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-212 Cristian Brinza
Verified:
1sing1.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 = IXKVIATGL UDASXHTWXNG GN. 22, RIXWWVG XG 1920 RQVG CIXVOZTG RTP28, ZDPW AV IVJTIOVO TP WQV ZNPW XZUNIWTGW PXGJSV UDASXHTWXNG XGHIFUWNSNJF. XW WNNL WQV PHXVGHV XGWN T GVR RNISO. VGWXWSVO WQV XGOVY NCHNXGHXOVGHV TGO XWP TUUSXHTWXNGP XG HIFUWNJITUQF, XW OVPHIXAVO WQVPNSDWXNG NC WRN HNZUSXHTWVO HXUQVI PFPWVZP. CIXVOZTG, QNRVKVI, RTP SVPPXGWVIVPWVO XG UINKXGJ WQVXI KDSGVITAXSXWF 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 T UIXZTIF HXUQVI TSUQTAVW RXWQNDW QTKXGJWN JDVPP TW T PXGJSV USTXGWVYW SVWWVI. ADW WQV NWQVI RTP UINCNDGO. CNI WQVCXIPW WXZV XG HIFUWNSNJF, CIXVOZTG WIVTWVO T CIVBDVGHF OXPWIXADWXNG TP TGVGWXWF, TP T HDIKV RQNPV PVKVITS UNXGWP RVIV HTDPTSSF IVSTWVO, GNW TP EDPWT HNSSVHWXNG NC XGOXKXODTS 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 TP UINZVWQVTG, CNICIXVOZTG'P PWINLV NC JVGXDP XGPUXIVO WQV GDZVINDP, KTIXVO, TGO KXWTSPWTWXPWXHTS WNNSP WQTW TIV XGOXPUVGPTASV WN WQV HIFUWNSNJF NC WNOTF.AVCNIV CIXVOZTG, HIFUWNSNJF VLVO NDW TG VYXPWVGHV TP T PWDOF DGWNXWPVSC, TP TG XPNSTWVO UQVGNZVGNG, GVXWQVI ANIINRXGJ CINZ GNIHNGWIXADWXGJ WN NWQVI ANOXVP NC LGNRSVOJV. CIVBDVGHF HNDGWP, SXGJDXPWXHHQTITHWVIXPWXHP, LTPXPLX VYTZXGTWXNGP—TSS RVIV UVHDSXTI TGO UTIWXHDSTI WNHIFUWNSNJF. XW ORVSW T IVHSDPV XG WQV RNISO NC PHXVGHV. CIXVOZTG SVOHIFUWNSNJF NDW NC WQXP SNGVSF RXSOVIGVPP TGO XGWN WQV AINTO IXHQ ONZTXG NCPWTWXPWXHP. QV HNGGVHWVO HIFUWNSNJF WN ZTWQVZTWXHP. WQV PVGPV NCVYUTGOXGJ QNIXMNGP ZDPW QTKV IVPVZASVO WQTW CVSW AF HQVZXPWP RQVGCIXVOIXHQ RNQSVI PFGWQVPXMVO DIVT, OVZNGPWITWXGJ WQTW SXCV UINHVPPVPNUVITWV DGOVI RVSSLGNRG HQVZXHTS STRP TGO TIV WQVIVCNIV PDAEVHW WNVYUVIXZVGWTWXNG TGO HNGWINS, TGO SVTOXGJ WN WNOTF'P KTPW PWIXOVP XGAXNHQVZXPWIF. RQVG CIXVOZTG PDAPDZVO HIFUWTGTSFPXP DGOVI PWTWXPWXHP, QV SXLVRXPV CSDGJ RXOV WQV ONNI WN TGTIZTZVGWTIXDZ WN RQXHQ HIFUWNSNJF QTO GVKVI AVCNIV QTO THHVPP. XWPRVTUNGP—ZVTPDIVP NC HVGWITS WVGOVGHF TGO OXPUVIPXNG, NC CXW TGOPLVRGVPP, NC UINATAXSXWF TGO PTZUSXGJ TGO PXJGXCXHTGHV—RVIV XOVTSSFCTPQXNGVO WN OVTS RXWQ WQV 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 SNNLXGJ ATHL NKVI QXP HTIVVI, WQTWWQV XGOVY NC HNXGHXOVGHV RTP QXP JIVTWVPW PXGJSV HIVTWXNG. XW TSNGV RNDSOQTKV RNG QXZ QXP IVUDWTWXNG. ADW XG CTHW XW RTP NGSF WQV AVJXGGXGJ. QV TGO ZIP. CIXVOZTG BDXW IXKVIATGL GVTI WQV VGO NC 1920. WQVPXWDTWXNG QTO AVHNZV XGWNSVITASV. CTAFTG QTO SDIVO QXZ ATHL TCWVI WQVRTI RXWQ ITXPVP TGO UINZXPVP NC TAPNSDWV CIVVONZ WN UINKV NI OXPUINKVWQV VYXPWVGHV NC HXUQVIP XG PQTLVPUVTIV. ADW QV QTO PBDVSHQVO VKVIFTWWVZUW WN ON PN TGO QTO VZATIITPPVO CIXVOZTG XGWN TUUTIVGWSFTHBDXVPHVGW PXSVGHV TW STGWVIG-PSXOV SVHWDIVP NG WQV PDAEVHW. NG ETGDTIF1, 1921, CIXVOZTG AVJTG T PXY-ZNGWQ HNGWITHW RXWQ WQV PXJGTS HNIUP WNOVKXPV HIFUWNPFPWVZP. RQVG XW VYUXIVO, QV RTP WTLVG NG WQV HXKXS-PVIKXHVUTFINSS NC WQV RTI OVUTIWZVGW TW $4,500 T FVTI.NGV NC QXP CXIPW TPPXJGZVGWP RTP WN WVTHQ T HNDIPV XG ZXSXWTIF HNOVPTGO HXUQVIP TW WQV PXJGTS PHQNNS, WQVG TW HTZU TSCIVO KTXS, GVR EVIPVF.CNI WQXP QV RINWV T WVYWANNL WQTW, CNI WQV CXIPW WXZV, XZUNPVO NIOVI DUNGWQV HQTNP NC HXUQVI PFPWVZP TGO WQVXI WVIZXGNSNJF. WQVPV QTO PUINDWVOXG T AVRXSOVIXGJ KTIXVWF, TGO RIXWVIP WIVTWVO VTHQ TP XGOXKXODTS TGOPUVHXTS HTPVP. CIXVOZTG PNIWVO WQVZ NDW NG WQV ATPXP NC PWIDHWDIVXGPWVTO NC TPUVHW, TGO PN SNJXHTS TGO DPVCDS RTP WQXP HSTPPXCXHTWXNG WQTW XWQTP AVHNZV PWTGOTIO. QV ZNOVSVO

QXP GNZVGHSTWDIV NG QXP HTWVJNIXVP, PNWQTW WQV GTZVP QV ZXGWVO QTKV WQV JIVTW ZVIXW NC ZTLXGJ WQV IVSTWXNGPAVWRVVG WQV KTIXNDP JVGVIT NC HXUQVIP VKXOVGW NG PXJQW. TG VYTZUSV XP WQVHNZUSVZVGWTIF UTXI "ZNGN-TSUQTAVW" TGO "UNSFTSUQTAVW"; WQV CIVGHQRVIV PWXSS HTSSXGJUNSFTSUQTAVWXH PFPWVZP AF WQV TSZNPW NACDPHTWNIF"ONDASV PDAPWXWDWXNG," RQXHQ WVSSP TAPNSDWVSF GNWQXGJ TW TSS TANDW WQVPFPWVZ. CIXVOZTG'P ZNPW XZUNIWTGW HNXGTJV RTP WQV RNIO"HIFUWTGTSFPXP," RQXHQ QV OVKXPVO XG 1920 WN HSVTI DU T HQINGXH PNDIHV NCHNGCDPXNG XG HIFUWNSNJF—WQV TZAXJDXWF NC WQV KVIA "OVHXUQVI," WQVG DPVOWN ZVTG ANWQ TDWQNIXMVO TGO DGTDWQNIXMVO IVODHWXNGP NC T HIFUWNJITZ WN USTXGWVYW.QV WXWSVO QXP ANNL VSVZVGWP NC HIFUWTGTSFPXP, TGO WQV WVIZ QTP PNUINPUVIVO WQTW WNOTF XW HXIHDSTWVP XG JVGVITS HNGKVIPTWXNG TGO UIXGW.

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

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

The frequencies of the intercept are:

V	W	T	X	P	G	N	I	Q	O	H	S	U	Z	D	C	F	R	A	J	K	L	Y	B	E	M
434	356	305	295	263	262	257	229	169	153	148	148	89	88	86	78	75	63	59	52	37	19	13	6	5	5
11.7	9.6	8.3	8.0	7.1	7.1	7.0	6.2	4.6	4.1	4.0	4.0	2.4	2.4	2.3	2.1	2.0	1.7	1.6	1.4	1.0	0.5	0.4	0.2	0.1	0.1
e	t	a	i	s	n	o	r	h	d	c	l	p	m	u	f	y	w	b	g	v	k	x	q	j	z

Figure 1 – Table of frequency

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

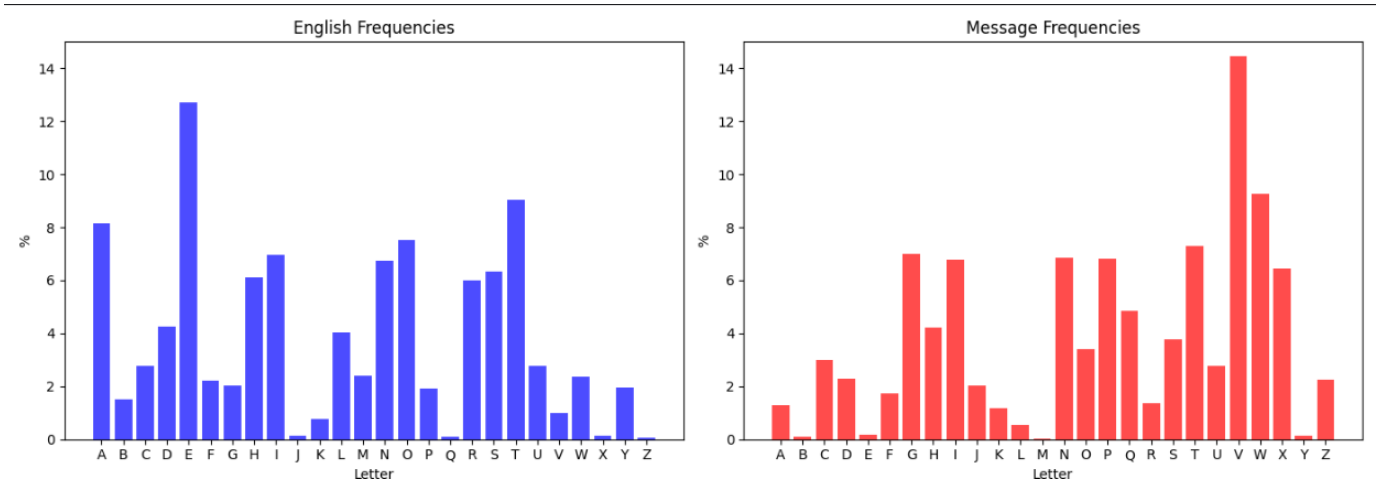

Figure 2 - Graphs of frequency

The first step is to find the frequencies of all letters that appear in the cryptogram, as shown in Table.

Above, 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).

Result

Riverbank Publication No. 22, written in 1920 when Friedman was28, must be regarded as the most important single publication incryptology. It took the science into a new world. Entitled The Index ofCoincidence and Its Applications in Cryptography, it described thesolution of two complicated cipher systems. Friedman, however, was lessinterested in proving their vulnerability 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 a primary cipher alphabet without havingto guess at a single plaintext letter. But the other was profound. For thefirst time in cryptology, Friedman treated a frequency distribution as anentity, as a curve whose several points were causally related, not as justa collection of individual 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 as Promethean, forFriedman's stroke of genius inspired the numerous, varied, and vitalstatistical tools that are indispensable to the cryptology of today.Before Friedman, cryptology eked out an existence as a study untoitself, as an isolated phenomenon, neither borrowing from norcontributing to other bodies of knowledge. Frequency counts, linguisticcharacteristics, Kasiski examinations—all were peculiar and particular tocryptology. It dwelt a recluse in the world of science. Friedman ledcryptology out of this lonely wilderness and into the broad rich domain ofstatistics. He connected cryptology to mathematics. The sense ofexpanding horizons must have resembled that felt by chemists whenFriedrich Wohler synthesized urea, demonstrating that life processesoperate under wellknown chemical laws and are therefore subject toexperimentation and control, and leading to today's vast strides inbiochemistry. When Friedman subsumed cryptanalysis under statistics, he likewise flung wide the door to anarmamentarium to which cryptology had never before had access. Itsweapons—measures of central tendency and dispersion, of fit andskewness, of probability and sampling and significance—were ideallyfashioned to deal with the statistical behavior of letters and words.Cryptanalysts, seizing them with alacrity, have wielded them withnotable success ever since.This is why Friedman has said, in looking back over his career, thatThe Index of Coincidence was his greatest single creation. It alone wouldhave won him his reputation. But in fact it was only the beginning. He and Mrs. Friedman quit Riverbank near the end of 1920. Thesituation had become intolerable. Fabyan had lured him back after thewar with raises and promises of absolute freedom to prove or disprovethe existence of ciphers in Shakespeare. But he had squelched everyattempt to do so and had embarrassed Friedman

into apparentlyacquiescent silence at lantern-slide lectures on the subject. On January1, 1921, Friedman began a six-month contract with the Signal Corps todevise cryptosystems. When it expired, he was taken on the civil-servicepayroll of the War Department at $4,500 a year.One of his first assignments was to teach a course in military codesand ciphers at the Signal School, then at Camp Alfred Vail, New Jersey.For this he wrote a textbook that, for the first time, imposed order uponthe chaos of cipher systems and their terminology. These had sproutedin a bewildering variety, and writers treated each as individual andspecial cases. Friedman sorted them out on the basis of structureinstead of aspect, and so logical and useful was this classification that ithas become standard. He modeled his nomenclature on his categories, sothat the names he minted have the great merit of making the relationsbetween the various genera of ciphers evident on sight. An example is thecomplementary pair "mono-alphabet" and "polyalphabet"; the Frenchwere still callingpolyalphabetic systems by the almost obfuscatory"double substitution," which tells absolutely nothing at all about thesystem. Friedman's most important coinage was the word"cryptanalysis," which he devised in 1920 to clear up a chronic source ofconfusion in cryptology—the ambiguity of the verb "decipher," then usedto mean both authorized and unauthorized reductions of a cryptogram to plaintext.He titled his book Elements of Cryptanalysis, and the term has soprospered that today it circulates in general conversation and print.

Conclusion

The inherent vulnerability of monoalphabetic ciphers lies in their predisposition to frequency analysis. Given the idiosyncratic letter distribution inherent to distinct languages — notably, the prevalence of letters such as 'e' and 't' in the English lexicon — a comprehensive examination of a substantive portion of encrypted text may elucidate patterns congruent with the established letter frequencies of the underlying language. Such discernible patterns afford cryptanalysts the opportunity to postulate with a degree of certitude the potential substitutions, thereby facilitating the decryption process.

While monoalphabetic ciphers once held a reputation for being robust, the emergence of frequency analysis techniques has undermined their efficacy, particularly when subjected to extensive encrypted passages. In contemporary contexts, these ciphers are predominantly relegated to pedagogical or enigmatic roles, rather than serving as formidable cryptographic instruments.

As the realm of cryptography has undergone significant advancements, the methodologies to safeguard communications have concomitantly evolved. Present-day cryptographic paradigms are characterized by their intricate designs and heightened resilience against a myriad of potential breaches.

Nevertheless, a profound comprehension of the virtues and limitations of foundational cryptographic systems, such as the monoalphabetic cipher, elucidates the trajectory and metamorphosis of cryptographic robustness within academia.