Ministerul Educaţiei şi Cercetării al Republicii Moldova

Universitatea Tehnică a Moldovei

Facultatea Calculatoare, Informatică şi Microelectronică

Report

Laboratory Work nr. 2

Cryptanalysis of monoalphabetic substitution

Elaborated by: Afteni Maria, FAF-213

Verified by: Cătălin Mîțu

Chișinău – 2023

Subject: Cryptanalysis of monoalphabetic substitution

Task:

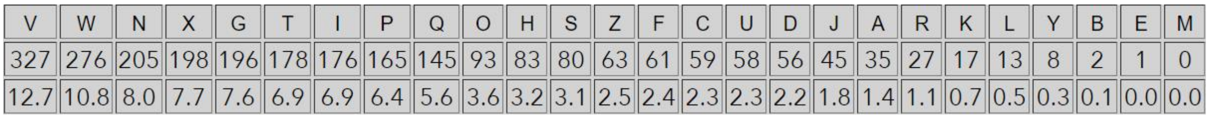
Having an encrypted message that has been intercepted which is known to have been obtained by using a monoalphabetic cypher. Apply the frequency analysis attack to find out the original message, assuming that it is a text written in English. Note that only the letters have been encrypted, the other characters remain unencrypted.

Result:

Having the cryptogram c, where

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 rixwxgj ″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.

First Step - Find the letter frequency of appearing in the cryptogram:



V	W	N	X	G	T	I	P	Q	O	H	S	Z	F	C	U	D	J	A	R	K	L	Y	B	E	M
327	276	205	198	196	178	176	165	145	93	83	80	63	61	59	58	56	45	35	27	17	13	8	2	1	0
12.7	10.8	8.0	7.7	7.6	6.9	6.9	6.4	5.6	3.6	3.2	3.1	2.5	2.4	2.3	2.3	2.2	1.8	1.4	1.1	0.7	0.5	0.3	0.1	0.0	0.0

Second step - Make substitution

Seing that the most common letters are ‘v’ and ‘w’, we can substitute them with the most common English letters ‘e’ and ‘t’. After the operation we obtain:

‶NG T OTF GeTISF 4,000 FeTIP TJN, XG T tNRG HTSSeO ZeGet LQDCDANIOeIXGJ **tQe** tQXG IXAANG NC **tQe**
GXSe, T ZTPteI PHIXAe PLetHQeO NDt tQeQXeINJSFUQP tQTt tNSO **tQe** PtNIF NC QXP SNIO'P SXCe…″

In this lines we can observe the occurrence of the “tQe” word. The most common English word that is similar to the pattern is “the”. Therefore the ‘Q’ is probably the letter ‘h’.

We get:

‶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″

After the substitution, we can observe a lot of stand-alone ‘T’, that can be ‘a’. The next frequently occurring pattern is “thTt”. This is more probable the word “that”, so ‘T’ should be the letter ‘a’.

〝PeHIetRIXtXGJ aP the ZNOeIG RNISO LGNRP **Xt;** he DPeO GN CDSSF OeKeSNUeO HNOe NC″

‘X’ is one of the next most common letters in this text, so it could be either ‘o’ or ‘i’. In the line above we can also see the combination “Xt”. If we take the choises from above, we can see that the ‘X’ is the letter ‘i’.

〝the PtNIF NC **hiP** SNIO'P SiCe— aGO iG PN ONiGJ heNUeGeO the IeHNIOeO hiPtNIF NC HIFUtNSNJF. **hiP** RaP GNt a PFPteZ NC PeHIetRIitiGJ aP the ZNOeIG RNISO LGNRP it; he DPeO GN CDSSF OeKeSNUeO HNOe NC″

In the text we can see the crypted word “hiP”. The most common word of this type is “his”, so the ‘P’ should be ‘s’.

One of the most encountered letters in this text is ‘N’. If we analyze the most common letters in English, this may be the encrypted version of the letter ‘o’.

〝**oG** a OaF GeaISF 4,000 FeaIs aJo, **iG** a toRG HaSSeO ZeGet LhDCDAoIOeIiGJ the **thiG** IiAAoG oC the
GiSe, a ZasteI sHIiAe sLetHheO **oDt** thehieIoJSFUhs that toSO the stoIF oC his SoIO's SiCe—″

After the substitution we can try to understand some words. From the “oG” and “iG” we can conclude that ‘G’ is either ‘n’ or ‘f’. To choose from them we analyze the word “thiG”. Taking in consideration our options, this word can only be ‘thin’, therefore ‘G’ is ‘n’.

We can also observe the frequent “oDt” pattern. This can be the word “out”, so ‘D’ is a substitution for ‘u’.

〝on a OaF neaISF 4,000 FeaIs **aJo,** in a toRn HaSSeO Zenet LhuCuAoIOeIinJ the thin IiAAon **oC** the niSe, a ZasteI sHIiAe sLetHheO out thehieIoJSFUhs that toSO the stoIF **oC** his SoIO's SiCe″

Here we can see the pattern “aJo” that can only be the word “ago”. It means that ‘J’ is actually ‘g’.

We can encounter and the “oC” combination. From this we can understand that ‘C’ is either ‘n’ or ‘f’. We already found the substitution for ‘n’, so ‘C’ must be ‘f’.

“toRn” is another word that we can understand. It could be either “torn” or “town”. The ‘R’ symbol is not as frequent in this text as the letter ‘r’ in English, so it should be a substitution for the ‘w’ letter, that has a lower frequency.

〝 **Lnows** it; he **useO** no fuSSF OeKeSoUeO HoOe of″

From this phrase we can understand, that ‘L’ is probably a substitution for ‘k’ and ‘O’ can be changed with ‘d’. This is because “knows” and “used” are the only suitable words for the “Lnows” and “useO” patterns.

‶ onethousand eight **hundIed** and siYtF **thIee"** instead of Eust **wIiting** "1863:″

After the substitution we observe that ‘I’ can be changed with ‘r’, because they’re frequency match and it makes sense in the context.

‶Zodern **worSd** knows it″

‶"in the Fear of our **Sord** onethousand eight hundred and **siYtF** three" instead of **Eust** writing "1863″

The phrases above show us some words that contain the ‘S’ character. If we look at it’s frequency, it could be either ‘l’ or ‘c’. It’s not ‘c’, because after substitution the results aren’t valid words, therefor it is ‘l’. If it’s ‘l’, from “worSd” and “Sord” we get “world” and “lord”, that are valid words.

‶onethousand eight hundred and siYtF three″ should be the spelling of the “1863”, so ‘Y’ and ‘F’ should be substituted by ‘x’ and ‘y’.

The ‘E’ can be substituted by ‘j’, as only valid word that it can be is “just”.

‶**systeZ** of seHretwriting as the **Zodern** world knows it″

The words “systeZ” and “Zodern” help us understand that ‘Z’ should be transformed in ‘m’.

‶the thin **riAAon** of the nile″

‶hieroglyUhiHsymAol **suAstitutions**″

From the context in the phrasses above, we can see that ‘A’ is a substitution for the letter ‘b’.

‶in a **seHtion reHording** the monumentsthat″demonsta

‶most **oHHur** in the last″

Observing the words “seHtion”, “reHording” and “oHHur” that appear in the text, we can conclude that ‘H’ is actually ‘c’. This is demonstrated by they’re frequency in the text and in English.

‶had erected in the **serKice** of the **Uharaoh** amenemhetii. the intention was not to make it hard to read the text. it was to**imUart** a dignity and authority to it, **UerhaUs** in the same way″

The words “Uharaoh” and “UerhUs” are most probable “pharaoh” and “perhaps”, so ‘U’ should be substituted by ‘p’.

‶ may also **haKe** been demonstrating his knowledgefor posterity.″

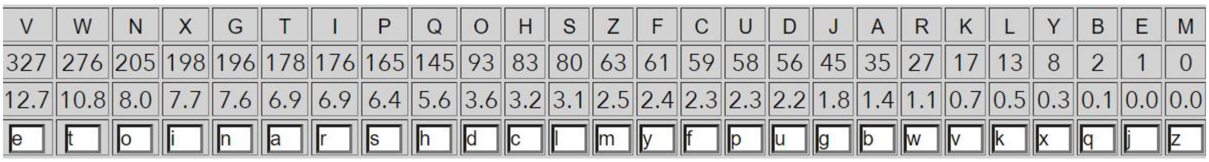
‶to **haKe** passersby read theirepitaphs and so confer″

We can see another pattern now, that is “haKe”. If we compare it with the remaining letters to decypher, it should be the letter ‘v’, so the word is “have”.

“madison avenue techniBue in the valley of the kings. butthe techniBue failed utterly.”

The word “techniBue” should be “technique”, so ‘B’ can be transformed in ‘q’.

The last letter that has no appearances in this text is ‘M’ and it should be ‘Z’, which matches with it’s frequency in English.



V	W	N	X	G	T	I	P	Q	O	H	S	Z	F	C	U	D	J	A	R	K	L	Y	B	E	M
327	276	205	198	196	178	176	165	145	93	83	80	63	61	59	58	56	45	35	27	17	13	8	2	1	0
12.7	10.8	8.0	7.7	7.6	6.9	6.9	6.4	5.6	3.6	3.2	3.1	2.5	2.4	2.3	2.3	2.2	1.8	1.4	1.1	0.7	0.5	0.3	0.1	0.0	0.0
e	t	o	i	n	a	r	s	h	d	c	l	m	y	f	p	u	g	b	w	v	k	x	q	j	z

Now that we find all the substitutions, we obtain the text:

“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 writing "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 from the 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."

Conclusion:

In conclusion, the laboratory work on the cryptanalysis of monoalphabetic substitution has provided me with valuable insights into the vulnerabilities of this classic encryption technique. Through a systematic approach involving frequency analysis, I was able to decipher the encoded messages and unveil the hidden information.

Throughout the course of this experiment, I learned that monoalphabetic substitution, despite its simplicity, can be easily cracked if not properly implemented. By analyzing the frequency distribution of letters in the ciphertext and comparing it to the known frequencies of letters in the English language, I was able to deduce the substitution key and decrypt the message successfully.

This laboratory work reinforced the importance of using strong encryption methods in real-world applications to protect sensitive information.