Ministry of Education and Research of Republic of Moldova

Technical University of Moldova

Faculty of Computers, Informatics and Microelectronics

Software Engineering Department

Cryptography and Security

Laboratory work #2

Cryptanalysis of monoalphabetic substitution

Author:
Mihai Gurduza
std. gr. FAF-233

Verified:
M. Zaica

Chișinău 2025

Purpose

The purpose of this lab is to learn how to break simple substitution ciphers using frequency analysis. We will study how often letters appear in encrypted text and compare this with normal English letter frequencies.

Theory Background

A monoalphabetic substitution cipher is a simple type of secret code where each letter in the original message is always replaced by the same different letter. For example, every "A" might become "X" and every "B" might become "Q" throughout the entire message. The problem with this type of cipher is that it keeps the same pattern of how often letters appear as in normal text.

Frequency analysis works because some letters appear much more often than others in any language. In English, the letter "E" appears about 12.7% of the time, making it the most common letter. The letter "T" comes second at about 9.06%, then "A" at 8.17%. This pattern stays pretty much the same in most English texts that are long enough.

To break these ciphers, we follow a step-by-step process. First, we count how many times each letter appears in the encrypted message. Then we compare these numbers with what we know about normal English. The most common encrypted letter probably represents "E", the second most common probably represents "T", and so on. After making these first guesses, we look for common letter combinations like "TH", "HE", and "AN". We also look for three-letter combinations like "THE", "AND", and "THA".

We pay attention to letters that appear twice in a row, since only certain letters do this often in English (like "SS", "EE", "TT"). We also look for single letters that stand alone, since in English only "A" and "I" can be words by themselves. The process requires human thinking because computers have trouble understanding the subtle patterns of language that help us figure out the correct letters.

The Tasks

Our main job is to decode a secret message that was encrypted using a simple substitution cipher. We need to figure out what the original message said by using frequency analysis.

We will use an online tool to help us with this process: Frequency Analysis Tool

The message to be decoded from V11 is:

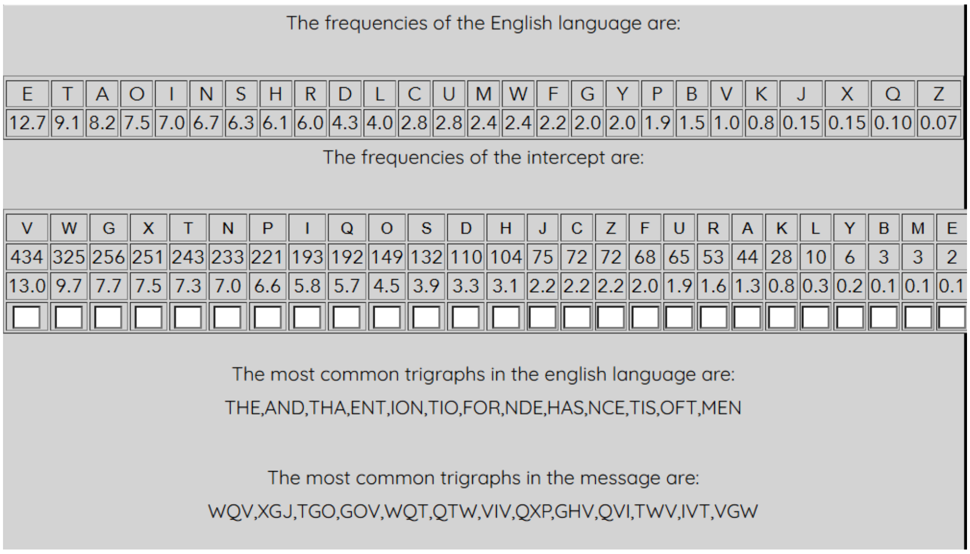
> IVTSZNGW RTP DGOVI PXVJV. WQV INFTS TIZF, DGOVI QVGIF XX NC ANDIANG,UIXGHV NC HNGOV, QTO XGKVPWVO XW TW OTRG RVOGVPOTF, TUIXS 19, 1628. ADWWQV QDJDVGNWP, XGPXOV WQV ATWWSVZVGWP NC WQV SXWWSV WNRG XG PNDWQVIGC-ITGHV, RVIV UDWWXGJ DU T PWXCC OVCVGPV. WQVF HTGGNG-TOVO HNGOV CINZ TWNRVI TGO HNGWVZUWDNDPSF IVEVHWVO QXP OVZTGOP WQTW WQVF PDIIVGOVI,PTFXGJ WQTW WQVF RNDSO OXV XGPWVTO. HNGOV AINDJQW DU CXKV AXJ HTG-GNGCINZ TSAX, T ONMVG ZXSVP TRTF, TGO NG PDGOTF ITGJVO WQVZ XG TGNZXGNDP SXGV CTHXGJ IVTSZNGW.WQTW PTZV OTF QXP PNSOXVIP HTUWDIVO TG XGQTAXWTGW NC WQV WNRG RQNRTP WIFXGJ WN HTIIF TG VGHXUQVIVO ZVPPTJV WN QD-JDVGNW CNIHVP NDWPXOV.GNGV NC HNGOV'P ZVG HNDSO DG-IXOOSV XW, ADW ODIXGJ WQV RVVL WQV UIXGHVSVTIGVO WQTW XW ZXJQW AV PNSKVO AF WQV PHXNG NC T SVTOXGJ CTZXSF NC TSAXRQN RTP LGNRG WN QTKV TG XGWVIVPW XG HXUQVIP.HNGOV PVGW QXZ WQV HIFUWNJITZ. WQV FNDGJ ZTG PNSKVO XW NG WQV PUNW.XW IVKVTSVO WQTW WQV QDJDVGNWP OVPUVITWVSF GVVOVO ZDGXWXNGP TGO WQTW, XCWQVF RVIV GNW PDUUSXVO, WQVF RNDSO QTKV WN FXVSO. WQXP RTP GVRP XGOVVO,CNI OVPUXWV WQV OVPWIDHWXNG NC T GDZAVI NC QNDPVP AF WQV HTWQNSXHATWWVIXVP, WQV WNRG RTP HNGWXGDXGJ WN IVPXPW PWNDWSF RXWQ GN PXJG NCPDIIVGOVI. HNGOV IVWDIGVO WQV HIFUWNJITZ WN WQV XGQTAXWTGWP, TGO NG-PDGOTF, TUIXS 30, 1628, WQNDJQ XWP CNIWXCXHTWXNGP RVIV PWXSS DGAIVTHQVOTGO XWP OVCVGPVP PWXSS TUUTIVGWSF TOVBDTWV CNI T SNGJ PXVJV, IVTSZNGWPDOOVGSF TGO DGVYU-VHWVOSF HTUXWDSTWVO. RXWQ WQXP OITZTWXH PDHHVP-PAVJTG WQV HTIVVI NC WQV ZTG RQN RTP WN AVHNZV CIT-GHV'P CXIPW CDSS-WXZVHIFUWNSNJXPW: WQV JIVTW TGWNXGV INPPXJGNS.RQVG RNIO NC WQV XGHXOVGW IVTHQVO HTIOXGTS IXHQVSXVD, WQV TPWDWV TGOTASV JITF VZXGVGHV NC CIT-GHV, QV TW NGHV TWWTHQVO WQXP DPVCDS WTSVGW WN-QXP PDXWV. INPPXJGNS UINKVO QXP RNIWQ TSZNPW XZZVOX-TWVSF. WQV HTWQNSXHTIZXVP DGOVI IXHQVSXVD PDIINDGOXGJ WQV HQXVC QDJDVGNW ATPWXNG NC STINHQVSSV XGWVIHVUWVO PNZV SVWWVIP XG HXUQVI, RQXHQ WQV FNDGJ HNOVAIVTLVINC

TSAX IVTO RXWQ VTPV. QV WNSO QXP VZXGVGHV WQTW WQV PWTIKXGJ HXWXMVGPRVIV VTJVISF TRTXWXGJ QVSU WQTW WQV VGJSXPQ QTO UINZXPVO WN PVGO AF PVT. RQVG WQV CSVVWTIIXKVO, WQV UIXZVO JDTIOPQXUP TGO CNIWP PN XG-WXZXOTWVO XW WQTW XW PWNNONCC WQV UNIW'P VGWIT-GHV TGO ZTOV GN PVIXNDP TWWVZUW WN CNIHV T UTPPTJV. TZNGWQ STWVI, WQV HXWF HTUXWDSTWVO XG CDSS PXJQW NC WQV VGJSXPQ KVPPVSP—TGOWQV JIVTW CIVGHQ WITOXWXNG NC VYUVIWXPV XG HIFUWNSNJF QTO AVVG CNDGOVO.INPPXJGNS KVIF BDXHLSF VPWTASXPQVO QXZPVSC XG WQV INFTS PVIKXHV. AF1630, QXP PNSDWXNGP QTO ZTOV QXZ IXHQ VGNDJQ WN ADXSO T PZTSS ADWVSVJTGW HQTWVTD TW EDKXPF, 12 ZXSVP PNDWQ NC UTIXP, STWVI PDIINDGOXGJ XWRXWQ T HQTIZXGJ XGCNIZTS JTIOVG OVPXJGVO AF SV GNWIV, WQV JTIOVGVI NCKVIPTXSSVP. QVIV SNDXP YXXX PWNUUVO WN KXPXW WQV FNDGJ HIFUW-TGTSFPW XG1634, 1635 TGO 1636 NG QXP IVWDIGP WN UTIXP CINZ CNGWTXGVASVTD.XG WQV PRTPQADHLSXGJ HNDIW NC WQTW ZNGTIHQ, TGO WQVG XG WQVIVPUSVGOVGW NGV NC SNDXP YXK, INPPXJGNS PVIKVO RXWQ TG VYWITNIOXGTIFCTHXSXWF. WQV PWINGJQNSO NC QVPOXG PDIIVGOVIVO T RVVL PNNGVI WQTG XWNWQVIRXPV RNDSO QTKV AVHTDPV QV PNSKVO TG VGHXUQVIVO USVT CNI QVSU, TGOWQVG HNZUNPVO T IVUSF XG WQV PTZV HXUQVI WVSSXGJ WQV WNRGPUVNUSV QNRCD-WXSV WQVXI QNUVP RVIV. QNR ZTGF NWQVI WNRGP QV HNZU-VSSVO WNPDIIVGOVI, QNR ZTGF OXUSNZTWXH HNDUP QV ZTOV UNPPXASV, QNR ZTGFAVWITFTSP QV DGHNKVIVO TZNGJ WQV JIVTW GNASVP XG WQNPV OTFP NC PQXCWXGJTSSVJXTGHVP, QV GVKVI OXPHDPPVO. WQXP IVWXHVGHV HTDPVO PNZV TW WQV HNDIWWN HQTIJV WQTW QV GVKVI THWDTSSF PNSKVO T PXGJSV HXUQVI, TGO WQTW WQVHTIOXGTS PUIVTO XGCST-WVO IDZNIP TANDW QXP TAXSXWXVP WN OXPHNDITJV RNDSO-AV HNGPUXITWNIP. ADW XG CTHW IXHQVSXVD RTP CIVBDVG-WSF WVSSXGJ QXPPDANIOXGTWVP PDHQ WQXGJP TP, "XW XP GVHVPPTIF WN ZTLV DPV, XG ZF NUXGXNG,NC WQV SVWWVIP NC WQV ZTG RQN QTP AVVG TIIVPWVO AF WQV HXKXS TD-WQNIXWXVP TWZVMXVIVP, WQTW XP WN PTF, QTKV WQVZ UDW XGWN INPPXJGNS'P QTGOP WN PVV XCWQVIV XP PNZVWQXGJ XZUNIWTGW XG WQVZ." NI, VXJQW FVTIP STWVI, XG 1642,RIXWXGJ

WN ZVPPXVDIP OV GNFVIP TGO OV HQTKXJGF: "X PTR, XG PNZVVY-WITHWP, WQTW INPPXJGNS PVGW ZV, T WIDHV GVJNWXTWXNG NC WQV LXGJ NCVGJSTGO RXWQ WQV UIXGHV NC NITGJV; X ON GNW WQXGL WQTW XW HTG QTKV TGFVCCVHW, ADW ... XW XP DU WN FND, JVGWSVZVG, WN LVVU FNDI VFVP UVVSVO."

Implementation

The decryption process began with analyzing the frequency distribution of letters in the cipher text, as shown in Figure 1. This initial step provided the foundation for our frequency analysis approach.



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	G	X	T	N	P	I	Q	O	S	D	H	J	C	Z	F	U	R	A	K	L	Y	B	M	E
434	325	256	251	243	233	221	193	192	149	132	110	104	75	72	72	68	65	53	44	28	10	6	3	3	2
13.0	9.7	7.7	7.5	7.3	7.0	6.6	5.8	5.7	4.5	3.9	3.3	3.1	2.2	2.2	2.2	2.0	1.9	1.6	1.3	0.8	0.3	0.2	0.1	0.1	0.1

The most common trigraphs in the english language are:
THE,AND,THA,ENT,ION,TIO,FOR,NDE,HAS,NCE,TIS,OFT,MEN

The most common trigraphs in the message are:
WQV,XGJ,TGO,GOV,WQT,QTW,VIV,QXP,GHV,QVI,TWV,IVT,VGW

Figure 1: Counting the frequency in the given text

Based on trigraph patterns, THE and THA were identified (Figure 2). These common three-letter combinations provided crucial starting points for our analysis.

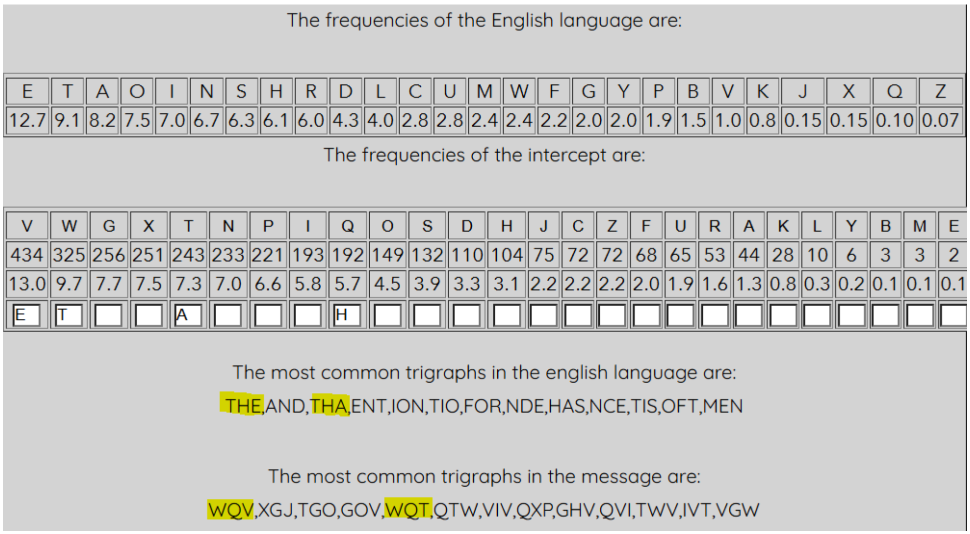


Figure 2: Analyzing trigraphs

The letter “I” was identified from the word “it” where no other vowel fits, as demonstrated in Figure 3.

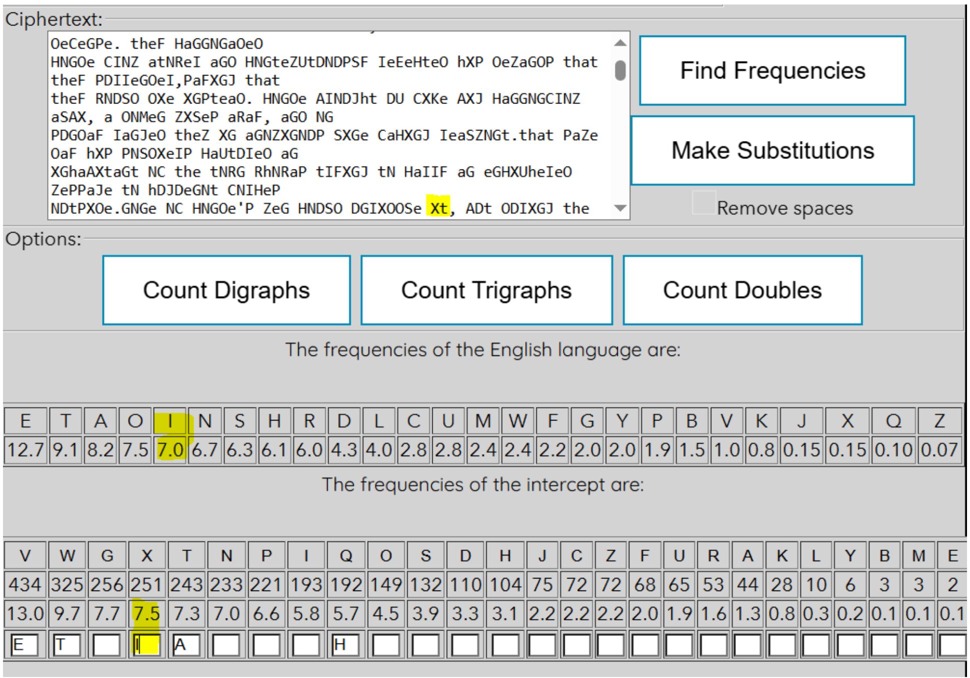


Figure 3: Finding “I”

Another common trigraph AND was analyzed, where A was already identified. This helped us determine the letters “N” and “D” as shown in Figure 4.

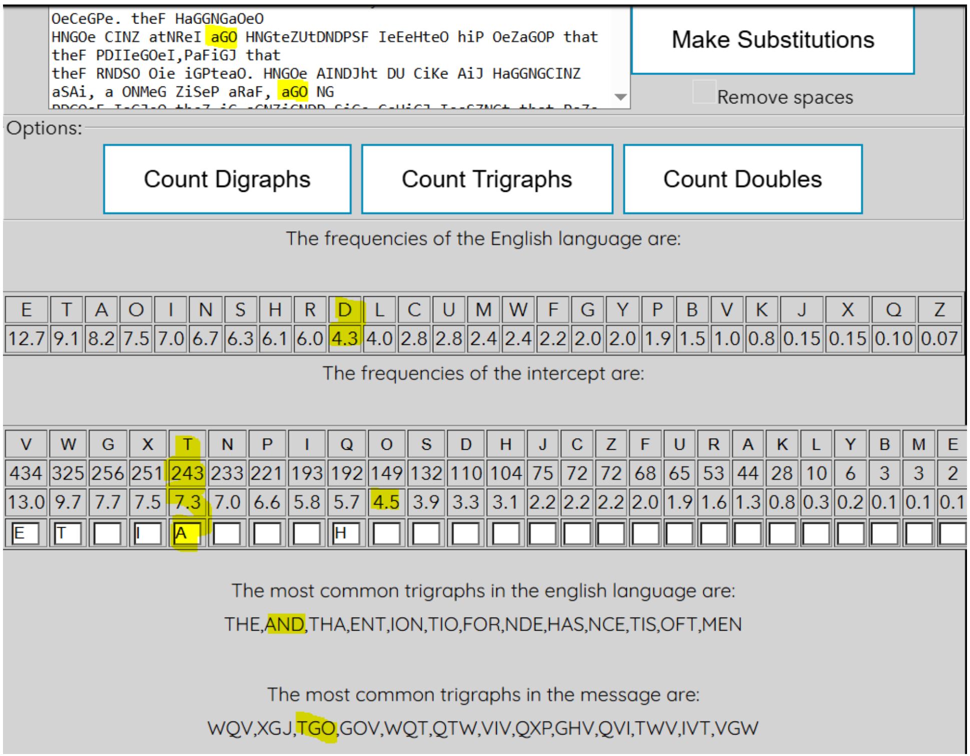


Figure 4: Finding "N" and "D"

The words "despite" and "this" stood out and were easy to identify which gave another 2 letters: "S" and "P". This process is illustrated in Figure 5.

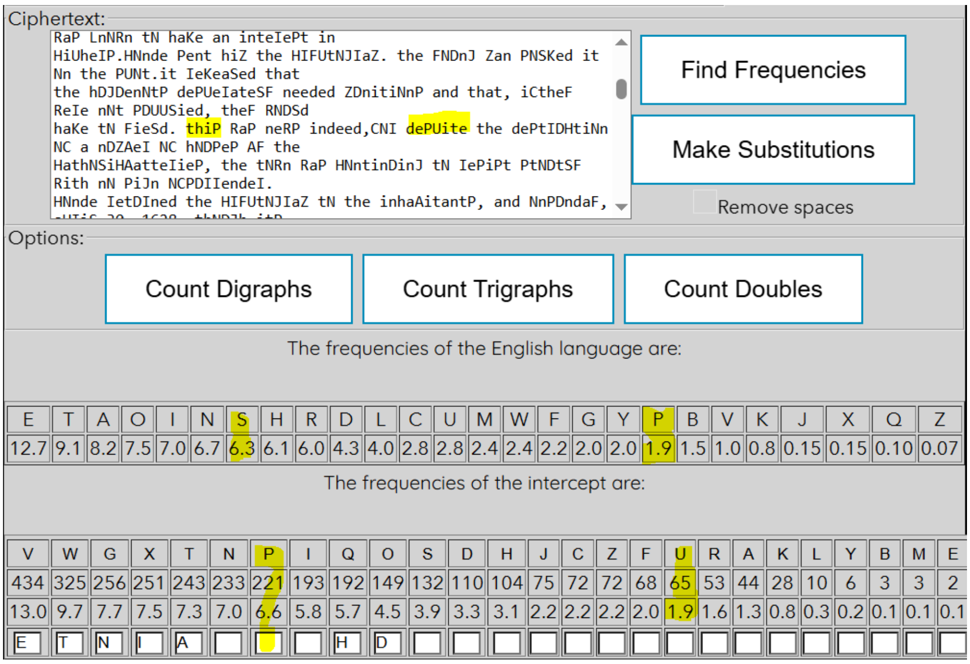


Figure 5: Finding "S" and "P"

To identify "O" - another popular vowel, the words "to" and "none" were used. Figure 6 shows this identification process.

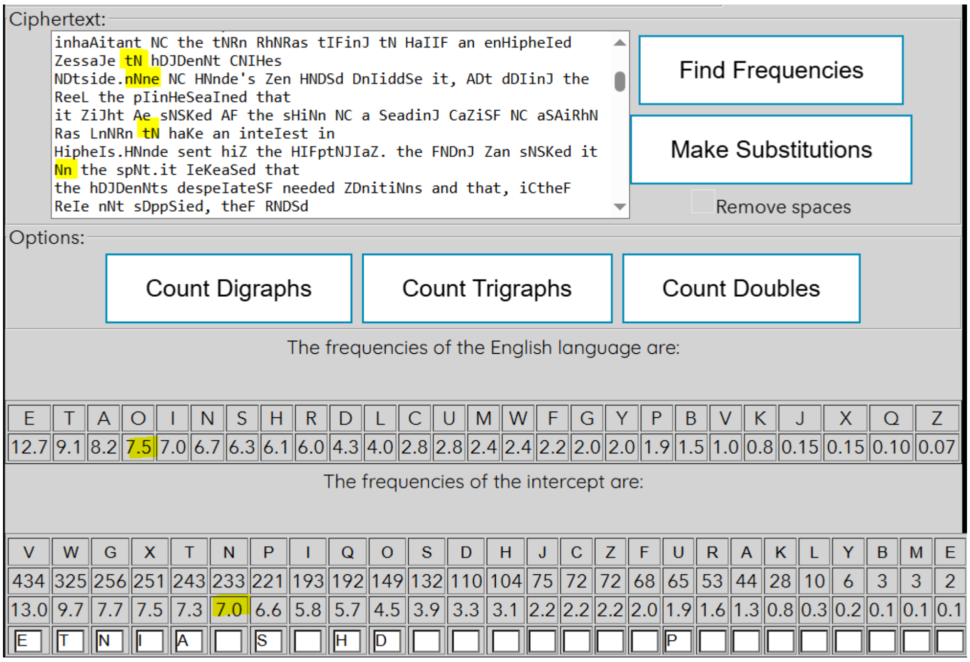


Figure 6: Identifying “O”

Another 2 words stood out from the phrase “have an interest in”. This analysis, shown in Figure 7, helped identify “V” and “R”.

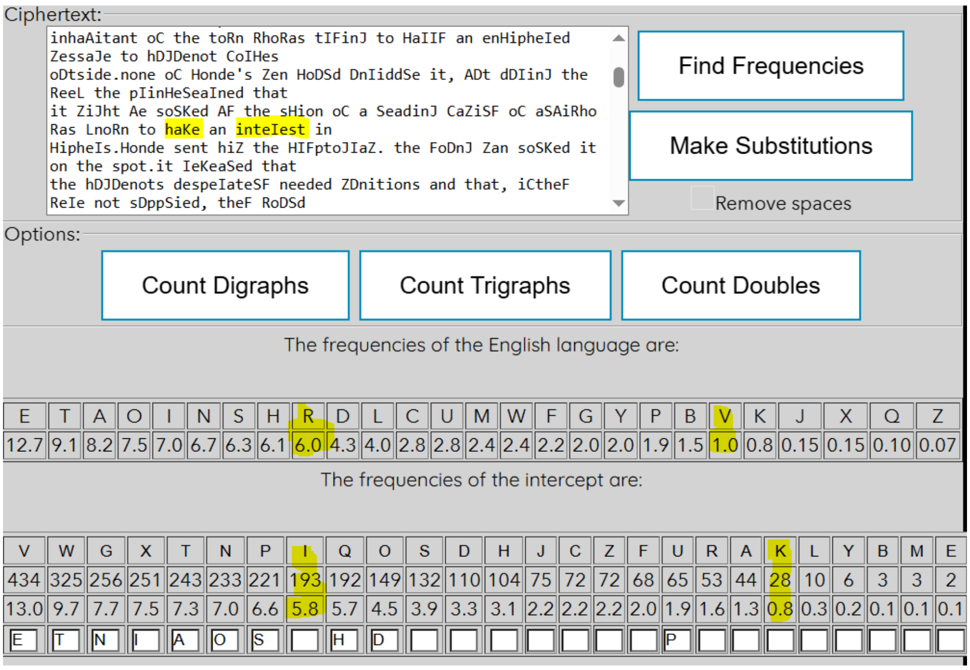


Figure 7: Finding “V” and “R”

The words “solved”, “revealed”, “desperately”, “they” confirmed the mapping for letter “L” and “Y”. Figure 8 demonstrates this identification process.

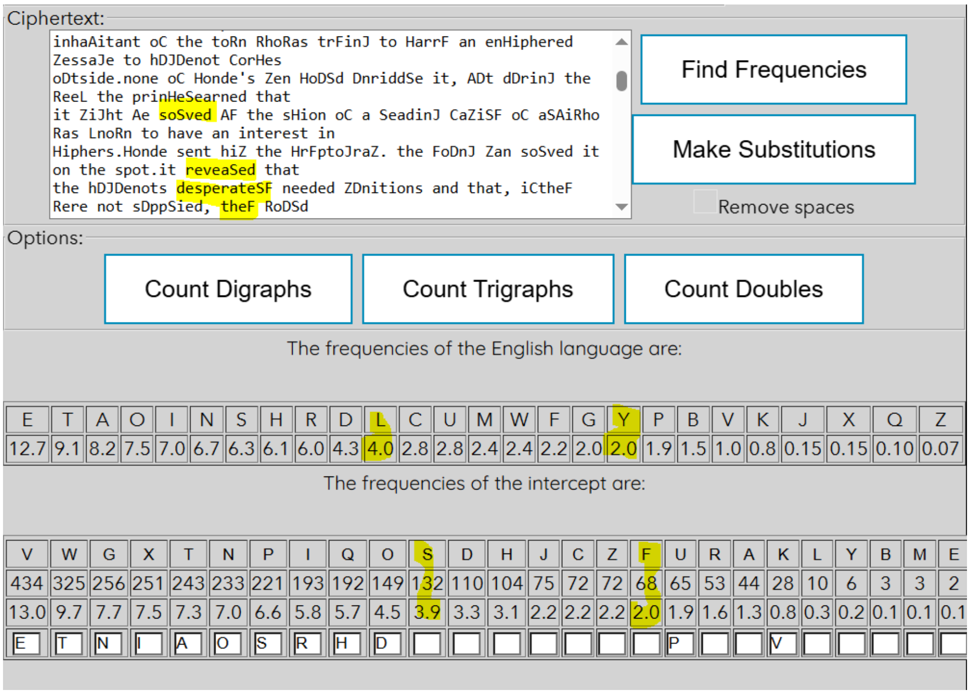


Figure 8: Finding “L” and “Y”

Words are easier to recognize now as most popular letters were identified. Expression “a dozen miles away and on sunday” stands out. Other words missing a letter are identified. Figure 9 shows the identification of multiple letters in this phase.

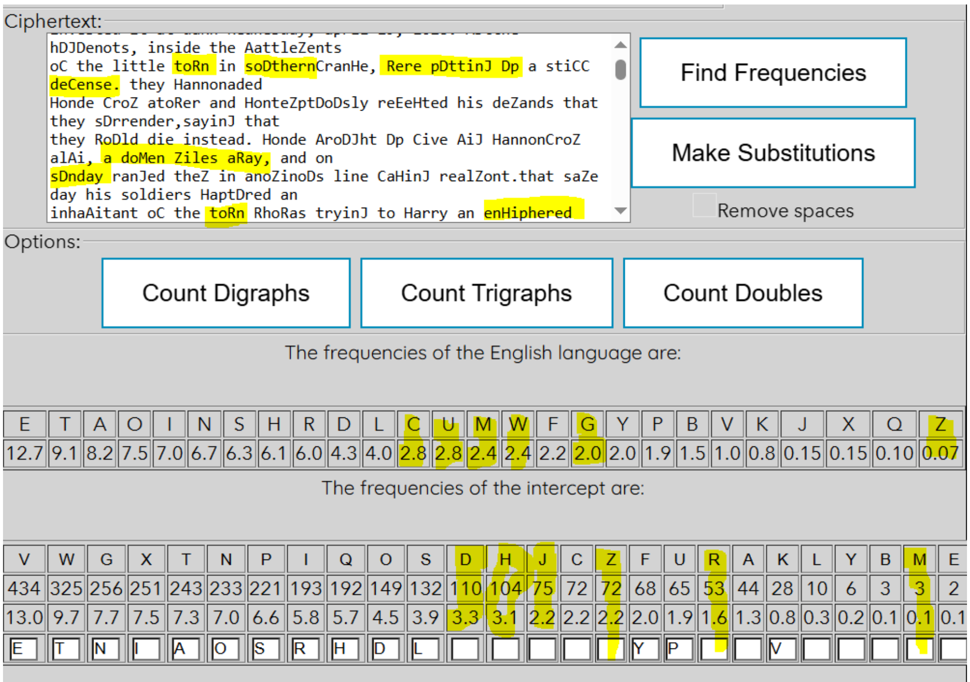


Figure 9: Identifying mapping for “C”, “U”, “M”, “W”, “G” and “Z”

“France”, “stiff” and “defense” confirm the mapping for letter “F”. “inhabitant” reveals mapping for letter “B”. This process is shown in Figure 10.

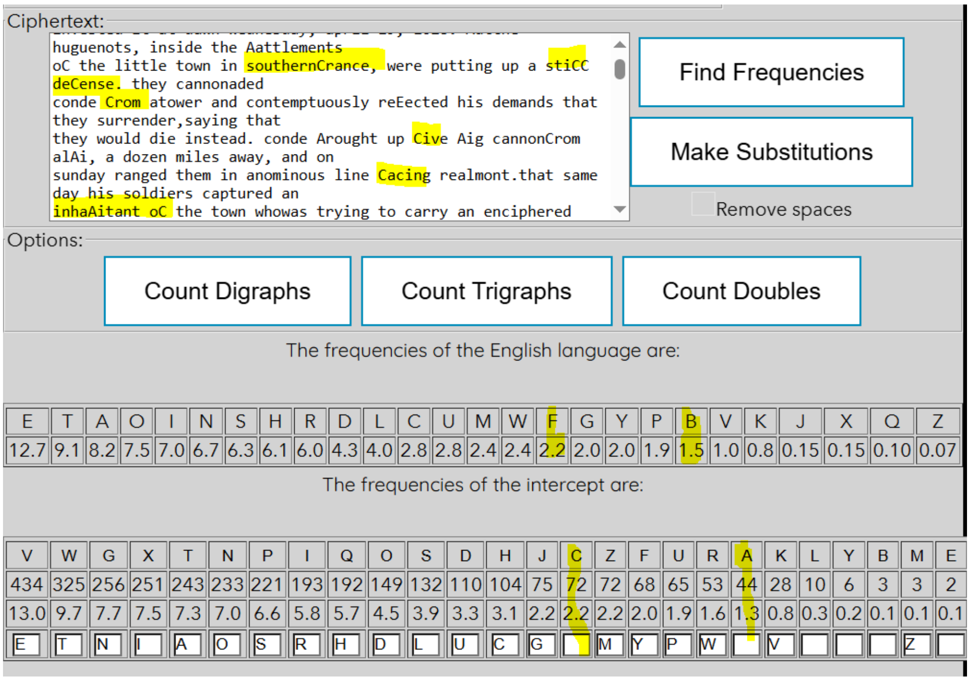


Figure 10: Finding mapping for “F”

More uncommon letters are identified as the majority of letters is mapped. “K” is found from word “known”, as demonstrated in Figure 11.

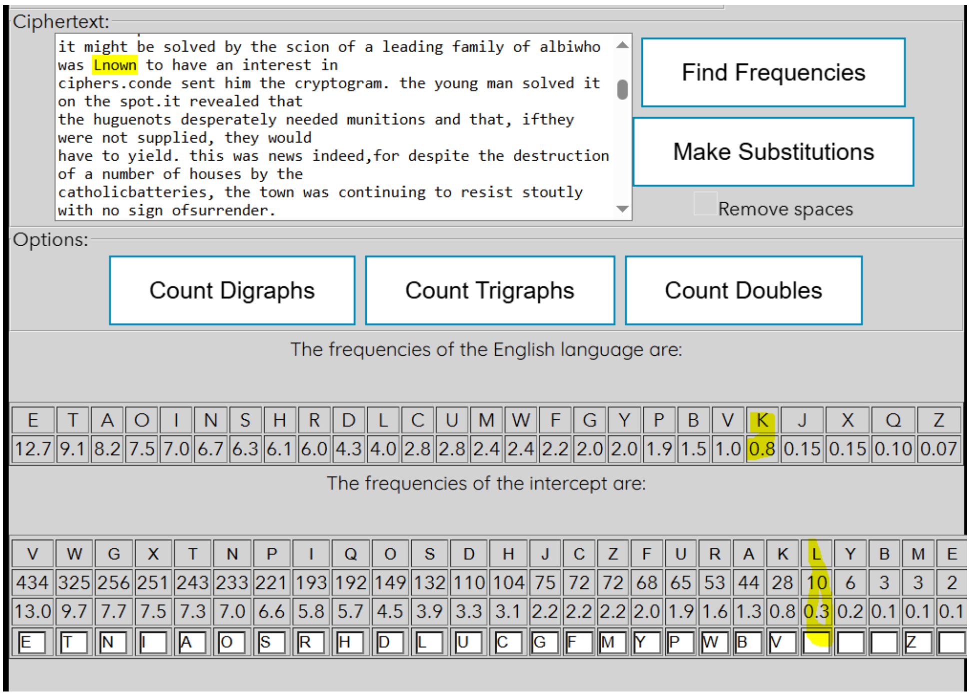


Figure 11: Finding “K”

“X” is found from word “unexpectedly”. Figure 12 illustrates this identification.

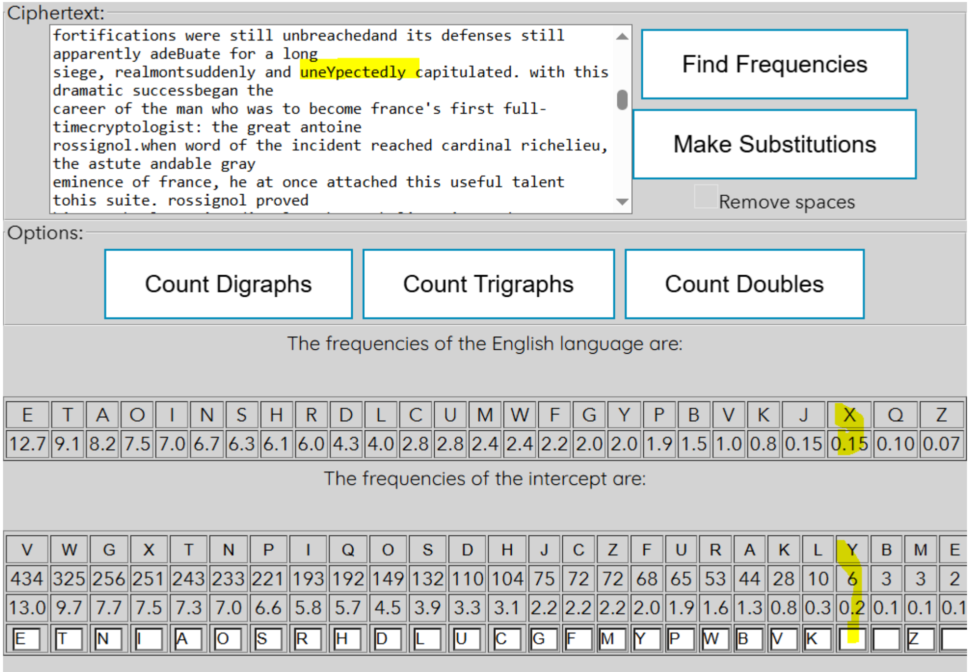


Figure 12: Finding “X”

“Q” is identified from word “quickly”, as shown in Figure 13.

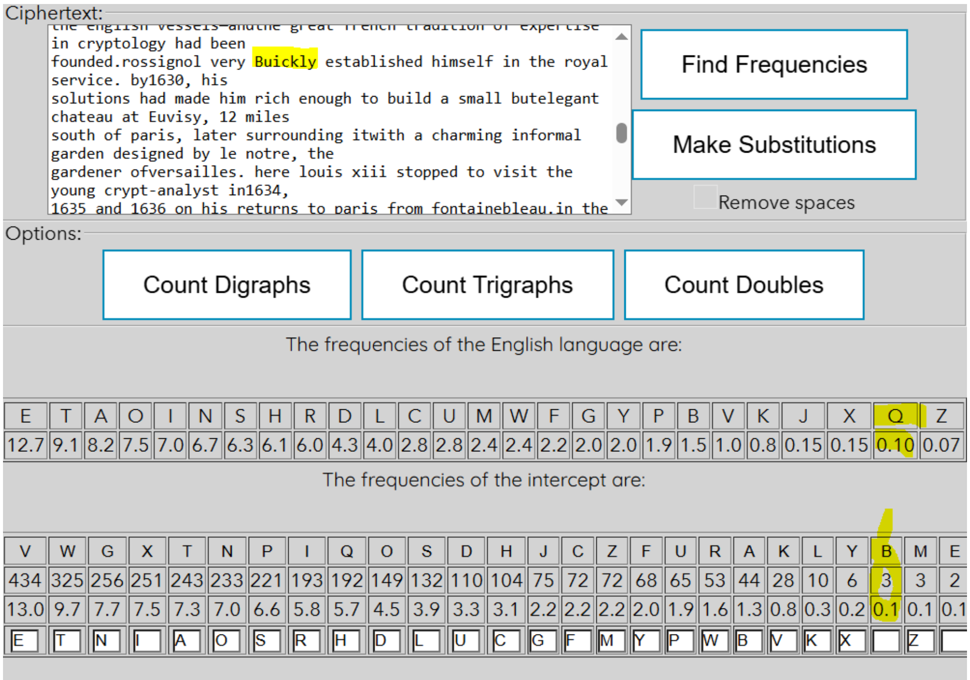


Figure 13: Finding “Q”

“J” is identified from word “rejected”. Figure 14 demonstrates this final identification step.

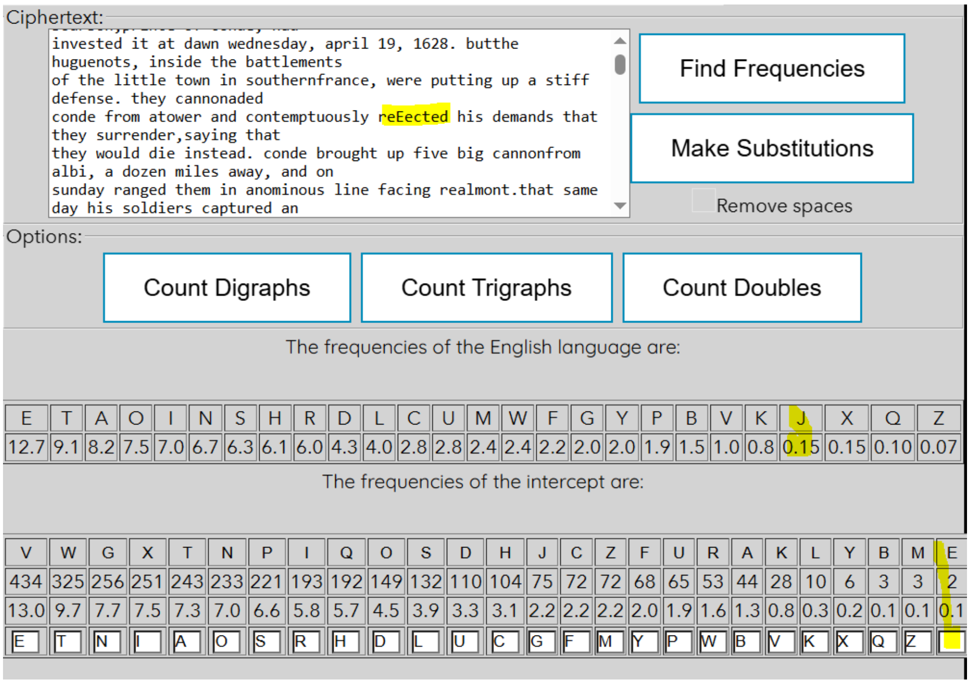


Figure 14: Finding “J”

Results

The decryption process was successful. Through systematic frequency analysis and pattern recognition, we were able to completely decode the cipher text. The final mapping revealed the following letter substitutions:

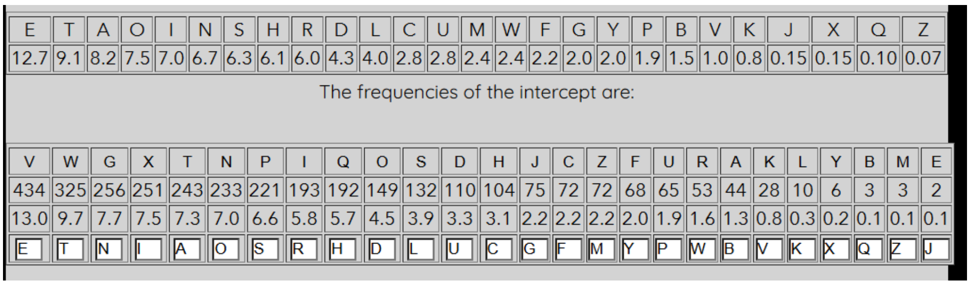


Figure 15: Final Mapping

The decrypted message reads:

realmont was under siege. the royal army, under henry ii of bourbon,prince of conde, had invested it at dawn wednesday, april 19, 1628. butthe huguenots, inside the battlements of the little town in southernfrance, were putting up a stiff defense. they cannonaded conde from atower and contemptuously reEected his demands that they surrender,saying that they would die instead. conde brought up five big cannonfrom albi, a dozen miles away, and on sunday ranged them in anominous line facing realmont.that same day his *soldiers captured an inhabitant of the town whowas trying to carry an enciphered message to huguenot forces outside.none of conde's men could unriddle it, but during the week the princelearned that it might be solved by the scion of a leading family of albiwho was known to have an interest in ciphers.conde sent him the cryptogram. the young man solved it on the spot.it revealed that the huguenots desperately needed munitions and that, ifthey were not supplied, they would have to yield. this was news indeed,for despite the destruction of a number of houses by the catholicbatteries, the town was continuing to resist stoutly with no sign ofsurrender. conde returned the cryptogram to the inhabitants, and onsunday, april 30, 1628, though its fortifications were still unbreachedand its defenses still apparently adequate for a long siege, realmontsuddenly and unexpectedly capitulated. with this dramatic successbegan the career of the man who was to become france's first full-timecryptologist: the great antoine rossignol.when word of the incident reached cardinal richelieu, the astute andable gray eminence of france, he at once attached this useful talent tohis suite. rossignol proved his worth almost immediately. the catholicarmies under richelieu surrounding the chief huguenot bastion of larochelle intercepted some letters in cipher, which the young codebreakerof albi read with ease. he told his eminence that the starving citizenswere eagerly awaiting help that the english had promised to send by sea. when the fleetarrived, the primed guardships and forts so intimidated it that it stood-off the port's entrance and made no serious attempt to force a passage. amonth later, the city capitulated in full sight of the english vessels—andthe great french tradition of expertise in cryptology had been founded.rossignol very quickly established himself in the royal service. by1630, his solutions had made him rich enough to build a small butelegant chateau at Euvisy, 12 miles south of paris, later surrounding itwith a charming informal garden designed by le notre, the gardener ofversailles. here louis xiii stopped to visit the young crypt-analyst in1634, 1635 and 1636 on his returns to paris from fontainebleau.in the swashbuckling court of that monarch, and then in theresplendent one of louis xiv, rossignol served with an extraordinaryfacility. the stronghold of hesdin surrendered a week sooner than itotherwise would have because he solved an enciphered plea for help, andthen composed a reply in the same cipher telling the townspeople howfutile their hopes were. how many other towns he compelled tosurrender, how many diplomatic coups he made possible, how manybetrayals he uncovered among the great nobles in those days of shiftingallegiances, he never discussed. this reticence caused some at the courtto charge that he never actually solved a single cipher, and that thecardinal spread inflated rumors about his abilities to discourage would-be conspirators. but in fact richelieu was frequently telling hissubordinates such things as, "it is necessary to make use, in my opinion,of the letters of the man who has been arrested by the civil authorities atmezieres, that is to say, have them put into rossignol's hands to see ifthere is something important in them." or, eight years later, in 1642,writing to messieurs de noyers and de chavigny: "i saw,* *in someextracts, that rossignol sent me, a truce negotiation of the king ofengland with the prince of orange; i do not think that it can have anyeffect, but ... it is up to you, gentlemen, to keep your eyes peeled."*

Conclusion

This lab successfully showed us how to use frequency analysis to break simple substitution ciphers. By counting letters and comparing them to normal English patterns, we were able to decode the secret message and figure out the complete cipher key.

Frequency analysis works very well against simple substitution ciphers, which explains why people had to invent better encryption methods throughout history. The weakness we exploited led to the development of more complex ciphers and eventually to the strong encryption methods we use today that resist statistical attacks.

The problems we found with simple substitution ciphers can be fixed in several ways. One approach is to use polyalphabetic ciphers like the Vigenère cipher, which use multiple substitution alphabets that change throughout the message. This makes frequency analysis much harder because the same letter gets encrypted differently each time.

Another method is homophonic substitution, where frequently used letters like "E" get assigned multiple different cipher symbols. This flattens out the frequency distribution so our counting method doesn't work as well. People can also add meaningless characters or padding to hide the true message length and confuse the frequency patterns.

The skills we gained from this lab include recognizing patterns, using statistical reasoning, and solving problems systematically. These are all important abilities in cybersecurity and information security. Understanding these old techniques helps us appreciate both the strengths and weaknesses of different cryptographic systems, and shows us why it's so important to use strong, well-tested encryption methods in modern applications.

Bibliography

1. GitHub Repository: https://github.com/m33ga/cs-laboratory-works
2. Frequency Analysis: Breaking the Code: https://crypto.interactive-maths.com/frequency-analysis-breaking-the-code.html