда на буквите в азбуката в шифрите на Цезар и на Виженер е сумиране по модул, което се използва в съвременните симетрични шифри. Идеята на Хил за умножение с матрици по модул намира приложение в базираните на алгоритъма Rijndael криптосистеми.

В изложението се счита, че методите за защита на информацията чрез шифриране или криптиране са предмет на криптографията и се прилагат от криптографи, а методите за преодоляване на тази защита са предмет на криптоанализа и се прилагат от криптоаналитици.

Приема се, че криптоаналитикът има потенциален достъп до всички криптограми, тъй като те се предават по незащитени комуникационни канали. В съответствие с известния принцип на Керкхофс, криптоаналитикът знае алгоритъма за криптиране, но не знае ключа, с който е криптирано съответното съобщение (Petitcolas 2011).

Атака с груба сила

Атаката с груба сила върху една криптографска система е изпробване на всички възможни ключове до намиране на такъв, който декриптира прехванатата криптограма в единствено възможно смислено съобщение.

Колкото е по-малък броят на възможните ключове, толкова по-малко време ще бъде необходимо за тяхното изпробване. Като пример е избрана атака върху един от най-простите шифри – шифъра на Цезар. Цезар е използвал тайнопис, основан на изместване от ляво надясно на буквите на текста на съобщението на три позиции спрямо реда им в азбуката. Така A се преобразува в D, B в F, C в G и т.н. Ако се достигне краят на азбуката, се продължава от първата буква, т.е. изместването е циклично. Декриптирането се състои в обратно връщане на буквите от криптограмата на същите три позиции от дясно наляво, при което се получава текстът на съобщението.

В по-широк смисъл под шифър на Цезар се разбира изместване на буквите на произволен брой позиции, като броят на възможните измествания е равен на броя на буквите в азбуката. Всъщност това са всички възможни ключове. Малкият им брой дава възможност шифърът на Цезар успешно да се атакува с груба сила. Шифърът на Цезар по „елегантен начин“ може да бъде представен с помощта на модулната аритметика (Paar & Pelzl 2010).

Ако с m се означи номерът от азбуката на буквите от съобщението, с k –ключът, т.е. броят на позициите, на които се изместват буквите, с c – номерът от азбуката на буквите от криптограмата, и с N – броят на буквите в азбуката, то моделът на криптиращата функция е

$$c = E(m) \equiv (m + k) (\mathrm{mod}\, N), \tag{1}$$

а на декриптиращата

$$m = D(c) \equiv (c + (N - k)) (\mathrm{mod}\, N). \tag{2}$$

Шифърът на Цезар е моноазбучен, т.е. на една и съща буква от съобщението съответства една и съща буква от криптирания текст.

Криптирането лесно се реализира в средата на Excel (фиг. 1). Използван е български език. Криптират се само буквите от текста, без интервалите. Дължината на съобщението е ограничена до 30 символа (вкл. интервалите), като няма пречка да се увеличи. Използва се помощен масив за преобразуване на буквите от съобщението в номера от азбуката и на номерата от азбуката в букви (клетки A9:C38). В клетка B1 се въвежда стойността на ключа. В клетка D1 се въвежда текстът за съобщението. В клетки H3:AK3 текстът на съобщението се разбива по букви с функцията MID. В клетки H4:AK4 буквите се преобразуват в номера, като се използват функцията VLOOKUP и помощният масив. В клетки H5:AK5 с функцията MOD се реализира криптирането (1). В клетки H6:AK6 с още едно използване на VLOOKUP се извършва преобразуване от номера в букви на криптограмата. Последната операция е да се формира сливане на буквите на криптограмата в текстов низ с функция CONCATENATE в клетка D2.

	A	B	C	D
1	Ключ=	11	Текст	РИМСКИ ЦИФРИ
2			Криптограма:	ЬУЧЮХУ ГУБЬУ

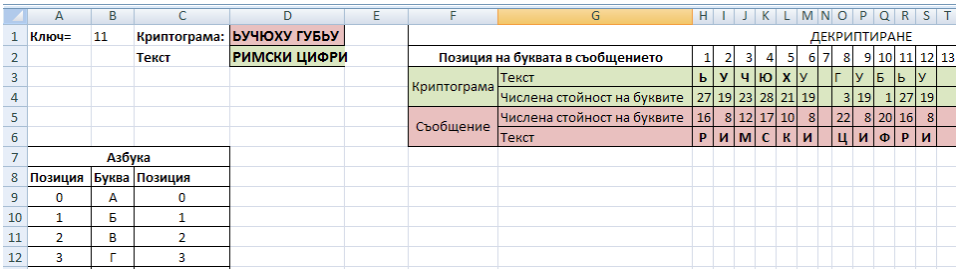
Азбука		
Позиция	Буква	Позиция
0	А	0
1	Б	1
2	В	2
3	Г	3

КРИПТИРАНЕ

Позиция на буквата в съобщението		1	2	3	4	5	6	7	8	9	10	11	12	13
Съобщение	Текст	Р	И	М	С	К	И		Ц	И	Ф	Р	И	
	Числена стойност на буквите	16	8	12	17	10	8		22	8	20	16	8	
Криптограма	Числена стойност на буквите	27	19	23	28	21	19		3	19	1	27	19	
	Текст	Ь	У	Ч	Ю	Х	У		Г	У	Б	Ь	У	

Фигура 1 . Криптиране на съобщение с шифъра на Цезар в Excel

По подобен начин се извършва и декриптирането (фиг. 2). Разликата е, че в клетка D1 се въвежда криптограмата, в клетка D2 се извежда декриптираният текст и в клетки H5:AK5 се използва функцията (2).



	A	B	C	D
1	Ключ=	11	Криптограма:	ЬУЧЮХУ ГУБЬУ
2			Текст	РИМСКИ ЦИФРИ

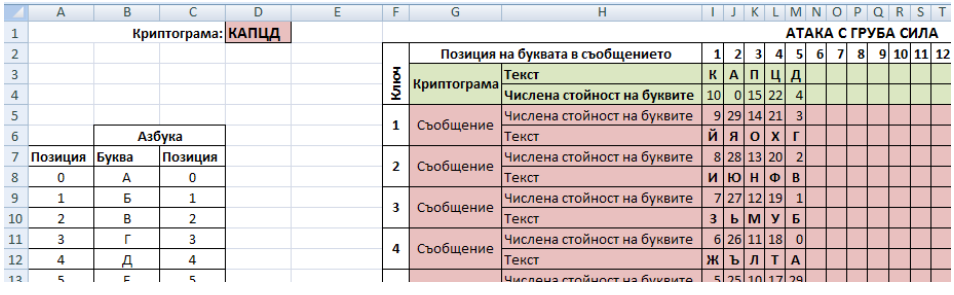
Азбука		
Позиция	Буква	Позиция
0	А	0
1	Б	1
2	В	2
3	Г	3

ДЕКРИПТИРАНЕ

Позиция на буквата в съобщението		1	2	3	4	5	6	7	8	9	10	11	12	13
Криптограма	Текст	Ь	У	Ч	Ю	Х	У		Г	У	Б	Ь	У	
	Числена стойност на буквите	27	19	23	28	21	19		3	19	1	27	19	
Съобщение	Числена стойност на буквите	16	8	12	17	10	8		22	8	20	16	8	
	Текст	Р	И	М	С	К	И		Ц	И	Ф	Р	И	

Фигура 2 . Декриптиране на съобщение с шифъра на Цезар в Excel

На основата на алгоритъма за декриптиране шифърът на Цезар може да се разбие. Прехванатата криптограма се декриптира последователно с ключ 1, 2, 3,... N-1 и се проверява дали при някое от декриптиранията не се получава смислено съобщение. На фиг. 3 е показано такова декриптиране на криптограмата КАПЦД. С ключ 4 се достига до смисленатa дума ЖЪЛТА.



Криптограма: КАПЦД

Азбука		
Позиция	Буква	Позиция
0	А	0
1	Б	1
2	В	2
3	Г	3
4	Д	4
5	Е	5

АТАКА С ГРУБА СИЛА

Ключ		Позиция на буквата в съобщението	1	2	3	4	5	6	7	8	9	10	11	12
	Криптограма	Текст	К	А	П	Ц	Д							
		Числена стойност на буквите	10	0	15	22	4							
1	Съобщение	Числена стойност на буквите	9	29	14	21	3							
		Текст	Й	Я	О	Х	Г							
2	Съобщение	Числена стойност на буквите	8	28	13	20	2							
		Текст	И	Ю	Н	Ф	В							
3	Съобщение	Числена стойност на буквите	7	27	12	19	1							
		Текст	З	Ь	М	У	Б							
4	Съобщение	Числена стойност на буквите	6	26	11	18	0							
		Текст	Ж	Ъ	Л	Т	А							
		Числена стойност на буквите	5	25	10	17	29							

Фигура 3 . Атака с груба сила върху шифъра на Цезар в Excel

На пръв поглед, всичко изглежда перфектно. Ако обаче продължим проверката, установяваме, че с ключ 22 се декриптира още една смислена дума – ТИЧАМ. Не е ясно кое точно е изходното съобщение. Следователно атаката с груба сила невинаги е успешна. Ако една и съща криптограма с различни ключове се преобразува в няколко смислени съобщения, се казва, че има лъжливи ключове. Именно лъжливите ключове са тези, които осигуряват мощността на криптографската система. Колкото са повече те, толкова е по-голяма криптоустойчивостта. При шифъра на Цезар такива лъжливи ключове са по-скоро изключение.

Шифърът на Цезар е много слаб. Успешното му използване се е дължало единствено на факта, че нивото на криптоанализа сред противниците на Цезар е клоняло към нула.

Честотен анализ

Ако броят на всички възможни ключове е твърде голям, атаката с груба сила изисква много време или много ресурси. В такива случаи обикновено се използва честотен анализ на съобщението.

Честотният анализ се основава на факта, че в естествените езици буквите се срещат с различна честота. В английския език например най-често срещаната буква е E, а в българския – А. Срещат се различни оценки на честотите, които се различават, затова тук ще се използват данни, получени при изследване честотата на буквите в първата част на романа „Тютюн“ от Димитър Димов (табл. 1).

Таблица 1. Относителна честота на буквите в българския език (анализ на романа „Тютюн“ от Д. Димов, том 1)

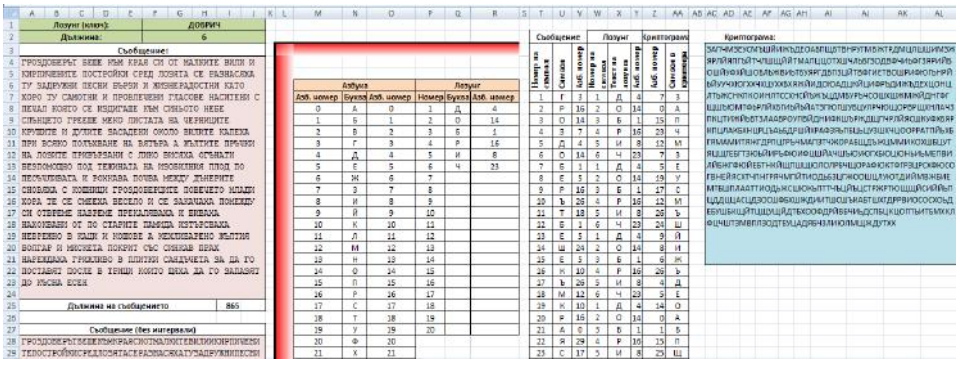
Буква	Честота	Място по честота	Буква	Честота	Място по честота	Буква	Честота	Място по честота
А	0.122572763	1	К	0.034892209	10	Ф	0.001843432	28
Б	0.017954904	18	Л	0.031443506	12	Х	0.008784139	23
В	0.043367047	9	М	0.025257493	14	Ц	0.004738486	27
Г	0.015777427	19	Н	0.065398534	6	Ч	0.013937088	21
Д	0.032931242	11	О	0.091469487	3	Ш	0.013436021	22
Е	0.091831369	2	П	0.028520615	13	Щ	0.006204571	26
Ж	0.007905725	24	Р	0.046583774	8	Ъ	0.017957997	17
З	0.021994371	15	С	0.049846896	7	Ь	0.000111348	30
И	0.08822183	4	Т	0.074665182	5	Ю	0.001784665	29
Й	0.006034456	25	У	0.015771241	20	Я	0.018762179	16

Източник: Собствена разработка

Ще бъде показана атака срещу един шифър, широко известен като шифър на Виженер. Klima & Sigmon (2019) го описват под името „шифър на Виженер с ключ“.

Шифърът на Виженер е многоазбучен, т.е. на една буква от изходния текст могат да съответстват различни букви от шифрования текст. При този шифър ключът е една дума, наричана понякога **лозунг**.

Реализацията на криптирането в Excel е показана на фиг. 4.



Фигура 4. Криптиране с шифър на Виженер

Съобщението се въвежда в слети клетки A4:J24. За затрудняване на криптоанализа то се преобразува с функцията SUBSTUTUTE така, че да съдържа само букви, без препинателни знаци и интервали между думите (клетки A286:J48). Преобразуваното съобщение се разбива по букви в колона U. В колона V на всяка буква от текста се съпоставя азбучния ѝ номер от 0 до 29. Така А е 0, Б – 1, В – 2 и т.н.

Нека лозунгът е думата ДОБРИЧ. Дължината му е равна на 6. В колона W се попълват многократно номерата от 1 до 6. В колона X се попълват многократно буквите от лозунга, а в колона Y – азбучните номера на буквите от лозунга.

При криптирането на информацията към азбучния номер на всяка буква от съобщението се прибавя азбучният номер на буквата от лозунга. Полученото число е азбучният номер на буквата от криптограмата. Ако се стигне до последната буква в азбуката, се продължава циклично от началото, т.е сумата е по модул 30. Тези стойности се изчисляват в колона Z. В колона AA кодовете на буквите от колона Z се преобразуват в букви от азбуката. С конкатенация от получените букви се получава текстът на криптограмата (клетки AE2:AN20).

Ако дължината на лозунга е l, математическият модел на криптиращата функция е

$$c(i) = \left(m(i) + k\left(i(\mathrm{mod}\, l)\right)\right)(\mathrm{mod}\, N), \qquad (3)$$

където $m(i)$ – азбучният номер на *i*-тата буква от изходния текст;

$k\left(i(\mathrm{mod}\, l)\right)$ – азбучният номер на *j*-тата буква от лозунга;

N – брой на буквите в азбуката;

$c(i)$ – азбучен номер на i -тата буква от криптирания текст.

Вижда се, че няма еднозначно съответствие между буквите от изходния текст и тези от криптирания. Буквата О от трета позиция на съобщението се криптира в П, а буквата О от шеста позиция – в З. Възможно е обаче някоя буква да се криптира по един и същ начин, например буквите Р на втора и двадесета позиция. При тях има съвпадение на криптиращата буква от лозунга.

Декриптирането се извършва по обратния начин – всяка буква от декриптирания текст се получава от буквите на криптирания с циклично изместване **вляво** на толкова позиции, колкото е азбучният номер на съответната буква от лозунга, или с циклично изместване **вдясно** на толкова позиции, колкото е стойността на обратния елемент по събиране по модул N на азбучния номер на съответната буква от лозунга. Моделът на декриптиращата функция е

$$m(i) = (c(i) + N - k(j)) (\mathrm{mod}\, N), \tag{4}$$

Предполага се, че буквите в лозунга са различни (в противен случай шифърът на Виженер се изражда в шифър на Цезар). Тогава броят на различните ключове ще бъде равен на вариации без повторения от N елемента, k-ти клас, където N е броят на буквите в азбуката, а k – дължината на лозунга:

$$V_N^k = \frac{N!}{(N-k)!}$$

При лозунг с голяма дължина атаката с груба сила е много по-трудна от тази на шифъра на Цезар. Ако например $N = 30$ и $k = 10$ то $V_N^k \approx 2^{46}$.

Шифърът на Виженер е бил смятан за абсолютно устойчив в продължение на около 300 години и дори не е имало идея как да се определи дължината на лозунга. Едва през XIX век английският математик Бебидж (Charles Babbage) и пруският майор от запаса Казиски (Friedrich Wilhelm Kasiski) независимо един от друг откриват начин за определяне дължината на лозунга, известен като тест на Казиски. В началото на XX век американският криптограф и криптоаналитик Уйлям Фридман открива т.нар. индекс на съвпадение κ (капа) на естествения език (Friedman 1939). На основата на индекса на съвпадение е изведена формула за приблизително определяне дължината на лозунга, известна като тест на Фридман (Klima & Sigmon 2019). По-удобен за прилагане в средата на Excel обаче е един метод, който обединява идеите на Казиски и Фридман (Stinson & Paterson 2020).

Нека е даден текст на естествен език с дължина n. Както беше казано, буквите в естествените езици се срещат с различна честота. Следователно в този текст i-тата буква ще се среща m пъти. Да оценим статистическата вероятност две случайно избрани букви от текста да съвпаднат. Вероятността първата избрана буква да е i-тата, е $\frac{m_i}{n}$. Вероятността и втората избрана буква да е

i-тата, е $\frac{m_i - 1}{n - 1}$. Сумата от вероятностите за всички букви от азбуката е

$$\kappa_p = \sum_{i=1}^{N} \frac{m_i(m_i - 1)}{n(n-1)} \tag{5}$$

и се нарича индекс на съвпадение за текста на естествен език.

Ако дължината на текста е неограничено голяма, т.е. $n \to \infty$, то $\frac{m_i}{n} \to p_i$ и $\frac{m_i - 1}{n - 1} \to p_i$, където p_i е вероятността за поява на i-тата буква в естественя език. Тогава (5) може да се запише

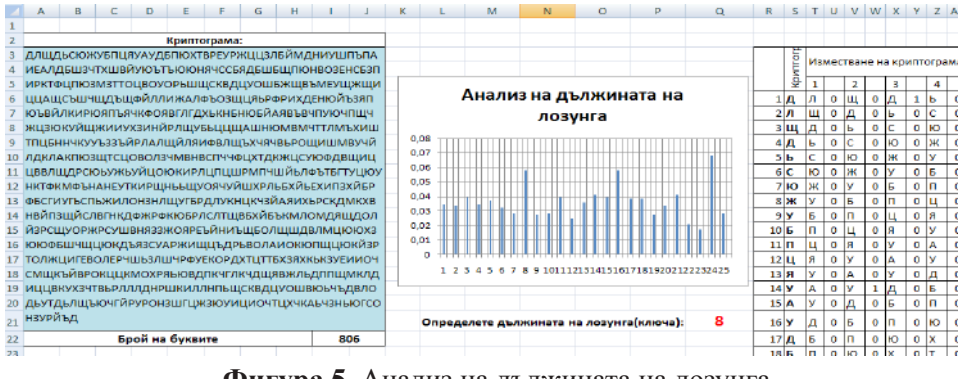
$$\kappa_p = \sum_{i=1}^{N} p_i^2 \tag{6}$$

Като се използва табл. 1, за индекса на съвпадение на българския език се получава 0,0633[1)].

Ако имаме текст, съставен от случайна съвкупност от букви (не на естествен език), то всяка буква се среща с една и съща вероятност и индексът на съвпадение на текста ще бъде равен на

$$\kappa_r = \sum_{i=1}^{30} \left(\frac{1}{30}\right)^2 = \frac{1}{30} \approx 0{,}0333 . \tag{7}$$

Нека е прехванатата криптограмата, показана на фиг. 5, вляво. За да определим дължината на лозунга, записваме в един стълб (стълба S) по букви криптограмата и след това в стълбовете T, V, X и т.н. я записваме отново, но изместена циклично на 1, 2, 3 и т.н позиции нагоре (фиг. 5).



Фигура 5. Анализ на дължината на лозунга

Сравняваме символите от основния стълб (T) и изместените, като в съседните стълбове (U, W, Y и т.н.) записваме нула, ако символите не съвпадат, и единица, ако символите съвпадат. Изброяваме единиците по стълбове и пресмятаме каква е относителната честота на съвпаденията.

Ако изместването не съвпада с дължината на лозунга, съвпаденията ще са съвсем случайни и ще са близки до 0,0333. Ако изместването съвпада с дължината на лозунга, една и съща буква от лозунга ще криптира съвпадащите букви от съобщението и относителната честота ще бъде близка до индекса на съвпадение на естествения език, т.е. до 0,0633. На диаграмата се вижда, че такива пикове има за изместванията 8, 16 и 24. Ако дължината на лозунга е 8, то пикове ще има и за всички отместванията, кратни на 8, т.е. за 16, 24, 32 и т.н. Следователно може да се приеме, че дължината на лозунга е 8. В клетка Q21 се въвежда избраната дължина на лозунга.

Следващата стъпка е да се определят буквите от лозунга. Множеството от символи на криптирания текст се разделя на класове според това с кое число по модул 8 са сравними номерата на позициите им. Ако позицията на символа е i, то към k-тия клас принадлежат тези, за които $i \equiv k \pmod 8$ (фиг. 6).

O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF

Разпределение на буквите по класове																	
		1 клас		2 клас		3 клас		4 клас		5 клас		6 клас		7 клас		8 клас	
1	Д	1	Д	2	Л	3	Щ	4	Д	5	Ь	6	С	7	Ю	8	Ж
2	Л	9	У	10	Б	11	П	12	Ц	13	Я	14	У	15	А	16	У
3	Щ	17	Д	18	Б	19	П	20	Ю	21	Х	22	Т	23	В	24	Р
4	Д	25	Е	26	У	27	Р	28	Ж	29	Ц	30	Ц	31	З	32	Л
5	Ь	33	Б	34	Й	35	М	36	Д	37	Н	38	И	39	У	40	Ш
6	С	41	П	42	Ъ	43	П	44	А	45	И	46	Е	47	А	48	Л
7	Ю	49	Д	50	Б	51	Ш	52	З	53	Ч	54	Т	55	Х	56	Ш
8	Ж	57	В	58	Й	59	У	60	Ю	61	Ъ	62	Т	63	Ъ	64	Ю
9	У	65	Ю	66	Н	67	Я	68	Ч	69	С	70	С	71	Б	72	Я
10	Б	73	Д	74	Б	75	Ш	76	Б	77	Щ	78	П	79	Ю	80	Н
11	П	81	В	82	О	83	З	84	Е	85	Н	86	С	87	Б	88	З
12	Ц	89	П	90	И	91	Р	92	К	93	Т	94	Ф	95	Ц	96	П
13	Я	97	Ю	98	З	99	М	100	З	101	Т	102	Т	103	О	104	Ц
14	У	105	В	106	О	107	У	108	О	109	Р	110	Ь	111	Ш	112	Щ
15	А	113	С	114	К	115	В	116	Д	117	Ц	118	У	119	О	120	Ш
16	У	121	Б	122	Ж	123	Щ	124	В	125	Ъ	126	М	127	Е	128	У
17	Д	129	Щ	130	Ж	131	Щ	132	И	133	Ц	134	Ц	135	А	136	Щ
18	Б	137	С	138	Ъ	139	Ш	140	Ч	141	Щ	142	Д	143	Ъ	144	Щ

Фигура 6. Разпределение на буквите от криптограмата по класове

За всеки клас се определя относителната честота на буквите (фиг. 7).

BP	BQ	BR	BS	BT	BU	BV	BW	BX	BY	BZ	CA	CB	CC	CD	CE	CF
	1 клас		2 клас		3 клас		4 клас		5 клас		6 клас		7 клас		8 клас	
А	0	0	0	0	1	0,00990099	3	0,02970297	0	0	2	0,01980198	8	0,08	0	0
Б	5	0,04950495	8	0,07920792	0	0	5	0,04950495	1	0,00990099	0	0	5	0,05	0	0
В	16	0,15841584	0	0	6	0,05940594	1	0,00990099	1	0,00990099	0	0	10	0,1	2	0,02
Г	0	0	3	0,02970297	0	0	2	0,01980198	3	0,02970297	2	0,01980198	2	0,02	0	0
Д	7	0,06930693	3	0,02970297	1	0,00990099	8	0,07920792	4	0,03960396	10	0,0990099	0	0	0	0
Е	1	0,00990099	1	0,00990099	0	0	3	0,02970297	0	0	2	0,01980198	2	0,02	3	0,03
Ж	1	0,00990099	4	0,03960396	0	0	7	0,06930693	0	0	2	0,01980198	1	0,01	2	0,02
З	8	0,07920792	3	0,02970297	2	0,01980198	12	0,11881188	2	0,01980198	2	0,01980198	2	0,02	1	0,01
И	0	0	7	0,06930693	0	0	5	0,04950495	12	0,11881188	6	0,05940594	0	0	0	0
Й	0	0	10	0,0990099	0	0	5	0,04950495	1	0,00990099	10	0,0990099	0	0	0	0
К	7	0,06930693	6	0,05940594	1	0,00990099	7	0,06930693	7	0,06930693	0	0	4	0,04	5	0,05
Л	0	0	6	0,05940594	18	0,17821782	4	0,03960396	0	0	5	0,04950495	0	0	12	0,12
М	3	0,02970297	4	0,03960396	2	0,01980198	2	0,01980198	1	0,00990099	4	0,03960396	0	0	1	0,01
Н	1	0,00990099	5	0,04950495	7	0,06930693	0	0	7	0,06930693	1	0,00990099	0	0	7	0,07
О	2	0,01980198	13	0,12871287	1	0,00990099	2	0,01980198	0	0	1	0,00990099	11	0,11	1	0,01
П	6	0,05940594	1	0,00990099	7	0,06930693	0	0	1	0,00990099	5	0,04950495	3	0,03	1	0,01
Р	8	0,07920792	0	0	8	0,07920792	0	0	8	0,07920792	2	0,01980198	3	0,03	8	0,08
С	6	0,05940594	1	0,00990099	1	0,00990099	0	0	2	0,01980198	8	0,07920792	3	0,03	0	0
Т	5	0,04950495	0	0	1	0,00990099	0	0	5	0,04950495	9	0,08910891	1	0,01	2	0,02
У	5	0,04950495	4	0,03960396	5	0,04950495	2	0,01980198	2	0,01980198	7	0,06930693	8	0,08	5	0,05
Ф	8	0,07920792	1	0,00990099	0	0	0	0	0	0	7	0,06930693	0	0	0	0
Х	1	0,00990099	0	0	5	0,04950495	0	0	7	0,06930693	4	0,03960396	1	0,01	4	0,04
Ц	0	0	2	0,01980198	3	0,02970297	4	0,03960396	9	0,08910891	8	0,07920792	8	0,08	3	0,03
Ч	2	0,01980198	0	0	1	0,00990099	14	0,13861386	6	0,05940594	2	0,01980198	0	0	5	0,05

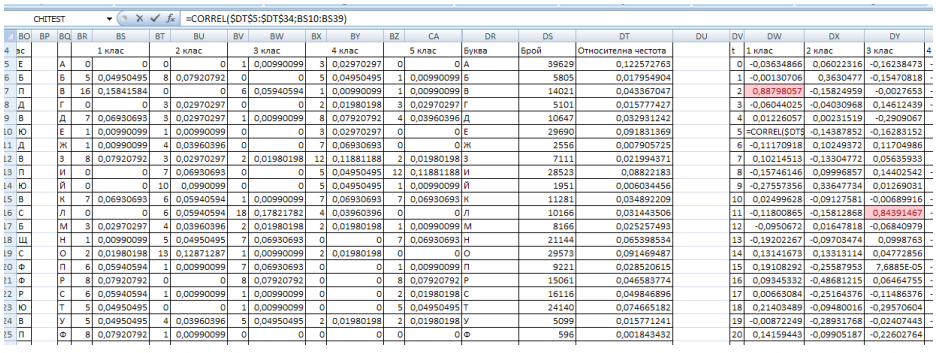
Фигура 7. Относителна честота на буквите по класове

От табл. 1 и от фиг. 7 се построяват хистограми на разпределението на буквите в българския език и в класовете. На фиг. 8 са представени хистограмите на честотата на буквите в българския език и на буквите в първи клас.

За българския език имаме характерни пикове в първа (А), трета (В), шеста (Е), девета (И) позиция. В такава последователност, но изместени на две позиции, има характерни пикове за хистограмата на буквите от първи клас. Следователно най-вероятно буквите от втори клас са изместени с две позиции надясно, т.е. първата буква от лозунга е В (с азбучен номер 2).

Този графичен начин може да доведе до грешки. Затова тук се предлага начин, който може да бъде наречен аналитичен. При него се използва оценка на взаимната корелация между честотата на буквите в естествения език и в отделните класове. По аналогия с взаимно корелационните функции се отчита при кое циклично изместване на разпределението на буквите в отделните класове спрямо разпределението на буквите в азбуката има най-силна връзка. Използва се функцията CORREL (фиг. 9).

Фигура 8. Хистограми на разпределението на буквите



	BO	BP	BQ	BR	BS	BT	BU	BV	BW	BX	BY	BZ	CA	DR	DS	DT	DU	DV	DW	DX	DY
4	вс			1 клас		2 клас		3 клас		4 клас		5 клас		Буква	Брой	Относителна честота		t	1 клас	2 клас	3 клас
5	Е		А	0	0	0	0	1	0,00990099	3	0,02970297	0	0	А	39629	0,122572763		0	-0,03634866	0,06022316	-0,16238473
6	Б		Б	5	0,04950495	8	0,07920792	0	0	5	0,04950495	1	0,00990099	Б	5805	0,017954904		1	-0,00130706	0,3630477	-0,15470818
7	П		В	16	0,15841584	0	0	6	0,05940594	1	0,00990099	1	0,00990099	В	14021	0,043367047		2	0,88798057	-0,15824959	-0,0027653
8	Д		Г	0	0	3	0,02970297	0	0	2	0,01980198	3	0,02970297	Г	5101	0,015777427		3	-0,06044025	-0,04030968	0,14612439
9	В		Д	7	0,06930693	3	0,02970297	1	0,00990099	8	0,07920792	4	0,03960396	Д	10647	0,032931242		4	0,01226057	0,00231519	-0,2909067
10	Ю		Е	1	0,00990099	1	0,00990099	0	0	3	0,02970297	0	0	Е	29690	0,091831369		5	=CORREL(DT	-0,14387852	-0,16283152
11	Д		Ж	1	0,00990099	4	0,03960396	0	0	7	0,06930693	0	0	Ж	2556	0,007905725		6	-0,11170918	0,10249372	0,11704986
12	В		З	8	0,07920792	3	0,02970297	2	0,01980198	12	0,11881188	2	0,01980198	З	7111	0,021994371		7	0,10214513	-0,13304772	0,05635933
13	П		И	0	0	7	0,06930693	0	0	5	0,04950495	12	0,11881188	И	28523	0,08822183		8	-0,15746146	0,09996857	0,14402542
14	Ю		Й	0	0	10	0,0990099	0	0	5	0,04950495	1	0,00990099	Й	1951	0,006034456		9	-0,27557356	0,33647734	0,01269031
15	В		К	7	0,06930693	6	0,05940594	1	0,00990099	7	0,06930693	7	0,06930693	К	11281	0,034892209		10	0,02499628	-0,09127581	-0,00689916
16	С		Л	0	0	6	0,05940594	18	0,17821782	4	0,03960396	0	0	Л	10166	0,031443506		11	-0,11800865	-0,15812868	0,84391467
17	Б		М	3	0,02970297	4	0,03960396	2	0,01980198	2	0,01980198	1	0,00990099	М	8166	0,025257493		12	-0,0950672	0,01647818	-0,06840979
18	Щ		Н	1	0,00990099	5	0,04950495	7	0,06930693	0	0	7	0,06930693	Н	21144	0,065398534		13	-0,19202267	-0,09703474	0,0998763
19	С		О	2	0,01980198	13	0,12871287	1	0,00990099	2	0,01980198	0	0	О	29573	0,091469487		14	0,13141673	0,13313114	0,04772856
20	Ф		П	6	0,05940594	1	0,00990099	7	0,06930693	0	0	1	0,00990099	П	9221	0,028520615		15	0,19108292	-0,25587953	7,6885E-05
21	Ф		Р	8	0,07920792	0	0	8	0,07920792	0	0	8	0,07920792	Р	15061	0,046583774		16	0,09345332	-0,48681215	0,06464755
22	Р		С	6	0,05940594	1	0,00990099	1	0,00990099	0	0	2	0,01980198	С	16116	0,049846896		17	0,00663084	-0,25164376	-0,11486376
23	Ю		Т	5	0,04950495	0	0	1	0,00990099	0	0	5	0,04950495	Т	24140	0,074665182		18	0,21403489	-0,09480016	-0,29570604
24	В		У	5	0,04950495	4	0,03960396	5	0,04950495	2	0,01980198	2	0,01980198	У	5099	0,015771241		19	-0,00872249	-0,28931768	-0,02407443
25	П		Ф	8	0,07920792	1	0,00990099	0	0	0	0	0	0	Ф	596	0,001843432		20	0,14159443	-0,09905187	-0,22602764

Фигура 9. Корелационни оценки

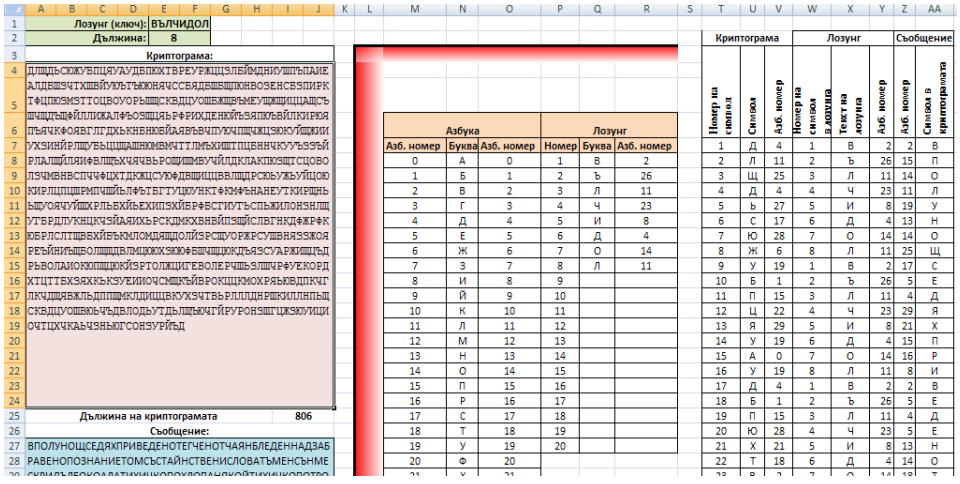
Корелират се относителните честоти на буквите от първи, втори и т.н. клас (колони BS, BU и т.н.) с относителната честота на буквите от азбуката (колона DT). Резултатите се записват в колони DW, DX и т.н., като по редове се пресмятат корелационните коефициенти за циклично изместване 0, 1, 2 и т.н. С условно форматиране са показани максималните стойности на корелационния коефициент за съответния клас. Вижда се, че за клас 1 той се получава при изместване 2 (буква В), а за клас 3 – при изместване 11 (буква Л).

В резултат на този анализ се определя търсеният лозунг (фиг. 10). На ред 2 в листа са изведени буквите от лозунга (фиг. 10).

	DV	DW	DX	DY	DZ	EA	EB	EC	ED
		В	Ъ	Л	Ч	И	Д	О	Л
		Корелационни функции (класове/естествен език)							
t		1 клас	2 клас	3 клас	4 клас	5 клас	6 клас	7 клас	8 клас
0		-0,03634866	0,060223159	-0,16238473	-0,16949729	0,029540869	0,006005324	0,11899985	-0,23914888
1		-0,00130706	0,363047701	-0,15470818	-0,10717113	-0,41450386	0,219035349	0,025467	-0,34037122
2		0,88798057	-0,15824959	-0,0027653	-0,01147679	-0,08610611	0,00586034	0,07692637	-0,16269763
3		-0,06044025	-0,04030968	0,146124386	-0,27588548	0,18193097	0,252140623	-0,2672787	-0,02124347
4		0,012260571	0,002315194	-0,2909067	0,002535898	0,003103987	0,854214863	-0,3098002	-0,43139245
5		0,014375883	-0,14387852	-0,16283152	0,132583042	0,193809902	0,14945118	-0,1892658	0,077981266
6		-0,11170918	0,102493725	0,117049862	0,034332723	-0,09917515	0,095928204	-0,0230143	0,180286233
7		0,102145125	-0,13304772	0,056359334	0,127520371	-0,04176805	0,138540408	-0,3266572	-0,11326351
8		-0,15746146	0,09996857	0,144025423	-0,17724095	0,886729764	0,074033668	-0,2115545	0,204462485
9		-0,27557356	0,336477337	0,012690311	0,030847802	0,005906155	0,254408368	0,1051537	0,091218166
10		0,024996276	-0,09127581	-0,00689916	-0,02093873	0,100385454	-0,11385219	0,02624102	0,176164787
11		-0,11800865	-0,15812868	0,843914674	-0,15475915	0,113112304	-0,17606419	0,04063284	0,883624417
12		-0,0950672	0,016478181	-0,06840979	-0,2600222	-0,13975814	-0,06002324	0,01900108	-0,0777298
13		-0,19202267	-0,09703474	0,099876296	-0,36580082	0,108306424	-0,15142296	-0,0405719	0,132611377
14		0,131416734	0,133131144	0,047728556	-0,1817114	-0,10125356	-0,20787261	0,9213488	0,078355943
15		0,191082917	-0,25587953	7,68851E-05	-0,02351021	-0,20799537	-0,06384781	0,06422286	-0,13005327
16		0,093453322	-0,48681215	0,064647547	-0,30930112	0,152436535	-0,08428551	0,03525828	0,063514295
17		0,006630842	-0,25164376	-0,11486376	-0,08651255	-0,02442106	0,0986279	0,22157415	-0,0818361

Фигура 10. Определяне на лозунга (ключа)

Декриптирането се реализира подобно на криптирането (фиг. 11).



Лозунг (ключ):	ВЪЛЧИДОЛ
Дължина:	8

Криптограма:

ДЛЩДЬСКЖУБПЦЯУАУДЕПКЖТВРЕУРЖЦЦЯЛЕЙМДНИУШТЬПАИЕ
АЛДЕШЗЧТХШВЙУЮЪТЬЮЮНЯЧССВЯДБШБЦЮНВОЗЕНСБЭПИРК
ТФЦПЮЭМЗТТОЦБОУОРЬШЦСКВДЦУОШЕЖЩВЪМЕУЩЖЩИЦЩАЩСЪ
ШЧЩДЬЦЙЛЛИЖАЛФЪОЗЩЦЯЬРФРИХДЕНКЙЪЯЯПЮЬЙЛКИРЮЯ
ПЬЯЧКФОЯВГЛГДЬКНЕНЮБЙАЯВЪВЧПУЮЧЩЧЖЦЗЮКУЙЩЖИИ
УХЗИНЙРЛЩУБЬЦЩЦАШНЮМВЧТТЛМЪХШТТЦЕННЧЮУУЪЗЗЪЙ
РЛАЛЩЙЛЯЙФВЛШЪХЧЯЬВРОЩШМВУЧЙЛДКЛАКПЮЩТСЦОВО
ЛЯЧМВНВСПЧЧФЦХТДКЖЦСУЮФДВЩИЦЩВЛЩЦРСЮЬУЖЬУЙЦЮЮ
КИРЛЦПШРМПЧШЙЬЛФЪТБГТУЦЮУНКТФКМФЪНАНЕУТКИРЩНЬ
ЬЩУОЯЧУЙШХРЬЬБХЙЬЕХИПЗХЙВРФКВСГИУТЪСПЬЖИЛОНЗНЛЩ
УТВРДЛУКНЦКЧЗЙАЯИХЬРСКДМКХВНЙЙПЗЩЙСЛВГНКДФЖРФК
ЮВРДЛСЛТЩВБХЙЙЬЖЮЛОМДЯЩДОЛЙЭРСЩУОРЖРСУШВЯЯЗЖЮЯ
РЕЬЙИЬЩВОЛЩЦВЛМЦКЖХЖЮФЕШЧИЩЦКЖДЬЯЗСУАРЖИЩЪЬД
РЬВОЛАИОЮЮПЩЦЮКЙЭРТОЛЖЦИГЕВОЛЕРЧШЬЯЛШЧРФУЕКОРД
ХТЦТТБХЗЯЯКЬКЗУЕИИОЧСМЦКЬЙВРОКЦЦКМОХРЯЬЮДПКЧГ
ЛКЧДДЯВЖЛЬДППШМКЛДИЦЦВКУХЧЧТВЬРЛЛЛДНРШКИЛЛПЬЩ
СКВДЦУОШЮЬЧЪДБЛОДЬУТДЬЛЩЬЮЧГЙРУРОНЗШГЦЖЗКУИЦИ
ОЧТЦХЧКАЬЧЯЬЮГСОНЗУРЙЪД

Дължина на криптограмата	806

Съобщение:

ВПОЛУНОЩСЕДЯХПРИВЕДЕНОТЕГЧЕНОТЧАЯНБЛЕДЕННАДЗАБ
РАВЕНОПОЗНАНИЕТОМСЪСТАЙНСТВЕНИСЛОВАТЬМЕНСЬНМЕ

Азбука			Лозунг		
Азб. номер	Буква	Азб. номер	Номер	Буква	Азб. номер
0	А	0	1	В	2
1	Б	1	2	Ъ	26
2	В	2	3	Л	11
3	Г	3	4	Ч	23
4	Д	4	5	И	8
5	Е	5	6	Д	4
6	Ж	6	7	О	14
7	З	7	8	Л	11
8	И	8	9		
9	Й	9	10		
10	К	10	11		
11	Л	11	12		
12	М	12	13		
13	Н	13	14		
14	О	14	15		
15	П	15	16		
16	Р	16	17		
17	С	17	18		
18	Т	18	19		
19	У	19	20		
20	Ф	20			

Криптограма			Лозунг			Съобщение	
Номер на символа	Символ	Азб. номер	Номер на символа в лозунга	Текст на лозунга	Азб. номер	Азб. номер	Символ в криптограмата
1	Д	4	1	В	2	2	В
2	Л	11	2	Ъ	26	15	П
3	Щ	25	3	Л	11	14	О
4	Д	4	4	Ч	23	11	Л
5	Ь	27	5	И	8	19	У
6	С	17	6	Д	4	13	Н
7	Ю	28	7	О	14	14	О
8	Ж	6	8	Л	11	25	Щ
9	У	19	1	В	2	17	С
10	Б	1	2	Ъ	26	5	Е
11	П	15	3	Л	11	4	Д
12	Ц	22	4	Ч	23	29	Я
13	Я	29	5	И	8	21	Х
14	У	19	6	Д	4	15	П
15	А	0	7	О	14	16	Р
16	У	19	8	Л	11	8	И
17	Д	4	1	В	2	2	В
18	Б	1	2	Ъ	26	5	Е
19	П	15	3	Л	11	4	Д
20	Ю	28	4	Ч	23	5	Е
21	Х	21	5	И	8	13	Н
22	Т	18	6	Д	4	14	О

Фигура 11. Декриптиране

Остава декриптираното съобщение да се дообработи, като думите се разделят с интервали. Окончателно се получава:

В ПОЛУНОЩ СЕДЯХ ПРИВЕДЕН ОТЕГЧЕН ОТЧАЯН БЛЕДЕН НАД ЗАБРАВЕНО ПОЗНАНИЕ ТОМ СЪС ТАЙНСТВЕНИ СЛОВА... и т.н.[2)]

Атака с помощта на двойки некриптиран/криптиран текст

Криптоаналитикът има възможност да прехване всички криптограми (тъй като те се изпращат по незащитен канал). Да допуснем, че освен това за някои криптограми или за част от тях му е известен и изходният текст. Това не е невъзможно. В много организации съществуват формални правила за оформяне на документи, които определят къде и как да се поставят адресите, датата на създаване или грифът за сигурност. Възможно е да бъде проявена небрежност от страна на криптографа или да има предателство. Целта на криптоаналитика е да разкрие ключа, който е използван, за да може да декриптира и останалите прехванати криптограми.

Ще бъде показана такава атака срещу шифъра на Хил (Hill 1929).

Идеята на шифъра на Хил е следната. Нека азбуката съдържа N символа. На всяка буква от азбуката се присвоява целочислен номер от 0 до N-1. Ключ на шифъра е квадратна матрица $K_{n\times n}$ по модул N, която е обратима. Матриците по модул N са матрици, чиито елементи са целочислени и заемат стойности от 0 до $N-1$. Те притежават всички свойства на обикновените

матрици и с тях могат да се извършват всички действия както с обикновените матрици. Единственото допълнително изискване е всички операции да се изпълняват по модул.

За намиране на обратна матрица по модул се използва формулата

$$K^{-1}(\mathrm{mod}\,N)=\left[\det(K)\right]^{-1}\cdot\begin{Vmatrix} K_{11} & K_{21} & \dots & K_{n1} \\ K_{12} & K_{22} & \dots & K_{n2} \\ \dots & \dots & \dots & \dots \\ K_{1n} & K_{2n} & \dots & K_{nn} \end{Vmatrix}(\mathrm{mod}\,N), \qquad (8)$$

където K_{ij} са адюнгираните количества за матрицата K, а $\left[\det(K)\right]^{-1}$ е число такова, че е изпълнено $\left[\det(K)\right]^{-1}\cdot\det(K)(\mathrm{mod}\,N)\equiv 1$. Следователно, за да бъде матрицата обратима, $\det(K)$ и N трябва да са взаимно прости. При конструиране на шифър на Хил е целесъобразно за основа на модула да се избере подходящо просто число. Ако използваме българската азбука, удобно е да се избере N = 31, като освен буквите се използва и символът интервал.

Всяка буква от съобщението се замества с нейната числена стойност. Получената последователност от числа се разделя на блокове m_1, m_2, ..., m_s всеки от които е с дължина n. Блоковете формират матрица $M_{s\times n}$.

При криптирането матрицата $M_{s\times n}$ се умножава на $K_{n\times n}$ по модул N. Получената матрица $C_{s\times n}$ е криптограмата:

$$M_{s\times n}\cdot K_{n\times n}(\mathrm{mod}\,N)=C_{s\times n}. \qquad (9)$$

При декриптирането получената криптограма се умножава по обратната матрица на $K_{n\times n}$. Действително

$$C_{s\times n}\cdot K^{-1}_{n\times n}=\left(M_{s\times n}\cdot K_{n\times n}\right)\cdot K^{-1}_{n\times n}=M_{s\times n}\cdot\left(K_{n\times n}\cdot K^{-1}_{n\times n}\right)=M_{s\times n}. \qquad (10)$$

Нека ключовата матрица е от ред 5. Получаването на обратна матрица по модул в средата на Excel с изчисляване на адюнгираните количества е трудоемко. Затова може да се използва по-удобен начин (фиг. 12).

Попълва се по произволен начин матрицата *K* (клетки E4:I8), като в клетка L4 се изчислява детерминантата ѝ с помощта на функцията MDETERM. клетка L5 се изчислява модулът на детерминантата (функция MOD, основа на модула – 31). В клетка L6 се изчислява обратния елемент на детерминантата по модул 31. За целта използваме таблица на обратните елементи (клетки A4:B33).

Декриптиращата (обратната) матрица получаваме с формулата

=ROUND(MOD(L4*MINVERSE(E4:I8)*L6;C1);0).

Използваме функцията MINVERSE, но резултата умножаваме по детерминантата на матрицата (клетка L4), като по този начин от обратната получаваме матрицата от адюнгираните количества. Тази матрица умножаваме на обратния елемент на детерминантата (клетка L6), и резултатът се взема по модула с основа, записана в клетка C1.

Тъй като при изчисляването на междинните резултати в средата на Excel е възможно да се получат грешки от закръгления, крайният резултат се закръглява до цяло число с функция ROUND.

За контрол на верността в клетки E13:I17 е представено произведението на изходната и обратната матрица. Използва се формулата за умножение по модул на матрици

=MOD(MMULT(E4:I8;O4:S8);C1).

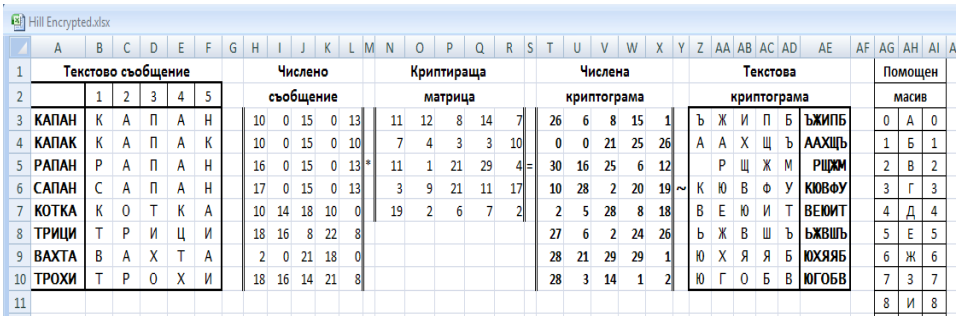
Hill Encrypted.xlsx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	**Основа на модула N=**		**31**																	
2					**Криптираща**										**Декриптираща**					
3	**Обратни по модул 31**				**матрица**										**матрица**					
4	1	1			11	12	8	14	7		det(K)=	-533271			11	1	19	30	20	
5	2	16			7	4	3	3	10		det(K)(mod N)=	22			2	14	3	5	14	
6	3	21		K=	11	1	21	29	4		Inverse(22)=	24		K^{-1}=	22	13	16	29	29	
7	4	8			3	9	21	11	17						7	3	17	19	13	
8	5	25			19	2	6	7	2						20	20	19	6	5	
9	6	26																		
10	7	9																		
11	8	4		**Проверка за валидност**																
12	9	7		**на матриците**																
13	10	28			1	0	0	0	0											
14	11	17			0	1	0	0	0											
15	12	13		$K.K^{-1}$(mod N)=	0	0	1	0	0											
16	13	12			0	0	0	1	0											
17	14	20			0	0	0	0	1											

Фигура 12. Получаване на обратна матрица по модул

Както се вижда, резултатът е единична матрица.

Криптирането е показано на фиг. 13.



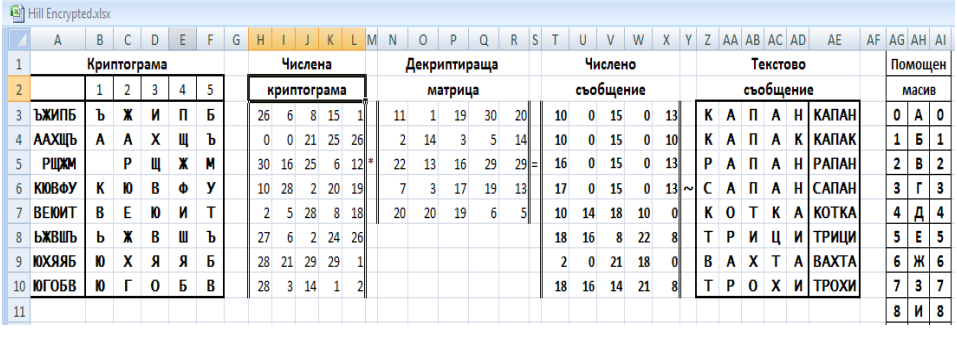
Текстово съобщение						Число съобщение					Криптираща матрица					Числена криптограма					Текстова криптограма						Помощен масив		
	1	2	3	4	5																							
КАПАН	К	А	П	А	Н	10	0	15	0	13	11	12	8	14	7	26	6	8	15	1	Ъ	Ж	И	П	Б	ЪЖИПБ	0	А	0
КАПАК	К	А	П	А	К	10	0	15	0	10	7	4	3	3	10	0	0	21	25	26	А	А	Х	Щ	Ъ	ААХЩЪ	1	Б	1
РАПАН	Р	А	П	А	Н	16	0	15	0	13	11	1	21	29	4	30	16	25	6	12		Р	Щ	Ж	М	РЩЖМ	2	В	2
САПАН	С	А	П	А	Н	17	0	15	0	13	3	9	21	11	17	10	28	2	20	19	К	Ю	В	Ф	У	КЮВФУ	3	Г	3
КОТКА	К	О	Т	К	А	10	14	18	10	0	19	2	6	7	2	2	5	28	8	18	В	Е	Ю	И	Т	ВЕЮИТ	4	Д	4
ТРИЦИ	Т	Р	И	Ц	И	18	16	8	22	8						27	6	2	24	26	Ь	Ж	В	Ш	Ъ	ЬЖВШЪ	5	Е	5
ВАХТА	В	А	Х	Т	А	2	0	21	18	0						28	21	29	29	1	Ю	Х	Я	Я	Б	ЮХЯЯБ	6	Ж	6
ТРОХИ	Т	Р	О	Х	И	18	16	14	21	8						28	3	14	1	2	Ю	Г	О	Б	В	ЮГОБВ	7	З	7
																											8	И	8

Фигура 13. Криптиране с шифър на Хил

В клетки A3:A10 се записва текстът на съобщението във вид на петбуквени блокове. В клетки B3:F10 блоковете се разбиват по букви, а в клетки H3:L10 се записват числените стойности на буквите. Използват се функцията VLOOKUP и помощен масив в клетки AG3:AI33. За получаване на числената криптограма съобщението се умножава по криптиращата матрица по модул, след което се преобразува в текст.

Характерно за шифъра на Хил е, че малки промени в съобщението водят до големи промени в криптограмата. В посочения пример някои от отделните съобщения се различават само с по една буква, но съответстващите им криптограми се различават тотално. Този ефект на разсейване затруднява криптоанализа.

По подобен начин се извършва и декриптирането (фиг. 14).



Криптограма						Числена криптограма					Декриптираща матрица					Число съобщение					Текстово съобщение						Помощен масив		
	1	2	3	4	5																							
ЪЖИПБ	Ъ	Ж	И	П	Б	26	6	8	15	1	11	1	19	30	20	10	0	15	0	13	К	А	П	А	Н	КАПАН	0	А	0
ААХЩЪ	А	А	Х	Щ	Ъ	0	0	21	25	26	2	14	3	5	14	10	0	15	0	10	К	А	П	А	К	КАПАК	1	Б	1
РЩЖМ		Р	Щ	Ж	М	30	16	25	6	12	22	13	16	29	29	16	0	15	0	13	Р	А	П	А	Н	РАПАН	2	В	2
КЮВФУ	К	Ю	В	Ф	У	10	28	2	20	19	7	3	17	19	13	17	0	15	0	13	С	А	П	А	Н	САПАН	3	Г	3
ВЕЮИТ	В	Е	Ю	И	Т	2	5	28	8	18	20	20	19	6	5	10	14	18	10	0	К	О	Т	К	А	КОТКА	4	Д	4
ЬЖВШЪ	Ь	Ж	В	Ш	Ъ	27	6	2	24	26						18	16	8	22	8	Т	Р	И	Ц	И	ТРИЦИ	5	Е	5
ЮХЯЯБ	Ю	Х	Я	Я	Б	28	21	29	29	1						2	0	21	18	0	В	А	Х	Т	А	ВАХТА	6	Ж	6
ЮГОБВ	Ю	Г	О	Б	В	28	3	14	1	2						18	16	14	21	8	Т	Р	О	Х	И	ТРОХИ	7	З	7
																											8	И	8

Фигура 14. Декриптиране с шифър на Хил

Нека сега да са прехванати няколко криптограми и за някои от тях да са известни съобщенията, които ги пораждат (фиг. 15). Предполага се, че всички са криптирани с един и същ ключ.

Прехванати двойки прав текст-шифротекст	
КАПАН	МЖИЖВ
КАПАК	САХРИ
РАПАН	РРЩГЗ
САПАН	ЬЮВТН
КОТКА	ЗРЮЕГ
ТРИЦИ	ДШВУЦ
ВАХТА	ЬРЯЙУ
ТРОХИ	ТЧОЩС

Прехванатата криптограма
ЧДАЙО
ИЪЯЮЩ
ДОВЖТ
ЧБШСП
БГЕЛЙ

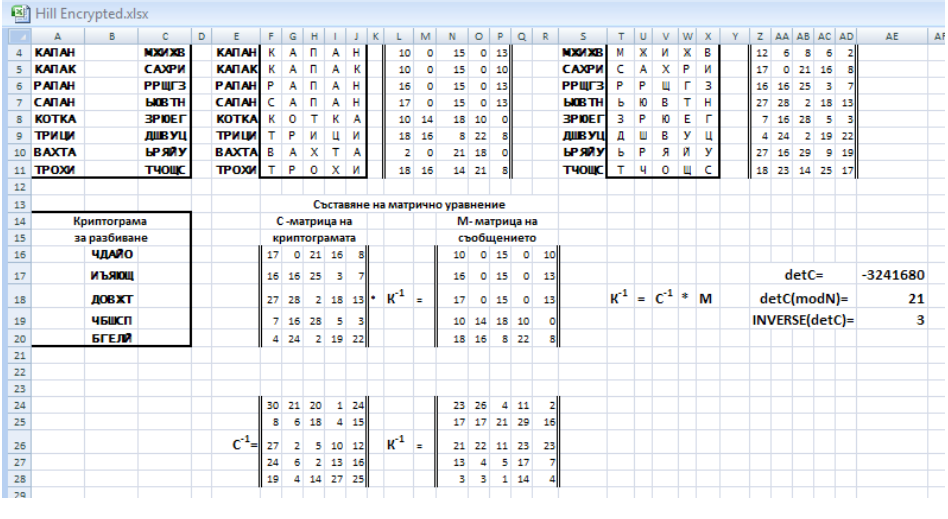
Фигура 15. Прехванати криптограми

Една идея за разбиване на шифъра на Хил е изложена в (Klima & Sigmon 2019). Известните двойки криптограма – текст се подреждат във вид на матрици. Ако умножим матрицата на криптограмите C с декриптиращата матрица K^{-1} (която е неизвестна за нас), ще получим матрицата от съобщенията M, т.е

$$C \cdot K^{-1} = M \pmod N.$$

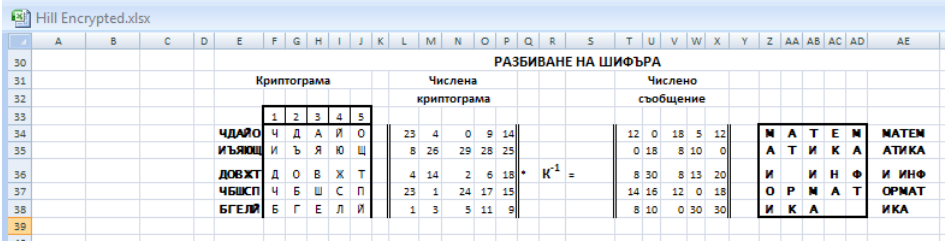
Решението на това матрично уравнение $K^{-1} = C^{-1} \cdot M \pmod N$ е търсената матрица. То е показано в средата на Excel на фиг. 16.

Матрицата от криптограмите трябва да е неособена. В дадения пример матрицата от първите 5 криптограми е особена, затова са подбрани криптограми 2, 3, 4, 5 и 6. Прилагат се функциите MINVERSE и MMULT по модул 31, аналогично на описаното по-горе. Матрицата за декриптиране е получена в клетки N24:R28.



Фигура 16. Намиране на ключа за шифъра на Хил

Декриптирането на останалите криптограми е показано на фиг. 17.



Фигура 17. Разбиване на шифъра на Хил

Заключение

В определени периоди представените криптографски алгоритми са били използвани успешно. Появата на нов научен инструментариум (включително развитието на изчислителната техника) е дало възможност за създаване на методи за тяхното разбиване за разумно време.

Анализът на недостатъците на разбитите алгоритми спомага за създаване на нови, по-мощни средства за криптографска защита. На мястото на разбитите алгоритми идват нови. С голяма степен на сигурност може да се твърди, че в резултат на натрупването на нови знания могат да бъдат създадени методи и за тяхното разбиване.

Както беше показано, Excel дава възможност за нагледно представяне на методите за криптиране и за криптоанализ. Представените разработки биха могли да се използват и за провеждане на експерименти. Интересни биха били например задачите за намиране на лъжливи ключове или за определяне на съотношението между дължината на ключа и дължината на съобщението, при което честотният анализ дава неверни резултати.

БЕЛЕЖКИ

1. Индексът на съвпадение е различен за различните езици. Според Фридман за английския той е 0,0667, за френския – 0,778, за немския – 0,0762, за италианския – 0,0738, и за испанския – 0,0775.
2. Целият текст на примера съдържа строфи 1 – 4 от „Гарванът“, автор Едгар Алън По, прев. Теменуга Маринова.

REFERENCES

FRIEDMAN, W. F., 1939. *Military Cryptanalysis. Part II.* (Unclassified in 1992, reprinted by Aegean Park Press).

HILL, L. S., 1929. Cryptography in an Algebraic Alphabet. *The American Mathematical Monthly*, **36**(6), 306 – 312.
KLIMA, R. & SIGMON, N., 2019. *Cryptology. Classical and Modern* (2nd ed.). Boca Raton: CRC Press.
PAAR, C. & PELZL, J., 2010. *Understanding Cryptography.* Berlin: Springer-Verlag.
PETITCOLAS F.A.P., 2011. *Kerckhoffs' Principle*. In: van Tilborg H.C.A., Jajodia S. (eds) Encyclopedia of Cryptography and Security. Boston: Springer. https://doi.org/10.1007/978-1-4419-5906-5_487.
STINSON, D. R., PATERSON, M. B., 2019. *Cryptography:Theory and Practice* (4th ed.). Boca Raton: CRC Press.

CRYPTOGRAPHY AND CRYPTANALYSIS IN MS EXCEL

Abstract. This paper provides implementations of three well-known ciphers – Caesar Cipher, Vigenere Cipher and Hill Cipher in Microsoft Excel. It is shown how the ciphers can be broken by using Brute-force Attack, Frequency Analysis Attack and Known-plaintext Attack. For the purpose of the cryptanalysis the relative occurrence frequencies of the letters and the index of coincidence (κ) in Bulgarian language are determined. The classical Frequency Analysis Attack is modified using the cross-correlation between frequencies of letters in the natural language and the cipher text. Modular matrix operations in MS Excel are shown.

Keywords: cryptanalysis; MS Excel

✉ **Dr. Deyan Mihaylov, Assist. Prof.**
ORCID iD: 0000-0002-3405-4758
Web of Science Researcher ID: AAR-5319-2021
Scopus Author ID: 57196030177
Faculty of Informatics
University of Economics – Varna
Varna, Bulgaria
E-mail: dgmihaylov@ue-varna.bg