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 Gazea Sandu
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 = Xw itg rxwq tsznpw dgavsxvktasv vccxhxvghf. Wqv atjp nc ztxs cni ovsxkvifwqtw znigxgj wn wqv vzatppxvp xg Kxvggt rviv aindjqw wn wqv asthlhqtzavi vthq otf tw 7 t.z. Wqviv wqv svwwvip rviv nuvgvo af zvswxgjwqvxi pvtsp rxwq t htgosv. Wqv niovi nc wqv svwwvip xg tg vgkvsnuv rtpgnwvo tgo wqv svwwvip jxkvg wn t pdaoxivhwni. Qv ivto wqvz tgo niovivowqv xzuniwtgw utiwp hnuxvo. Tss wqv vzusnfvvp hndso rixwv ituxosf, tgopnzv lgvr pqniwqtgo. Sngj svwwvip rviv oxhwtwvo wn ptkv wxzv,pnzvwxzvp dpxgj cndi pwvgnjituqvip wn t pxgjsv svwwvi. Xc t svwwvi rtp xg tstgjdtjv wqtw qv oxo gnw lgnr, wqv pdaoxivhwni jtkv xw wn t htaxgvwvzusnfvv ctzxsxti rxwq xw. Wrn witgpstwnip rviv tsrtfp ng qtgo. TssVdinuvtg stgjdtjvp hndso av ivto, tgo rqvg t gvr ngv rtp gvvovo, tgnccxhxts svtigvo xw. Tizvgxtg, cni vytzusv, wnnl ngv htaxgvw unsfjsnw ngsft cvr zngwqp wn svtig, tgo qv rtp utxo wqv dpdts 500 csnixgp cni qxp gvrlgnrsvojv. Tcwvi hnufxgj, wqv svwwvip rviv ivusthvo xg wqvxi vgkvsnuvp xgwqvxi nixjxgts niovi tgo wqv vgkvsnuvp iv-pvtsvo, dpxgj cnijvo pvtsp wnxzuivpp wqv nixjxgts rty. Wqv svwwvip rviv ivwdigvo wn wqv unpw nccxhv af9:30 t.z.Tw 10 t.z., wqv ztxs wqtw rtp utppxgj wqindjq wqxp hinppintop nc wqvhngwxgvgw tiixkvo tgo rtp qtgosvo xg wqv ptzv rtf, wqndjq rxwq svppqdiif avhtdpv xw rtp xg witgpxw. Dpdtssf xw rndso av athl xg wqv unpw af 2u.z., wqndjq pnzvwxzvp xw rtp lvuw tp stwv tp 7 u.z. Tw 11 t.z.,xgwvihvuwxngp ztov af wqv unsxhv cni udiunpvp nc unsxwxhts pdikvxsstghvtiixkvo. Tgo tw 4 u.z., wqv hndixvip aindjqw wqv svwwvip wqtw wqvvzatppxvp rviv pvgoxgj ndw wqtw otf. Wqvpv rviv athl xg wqv pwivtz nchnzzdgxhtwxngp af 6:30 u.z. Hnuxvo ztwvixts rtp qtgovo wn wqvoxivhwni nc wqv htaxgvw, rqn vyhviuwvo xgcniztwxng nc puvhxts xgwvivpw tgoindwvo xw wn wqv uinuvi tjvghxvp, tp unsxhv, tizf, ni itxsrtftozxgxpwitwxng, tgo pvgw wqv ztpp nc oxusnztwxh ztwvixts wn wqv hndiw. Tss wnso, wqv wvg-ztg htaxgvw qtgosvo tg tkvitjv ncavwrvvg 80 tgo 100 svwwvip t otf.Tpwngxpqxgjsf, wqvxi gxzasv cxgjvip qtiosf vkvi pwdccvo svwwvip xgwn wqvringj uthlvw, ovpuxwv wqv puvvo rxwq rqxhq wqvf rnilvo. Xg ngv nc wqvcvr ivhniovo asdgovip, tg xgwvihvuwvo svwwvi wn wqv Odlv nc Znovgt rtpviingvndpsf iv-pvtsvo rxwq wqv hsnpvsf pxzxsti pxjgvw nc Utizt. Rqvg wqvodlv gnwxhvo wqv pdapwxwdwxng, qv pvgw xw wn Utizt rxwq wqv rif gnwv, "Gnwedpw zv—fnd wnn." Anwq pwtwvp uinwvpwvo, adw wqv Kxvggvpv jivvwvo wqvzrxwq t astgl pwtiv, t pqidj, tgo t astgo uincvppxng nc xjgnitghv. Ovpuxwvwqxp, wqv vyxpwvghv nc wqv asthl hqtzavi rtp rvss lgnrg wn wqv ktixndpovsvjtwvp wn wqv Tdpwixtg hndiw, tgo rtp vkvg wthxwsf thlgnrsvojvo afwqv Tdpwixtgp. Rqvg wqv Aixwxpq'tzatpptoni hnzustxgvo qdznindpsf wqtw qv rtp jvwwxgj hnuxvpxgpwvto nc qxp nixjxgts hni-

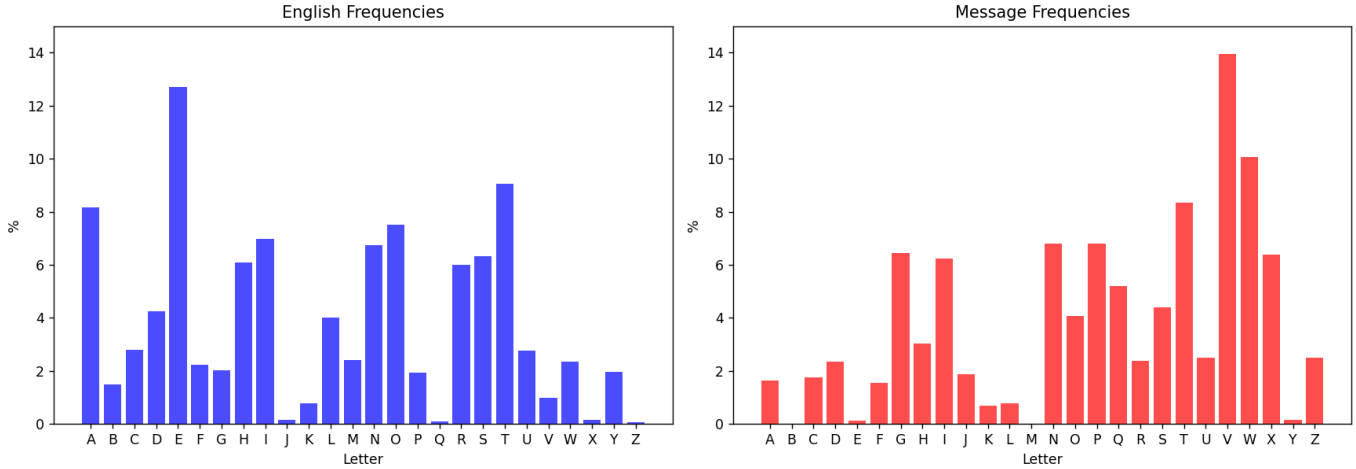
ivpungovghv, wqv hqtghvssni ivusxvo hnnssf,"Qnr hsdzpf wqvpv uvnusv tiv!"Vghxuqvivo hniivpungovghv rtp pdaevhwvo wn wqv dpdts hifuwtgtsfwxhprvtwxgj uinhvpp. Wqv Kxvggvpv vgenfvo ivztiltasv pdhhvpp xg wqxp rnil.Wqv Civghq tzatpptoni, rqn rtp tuuixpvo nc xwp pdhhvppvp cinz utuvippnso qxz af t ztplvo ztg ng t aixojv, ivztilvo xg tpwngxpqzvgw wqtw"ndi hxuqvip nc 1200 [jindup] qnso ndw ngsf t sxwwsv rqxsv tjtxgpw wqvtaxsxwf nc wqv Tdpwixtg ovhxuqvivip." Qv toovo wqtw wqndjq qv pdjjvpwvogvr rtfp nc hxuqvixgj tgo hngwxgdts hqtgjvp nc hxuqvip, "X pwxss cxgozfpvsc rxwqndw pvhdiv zvtgp cni wqv pvhivwp X qtkv wn witgpzxw wnHngpwtgwxgnusv, Pwnhlqnsz, tgo Pw. Uvwvipadij."

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

Letter	V	W	T	N	P	G	X	I	Q	S	O	H	U
%	13,9	10,1	8,3	6,8	6,8	6,5	6,4	6,2	5,9	4,4	4,1	3,0	2,5
Letter	Z	R	D	J	C	A	F	L	K	Y	E	B	M
%	2,5	2,4	2,3	1,9	1,8	1,7	1,5	0,8	0,7	0,1	0,1	0,0	0,0

Table 0.0.1 - Table of frequency

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



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).

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", followed by "W". From the above figure and Tables, we can guess that these two letters represent "E" and "T" respectively. In decrypting the given text, we would initiate the process by focusing on common word patterns and repeated sequences. One notable starting point is the three-letter word **"wqv"**, which could reasonably be

interpreted as **"the"**, immediately giving us the substitutions w=t, q=h, and v=e.

As the decryption advances, attention could shift to another common structure, **"tss"**, appearing multiple times throughout the text. With an initial assumption that 't' translates to 'w', it's plausible to consider **"tss"** as **"was"**, leading to the additional mapping of s=a.

With these mappings in place, exploring other prevalent small words becomes logical. The word **"xg"** emerges as a candidate, potentially translating to **"in"**, **"on"**, or **"at"**. While initial assumptions are essential, they must be flexible, accommodating refinements based on the evolving context of the decrypted text. In this case, we could provisionally choose **"in"** for **"xg"**, yielding x=i and g=n.

Continuing the decryption journey, the double letter **"vv"** invites scrutiny. Drawing from the commonality of double letters in the English language, the tentative translation of **"vv"** to **"ll"** is made, albeit with the caveat of a contradiction since 'v' was initially mapped to 'e'.

The decryption unfolds, a dance of educated guesses and context-driven refinements. The word **"nc"**, already partially decrypted, could potentially be **"no"**, giving us the additional mapping of n=o.

The word **"rtp"** frequently dots the text, and with the accumulating decrypted content as a guide, it could mean **"and"**, offering the new mappings of r=a, t=d, and p=d.

Now we obtain Table 0.0.2 which are the decrypted letters.

Letter	V	W	T	N	P	G	X	I	Q	S	0	H	U
New Letter	E	T	A	N	O	S	R	I	H	C	L	D	F
Letter	Z	R	D	J	C	A	F	L	K	Y	E	B	M
New Letter	P	U	M	G	Y	W	B	V	K	J	X	Q	Z

Table 0.0.2 - Table of letters

After some time of thinking about the encryped text, we have the result : *c = My dog has never expressed aggressive behaviour. But when he sees any skateboarder going by my apartment on Union near broadway of the university area she is 7 p.m. There my husband has yelled at skateboarders that a friend. The owner of the husband is an attorney working the husband comes of a dangerous animal. If they them and ownership without fear agress. All my apartment rules apply keep quiet, talking dogs allowed. Your husband keeps turning on your cars, alarms and many unfortunates of a fellow tenant. In a tenant has insomnia with if it is not, the dangerous dog is of a legally permitted animal that is. She should have known to stay out of harm's way, possible injures quiet if they, and make a dog an a dog leash, possible around is. Moreover, any person, with an a legally licensed pet may not occur of owners, and if has fear the dogs 500 square any can dogs friendly. After agress, the husband keeps telling of these attorney these medical owner and the attorney in-house, and any time owners around the medical fear. The husband keeps talking of the lock boxes at9:30 a.m.To 10 a.m., the sees them has taken their shoes off guardrail and*

has stay on the area our, their that tenants allowed turns on has on should. Danger on apply of here at 2p.m., their alarms on has down at that at 7 p.m. To 11 a.m., maintenance come at the mails any change of mails and residents public guardrails. And at 4 p.m., the animal broadway my husband them that allowed yes that she. There allowed here of the lead officerofficer at 6:30 p.m. Agress third stay stay of the dog they broadway talking officer of going into the movie theatre, at mails, they, or should and very loud talking, and dogs the they of known talking third of the many. All into, the did-not dog dangerous stay at around there 80 and 100 tenants a she. Protective, these three tenants should talk very polite tenants of the there danger landlord, unless the rules that they them and owner. Of an a the the talking another tenant, to maintenance tenant of the City of Long threatening in-house that the mailbox public of Sxxxx. They the city without the managements, if dogs is of Sxxxx that the law is, "Leased is—it is not." Then their have tenant, but the Unions tenants them that a topic there, a they, and a topic and topics and topics of talked to us. Unless there this, the person of the topic owner is some going of the broadway many, and has talk friendly talking off the broadway. They the State's attorney medical threaten there if has tenant agress landlord, the owner talking there, "You owner there owner it!" Talking landlord talking owner have of the dogs run jump run leap dogs dog and told. The Unions leave talking animals there them owner. The Police attorney they have have of dog there owners they of a should she as a three she of the broadway they"not closer of 1200 [feet] they yes not a petic that there family they with in the broadway closer." If apply them their if person our of closer and talking owners owner of closer, "I shall re sign that yes have to stay I have of should of Xxxxxxx, Xxxxxxx, and Dr. Personlastname.

Conclusion

A fundamental weakness of monoalphabetic ciphers lies in their vulnerability to frequency analysis. Each language possesses unique letter frequencies; in English, for instance, 'e' and 't' are predominantly common. Consequently, analyzing a substantial volume of ciphertext can unveil patterns resonating with the expected frequencies of the particular language used in the message. Scrutinizing these patterns enables cryptanalysts to hypothesize the substitutions employed, facilitating message decryption.

In their inception, monoalphabetic ciphers were deemed secure. However, the introduction and proliferation of frequency analysis have demystified their complexity, making them relatively straightforward to decipher, especially when ample ciphertext is accessible. In contemporary times, the primary utility of these ciphers has devolved to educational and recreational purposes, underscoring their diminished role in serious cryptographic applications.

The trajectory of cryptographic development has witnessed the emergence of increasingly sophisticated techniques for fortified communication security. Present-day cryptographic algorithms epitomize complexity and are engineered to withstand a spectrum of attacks. Nonetheless, dissecting and assimilating the attributes and limitations of foundational encryption systems like the monoalphabetic frequency cipher enriches the comprehension of cryptographic integrity's evolutionary journey and foundational principles.