MINISTRY OF EDUCATION, CULTURE AND RESEARCH OF REPUBLIC OF MOLDOVA
TECHNICAL UNIVERSITY OF MOLDOVA
FACULTY OF COMPUTERS, INFORMATICS AND MICROELECTRONICS
DEPARTMENT OF SOFTWARE ENGINEERING AND AUTOMATICS

Criptography and Security

Laboratory work 2: Cryptanalysis of monoalphabetic substitution

Elaborated:
st.gr. FAF-211 Grama Alexandru
Verified:
asist.univ. Cătălin Mîțu

Chișinău, 2023

Introduction

It was intercepted a encrypted message which is known to have been obtained using a monoalphabetic cipher. By applying the frequency analysis attack, determine the original message, assuming it is a text written in English. Keep in mind that only the letters were encrypted, with the other characters remaining unencrypted.

*c = Pvhivw Nccxhv tgo wqvxi onzvpwxh ngvp cinz wqv Uixktwv - Nccxhv, anwqpdaoxkxpxngp nc wqv Unpw Nccxhv. Wqv Pvhivw Nccxhv rtp bdtiwvivo xg wqivvinnzp toenxgxgj wqv Cnivxjg Nccxhv tgo vgwvivo uixktwvsf cinz TahqdihqStgv. Cxiv tgo htgosvp adigvo hngpwtgwsf xg ngv innz; wqv pwtcc snojvo xgwqv nwqvip. Xw xghsdovo zvg rqn ztov wqvxi sxcv'p rnil wqv puvhxtswf ncdgpvtsxgj oxusnztwxh uthlvwp rxwq pdhq, ovcwgvpp wqtw wqvf hndso avivpvtsvo rxwqndw vkxovghv nc wtzuvixgj; ngv pdhq nuvgvi rtp E. V. Anov,ctwqvi nc Enqg Anov, Ei. Qv ivjdstisf puvgw wqivv qndip ng wqv oxputwhqvpnc wqv Lxgj nc Uidppxt, nuvgxgj wqvz tgo wqvg iv-pvtsxgj wqvz rxwqpuvhxts rty tgo htivcdssf hndgwvicvxwvo pvtsp. Uviqtup pdiuixpxgjsf xg tatpwxng nc qdztg ixjqwp, xwp xgwvihvuwxngp vgenfvo cdss svjtsxwf. Wqvpwtwdwv nc 1657 wqtw vpwtasxpqvo wqv unpwts pvikxhv ovhstivo ndwixjqw wqtwwqv ztxsp rviv wqv avpw zvtgp nc oxphnkvixgj otgjvindp tgo rxhlvoovpxjgp tjtxgpw wqv hnzzngrvtswq. Svtpvp nc 1660 tgo 1663, hngcxizvoaf wqv Unpw Nccxhv Thw nc 1711, uvizxwwvo jnkvigzvgw nccxhxtsp wn nuvgztxs dgovi rtiitgwp wqtw wqvf wqvzpvskvp xppdvo. Wqvf pxovpwvuuvo wqxpanwqvipnzv uinhvodiv af uinzdsjtwxgj tss-xghsdpxkv jvgvits rtiitgwp.*Wqv Pvhivw Nccxhv pvgw xgwvihvuwxngp vg hstxi wn wqv lxgj tgo wqnpv xghxuqvi wn wqv hifuwtgtsfpwp.Wqvf rviv lgnrg hnssvhwxkvsf tp wqv Ovhfuqvixgj Aitghq. Dgsxlv wqvPvhivw Nccxhv, wqv aitghq qto gn puvhxcxh snhtwxng. Xwp wxgf pwtcc nc vyuviwprnilvo stijvsf tw qnzv, ivhvxkxgj wqvxi ztwvixts af puvhxts zvppvgjvi.Gni qto xw tgf cnizts nijtgxmtwxng, wqv pvgxni Ovhfuqvivi avxgj zvivsfcxipw tzngj vbdtsp. Zniv pvhivw wqtg wqv Pvhivw Nccxhv, wqv aitghq'pcdgop htzv cinz pvhivw-pvikxhv zngvf xppdvo wn wqv Pvhivwtif nc wqv UnpwNccxhvcinz Utisxtzvgw'p pdiusdp ivkvgdv. Pvhdixwf rtp wxjqw—xg tss ncVgjstgo uinatasf ngsf 30 uvnusv lgvr rqtw oxusnztwxh hniivpungovghvrtp avxgj ivto tw tgf jxkvg znzvgw. Gvkviwqvsvpp, znpw zvg nc tcctxiprviv trtiv nc wqv uithwxhv nc nuvgxgj*Wqxp thwxkxwf cnizp wqv svjts uivhvovgw cni wqv znovig wtuuxgj ncwvsvuqngvp, tw svtpw xg Aixwtxg. Pxjgxcxhtgwsf, qnrvkvi, wqv pndihv nc wqvunrvi wn xgwvihvuw hnzzdgxhtwxngp qtp gvkvi avvg ovwvizxgvo. WqvHinrg pxzusf vyvihxpvo xw tgo, ovpuxwv nhhtpxngts ovatwv, qtp hngwxgdvown on pn, uivpdztasf rxwq wqv wthxw tuuinkts nc wqv udasxh tp gvhvpptifcni wqv ptcvwf nc wqv pwtwv.uixktwv svwwvip, tgo wqvf ncwvg vghxuqvivo wqvxi hniivpungovghv nivgwidpwvo xw wn uixktwv zvppvgjvip rqvg pvhivhf rtp vppvgwxts.Tcwvi wqv Vsvhwni nc Qtgnkvi pdhhvvovo wn wqv Vgjsxpq wqingv tpJvnijv X xg 1714, ivwtxgxgj wqv idsv nc wqv Jviztg pwtwv, wqv OvhfuqvixgjAitghq hnsstanitwvo rxwq wqv asthl hqtzavi ztxgwtxgvo tw Gxvgadij*

afwqv Qtgnkvixtg jnkvigzvgw. Hifuwtgtsfpwp Anov, Stzuv, tgo Gvdandijqto vkvg avvg xzuniwvo cinz wq-viv—tg xingxh ovkvsnuzvgw xg kxvr nc tivcdpts nc Rtssxp wn oxkdsjv qxp wvhqgxbdv wn Qtgnkvi t cvr fvtip vtisxvi.Ztxs nuvgxgj avhtzv qtaxwdts. Jvnijv tgo qxp pdhhvppnip wnnl thngpwtgw uvipngts xgwvivpw xg wqv rnil, ncwvg vghnditjxgj wtsvgw rxwqinfts andgwf. Hniivpungovghv rtp hsnpvsf rtwhqvo cni hixap wqtw rvivutppvo wn wqv Ovhfuqvixgj Aitghq.Odixgj wqv 1700p, wqv aitghq'p ndwudw tkvitjvo wrn ni wqivvoxputwhqvp t rvvl, tgo pnzvwxzvp ngv t otf. Xwp hifuwtgtsfpwp pnskvo wqvoxputwhqvp nc Citghv, Tdpwixt, Ptyngf tgo nwqvi Jviztg pwtwvp, Unstgo,Putxg, Uniwdjts, Qnsstgo, Ovgztil, Prvovg, Ptioxgxt, Gtusvp tgo nwqviXwtsxtg pwtwvp, Jivvhv, Wdilvf, Idppxt, tgo, stwvi, wqv Dgxwvo Pwtwvp. Wqvivhnio nc Civghq xgwvihvuwxngp hnkvip wrn hvgwdixvp tgo hnzuixpvp cxkvknsdzvp nc xgwvihvuwp wnwtsxgj 2,020 utjvp usdp wqivv knsdzvp nc lvfp.Uviqtup zniv wfuxhts xp wqv Putgxpq onppxvi—wqivv knsdzvp nc xgwvi-hvuwpcinz 1719 wn 1839 wnwtsxgj 872 utjvp. Gnw tss nc wqv zvpptjvp rvivpnskvo tw wqv wxzv nc wqvxi xgwvihvuwxng. Ztgf rviv qvso vxwqvi dgwxsvgndjq qto thhdzdstwvo cni t pdhhvppcds twwthl ni dgwxs t gvvo tinpv cniwqvxi pnsdwxng.

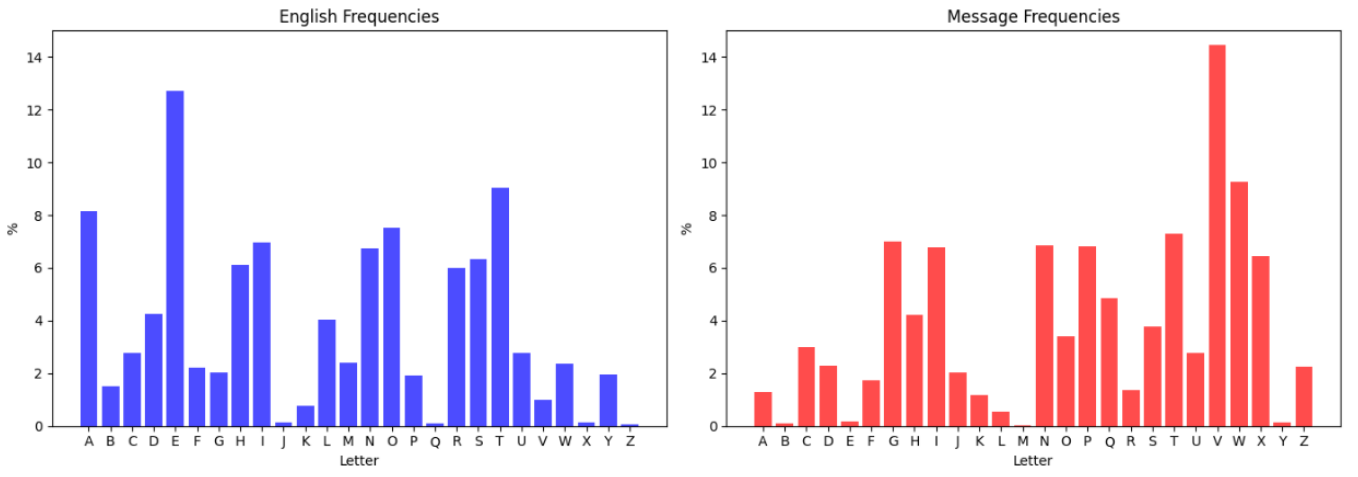
After ussing the site: *https://crypto.interactive-maths.com/frequency-analysis-breaking-the-code.html* , I obtained this frequency of letters:

Table 1.1 - Table of frequency

Letter	V	W	T	G	N	P	I	X	Q	H	S	O	C
%	14,5	9,3	7,3	7,0	6,9	6,8	6,8	6,5	4,9	4,2	3,8	3,4	3,0
Letter	U	D	Z	J	F	R	A	K	L	E	Y	B	M
%	2,8	2,3	2,3	2,0	1,7	1,3	1,3	1,2	0,50	0,2	0,1	0,1	0,0

And the graphics of the encrypted text are in this way:

Figure 1.1 - Graphs of frequency



The first step is to find the frequencies of all letters that appear in the cryptogram, as shown in Table. Below, we can observe the graphical representation of the letter frequencies in the English language (figure

on the left) and the frequencies of letters in the intercepted message (figure on the right). Now that we have all the letter frequencies from the encrypted text, we can start making some substitutions. We see that the most frequent letter in the encrypted text is ”V”, closely followed by ”W”. From the above Figure 1.1 and Table 1.1, we can guess that these two letters represent ”E” and ”T” respectively. After making these substitutions, we obtain:

The next step we look to words that consist of two, three letters and where we have double letters such as "***ss***","**oo**".I recognized first word that was "***wqv***", which is "***the***", then we looked at pronouns near years such as "***xg 1714***", which could be "***in 1714***", "***nc 1657***" is "***of 1657***",the word "***onep***" is "***ones***", the obtained word "***zost***" is crearly the word "***most***", the word "***offihe***" is "***office***" , is "***tassinj***" is"***tapping***","***inteicept***" is "***intercept***".

After some time of thinking about the encryped text, we have the result : *c = secret office and their domestic ones from the private - office, bothsubdivisions of the post office. the secret office was quartered in threerooms adjoining the foreign office and entered privately from abchurchlane. fire and candles burned constantly in one room; the staff lodged inthe others. it included men who made their life’s work the specialty ofunsealing diplomatic packets with such, deftness that they could beresealed without evidence of tampering; one such opener was j. e. bode,father of john bode, jr. he regularly spent three hours on the dispatchesof the king of prussia, opening them and then re-sealing them withspecial wax and carefully counterfeited seals. perhaps surprisingly in abastion of human rights, its interceptions enjoyed full legality. thestatute of 1657 that established the postal service declared outright thatthe mails were the best means of discovering dangerous and wickeddesigns against the commonwealth. leases of 1660 and 1663, confirmedby the post office act of 1711, permitted government officials to openmail under warrants that they themselves issued. they sidestepped thisbothersome procedure by promulgating all-inclusive general warrants.*the secret office sent interceptions en clair to the king and those incipher to the cryptanalysts.they were known collectively as the decyphering branch. unlike thesecret office, the branch had no specific location. its tiny staff of expertsworked largely at home, receiving their material by special messenger.nor had it any formal organization, the senior decypherer being merelyfirst among equals. more secret than the secret office, the branch’sfunds came from secret-service money issued to the secretary of the postoffice from parliament’s surplus revenue. security was tight—in all ofengland probably only 30 people knew what diplomatic correspondencewas being read at any given moment. nevertheless, most men* of affairswere aware of the practice of opening*this activity forms the legal precedent for the modern tapping oftelephones, at least in britain. significantly, however, the source of thepower to intercept communications has never been determined. thecrown simply exercised it and, despite occasional debate, has continuedto do so, presumably with the tacit approval of the public as necessaryfor the safety

of the state. private letters, and they often enciphered their correspondence orentrusted it to private messengers when secrecy was essential.after the elector of hanover succeeded to the english throne asgeorge i in 1714, retaining the rule of the german state, the decypheringbranch collaborated with the black chamber maintained at nienburg bythe hanoverian government. cryptanalysts bode, lampe, and neubourghad even been imported from there—an ironic development in view of arefusal of wallis to divulge his technique to hanover a few years earlier.mail opening became habitual. george and his successors took aconstant personal interest in the work, often encouraging talent withroyal bounty. correspondence was closely watched for cribs that werepassed to the decyphering branch.during the 1700s, the branch's output averaged two or threedispatches a week, and sometimes one a day. its cryptanalysts solved thedispatches of france, austria, saxony and other german states, poland,spain, portugal, holland, denmark, sweden, sardinia, naples and otheritalian states, greece, turkey, russia, and, later, the united states. therecord of french interceptions covers two centuries and comprises fivevolumes of intercepts totaling 2,020 pages plus three volumes of keys.perhaps more typical is the spanish dossier—three volumes of interceptsfrom 1719 to 1839 totaling 872 pages. not all of the messages weresolved at the time of their interception. many were held either untilenough had accumulated for a successful attack or until a need arose fortheir solution. Now we obtain Table 1.2 which are the decrypted letters.

Table 1.2 - Table of decryped letters

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

Conclusion

One of the primary vulnerabilities of monoalphabetic ciphers is their susceptibility to frequency analysis. Since languages have distinct letter frequencies – for example, in English, letters like ’e’ and ’t’ are quite common – a significant amount of ciphertext can reveal patterns that align with the known frequencies of the language in which the message is written. By examining these patterns, a cryptanalyst can often make educated guesses about the substitutions used and thereby decrypt the message. While monoalphabetic ciphers were once considered secure, the advent of frequency analysis has rendered them relatively easy to break, especially with larger amounts of ciphertext available. Today, these ciphers serve primarily as educational tools or puzzles rather than serious cryptographic methods for ensuring confidentiality.

As the world of cryptography has advanced, so too have the techniques for ensuring secure communication. Modern cryptographic algorithms are vastly more complex and resilient against various forms of attack. However, understanding the strengths and weaknesses of foundational ciphers like the monoalphabetic frequency cipher offers valuable insight into the principles and evolution of cryptographic security.