Ministry of Education, Culture and Research of the Republic of Moldova Technical University of Moldova
Faculty of Computers, Informatics and Microelectronics
Department of Software Engineering and Automation

Report
for laboratory work no. 2
in the course "Cryptanalysis of monoalphabetic ciphers"

A efectuat:Ceban Andrei, gr. FAF-211

A verificat: Cătălin MÎȚU

Chișinău – 2023

Subject: Cryptanalysis of monoalphabetic ciphers

Sarcini:

1. Either an encrypted message has been intercepted which is known to have been obtained using a monoalphabetic cipher. Applying the frequency analysis attack to find out the original message, if it is supposed to be a text written in English. Note that only the letters have been encrypted, the other characters remain unencrypted.

Cryptanalysis of monoalphabetic ciphers

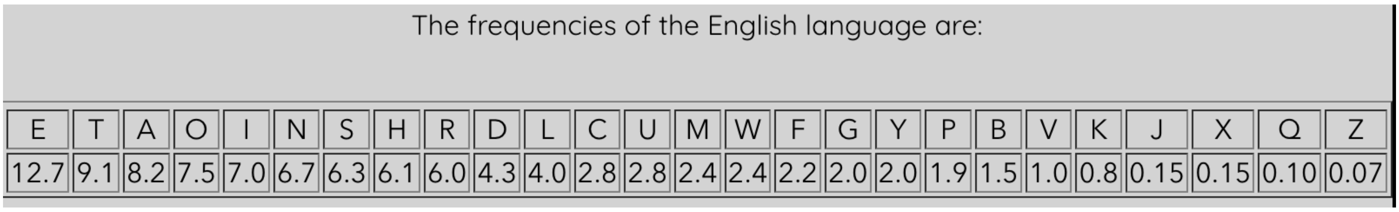
The weak point of monoalphabetic encryption systems is the frequency of occurrence of a characters in the text. If an encrypted text is long enough and the language in which it is written is known clear text, the system can be broken by an attack based on the frequency of occurrence of letters in a language (attack by frequency analysis), this frequency being an intensively studied problem (not necessarily in cryptographic purposes) and as a result various order structures were built relative to the frequency the appearance of letters in every European language and in other languages.

Usually, the longer a cipher text is, the closer the frequency of the letters used is this general order. A comparison between the two order relationships (that of the characters in the text encrypted and that of the letters from the alphabet of the current language) leads to the realization of several correspondences (letter clear text – encrypted text letter), which uniquely establishes the encryption key.

Since the set of possible keys is the set of all possible permutations of the alphabet, Monoalphabetic Substitution Ciphers have a keyspace of 26!, which is over 403 septillion. If someone were able to check 1,000,000 keys per second, it would still take over 12 trillion years to check all possible keys, so cryptanalysis by brute force is infeasible. If Eve were to intercept an encrypted ciphertext C from Alice to Bob, she could rely on her knowledge of the language the message was written in and use frequency analysis. Knowing that the most common English letter, E, occurs 12.7% of the time would allow Eve to assume the most common letter in C is mapped to by E. The next most-common letters according to [Bek82] are T at 9.1%, A at 8.2%, O at 7.5%, I at 7%, N at 6.7%, and S at 6.3%. These single-letter frequencies, generally referred to as unigram frequencies, are well-known for the English language, and illustrated in Figure 1.

The result of performing the tasks:

The frequencies of the English language are:

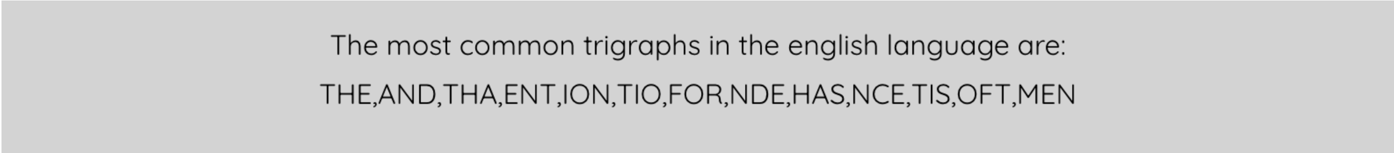


The frequencies of the English language are:

E	T	A	O	I	N	S	H	R	D	L	C	U	M	W	F	G	Y	P	B	V	K	J	X	Q	Z
12.7	9.1	8.2	7.5	7.0	6.7	6.3	6.1	6.0	4.3	4.0	2.8	2.8	2.4	2.4	2.2	2.0	2.0	1.9	1.5	1.0	0.8	0.15	0.15	0.10	0.07

Fig 1 – Frequencies

The most common trigraphs in the english language :



The most common trigraphs in the english language are:

THE,AND,THA,ENT,ION,TIO,FOR,NDE,HAS,NCE,TIS,OFT,MEN

Fig 2 – Trigraphs in english language

Intercept V - 8 :

NGSF TCWVI QV QTO VYUSTXGVO QNR HXUQVIP TIV PNSKVO OXO QV UINHVVO WNRTFP NC UIVKVGWXGJ

PNSDWXNG. QV HTUUVO QXP RNIL RXWQ T HXUQVI NC QXP NRGXGKVGWXNG WQTW QV HTSSVO "RNIWQF NC LXGJP"

TGO, SXLV TSS XGKVGWNIP, HSTXZVORTP DGAIVTLTASV. WQXP RTP WQV HXUQVI OXPL WQTW CNDGOVO UNSFTSUQTAVWXHXWF. RXWQ WQXP XGKVGWXNG, WQV RVPW, RQXHQ DU WN WQXP UNXGW QTOVBDTSVO ADW QTO

GVKVI PDIUTPPVO WQV VTPW XG HIFUWNSNJF, WNNL WQV SVTOWQTW XW QTP GVKVI SNPW."X ZTLV WRN

HXIHSVP NDW NC HNUUVI USTWVP. NGV, WQV STIJVI, XP HTSSVOPWTWXNGTIF, WQV PZTSSVI XP HTSSVO ZNKTASV.

WQV OXTZVWVI NC WQV PWTWXNGTIFUSTGW XP NGV-GXGWQ JIVTWVI WQTG WQTW NC WQV ZNKTASV USTWV. X

OXKXOV WQVHXIHDZCVIVGHV NC VTHQ HXIHSV XGWN 24 VBDTS UTIWP. WQVPV UTIWP TIV HTSSVOHVSSP. XG

WQV KTIXNDP HVSSP NC WQV STIJVI HXIHSV X RIXWV WQV HTUXWTS SVWWVIP, NGVTW T WXZV XG IVO, XG WQV

DPDTS NIOVI NC WQV SVWWVIP, T CXIPW, A PVHNGO, H WQXIO,TGO WQVG WQV IVPW, NZXWWXGJ Q TGO L

[TGO F] AVHTDPV WQVF TIV GNWGVHVPPTIF." WQXP JTKV QXZ 20 SVWWVIP, PXGHV E, D, TGO R RVIV GNW XG

QXPTSUQTAVW, TGO XG WQV IVZTXGXGJ CNDI PUTHVP QV XGPHIXAVO WQV GDZAVIP 1TGO 4 XG ASTHL. (WQV IVO

TGO ASTHL PVVZ WN PXJGXCF NGSF WQTW TSAVIWX SXLVOHNSNIP.) XG VTHQ NC WQV 24 HVSSP NC WQV ZNKTASV

HXIHSV QV XGPHIXAVO "T PZTSSSVWWVI XG ASTHL, TGO GNW XG IVJDSTI NIOVI SXLV WQV PWTWXNGTIF

HQTITHWVIP,ADW PHTWWVIVO TW ITGONZ. WQDP RV ZTF PDUUNPV WQV CXIPW NC WQVZ

WN AV T, WQV PVHNGO

J, WQV WQXIOB, TGO PN NG RXWQ WQV IVPW DGWXS WQV 24 HVSSP NC WQV HXIHSV TIV CDSS; CNI WQVIVTIV

24 HQTITHWVIP XG WQV STWXG TSUQTAVW, WQV STPW AVXGJ VW [UINATASFZVTGXGJ "&"]. TCWVI HNZUSVWXGJ

WQVPV TIITGJVZVGWP RV USTHV WQV PZTSSVIHXIHSV DUNG WQV STIJVI PN WQTW T GVVOSV OIXKVG WQINDJQ

WQV HVGWVIP NC ANWQZTF PVIKV TP WQV TYXP NC ANWQ TGO WQV ZNKTASV USTWV ZTF AV IVKNSKVOTINDGO

XW."WQV WRN HNIIVPUNGOVGWP—RQN, TSAVIWX HTIVCDSSF UNXGWVO NDW, ZDPWVTHQ QTKV XOVGWXHTS

OXPLP—TJIVV DUNG TG XGOVY SVWWVI XG WQV ZNKTASVOXPL, PTF L. WQVG, WN VGHXUQVI, WQV PVGOVI

USTHVP WQXP UIVTIITGJVO XGOVYSVWWVI TJTXGPW TGF SVWWVI NC WQV NDWVI OXPL. QV XGCNIZP QXP

HNIIVPUNGOVGWNC WQXP UNPXWXNG NC WQV OXPL AF RIXWXGJ, TP WQV CXIPW SVWWVI NC WQV

HXUQVIWVYW,WQXP SVWWVI NC WQV NDWVI IXGJ. TSAVIWX JTKV WQV VYTZUSV NC L AVXGJ USTHVOTJTXGPW

A. "CINZ WQXP TP T PWTIWXGJ UNXGW TSS WQV NWQVI HQTITHWVIP NC WQVZVPPTJV RXSS THBDXIV WQV CNIHV

TGO PNDGOP NC WQV PWTWXNGTIF HQTITHWVIPTANKV WQVZ."* PN CTI GNWQXGJ IVZTILTASV QTO QTUUVGVO.

ADW XG QXP GVYWPVGWVGHV TSAVIWX USTHVO HIFUWNJITUQF'P CVVW NG WQV INTO WN XWP

ZNOVIGHNZUSVYXWF. "TCWVI RIXWXGJ WQIVV NI CNDI RNIOP, X PQTSS HQTGJV WQV UNPXWXNGNC WQV XGOVY

XG NDI CNIZDST AF WDIGXGJ WQV HXIHSV, PN WQTW WQV XGOVY L ZTFAV, PTF, DGOVI O. PN XG ZF ZVPPTJV X

PQTSS RIXWV T HTUXWTS O, TGO CINZ WQXPUNXGW NG [HXUQVIWVYW] L RXSS PXJGXCF GN SNGJVI A ADW O,

TGO TSS WQV NWQVIPWTWXNGTIF SVWWVIP TW WQV WNU RXSS IVHVXKV GVR ZVTGXGJP."WQVIV XP WQV

HIDHXTS UNXGW: "GVR ZVTGXGJP." VTHQ GVR UNPXWXNG NC WQVXGGVI OXPL AIXGJP OXCCVIVGW SVWWVIP

NUUNPXWV NGV TGNWQVI XG WQV XGGVI TGONDWVI IXGJP. HNGPVBDVGWSF, VTHQ PQXCW ZVTGP WQTW

USTXGWVYW SVWWVIP RNDSOAV IVUSTHVO RXWQ OXCCVIVGW HXUQVIWVYW VBDXKTSVGWP. CNI VYTZUSV,

WQVUSTXGWVYW RNIO GN ZXJQW AV VGHXUQVIVO WN CH TW NGV PVWWXGJ TGO WN MV TWTGNWQVI.

VBDTSSF, TW VTHQ PQXCW T JXKVG HXUQVI-WVYW SVWWVI RNDSO PWTGO CNI TOXCCVIVGW USTXGWVYW

SVWWVI WQTG XW OXO TW WQV UIVKXNDP PVWWXGJ. WQDP, WQV CHWQTW CNIZVISF IVUIVPVGWVO GN

ZXJQW, TW WQV GVR PVWWXGJ, PWTGO CNIUSTXGWVYW WD. WQXP PQXCW XG ANWQ USTXG TGO HXUQVI

VBDXKTSVGWP OXC-*XG TSAVIWX'P OXPL, WQV NDWVI HTUXWTS SVWWVIP TIV WQV USTXGWVYW TGO

WQVXGGVI SNRVI-HTPV SVWWVIP TIV WQV HXUQVIWVYW. WQXP HNGWITOXHWP WQVHNGKVGWXNG NC WQXP

ANNL, TGO XP AVXGJ DPVO XG WQV PVHWXNG NG TSAVIWX NGSFWN TKNXO TSWVIXGJ QXP WVYW. WQV

OXCCVIVGHV XP PXJGTSXMVO AF GNW DPXGJ XWTSXHCNI WQV SNRVI HTPV.

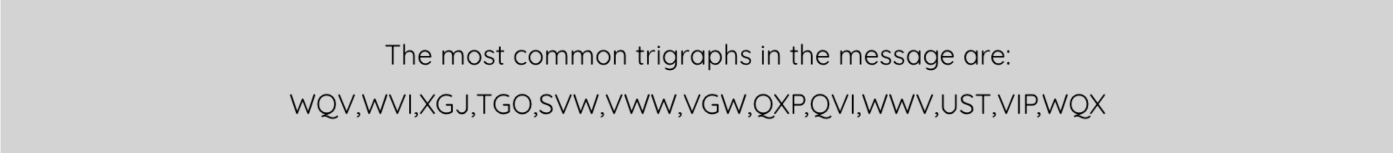
Cipher :

only after he had explained how ciphers are solved did he proceed toways of preventing solution. he capped his work with a cipher of his owninvention that he called "worthy of kings" and, like all inventors, claimedwas unbreakable. this was the cipher disk that founded polyalphabeticity. with this invention, the west, which up to this point hadequaled but had never surpassed the east in cryptology, took the leadthat it has never lost."i make two circles out of copper plates. one, the larger, is calledstationary, the smaller is called movable. the diameter of the stationaryplant is one-ninth greater than that of the movable plate. i divide thecircumference of each circle into 24 equal parts. these parts are calledcells. in the various cells of the larger circle i write the capital letters, oneat a time in red, in the usual order of the letters, a first, b second, c third,and then the rest, omitting h and k [and y] because they are notnecessary." this gave him 20 letters, since j, u, and w were not in hisalphabet, and in the remaining four spaces he inscribed the numbers 1and 4 in black. (the red and black seem to signify only that alberti likedcolors.) in each of the 24 cells of the movable circle he inscribed "a smallletter in black, and not in regular order like the stationary characters,but scattered at random. thus we may suppose the first of them to be a, the second g, the thirdq, and so on with the rest until the 24 cells of the circle are full; for thereare 24 characters in the latin alphabet, the last being et [probablymeaning "&"]. after completing these arrangements we place the smallercircle upon the larger so that a needle driven through the centers of bothmay serve as the axis of both and the movable plate may be revolvedaround it."the two correspondents—who, alberti carefully pointed out, musteach have identical disks—agree upon an index letter in the movabledisk, say k. then, to encipher, the sender places this prearranged indexletter against any letter of the outer disk. he informs his correspondentof this position of the disk by writing, as the first letter of the ciphertext,this letter of the outer ring. alberti gave the example of k being placedagainst b. "from this as a starting point all the other characters of themessage will acquire the force and sounds of the stationary charactersabove them."* so far nothing remarkable had happened. but in his nextsentence alberti placed cryptography's feet on the road to its moderncomplexity. "after writing three or four words, i shall change the positionof the index in our formula by turning the circle, so that the index k maybe, say, under d. so in my message i shall write a capital d, and from thispoint on [ciphertext] k will signify no longer b but d, and all the otherstationary letters at the top will receive new meanings."there is the *crucial point: "new meanings." each new position of theinner disk brings different letters opposite one another in the inner andouter rings. consequently, each shift means that plaintext letters wouldbe replaced with different ciphertext equivalents. for example, theplaintext word no might be enciphered to fc at one setting and to me atanother. equally, at each shift a given cipher-text letter would stand for adifferent plaintext letter than it did at the previous setting. thus, the fcthat formerly represented no might, at the new setting, stand forplaintext tu. this shift in both plain and cipher equivalents dif-*in alberti's disk, the outer capital letters are the plaintext and theinner lower-case letters are the ciphertext. this contradicts theconvention of this book, and is being used in the section on alberti onlyto avoid altering his text. the difference is signalimed by not using italicfor the lower case.*

Explanation :

First step was to find the frequencies for all letters which appear in the cryptogram. Then I started to make substitutions with the first 3 most common letters which for me were : E, T, A. Then we can observe that word: <u>then</u>, appeared very often in the text. Then we go for the next letter thich is I, and substitute in the interceptor. Then we look through all the word combinations and try to find a pattern , try to make words, for example if we have this combination likA , it is evident that A would go in ‘e’ , and substitute A with e   and so on.

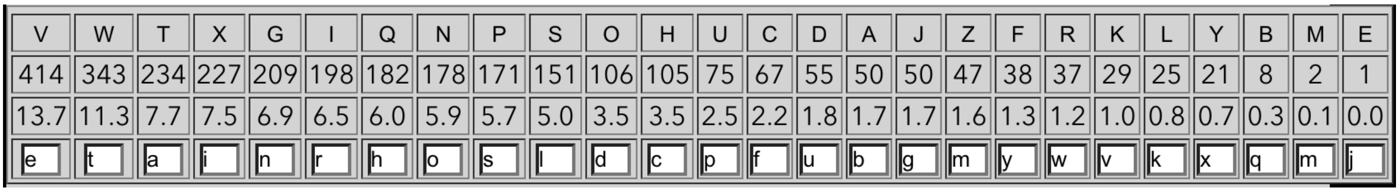
The most common trigraphs in the message :



The most common trigraphs in the message are:
WQV,WVI,XGJ,TGO,SVW,VWW,VGW,QXP,QVI,WWV,UST,VIP,WQX

Fig 3 – Trigraphs in message

The frequencies of the intercept are :



V	W	T	X	G	I	Q	N	P	S	O	H	U	C	D	A	J	Z	F	R	K	L	Y	B	M	E
414	343	234	227	209	198	182	178	171	151	106	105	75	67	55	50	50	47	38	37	29	25	21	8	2	1
13.7	11.3	7.7	7.5	6.9	6.5	6.0	5.9	5.7	5.0	3.5	3.5	2.5	2.2	1.8	1.7	1.7	1.6	1.3	1.2	1.0	0.8	0.7	0.3	0.1	0.0
e	t	a	i	n	r	h	o	s	l	d	c	p	f	u	b	g	m	y	w	v	k	x	q	m	j

Fig 4 – Frequencies of the intercept

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

Conclusion: In this lab activity, I understood how Cryptanalysis of monoalphabetic ciphers works, and how to use them. I used it to substitute each letter of the plaintext with a corresponding letter in the ciphertext, creating a one-to-one mapping. This simplicity makes them highly susceptible to frequency analysis. Attackers can exploit the fact that certain letters or combinations of letters in the plaintext will consistently map to specific letters in the ciphertext. By analyzing the frequency of these mappings, we can deduce the cipher's key and decrypt the message.