MINISTRY OF EDUCATION AND RESEARCH

OF THE REPUBLIC OF MOLDOVA

Technical University of Moldova

Faculty of Computers, Informatics and Microelectronics

Department of Software and Automation Engineering

MAXIM ALEXEI, FAF-232

Report

Laboratory work n.2

Cryptanalysis of monoalphabetic substitution

on Cryptography and Security

Checked by:

Maia Zaica, *university assistant*

FCIM, UTM

Chișinău – 2025

1. Objective:

Learn how to decrypt monoalphabetic substitution ciphers from a given encrypted message, applying the frequency analysis method, involving identifying letter frequencies in the ciphertext, comparing them to known English letter frequencies, and making substitutions to reconstruct the original message.

2. Theoretical Background:

Monoalphabetic substitution ciphers, while seemingly secure, possess a fundamental weakness: they preserve the relative frequencies of letters from the original plaintext language. This vulnerability allows for cryptanalysis using frequency analysis, a technique that exploits these statistical patterns to reveal the underlying plaintext.

The core principle involves comparing the observed frequency of letters in the ciphertext with the known frequency distribution of letters in the language of the original message (e.g., English). For a sufficiently long ciphertext, the frequencies of encrypted letters will closely mirror those of their corresponding plaintext letters. For example, if E is the most common letter in English, and X is the most frequent letter in the ciphertext, it is highly probable that X represents E.

Beyond individual letter frequencies, cryptanalysts employ several advanced techniques to enhance the attack:

- **Common Word Patterns:** Analyzing the lengths and patterns of words in the ciphertext can provide clues. For instance, frequently appearing short words like "the", "a", and "I" in English can be identified, leading to quick substitutions.

- **Digraph and Trigraph Frequencies:** Examining the occurrence of two-letter combinations, digraphs like "TH", "HE", "AN", and three-letter combinations, trigraphs like "THE", "AND", "ENT", in both the ciphertext and the target language can further refine substitutions. These combinations have distinct frequencies that aid in pattern recognition.
- **Double Letters:** The presence of double letters, such as "SS", "EE", "TT", "OO", "FF", in English can also be a valuable indicator.
- **Contextual Analysis and Human Insight:** While frequency tables provide a statistical starting point, human intuition and contextual understanding are crucial for success. Cryptanalysts look for emerging partial words, common prefixes and suffixes, and grammatical structures to make informed guesses and iterate on substitutions. This iterative process, often involving an "educated guess and check" approach, gradually reveals the plaintext.

3. Task, Variant I:

Given the cipher text, which was encrypted via a monoalphabetic cipher algorithm. Apply the frequency analysis method, in order to obtain the message in plaintext, knowing that the text is written in English. Also, to be kept in mind that only the letters were encrypted, the other characters must remain the same.

c = *NG T OTF gvtisf 4,000 fvtip tjn, xg t wnrg htssvo Zvgvw Lqdcdaniovixgj wqv wqxg ixaang nc wqv Gxsv, t ztpwvi phixav plvwhqvo ndw wqvqxvinjsfuqp wqtw wnso wqv pwnif nc qxp snio'p sxcv—tgo xg pn onxgj qvnuvgvo wqv ivhniovo qxpwnif nc hifuwnsnjf. Qxp rtp gnw t pfpwvz nc pvhivwrixwxgj tp wqv znovig rniso lgnrp xw; qv dpvo gn cdssf ovkvsnuvo hnov nc qxvinjsfuqxhpfzans pdapwxwdwxngp. Qxp xgphixuwxng, htikvo tandw 1900 A.H. xgwn wqvsxkxgj inhl xg wqv ztxg hqtzavi nc wqv wnza nc wqv gnasvztgLqgdzqnwvu XX, zvivsf dpvp pnzv dgdpdts qxvinjsfuqxh pfzansp qvivtgo wqviv xg usthv nc wqv zniv nioxgtif ngvp. Znpw nhhdi xg wqv stpw 20hnsdzgp nc wqv xgphixuwxng'p 222, xg t pvhwxng ivhnioxgj wqv zngdzvgwpwqtw*

Lqgdzqnwvu qto vivhwvo xg wqv pvikxhv nc wqv uqtitnq TzvgvzqvwXX. Wqv xgwvgwxng rtp gnw wn ztlv xw qtio wn ivto wqv wvyw. Xw rtp wnxzutiw t oxjgxwf tgo tdwqnixwf wn xw, uviqtup xg wqv ptzv rtf wqtw tjnkvigzvgw uinhstztwxng rxss puvss ndw "Xg wqv fvti nc Ndi Snio Ngvwqndptgo vxjqw qdgoivo tgo pxywf wqivv" xgpwvto nc edpw ri wxgj "1863."Wqv tgngfzndp phixav ztf tspn qtkv avvg ovzngpwitwxgj qxp lgnrsvojvcni unpwvixwf. Wqdp wqv xgphixuwxng rtp gnw pvhivw rixwxgj, adw xwxghniunitwvo ngv nc wqv vppvgwxts vsvzvgwp nc hifuwnjituqf: t ovsxavitwvwitgpcniztwxng nc wqv rixwxgj. Xw xp wqv nsovpw wvyw lgnrg wn on pn.Wqv witgpcniztwxngp nhhdi xg cdgvitif cnizdstp, xg t qfzg wn Wqnwq,xg t hqtuwvi nc wqv Annl nc wqv Ovto, ng wqv ptihnuqtjdp nc wqv uqtitnqPvwx X, xg infts wxwsvp oxpustfvo xg Sdyni, ng wqv tihqxwitkv nc wqv Wvzusv nc Sdyni, ng pwvsv, xgstdotwnif axnjituqxh xgphixuwxngp. Wqviv xp gnwqxgj zvtgw wn avhnghvtsvo xg tss wqxp; xgovvo, ztgf nc wqv pwtwvzvgwp tiv ivuvtwvo xgnioxgtif cniz ixjqw gvyw wn wqv tswvivo ngvp. Rqf, wqvg, wqvwitgpcniztwxngp? Pnzvwxzvp cni vppvgwxtssf wqv ptzv ivtpng tp xgLqgdzqnwvu'p wnza: wn xzuivpp wqv ivtovi. Nhhtpxngtssf cni thtssxjituqxh ni ovhnitwxkv vccvhw; itivsf, wn xgoxhtwv t hngwvzunitifuingdghxtwxng; uviqtup vkvg cni t ovsxavitwv tihqtxpz tp t ivthwxngtjtxgpw cnivxjg xgcsdvghv.Adw ztgf xgphixuwxngp tiv wxghwdivo, cni wqv cxipw wxzv, rxwq wqvpvhngo vppvgwxts cni hifuwnsnjf—pvhivhf. Xg t cvr htpvp, wqv pvhivhf rtpxgwvgovo wn xghivtpv wqv zfpwvif tgo qvghv wqv tihtgv ztjxhts unrvip nchviwtxg ivsxjxndp wvywp. Adw wqv pvhivhf xg ztgf zniv htpvp ivpdswvo cinzwqv dgovipwtgotasv ovpxiv nc wqv Vjfuwxtgp wn qtkv utppvipaf ivto wqvxivuxwtuqp tgo pn hngcvi dung wqv ovutiwvo wqv asvppxgjp rixwwvg wqvivxg.Xg Vjfuw, rxwq xwp hnghvgwitwxng dung wqv tcwvisxcv, wqv gdzavi nc wqvpvxgphixuwxngp pnng • uinsxcvitwvo wn pdhq tg vywvgw wqtw wqv twwvgwxng tgowqv jnnorxss nc kxpxwnip cstjjvo. Wn ivkxkv wqvxi xgwvivpw, wqv phixavpovsxavitwvsf ztov wqv xgphixuwxngp t axw naphdiv. Wqvf xgwinodhvo wqvhifuwnjituqxh pxjgp wn htwhq wqv ivtovi'p vfv, ztlv qxz rngovi, tgowvzuw qxz xgwn dgixoosxgj wqvz — tgo pn xgwn ivtoxgj wqv asvppxgjp. Xwrtp t pniw nc Ztoxpng Tkvgdv wvhqgxbdv xg wqv Ktssvf nc wqv Lxgjp. Adwwqv wvhqgxbdv ctxsvo dwwvisf. Xgpwvto nc xgwvivpwxgj wqv ivtovip, xwvkxovgwsf ovpwinfvo vkvg wqv psxjqwvpw ovpxiv wn ivto wqv vuxwtuqp, cnipnng tcwvi wqv cdgvitif hifuwnjituqf rtp avjdg, xw rtp tatgongvo.

4. Cryptanalysis of monoalphabetic substitution:

Firstly, I have taken the cipher ***c*** from the tasks section, and I placed it on this website[1], in order to obtain the frequency of each letter and the number of digraphs, trigraphs, and doubles.

Rank	Cipher Letter	Count	Cipher %	English Letter	Expected %
1	V	327	12.7	E	12.7
2	W	276	10.8	T	9.1
3	N	205	8.0	A	8.2
4	X	197	7.7	O	7.5
5	G	196	7.6	I	7.0
6	T	178	6.9	N	6.7
7	I	176	6.9	S	6.3
8	P	165	6.4	H	6.1
9	Q	145	5.7	R	6.0
10	O	93	3.6	D	4.3
11	H	83	3.2	L	4.0
12	S	80	3.1	C	2.8
13	Z	63	2.5	U	2.8
14	F	61	2.4	M	2.4
15	C	59	2.3	W	2.4
16	U	58	2.3	F	2.2
17	D	56	2.2	G	2.0
18	J	45	1.8	Y	2.0
19	A	35	1.4	P	1.9
20	R	27	1.1	B	1.5

21	K	17	0.7	V	1.0
22	L	13	0.5	K	0.8
23	Y	8	0.3	J	0.15
24	B	2	0.1	X	0.15
25	E	1	0.0	Q	0.1
26	M	0	0.0	Z	0.07

Table 1 – Frequency of letters in the cipher and in English

Cipher	**Count Cipher**	**English**
WQ	87	TH
QV	78	HE
XG	68	AN
NG	45	IN
IV	44	ER
WV	43	ON
VI	42	RE
WX	37	ED

Table 2 – Digraphs frequency in the cipher and the most common ones in English

Cipher	Count Cipher	English
WQV	72	THE
XNG	19	AND
WXN	18	THA
XGJ	16	ENT
CNI	13	ION
TGO	12	TIO
VGW	12	FOR
IVT	11	NDE

Table 3 – Trigraphs frequency in the cipher and the most common ones in English

Cipher	Count Cipher	English
SS	10	SS
PP	7	EE
NN	4	TT
WW	4	FF
HH	3	LL

Table 4 – Doubles frequency in the cipher and the most common ones in English

After gathering all the insights about the frequencies that occur within the given cipher, I have assumed that the letter V from my cipher is actually E in plaintext, as its frequency within my cipher is equal, or very close, to E, in terms of letters frequency in English.

On top of that, I have also realized that the digraph WQ is the most frequent one within my cipher, and QV is the second most frequent one. Therefore, knowing that V is actually E, and the fact that first two most frequent digraphs in English are TH

and HE, we observe the overlapping of the Q within those top two digraphs from my cipher, which means that Q is actually H in plaintext.

Moreover, we can also assume that W is T in plaintext, based on the theory of those two digraphs. In addition, I can enforce these assumptions, with the fact that WQV is the most common trigraph within my cipher, which would make it equivalent to *THE* in English. Since we know what Q and V are, that leaves us with W being T in the original message, as mentioned before.

So, the substitutions for this step were:

- $V \rightarrow E$
- $Q \rightarrow H$
- $W \rightarrow T$

NG T OTF getisf 4,000 fetip tjn, xg t tnrg htsseo Zeget Lhdcdanioeixgj the thxg ixaang nc the Gxse, t ztptei phixae plethheo ndt thehxeinjsfuhp thtt tnso the ptnif nc hxp snio'p sxce—tgo xg pn onxgj henuegeo the iehnioeo hxptnif nc hifutnsnjf. Hxp rtp gnt t pfptez nc pehietrixtxgj tp the znoeig rniso lgnrp xt; he dpeo gn cdssf oekesnueo hnoe nc hxeinjsfuhxhpfzans pdaptxtdtxngp. Hxp xgphixutxng, htikeo tandt 1900 A.H. xgtn thesxkxgj inhl xg the ztxg hhtzaei nc the tnza nc the gnaseztgLhgdzhnteu XX, zeiesf dpep pnze dgdpdts hxeinjsfuhxh pfzansp heietgo theie xg usthe nc the znie nioxgtif ngep. Znpt nhhdi xg the stpt 20hnsdzgp nc the xgphixutxng'p 222, xg t pehtxng iehnioxgj the zngdzegtpthtt Lhgdzhnteu hto eiehteo xg the peikxhe nc the uhtitnh TzegezhetXX. The xgtegtxng rtp gnt tn ztle xt htio tn ieto the teyt. Xt rtp tnxzutit t oxjgxtf tgo tdthnixtf tn xt, ueihtup xg the ptze rtf ***thtt*** *tjnkeigzegt uinhsttxng rxss puess ndt "Xg the feti nc Ndi Snio Ngethndptgo exjht hdgoieo tgo pxytf* ***thiee*** *" xgpteto nc edpt ri txgj "1863."The tgngfzndp phixae ztf tspn htke aeeg oezngptittxgj hxp lgnrseojecni unpteixtf. Thdp the xgphixutxng rtp gnt pehiet rixtxgj, adt xtxghniunitteo nge nc the eppegtxts esezegtp nc hifutnjituhf: t oesxaeittetitgpcnizttxng nc the rixtxgj. Xt xp the nsoept teyt lgnrg tn on pn.The titgpcnizttxngp nhhdi xg cdgeitif cnizdstp, xg t hfzg tn Thnth,xg t hhtutei nc the Annl nc the Oeto, ng the ptihnuhtjdp nc the uhtitnhPetx X, xg infts txtsep oxpustfeo xg*

Sdyni, ng the tihhxtitke nc the Tezuse nc Sdyni, ng ptese, xgstdottnif axnjituhxh xgphixutxngp. Theie xp gnthxgj zetgt tn aehnghetseo xg tss thxp; xgoeeo, ztgf nc the ptttezegtp tie ieuetteo xgnioxgtif cniz ixjht geyt tn the tsteieo ngep. Rhf, theg, thetitgpcnizttxngp? Pnzetxzep cni eppegtxtssf the ptze ietpng tp xgLhgdzhnteu'p tnza: tn xzuiepp the ietoei. Nhhtpxngtssf cni thtssxjituhxh ni oehnittxke ecceht; itiesf, tn xgoxhtte t hngtezunitifuingdghxttxng; ueihtup ekeg cni t oesxaeitte tihhtxpz tp t iethtxngtjtxgpt cniexjg xgcsdeghe.Adt ztgf xgphixutxngp tie txghtdieo, cni the cxipt txze, rxth thepehngo eppegtxts cni hifutnsnjf—pehiehf. Xg t cer htpep, the pehiehf rtpxgtegoeo tn xghietpe the zfpteif tgo heghe the tihtge ztjxhts unreip ncheittxg iesxjxndp teytp. Adt the pehiehf xg ztgf znie htpep iepdsteo cinzthe dgoeipttgotase oepxie nc the Ejfutxtgp tn htke utppeipaf ieto thexieuxttuhp tgo pn hngcei dung the oeutiteo the aseppxgjp rixtteg theiexg.Xg Ejfut, rxth xtp hnghegtittxng dung the tcteisxce, the gdzaei nc thepexgphixutxngp pnng • uinsxceitteo tn pdhh tg eytegt thtt the tttegtxng tgothe jnnorxss nc kxpxtnip cstjjeo. Tn iekxke thexi xgteiept, the phixaepoesxaeittesf ztoe the xgphixutxngp t axt naphdie. ***Thef*** *xgtinodheo thehifutnjituhxh pxjgp tn htthh the ietoei'p efe, ztle hxz rngoei, tgotezut hxz xgtn dgixoosxgj thez — tgo pn xgtn ietoxgj the aseppxgjp. Xtrtp t pnit nc Ztoxpng Tkegde tehhgxbde xg the Ktssef nc the Lxgjp. Adtthe tehhgxbde ctxseo dtteisf. Xgpteto nc xgteieptxgj the ietoeip, xtekxoegtsf oeptinfeo ekeg the psxjhtept oepxie tn ieto the euxttuhp, cnipnng tctei the cdgeitif hifutnjituhf rtp aejdg, xt rtp tatgongeo.*

Next, after analyzing the cipher with the replacements, we can identify some words that have only one letter that is encrypted, such as ***thtt***, ***thiee*** and ***thef***.

Therefore, after analyzing the top 60000 “lemmas”[2], in case of ***thtt*** we can observe that the word *th**a**t* is the most common word in English that matches this pattern. Moreover, in the cipher, we can also observe that we have got lots of single letter words, in this case mapped as *T*s, knowing that in English the most single letter words are *A* and *I,* we can safely assume that *T* is actually *A*.

In continuity, we have got ***thiee***, so knowing, according to the spreadsheet[2], that *three* is the most common word in English matching this pattern, we can specify that *I* is *R* in the case of this cipher.

On top of that, there is also ***thef***, and, again, the most common word in English matching this pattern is actually *the**y***, so we've realized that *F* is *Y*.

So, the substitutions for this step were:

- $T \rightarrow A$
- $I \rightarrow R$
- $F \rightarrow Y$

NG A OAY gearsy 4,000 ***yearp*** *ajn, xg a tnrg hasseo Zeget Lhdcdanroerxgj the thxg rxaang nc the Gxse, a zapter phrxae plethheo ndt thehxernjsyuhp that tnso the ptnry nc hxp snro'p sxce—ago xg pn onxgj henuegeo the rehnroeo hxptnry nc hryutnsnjy. Hxp rap gnt a pyptez nc pehretrrxtxgj ap the znoerg rnrso lgnrp xt; he dpeo gn cdssy oekesnueo hnoe nc hxernjsyuhxhpyzans pdaptxtdtxngp. Hxp xgphrxutxng, harkeo aandt 1900 A.H. xgtn thesxkxgj rnhl xg the zaxg hhazaer nc the tnza nc the gnasezagLhgdzhnteu XX, zeresy dpep pnze dgdpdas hxernjsyuhxh pyzansp hereago there xg usahe nc the znre nroxgary ngep. Znpt nhhdr xg the sapt 20hnsdzgp nc the xgphrxutxng'p 222, xg a pehtxng rehnroxgj the zngdzegtpthat Lhgdzhnteu hao erehteo xg the perkxhe nc the uharanh AzegezhetXX. The xgtegtxng rap gnt tn zale xt haro tn reao the teyt. Xt rap tnxzuart a oxjgxty ago adthnrxty tn xt, uerhaup xg the paze ray that ajnkergzegt urnhsazatxng rxss puess ndt "Xg the year nc Ndr Snro Ngethndpago exjht hdgoreo ago pxyty three" xgpteao nc edpt rr txgj "1863."The agngyzndp phrxae zay aspn hake aeeg oezngptratxgj hxp lgnrseojecnr unpterxty. Thdp the xgphrxutxng rap gnt pehret rrxtxgj, adt xtxghnrunrateo nge nc the eppegtxas esezegtp nc hryutnjrauhy: a oesxaeratetragpcnrzatxng nc the rrxtxgj. Xt xp the nsoept teyt lgnrg tn on pn.The tragpcnrzatxngp nhhdr xg cdgerary cnrzdsap, xg a hyzg tn Thnth,xg a hhauter nc the Annl nc the Oeao, ng the parhnuhajdp nc the uharanhPetx X, xg rnyas txtsep oxpusayeo xg Sdynr, ng the arhhxtrake nc the Tezuse nc Sdynr, ng ptese, xgsadoatnry axnjrauhxh xgphrxutxngp. There xp gnthxgj zeagt tn aehngheaseo xg ass thxp; xgoeeo, zagy nc the ptatezegtp*

*are reueateo xgnroxgary cnrz rxjht geyt tn the astereo ngep. Rhy, theg, thetragpcnrzatxngp? Pnzetxzep cnr eppegtxassy the paze reapng ap xgLhgdzhnteu'p tnza: **tn** xzurepp the reaoer. Nhhapxngassy cnr ahassxjrauhxh **nr** oehnratxke ecceht; raresy, **tn** xgoxhate a hngtezunraryurngdghxatxng; uerhaup ekeg cnr a oesxaerate arhhaxpz ap a reahtxngajaxgpt cnrexjg xgcsdeghe.Adt zagy xgphrxutxngp are txghtdreo, cnr the cxrpt txze, rxth thepehngo eppegtxas cnr hryutnsnjy—pehrehy. Xg a cer hapep, the pehrehy rapxgtegoeo tn xghreape the zyptery ago heghe the arhage zajxhas unrerp nchertaxg resxjxndp teytp. Adt the pehrehy xg zagy znre hapep repdsteo crnzthe dgoerptagoaase oepxre nc the Ejyutxagp tn hake uapperpay reao thexreuxtauhp ago pn hngcer dung the oeuarteo the aseppxgjp rrxtteg therexg.Xg Ejyut, rxth xtp hnghegtratxng dung the actersxce, the gdzaer nc thepexgphrxutxngp pnng • urnsxcerateo tn pdhh ag eytegt that the attegtxng agothe jnnorxss nc kxpxtnrp csajjeo. Tn rekxke thexr xgterept, the phrxaepoesxaeratesy zaoe the xgphrxutxngp a axt naphdre. They xgtrnodheo thehryutnjrauhxh pxjgp tn hathh the reaoer'p eye, zale hxz rngoer, agotezut hxz xgtn dgrxoosxgj thez — ago pn xgtn reaoxgj the aseppxgjp. Xtrap a pnrt nc Zaoxpng Akegde tehhgxbde xg the Kassey nc the Lxgjp. Adtthe tehhgxbde caxseo dttersy. Xgpteao nc xgtereptxgj the reaoerp, xtekxoegtsy oeptrnyeo ekeg the psxjhtept oepxre tn reao the euxtauhp, cnrpnng acter the cdgerary hryutnjrauhy rap aejdg, xt rap aaagongeo.*

Next, analyzing the cipher after the new substitution, we have the word ***yearp,*** which is clearly *year**s*** in terms of frequency in English, and based on the Table 1 we can also see that the frequency of *P* in our cipher is of 6.4% and *S* has a frequency of 6.3% in English.

Furthermore, we also got words like ***tn*** and ***nr***, where we can assume that *N* is *O*, according to the fact that ***to*** and ***or*** are most common words in English with that structure. Moreover, the frequency of *N* in our cipher is of 8% and 7.5% of *O* in English, which makes sense to assume that *N* is actually *O*.

So, the substitutions for this step were:

- $P \rightarrow S$

➢ $N \rightarrow O$

*OG A OAY **gearsy** 4,000 years ajo, xg a torg hasseo Zeget Lhdcdaoroerxgj the thxg rxaaog oc the Gxse, a zaster shrxae slethheo odt thehxerojsyuhs that toso the story oc hxs soro's sxce—ago **xg** so ooxgj heouegeo the rehoroeo **hxstory** oc hryutosojy. Hxs ras got a systez oc sehretrrxtxgj as the zooerg rorso lgors xt; he dseo go cdssy oekesoueo hooe oc hxerojsyuhxhsyzaos sdastxtdtxogs. **Hxs** xgshrxutxog, harkeo aaodt 1900 A.H. xgto thesxkxgj rohl xg the zaxg hhazaer oc the toza oc the goasezagLhgdzhoteu XX, zeresy dses soze dgdsdas hxerojsyuhxh syzaoss hereago there xg usahe oc the zore oroxgary oges. Zost ohhdr xg the sast 20hosdzgs oc the xgshrxutxog's 222, xg a sehtxog rehoroxgj the zogdzegtsthat Lhgdzhoteu hao erehteo xg the serkxhe oc the uharaoh AzegezhetXX. The xgtegtxog ras got to zale xt haro to reao the teyt. Xt ras toxzuart a oxjgxty ago adthorxty to xt, uerhaus xg the saze ray that ajokergzegt urohsazatxog rxss suess odt "Xg the year oc Odr Soro Ogethodsago exjht hdgoreo ago sxyty three" xgsteao oc edst rr txgj "1863."The agogyzods shrxae zay asso hake aeeg oezogstratxgj hxs lgorseojecor uosterxty. Thds the xgshrxutxog ras got sehret rrxtxgj, adt xtxghoruorateo oge oc the essegtxas esezegts oc hryutojrauhy: a oesxaeratetragscorzatxog oc the rrxtxgj. Xt xs the osoest teyt lgorg to oo so.The tragscorzatxogs ohhdr xg cdgerary corzdsas, xg a hyzg to Thoth,xg a hhauter oc the Aool oc the Oeao, og the sarhouhajds oc the uharaohSetx X, xg royas txtses oxsusayeo xg Sdyor, og the arhhxtrake oc the Tezuse oc Sdyor, og stese, xgsadoatory axojrauhxh xgshrxutxogs. There xs gothxgj zeagt to aehogheaseo xg ass thxs; xgoeeo, zagy oc the statezegts are reueateo xgoroxgary corz rxjht geyt to the astereo oges. Rhy, theg, thetragscorzatxogs? Sozetxzes cor essegtxassy the saze reasog as xgLhgdzhoteu's toza: to xzuress the reaoer. Ohhasxogassy cor ahassxjrauhxh or oehoratxke ecceht; raresy, to xgoxhate a hogtezuoraryurogdghxatxog; uerhaus ekeg cor a oesxaerate arhhaxsz as a reahtxogajaxgst corexjg xgcsdeghe.Adt zagy xgshrxutxogs are txghtdreo, cor the cxrst txze, rxth thesehogo essegtxas cor hryutosojy—sehrehy. **Xg** a cer hases, the sehrehy rasxgtegoeo to xghrease the zystery ago heghe the arhage zajxhas uorers ochertaxg resxjxods teyts. Adt the sehrehy **xg** zagy zore hases resdsteo crozthe dgoerstagoaase oesxre oc the Ejyutxags to hake uassersay reao thexreuxtauhs ago so hogcer duog the oeuarteo the asessxgjs rrxtteg therexg.Xg Ejyut, rxth xts hoghegtratxog duog the actersxce, the gdzaer **oc** thesexgshrxutxogs **soog** • urosxcerateo to sdhh ag eytegt that the attegtxog agothe jooorxss oc kxsxtors csajjeo. To rekxke thexr xgterest, the shrxaesoesxaeratesy zaoe the xgshrxutxogs a axt oashdre. They xgtroodheo thehryutojrauhxh sxjgs to hathh the reaoer's eye, zale hxz rogoer, agotezut hxz xgto dgrxoosxgj thez — ago so xgto reaoxgj the asessxgjs. Xtras a sort oc Zaoxsog Akegde tehhgxbde xg the Kassey **oc** the Lxgjs. Adtthe tehhgxbde caxseo dttersy. Xgsteao oc xgterestxgj the reaoers, xtekxoegtsy oestroyeo*

*ekeg the ssxjhtest oesxre to reao the euxtauhs, corsoog **acter** the cdgerary hryutojrauhy ras aejdg, xt ras aaagoogeo.*

Furthermore, we have obtained a word like ***hxstory***, which clearly tells us that *X* is *I*, which is confirmed by the similarity of their frequency represented in Table 1. Also, if we have a look at the word ***soo**g,* which according to the most common words list, *soo**n*** is the word that matches the pattern. Now, assuming that *X* is *I* and *G* is *N*, can also be backed up by the word *xg* that appears several times within the cipher, as if we tried to substitute the letters we would obtain *in* which makes sense, and is also quite frequent in English.

In continuity, we have also got ***acter*** that leaves us two choices, either *a**f**ter*, which is more common, or *a**l**ter*, which is less common. Looking at *Table 1*, we can observe that the frequency of *C* in our cipher, 2.3%, is more similar to the one of *F*, 2.2%, in English than *L*, 4%, so *F* is more suitable, and that can also be backed up by ***o**c* which can be mapped to ***of*** while maintaining the sense. Whereas *L* would be better placed here ***g**ear**s**y* instead of *R*, as we know that *G* is *N*, so we get ***n**ear**l**y*, which makes sense as the frequency of *S* in our cipher text is similar to the one of *L* in English.

So, the substitutions for this step were:

- $X \rightarrow I$
- $G \rightarrow N$
- $S \rightarrow L$
- $C \rightarrow F$

*ON A **OAY** nearly 4,000 years ajo, in a torn halleo Zenet Lhdfdaoroerinj the thin riaaon of the Nile, a zaster shriae slethheo odt thehierojlyuhs that tolo the story of his loro's life—**ano** in so **ooinj** heoueneo the rehoroeo history of hryutolojy. His*

__ras__ not a systez of sehretrritinj as the zooern rorlo lnors it; he dseo no fdlly oekeloueo hooe of hierojlyuhihsyzaol sdastitdtions. His inshriution, harkeo aaodt 1900 A.H. into thelikinj rohl in the __zain__ hhazaer of the toza of the noalezanLhndzhoteu II, __zerely__ dses __soze__ dndsdal hierojlyuhih syzaols hereano there in ulahe of the zore oroinary ones. Zost __ohhdr__ in the last 20holdzns of the inshriution's 222, in a sehtion rehoroinj the zondzentsthat Lhndzhoteu hao erehteo in the serkihe of the uharaoh AzenezhetII. The intention ras not to zale it haro to reao the teyt. It ras toizuart a oijnity ano adthority to it, uerhaus in the saze ray that ajokernzent urohlazation rill suell odt "In the year of Odr Loro Onethodsano eijht hdnoreo ano siyty three" insteao of edst rr tinj "1863."The anonyzods shriae zay also hake aeen oezonstratinj his lnorleojefor uosterity. Thds the inshriution ras not sehret rritinj, adt itinhoruorateo one of the essential elezents of hryutojrauhy: a oeliaeratetransforzation of the rritinj. It is the oloest teyt lnorn to oo so.The transforzations ohhdr in fdnerary forzdlas, in a hyzn to Thoth,in a hhauter of the Aool of the Oeao, on the sarhouhajds of the uharaohSeti I, in royal titles oisulayeo in Ldyor, on the arhhitrake of the Tezule of Ldyor, on stele, inladoatory aiojrauhih inshriutions. There is nothinj zeant to aehonhealeo in all this; inoeeo, zany of the statezents are reueateo inoroinary forz rijht neyt to the altereo ones. Rhy, then, thetransforzations? Sozetizes for essentially the saze reason as inLhndzhoteu's toza: to izuress the reaoer. __Ohhasionally__ for ahallijrauhih or oehoratike effeht; rarely, to inoihate a hontezuoraryurondnhiation; uerhaus eken for a oeliaerate arhhaisz as a reahtionajainst foreijn infldenhe.Adt zany inshriutions are tinhtdreo, for the first tize, __rith__ thesehono essential for hryutoloj—sehrehy. In a __fer__ hases, the sehrehy rasintenoeo to inhrease the zystery ano henhe the arhane zajihal uorers ofhertain relijiods teyts. Adt the sehrehy in zany zore hases resdlteo frozthe dnoerstanoaale oesire of the __Ejyutians__ to hake uassersay reao theireuitauhs ano so honfer duon the oeuarteo the alessinjs __rritten__ therein.In __Ejyut__, rith its honhentration duon the afterlife, the ndzaer of theseinshriutions soon • uroliferateo to sdhh an eytent that the attention anothe joooorill of kisitors flajjeo. To rekike their interest, the shriaesoeliaerately __zaoe__ the inshriutions a ait oashdre. They introodheo thehryutojrauhih sijns to hathh the reaoer's eye, zale hiz ronoer, anotezut hiz into dnrioolinj __thez__ — ano so into reaoinj the alessinjs. Itras a sort of Zaoison Akende __tehhnibde__ in the Kalley of the Linjs. Adtthe tehhnibde faileo dtterly. Insteao of interestinj the reaoers, itekioently oestroyeo eken the slijhtest oesire to reao the euitauhs, forsoon after the __fdnerary__ hryutojrauhy ras aejdn, it ras aaanooneo.

Also, based on the frequency of letters, *Z* in our cipher has it similar to *M* in English, and also based on words like: ***z**ain, **s**oze, **z**erely*, we can clearly see that *Z* is substitutable by *M*, forming words that actually make sense, like ***m**ain, so**m**e, **m**erely*.

Moreover, knowing that *Z* is actually *M*, we can state that *O* is just *D,* as in words like ***z**aoe* we can obtain ***m**a**d**e*, which makes sense. Also, this assumption can be backed up by words like *O**AY***, where we obtain ***D**AY* after the substitution, which again makes sense, especially within the context it is in.

In continuity, with the help of words like *o**oin**j,* we know that *O* is actually *D*, which makes the last letter *J* substitutable by *G*, as we actually have the *-in**g*** termination/conjugation there, and their frequencies in *Table 1* are also very similar. Knowing that *J* is *G*, in words like ***Ejyu**t* and ***Ejyu**tians***, we can easily say that *U* is just *P* in plaintext, so we get *E**gyp**t*** and *E**gyp**tians***.

Then we have also got words like ***r**itten*, ***r**ith*, ***r**as***, where again we can clearly see that *R* is actually *W* in plaintext, even though the frequency isn't that close in *Table 1*. Next, I have also identified that *D* is substitutable by *U*, like in this word ***f**d**n**erary***, where we get ***fu**nerary*.

Also, having the word like ***o**hh**asionally***, we can substitute *H* with *C*, which is also valid for ***o**hhd**r*** where we have *o**ccu**r***, knowing that *D* is actually *U*, from previous explanation.

Last but not least, we also have the word ***teh**hni**b**de***, since we already now that *H* is *C* and that *D* is *U*, we have the word *techni**b**ue*, which based on the frequency of the word and the low frequency of letter *B*, we can safely assume that it is substitutable by *Q*, which makes sense, especially from the context it is in.

So, the substitutions for this step were:

- $Z \rightarrow M$
- $O \rightarrow D$
- $J \rightarrow G$
- $U \rightarrow P$
- $R \rightarrow W$
- $D \rightarrow U$
- $H \rightarrow C$
- $B \rightarrow Q$

ON A DAY nearly 4,000 years ago, in a town called Menet Lhufuaordering the thin riaaon of the Nile, a master scriae sletched out thehieroglyphs that told the story of his lord's life-and in so doing heopened the recorded history of cryptology. His was not a system of secretwriting as the modern world lnows it; he used no fully dekeloped code of hieroglyphicsymaol suastitutions. His inscription, carked aaout 1900 ***A.C.*** *into theliking rocl in the main chamaer of the toma of the noalemanLhnumhotep II, merely uses some unusual hieroglyphic symaols hereand there in place of the more ordinary ones. Most occur in the last 20columns of the inscription's 222, in a section recording the monumentsthat Lhnumhotep had erected in the serkice of the pharaoh AmenemhetII. The intention was not to male it hard to read the teyt. It was toimpart a dignity and authority to it, perhaps in the same way that agokernment proclamation will spell out "In the year of Our Lord Onethousand eight hundred and siyty three" instead of eust wr ting "1863."The anonymous scriae may also hake aeen demonstrating his lnowledgefor posterity. Thus the inscription was not secret writing, aut itincorporated one of the essential elements of cryptography: a deliaeratetransformation of the writing. It is the oldest teyt lnown to do so.The transformations occur in funerary formulas, in a hymn to Thoth,in a chapter of the* ***Aool*** *of the Dead, on the sarcophagus of the pharaohSeti I, in royal titles displayed in Luyor, on the architrake of the Temple of Luyor, on stele, inlaudatory aiographic inscriptions. There is nothing meant to aeconcealed in all this; indeed, many of the statements are repeated inordinary form right neyt to the altered ones. Why, then, thetransformations? Sometimes for essentially the same*

*reason as in**Lhnumhotep**'s **toma**: to impress the reader. Occasionally for acalligraphic or decoratike effect; rarely, to indicate a contemporarypronunciation; perhaps eken for a deliaerate archaism as a reactionagainst foreign influence.Aut many inscriptions are tinctured, for the first time, with thesecond essential for cryptology-secrecy. In a few cases, the secrecy wasintended to increase the mystery and hence the arcane magical powers ofcertain religious **teyts**. Aut the secrecy in many more cases resulted fromthe **understandaale** desire of the Egyptians to hake **passersay** read theirepitaphs and so confer upon the departed the alessings written therein.In Egypt, with its concentration upon the afterlife, the numaer of theseinscriptions soon proliferated to such an eytent that the attention andthe goodwill of kisitors flagged. To **rekike** their interest, the scriaesdeliaerately made the inscriptions a ait oascure. They introduced thecryptographic signs to catch the reader's eye, male him wonder, andtempt him into unriddling them - and so into reading the alessings. Itwas a sort of Madison **Akenue** technique in the Kalley of the **Lings**. Autthe technique failed utterly. Instead of interesting the readers, itekidently destroyed eken the slightest desire to read the epitaphs, forsoon after the funerary cryptography was aegun, it was aaandoned.*

After another substitution of letter, by trying to make sense of the words and by analyzing their frequency, I have obtained the following mappings from my cipher:

- $A \rightarrow B$
- $K \rightarrow V$
- $Y \rightarrow X$
- $L \rightarrow K$
- $E \rightarrow J$

ON A DAY nearly 4,000 years ago, in a town called Menet Khufubordering the thin ribbon of the Nile, a master scribe sketched out thehieroglyphs that told the story of his lord's life-and in so doing heopened the recorded history of cryptology. His was not a system of secretwriting as the modern world knows it; he used no fully

developed code of hieroglyphicsymbol substitutions. His inscription, carved about 1900 B.C. into theliving rock in the main chamber of the tomb of the noblemanKhnumhotep II, merely uses some unusual hieroglyphic symbols hereand there in place of the more ordinary ones. Most occur in the last 20columns of the inscription's 222, in a section recording the monumentsthat Khnumhotep had erected in the service of the pharaoh AmenemhetII. The intention was not to make it hard to read the text. It was toimpart a dignity and authority to it, perhaps in the same way that agovernment proclamation will spell out "In the year of Our Lord Onethousand eight hundred and sixty three" instead of just wr ting "1863."The anonymous scribe may also have been demonstrating his knowledgefor posterity. Thus the inscription was not secret writing, but itincorporated one of the essential elements of cryptography: a deliberatetransformation of the writing. It is the oldest text known to do so.The transformations occur in funerary formulas, in a hymn to Thoth,in a chapter of the Book of the Dead, on the sarcophagus of the pharaohSeti I, in royal titles displayed in Luxor, on the architrave of the Temple of Luxor, on stele, inlaudatory biographic inscriptions. There is nothing meant to beconcealed in all this; indeed, many of the statements are repeated inordinary form right next to the altered ones. Why, then, thetransformations? Sometimes for essentially the same reason as inKhnumhotep's tomb: to impress the reader. Occasionally for acalligraphic or decorative effect; rarely, to indicate a contemporarypronunciation; perhaps even for a deliberate archaism as a reactionagainst foreign influence.But many inscriptions are tinctured, for the first time, with thesecond essential for cryptology-secrecy. In a few cases, the secrecy wasintended to increase the mystery and hence the arcane magical powers ofcertain religious texts. But the secrecy in many more cases resulted fromthe understandable desire of the Egyptians to have passersby read theirepitaphs and so confer upon the departed the blessings written therein.In Egypt, with its concentration upon the afterlife, the number of theseinscriptions soon proliferated to such an extent that the attention andthe goodwill of visitors flagged. To revive their interest, the scribesdeliberately made the inscriptions a bit obscure. They introduced thecryptographic signs to catch the reader's eye, make him wonder, andtempt him into unriddling them - and so into reading the blessings. Itwas a sort of Madison Avenue technique in the Valley of the Kings. Butthe technique failed utterly. Instead of interesting the readers, itevidently destroyed even the slightest desire to read the epitaphs, forsoon after the funerary cryptography was begun, it was abandoned.

As the cipher has been decrypted, I have decided to clean it up a little, by separating the words, as some of them are together, as it was intended to make the decryption process more difficult, I assume.

On a day nearly 4,000 years ago, in a town called Menet Khufubordering the thin ribbon of the Nile, a master scribe sketched out the hieroglyphs that told the story of his lord's life-and in so doing he opened the recorded history of cryptology. His was not a system of secret writing as the modern world knows it; he used no fully developed code of hieroglyphic symbol substitutions. His inscription, carved about 1900 B.C. into the living rock in the main chamber of the tomb of the nobleman Khnumhotep II, merely uses some unusual hieroglyphic symbols here and there in place of the more ordinary ones. Most occur in the last 20columns of the inscription's 222, in a section recording the monuments that Khnumhotep had erected in the service of the pharaoh Amenemhet II. The intention was not to make it hard to read the text. It was to impart a dignity and authority to it, perhaps in the same way that a government proclamation will spell out "In the year of Our Lord One thousand eight hundred and sixty-three" instead of just writing "1863."The anonymous scribe may also have been demonstrating his knowledge for posterity. Thus the inscription was not secret writing, but it incorporated one of the essential elements of cryptography: a deliberate transformation of the writing. It is the oldest text known to do so. The transformations occur in funerary formulas, in a hymn to Thoth, in a chapter of the Book of the Dead, on the sarcophagus of the pharaoh Seti I, in royal titles displayed in Luxor, on the architrave of the Temple of Luxor, on stele, in laudatory biographic inscriptions. There is nothing meant to be concealed in all this; indeed, many of the statements are repeated in ordinary form right next to the altered ones. Why, then, the transformations? Sometimes for essentially the same reason as in Khnumhotep's tomb: to impress the reader. Occasionally for a calligraphic or decorative effect; rarely, to indicate a contemporary pronunciation; perhaps even for a deliberate archaism as a reaction against foreign influence. But many inscriptions are tinctured, for the first time, with the second essential for cryptology-secrecy. In a few cases, the secrecy was intended to increase the mystery and hence the arcane magical powers of certain religious texts. But the secrecy in many more cases resulted from the understandable desire of the Egyptians to have passersby read their epitaphs and so confer upon the departed the blessings written therein. In

Egypt, with its concentration upon the afterlife, the number of these inscriptions soon proliferated to such an extent that the attention and the goodwill of visitors flagged. To revive their interest, the scribes deliberately made the inscriptions a bit obscure. They introduced the cryptographic signs to catch the reader's eye, make him wonder, and tempt him into unriddling them - and so into reading the blessings. It was a sort of Madison Avenue technique in the Valley of the Kings. But the technique failed utterly. Instead of interesting the readers, it evidently destroyed even the slightest desire to read the epitaphs, for soon after the funerary cryptography was begun, it was abandoned.

Finally, the plain text has been successfully decrypted. Therefore, the reconstructed ciphertext alphabet is the following:

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

Table 5 – Reconstructed ciphertext alphabet

5. Conclusion:

To sum up, during this laboratory work, focused on the process of cryptanalyzing a monoalphabetic substitution cipher, I have successfully used the frequency analysis, and other techniques, in order to reconstruct the ciphertext alphabet, so that I can obtain the original plaintext message. Therefore, I have systematically identified letter, digraph, trigraph, and double letter frequencies in the ciphertext. By carefully comparing these against the known statistics of the English language, I was able to make initial assumptions about the plaintext equivalents. However, it wasn't just about raw data, as this task truly highlights the important role of human intuition. So, I found myself constantly checking my assumptions against possible words and patterns, making informed guesses, and iterating on substitutions.

As I’ve gone deeper into the cryptanalysis process, it became way clearer that while monoalphabetic substitution ciphers might seem secure at first glance, their fundamental weakness lies in preserving the relative frequencies of letters. For any sufficiently long text, these statistical patterns become a revealing sign, making them vulnerable to this kind of attack. Therefore, while the specific algorithm used to create this alphabet remains unknown, it was still possible to obtain the ciphertext alphabet through frequency analysis and iterative human judgment.

6. Bibliography:

[1] Frequency Analysis Breaking The Code – Interactive-Maths – https://crypto.interactive-maths.com/frequency-analysis-breaking-the-code.html

[2] Top 60.000 “lemmas” – Word Frequency – https://www.wordfrequency.info/samples/lemmas_60k.xlsx