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

THE SUBJECT

LABORATORY WORK #2

Cryptanalysis of Monoalphabetic Ciphers

Author:

Andrei Chicu

std. gr. faf-233

Verified:

Zaica, asist. univ.

Department of SEA, FCIM UTM

Chișinău, 2025

1 Introduction

github url: `https://github.com/andyp1xe1/crypto_labs`

1.1 Objective

To analyze and decrypt a monoalphabetic substitution cipher using frequency analysis techniques, demonstrating the practical application of statistical cryptanalysis methods.

1.2 Tasks

An encrypted message was intercepted that is known to have been obtained using a monoalphabetic cipher. Apply the frequency analysis attack to find the original message, assuming that it is a text written in English. Only the letters were encrypted, while other characters remained unencrypted.

1.3 Theoretical Notes

Monoalphabetic substitution ciphers represent one of the oldest forms of cryptographic systems, where each letter of the plaintext is consistently replaced with another letter throughout the entire message. Despite their historical significance, these ciphers possess fundamental weaknesses that make them vulnerable to systematic cryptanalytic attacks.

The primary weakness lies in the preservation of letter frequency patterns from the original language. In English, certain letters appear with characteristic frequencies - 'E' being the most common at approximately 12.7%, followed by 'T' at 9.06%, and 'A' at 8.17%. This frequency distribution remains relatively stable across different texts of sufficient length.

Frequency analysis exploits this statistical property by examining the distribution of symbols in the ciphertext and correlating them with expected English letter frequencies. The methodology involves several key principles:

1. **Statistical correlation**: Mapping the most frequent cipher symbols to the most common English letters
2. **Pattern recognition**: Identifying common digraphs (TH, HE, AN, IN, ER) and trigraphs (THE, AND, THA, ENT, ION)
3. **Contextual analysis**: Using partial decryption results to identify complete words and validate substitutions

4. **Double letter identification**: Recognizing patterns where the same letter appears consecutively (SS, EE, TT, OO, FF)

The process requires both computational analysis and human intuition, as perfect frequency matches are rare in practice. The cryptanalyst must consider context, common English words, and linguistic patterns to make educated substitution decisions. This combination of statistical analysis and linguistic knowledge makes frequency analysis a powerful tool against monoalphabetic ciphers.

2 Process

The cryptanalysis process followed the established methodology for frequency analysis attacks, combining computational analysis with pattern recognition techniques.

2.1 Encrypted Message

The intercepted message contained the following ciphertext:

Ixkviatgl Udasxhtwxng Gn. 22, rixwwvg xg 1920 rqvg Cixvoztg rtp28, zdpw av ivjtiovo tp wqv znpwxzuniwtgw pxgjsv udasxhtwxng xghifuwnsnjf. Xw wnnl wqv phxvghv xgwn t gvr rniso. VgwxwsvoWqv Xgovy ncHnxghxovghv tgo Xwp Tuusxhtwxngp xg Hifuwnjituqf, xw ovphixavo wqvpnsdwxngnc wrn hnzusxhtwvo hxuqvi pfpwvzp. Cixvoztg, qnrvkvi, rtp svppxgwvivpwvo xg uinkxgj wqvxikdsgvitaxsxwf wqtg qv rtp xg dpxgj wqvz tp tkvqxhsv cni gvr zvwqnop nc hifuwtgtsfpxp.Xg xw,Cixvoztg ovkxpvo wrn gvr wvhqgxbdvp. Ngv rtp aixssxtgw. Xwuvizxwwvo qxz wn ivhngpwidhw tuixztif hxuqvi tsuqtavw rxwqndw qtkxgjwn jdvpp tw t pxgjsv ustxgwvyw svwwvi. Adw wqv nwqvirtp uincndgo. Cni wqvcxipw wxzv xg hifuwnsnjf, Cixvoztg wivtwvo t civbdvghf oxpwixadwxng tptgvgwxwf, tp t hdikv rqnpv pvkvits unxgwp rviv htdptssf ivstwvo, gnw tp edpwt hnssvhwxng ncxgoxkxodts svwwvip wqtw qtuuvg wn pwtgo xg t hviwtxg niovicni gnghtdpts (qxpwnixhts) ivtpngp,tgo wn wqxp hdikv qv tuusxvo pwtwxpwxhtshnghvuwp. Wqv ivpdswp htg ngsf av ovphixavo tpUinzvwqvtg, cniCixvoztg'p pwinlv nc jvgxdp xgpuxivo wqv gdzvindp, ktixvo, tgokxwtspwtwxpwxhts wnnsp wqtw tiv xgoxpuvgptasv wn wqv hifuwnsnjf nc wnotf.Avcniv Cixvoztg,hifuwnsnjf vlvo ndw tg vyxpwvghv tp t pwdof dgwnxwpvsc, tp tg xpnstwvo uqvgnzvgng, gvxwqvianiinrxgj cinz gnihngwixadwxgj wn nwqvi anoxvp nc lgnrsvojv. Civbdvghf hndgwp,sxgjdxpwxhhqtithwvixpwxhp, Ltpxplx vytzxgtwxngp—tss rviv uvhdsxti tgo utiwxhdstiwnhifuwnsnjf. Xw orvsw t ivhsdpv xg wqv rniso nc phxvghv. Cixvoztg svohifuwnsnjf ndw nc wqxpsngvsf rxsovigvpp tgo xgwn wqv ainto ixhq onztxg ncpwtwxpwxhp. Qv hnggvhwvo hifuwnsnjf wnztwqvztwxhp. Wqv pvgpv ncvyutgoxgj qnixmngp zdpw qtkv ivpvzasvo wqtw cvsw af hqvzxpwprqvgCixvoixhq Rnqsvi pfgwqvpxmvo divt, ovzngpwitwxgj wqtw sxcv uinhvppvpnuvitwv dgovi rvss-lgnrg hqvzxhts strp tgo tiv wqvivcniv pdaevhw wn-

vyuvixzvgwtwxng tgo hngwins, tgo svtoxgj wnwnotf'p ktpw pwixovp xgaxnhqvzxpwif. Rqvg Cixvoztg pdapdzvo hifuwtgtsfpxp dgovipwtwxpwxhp, qv sxlvrxpv csdgj rxov wqv onni wn tgtiztzvgwtixdz wn rqxhq hifuwnsnjf qto gvkviavcniv qto thhvpp. Xwprvtungp—zvtpdivp nc hvgwits wvgovghf tgo oxpuvipxng, nc cxwtgoplvrgvpp, nc uinataxsxwf tgo ptzusxgj tgo pxjgxcxhtghv—rviv xovtssfctpqxngvo wn ovts rxwqwqv pwtwxpwxhts avqtkxni nc svwwvip tgo rniop.Hifuwtgtsfpwp, pvxmxgj wqvz rxwq tsthixwf,qtkv rxvsovo wqvz rxwqgnwtasv pdhhvpp vkvi pxghv.Wqxp xp rqf Cixvoztg qtp ptxo, xg snnlxgjathl nkvi qxp htivvi, wqtwWqv Xgovy nc Hnxghxovghv rtp qxp jivtwvpw pxgjsv hivtwxng. Xw tsngvrndsoqtkv rng qxz qxp ivudwtwxng. Adw xg cthw xw rtp ngsf wqv avjxggxgj. Qv tgo Zip. Cixvoztgbdxw Ixkviatgl gvti wqv vgo nc 1920. Wqvpxwdtwxng qto avhnzv xgwnsvitasv. Ctaftg qto sdivo qxzathl tcwvi wqvrti rxwq itxpvp tgo uinzxpvp nc tapnsdwv civvonz wn uinkv ni oxpuinkvwqvvyxpwvghv nc hxuqvip xg Pqtlvpuvtiv. Adw qv qto pbdvshqvo vkviftwwvzuw wn on pn tgo qtovzatiitppvo Cixvoztg xgwn tuutivgwsfthbdxvphvgw pxsvghv tw stgwvig-psxov svhwdivp ng wqvpdaevhw. Ng Etgdtif1, 1921, Cixvoztg avjtg t pxy-zngwq hngwithw rxwq wqv Pxjgts Hniupwnovkxpv hifuwnpfpwvzp. Rqvg xw vyuxivo, qv rtp wtlvg ng wqv hxkxs-pvikxhvutfinss nc wqv RtiOvutiwzvgw tw $4,500 t fvti.Ngv nc qxp cxipw tppxjgzvgwp rtp wn wvthq t hndipv xg zxsxwtifhnovptgo hxuqvip tw wqv Pxjgts Phqnns, wqvg tw Htzu Tscivo Ktxs, Gvr Evipvf.Cni wqxp qv rinwvt wvywannl wqtw, cni wqv cxipw wxzv, xzunpvo niovi dungwqv hqtnp nc hxuqvi pfpwvzp tgo wqvxiwvizxgnsnjf. Wqvpv qto puindwvoxg t avrxsovixgj ktixvwf, tgo rixwvip wivtwvo vthq tp xgoxkxodtstgopuvhxts htpvp. Cixvoztg pniwvo wqvz ndw ng wqv atpxp nc pwidhwdivxgpwvto nc tpuvhw, tgopn snjxhts tgo dpvcds rtp wqxp hstppxcxhtwxng wqtw xwqtp avhnzv pwtgotio. Qv znovsvo qxpgnzvghstwdiv ng qxp htwvjnixvp, pnwqtw wqv gtzvp qv zxgwvo qtkv wqv jivtw zvixw nc ztlxgj wqvivstwxngpavwrvvg wqv ktixndp jvgvit nc hxuqvip vkxovgw ng pxjqw. Tg vytzusv xpwqvhnzusvzvgwtif utxi "zngn-tsuqtavw" tgo "unsftsuqtavw"; wqv Civghqrviv pwxss htssxgjunsftsuqtavwxh pfpwvzp af wqv tsznpw nacdphtwnif"ondasv pdapwxwdwxng," rqxhq wvssptapnsdwvsf gnwqxgj tw tss tandw wqvpfpwvz. Cixvoztg'p znpw xzuniwtgw hnxgtjv rtp wqvrnio"hifuwtgtsfpxp," rqxhq qv ovkxpvo xg 1920 wn hsvti du t hqingxh pndihv nchngcdpxng xghifuwnsnjf—wqv tzaxjdxwf nc wqv kvia "ovhxuqvi," wqvg dpvown zvtg anwq tdwqnixmvo tgodgtdwqnixmvo ivodhwxngp nc t hifuwnjitz wn ustxgwvyw.Qv wxwsvo qxp annl Vsvzvgwp ncHifuwtgtsfpxp, tgo wqv wviz qtp pnuinpuvivo wqtw wnotf xw hxihdstwvp xg jvgvits hngkviptwxngtgo uixgw.

2.2 Initial Analysis

The encrypted message contained 3,694 letters total. Using a Go program, I calculated the frequency distribution of each letter in the ciphertext and compared it with standard English letter frequencies.

Rank	Cipher Letter	Count	Message %	English Letter	Expected %	Difference
1	V	434	11.75	E	12.70	-0.95
2	W	356	9.64	T	9.06	+0.58
3	T	305	8.26	A	8.17	+0.09
4	X	295	7.99	I	6.97	+1.02
5	P	263	7.12	S	6.33	+0.79
6	G	262	7.09	N	6.75	+0.34
7	N	257	6.96	R	5.99	+0.97
8	I	229	6.20	H	6.09	+0.11
9	Q	169	4.57	L	4.03	+0.54
10	O	153	4.14	D	4.25	-0.11
11	S	148	4.01	C	2.78	+1.23
12	H	148	4.01	U	2.76	+1.25
13	U	89	2.41	M	2.41	0.00
14	Z	88	2.38	F	2.23	+0.15
15	D	86	2.33	W	2.36	-0.03
16	C	78	2.11	Y	1.97	+0.14
17	F	75	2.03	P	1.93	+0.10
18	R	63	1.71	B	1.49	+0.22
19	A	59	1.60	V	0.98	+0.62
20	J	52	1.41	G	2.01	-0.60
21	K	37	1.00	K	0.77	+0.23
22	L	19	0.51	J	0.15	+0.36
23	Y	13	0.35	X	0.15	+0.20
24	B	6	0.16	Q	0.09	+0.07
25	E	5	0.14	Z	0.07	+0.07
26	M	5	0.14			

2.3 Pattern Analysis

The analysis revealed significant patterns that aided in the cryptanalysis:

Double Letter	Occurrences
WW	18
PP	15
SS	13
NN	8
VV	5
GG	4
UU	4
HH	3
II	2
CC	1
OO	1
QQ	1
RR	1

Digraph	Occurrences
WQ	89
QV	84
TG	68
XG	64
VI	61
VO	60
PW	59
NG	51
WN	50
TW	48

Trigraph	Occurrences
WQV	58
TGO	30
XGJ	21
XNG	19
TWX	19
IFU	18
IXV	18
HIF	17
FUW	17
XVO	17

The trigraph WQV appearing 58 times was particularly significant, as it strongly suggested the English word "THE", confirming the initial frequency-based mapping of V$\rightarrow$E, W$\rightarrow$T, and indicating Q$\rightarrow$H.

2.4 Substitution Process

Following the frequency analysis methodology, I began with the most frequent cipher letters:

1. **First substitution**: V $\rightarrow$ E (most frequent cipher letter to most frequent English letter). The high frequency of V (11.75%) closely matched English E (12.70%)
2. **Second substitution**: W $\rightarrow$ T. W's frequency (9.64%) approximated English T (9.06%)
3. **Trigraph analysis**: The pattern WQV appeared 58 times, suggesting this represents "THE". This confirmed W $\rightarrow$ T and V $\rightarrow$ E, while indicating Q $\rightarrow$ H
4. **Contextual validation**: With partial substitutions, words began forming:
    - "WQV" became "THE"
    - Common patterns like "tgo" suggested T $\rightarrow$ A, G $\rightarrow$ N, O $\rightarrow$ D
5. **Progressive refinement**: Each successful substitution revealed more patterns:
    - X frequently appeared before G, suggesting X $\rightarrow$ I (forming "ING" endings)
    - P's high frequency and positioning suggested P $\rightarrow$ S
    - N's frequency matched English N, confirming N $\rightarrow$ R

2.5 Verification and Completion

The substitution pattern that emerged was systematically verified by:

- Checking against common English words and phrases
- Validating digraph and trigraph frequencies
- Ensuring linguistic coherence in the decrypted text

The final substitution mapping revealed a text about cryptography and William Friedman's contributions to the field, confirming the accuracy of the frequency analysis approach.

3 Results

The frequency analysis successfully decrypted the monoalphabetic substitution cipher, revealing the complete substitution key and plaintext.

3.1 Substitution Key

The final mapping between plaintext and cipher alphabets:

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

3.2 Decrypted Message

The complete decrypted plaintext:

RIVERBANK PUBLICATION NO. 22, WRITTEN IN 1920 WHEN FRIEDMAN WAS28, MUST BE REGARDED AS THE MOSTIMPORTANT SINGLE PUBLICATION INCRYPTOLOGY. IT TOOK THE SCIENCE INTO A NEW WORLD. ENTITLEDTHE INDEX OFCOINCIDENCE AND ITS APPLICATIONS IN CRYPTOGRAPHY, IT DESCRIBED THESOLUTIONOF TWO COMPLICATED CIPHER SYSTEMS. FRIEDMAN, HOWEVER, WAS LESSINTERESTED IN PROVING THEIRVULNERABILITY THAN HE WAS IN USING THEM AS AVEHICLE FOR NEW METHODS OF CRYPTANALYSIS.IN IT,FRIEDMAN DEVISED TWO NEW TECHNIQUES. ONE WAS BRILLIANT. ITPERMITTED HIM TO RECONSTRUCT APRIMARY CIPHER ALPHABET WITHOUT HAVINGTO GUESS AT A SINGLE PLAINTEXT LETTER. BUT THE OTHERWAS PROFOUND. FOR THEFIRST TIME IN CRYPTOLOGY, FRIEDMAN TREATED A FREQUENCY DISTRIBUTION ASANENTITY, AS A CURVE WHOSE SEVERAL POINTS WERE CAUSALLY RELATED, NOT AS ZUSTA COLLECTION OFINDIVIDUAL LETTERS THAT HAPPEN TO STAND IN A CERTAIN ORDERFOR NONCAUSAL (HISTORICAL) REASONS,AND TO THIS CURVE HE APPLIED STATISTICALCONCEPTS. THE RESULTS CAN ONLY BE DESCRIBED ASPROMETHEAN, FORFRIEDMAN'S STROKE OF GENIUS INSPIRED THE NUMEROUS, VARIED, ANDVITALSTATISTICAL TOOLS THAT ARE INDISPENSABLE TO THE CRYPTOLOGY OF TODAY.BEFORE FRIEDMAN,CRYPTOLOGY EKED OUT AN EXISTENCE AS A STUDY UNTOITSELF, AS AN ISOLATED PHENOMENON, NEITHERBORROWING FROM NORCONTRIBUTING TO OTHER BODIES OF KNOWLEDGE. FREQUENCY COUNTS,LINGUISTICCHARACTERISTICS, KASISKI EXAMINATIONS–ALL WERE PECULIAR AND PARTICULARTOCRYPTOLOGY. IT DWELT A RECLUSE IN THE WORLD OF SCIENCE. FRIEDMAN LEDCRYPTOLOGY OUT OF THISLONELY WILDERNESS AND INTO THE BROAD RICH DOMAIN OFSTATISTICS. HE CONNECTED CRYPTOLOGY TOMATHEMATICS. THE SENSE OFEXPANDING HORIJONS MUST HAVE RESEMBLED THAT FELT BY CHEMISTSWHENFRIEDRICH WOHLER SYNTHESIJED UREA, DEMONSTRATING THAT LIFE PROCESSESOPERATE UNDER WELL-KNOWN CHEMICAL LAWS AND ARE THEREFORE SUBZECT TOEXPERIMENTATION AND CONTROL, AND LEADING TOTO-

DAY'S VAST STRIDES INBIOCHEMISTRY. WHEN FRIEDMAN SUBSUMED CRYPTANALYSIS UNDERSTATISTICS, HE LIKEWISE FLUNG WIDE THE DOOR TO ANARMAMENTARIUM TO WHICH CRYPTOLOGY HAD NEVERBEFORE HAD ACCESS. ITSWEAPONS–MEASURES OF CENTRAL TENDENCY AND DISPERSION, OF FITANDSKEWNESS, OF PROBABILITY AND SAMPLING AND SIGNIFICANCE–WERE IDEALLYFASHIONED TO DEAL WITHTHE STATISTICAL BEHAVIOR OF LETTERS AND WORDS.CRYPTANALYSTS, SEIJING THEM WITH ALACRITY,HAV WIELDED THEM WITHNOTABLE SUCCESS EVER SINCE.THIS IS WHY FRIEDMAN HAS SAID, IN LOOKINGBACK OVER HIS CAREER, THATTHE INDEX OF COINCIDENCE WAS HIS GREATEST SINGLE CREATION. IT ALONEWOULDHAVE WON HIM HIS REPUTATION. BUT IN FACT IT WAS ONLY THE BEGINNING. HE AND MRS. FRIEDMANQUIT RIVERBANK NEAR THE END OF 1920. THESITUATION HAD BECOME INTOLERABLE. FABYAN HAD LURED HIMBACK AFTER THEWAR WITH RAISES AND PROMISES OF ABSOLUTE FREEDOM TO PROVE OR DISPROVETHEEXISTENCE OF CIPHERS IN SHAKESPEARE. BUT HE HAD SQUELCHED EVERYATTEMPT TO DO SO AND HADEMBARRASSED FRIEDMAN INTO APPARENTLYACQUIESCENT SILENCE AT LANTERN-SLIDE LECTURES ON THESUBZECT. ON ZANUARY1, 1921, FRIEDMAN BEGAN A SIX-MONTH CONTRACT WITH THE SIGNAL CORPSTODEVISE CRYPTOSYSTEMS. WHEN IT EXPIRED, HE WAS TAKEN ON THE CIVIL-SERVICEPAYROLL OF THE WARDEPARTMENT AT $4,500 A YEAR.ONE OF HIS FIRST ASSIGNMENTS WAS TO TEACH A COURSE IN MILITARYCODESAND CIPHERS AT THE SIGNAL SCHOOL, THEN AT CAMP ALFRED VAIL, NEW ZERSEY.FOR THIS HE WROTEA TEXTBOOK THAT, FOR THE FIRST TIME, IMPOSED ORDER UPONTHE CHAOS OF CIPHER SYSTEMS AND THEIRTERMINOLOGY. THESE HAD SPROUTEDIN A BEWILDERING VARIETY, AND WRITERS TREATED EACH AS INDIVIDUALANDSPECIAL CASES. FRIEDMAN SORTED THEM OUT ON THE BASIS OF STRUCTUREINSTEAD OF ASPECT, ANDSO LOGICAL AND USEFUL WAS THIS CLASSIFICATION THAT ITHAS BECOME STANDARD. HE MODELED HISNOMENCLATURE ON HIS CATEGORIES, SOTHAT THE NAMES HE MINTED HAVE THE GREAT MERIT OF MAKING THERELATIONSBETWEEN THE VARIOUS GENERA OF CIPHERS EVIDENT ON SIGHT. AN EXAMPLE ISTHECOMPLEMENTARY PAIR "MONO-ALPHABET" AND "POLYALPHABET"; THE FRENCHWERE STILL CALLINGPOLYALPHABETIC SYSTEMS BY THE ALMOST OBFUSCATORY"DOUBLE SUBSTITUTION," WHICH TELLSABSOLUTELY NOTHING AT ALL ABOUT THESYSTEM. FRIEDMAN'S MOST IMPORTANT COINAGE WAS THEWORD"CRYPTANALYSIS," WHICH HE DEVISED IN 1920 TO CLEAR UP A CHRONIC SOURCE OFCONFUSION INCRYPTOLOGY–THE AMBIGUITY OF THE VERB "DECIPHER," THEN USEDTO MEAN BOTH AUTHORIJED ANDUNAUTHORIJED REDUCTIONS OF A CRYPTOGRAM TO PLAINTEXT.HE TITLED HIS

BOOK ELEMENTS OFCRYPTANALYSIS, AND THE TERM HAS SOPROSPERED THAT TODAY IT CIRCULATES IN GENERAL CONVERSATIONAND PRINT.

4 Conclusions

This laboratory work successfully demonstrated the practical application of frequency analysis against monoalphabetic substitution ciphers, revealing both the power and limitations of statistical cryptanalysis.

The attack proved successful against the given message. The correlation between ciphertext letter frequencies and expected English frequencies provided a good start in the initial substitutions, followed by pattern analysis of digraphs and trigraphs. The 3,694-letter ciphertext provided enough statistical data for a meaningful frequency and pattern analysis. shorter texts would likely have shown greater variance from expected frequencies.

The exercise also revealed some limitations of frequency analysis:

- Not all cipher letters matched their expected plaintext frequencies perfectly (V, W, T to E, T, A for example) requiring pattern & contextual analysis as assistance.
- Computational frequency analysis alone proved insufficient without human interpretation.
- The effectiveness of frequency analysis depends heavily on the text content and length. Specialized vocabulary or deliberately skewed letter distributions could significantly complicate the attack.