MINISTERUL EDUCAŢIEI ŞI CERCETĂRII AL REPUBLICII MOLDOVA

Universitatea Tehnică a Moldovei

Facultatea Calculatoare, Informatică și Microelectronică

Departamentul Inginerie Software și Automatică

Report

Laboratory work №2

Cryptography and Security

Subject: Cryptanalysis of Monoalphabetic Ciphers

Author:
Mihaela Untu,
std. gr. FAF-232

Verified:
Maia Zaica,
university assistant

Chișinău, 2025

Contents

1 Purpose of the laboratory work

The purpose of this laboratory work is to understand and apply the method of frequency analysis in order to break a monoalphabetic substitution cipher. By studying the statistical distribution of letters and common patterns in the English language, students learn how cryptanalysis can exploit language regularities to recover the original plaintext from an encrypted message. The intercepted ciphertext (Variant 2) is analyzed using statistical properties of the English language to reconstruct the original plaintext message without knowledge of the encryption key.

2 The Task

Suppose an encrypted message was intercepted which is known to have been produced using a monoalphabetic cipher. Recover the original message by applying a frequency-analysis attack, assuming it is a text written in English. Take into account that only the letters were encrypted; all other characters remain unencrypted.

Note: Use the online service at

`https://crypto.interactive-maths.com/frequency-analysis-breaking-the-code.html`.

The report should contain a description of the cracking process, exactly as presented in Section 2.3 ("Example of a frequency-analysis attack").Each student will take the variant corresponding to their position number on the group list.

3 Theoretical Considerations

The weakness of monoalphabetic cipher systems lies in the frequency of occurrence of characters in the text. If a ciphertext is sufficiently long and the language of the plaintext is known, the system can be broken by an attack based on the frequency of letter occurrence in that language (the frequency-analysis attack). This frequency has been intensively studied (not necessarily for cryptographic purposes) and, as a result, various ordered structures were built that rank letters by frequency for each European language and other languages. Usually, the longer the ciphertext, the closer the frequency of letters used approaches this general ordering. Comparing the two orderings (the ordering of characters in the ciphertext and the ordering of letters in the alphabet of the language assumed) leads to a number of correspondences (plaintext letter – ciphertext letter), which in many cases uniquely determine the encryption key.

We can use information about the frequency of letters in a language to try to break a monoalphabetic substitution cipher. This is possible because, for example, if in an English message the letter "E", which has the highest frequency, was encrypted as "X", then every "X" in the ciphertext corresponds to an "E" in the plaintext. Therefore, the

most frequent letter in the ciphertext should be "X". Thus, if we intercept a ciphertext and the most frequent letter in it is "P", we may assume that "P" was used to encrypt "E", and consequently we can replace every "P" with "E". Of course, not every text has exactly the same frequency distribution and, as noted above, letters like "T" and "A" also have high frequencies, so "P" might correspond to one of these instead. However, it is unlikely to be "Z", which is rare in English. Repeating this process with the next most frequent letter allows us to make progress in breaking the message.

If we were to simply order all letters by frequency and substitute according to the frequency table, we would most likely not obtain the expected result. The cryptanalyst must use other "personality traits" of letters to break the cryptogram. These include examining pairs of letters (digraphs), the most frequent of which in English are: TH, HE, AN, IN, ER, ON, RE, ED, ND, HA, AT, EN. Triplets of letters (trigraphs) are also useful; the most frequent in English include: THE, AND, THA, ENT, ION, TIO, FOR, NDE, HAS, NCE, TIS, OFT, MEN. Additionally, in English only a few letters commonly appear doubled (SS, EE, TT, OO and FF are the most frequent). There are only two single-letter English words with meaning: "A" and "I". Other common words begin to appear as we make substitutions. For example, "T*E" may appear frequently after substituting for "T" and "E". In that case "T*E" is very likely "THE", a very common English word. The frequency-analysis process exploits various subtle properties of language and for this reason it is almost impossible for a computer to do all the work by itself. Inevitably, human judgment is required in the process to make informed decisions about which letters should be substituted.

Variant 2:

> "WQV TOOXWXNG NC PVHIVHF WN WQV WITGPCNIZTWXNGP UINODHVOHIFUWNJITUQF. WIDV, XW RTP ZNIV NC T JTZV WQTG TGFWQXGJ VSPV—XW PNDJQWWN OVSTF HNZUIVQVGPXNG CNI NGSF WQV PQNIWVPW UNPPXASV WXZV, GNW WQVSNGJVPW—TGO WQV HIFUWTGTSFPXP RTP, SXLVRXPV, EDPW T UDMMSV. VJFUW'P RTPWQDP T BDTPX HIFUWNSNJF XG HNGWITPW WN WQV OVTOSF PVIXNDP PHXVGHV NC WNOTF.FVW JIVTW WQXGJP QTKV PZTSS AVJXGGXGJP, TGO WQVPV QXVINJSFUQP OXOXGHSDOV, WQNDJQ XG TG XZUVICVHW CTPQXNG, WQV WRN VSVZVGWP NC PVHIVHF TGOWITGPCNIZTWXNG WQTW HNZUIXPV WQV VPPVGWXTS TWWIXADWVP NC WQV PHXVGHV. TGOPN HIFUWNSNJF RTP ANIG. XG XWP CXIPW 3,000 FVTIP, XW OXO GNW JINR PWVTOXSF. HIFUWNSNJF TINPVXGOVUVGOVGWSF XG ZTGF USTHVP, TGO XG ZNPW NC WQVZ XW OXVO WQV OVTWQP NCXWP HXKXSXMTWXNGP. XG NWQVI USTHVP, XW PDIKXKVO, VZAVOOVO XG T SXWVITWDIV,TGO

CINZ WQXP WQV GVYW JVGVITWXNG HNDSO HSXZA WN QXJQVI SVKVSP.ADW UINJIVPP RTP PSNR TGO EVILF. ZNIV RTP SNPW WQTG IVWTXGVO. ZDHQ NC WQVQXPWNIF NC HIFUWNSNJF NC WQXP WXZV XP T UTWHQRNIL, T HITMF BDXSW NCDGIVSTWVO XWVZP, PUINDWXGJ, CSNDIXPQXGJ, RXWQVIXGJ. NGSF WNRTIO WQVRVPWVIG IVGTXPPTGHV ONVP WQV THHIVWXGJ LGNRSVOJV AVJXG WN ADXSO DU TZNZVGWDZ. WQV PWNIF NC HIFUWNSNJF ODIXGJ WQVPV FVTIP XP, XG NWQVI RNIOP,VYTHWSF WQV PWNIF NC ZTGLXGO. HQXGT, WQV NGSF QXJQ HXKXSXMTWXNG NC TGWXBDXWF WN DPV XOVNJITUQXHRIXWXGJ, PVVZP GVKVI WN QTKV OVKVSNUVO ZDHQ IVTS HIFUWNJITUQF —UVIQTUP CNI WQTW IVTPNG. XG NGV HTPV LGNRG CNI ZXSXWTIF UDIUNPVP, WQV11WQ-HVGWDIF HNZUXSTWXNG, RD-HQXGJ WPDGJFTN (”VPPVGWXTSP CINZ ZXSXWTIFHSTPPXHP”), IVHNZZVGOVO T WIDV XC PZTSS HNOV. WN T SXPW NC 40 USTXGWVYWXWVZP, ITGJXGJ CINZ IVBDVPWP CNI ANRP TGO TIINRP WN WQV IVUNIW NC TKXHWNIF, WQV HNIIVPUNGOVGWP RNDSO TPPXJG WQV CXIPW 40 XOVNJITZP NC TUNVZ. WQVG, RQVG T SXVDWVGTGW RXPQVO, CNI VYTZUSV, WN IVBDVPW ZNIVTIINRP, QV RTP WN RIXWV WQV HNIIVPUNGOXGJ XOVNJITZ TW T PUVHXCXVO USTHVNG TG NIOXGTIF OXPUTWHQ TGO PWTZU QXP PVTS NG XW.XG HQXGT’P JIVTW GVXJQANI WN WQV RVPW, XGOXT, RQNPV HXKXSXMTWXNGSXLVRXPV OVKVSNUVO VTISF TGO WN QXJQ VPWTWV, PVKVITS CNIZP NC PVHIVWHNZZDGXHTWXNGP RVIV LGNRG TGO, T UUTIVGWSF, UITHWXHVO. WQV TIWQT-PTPWIT, T HSTPPXH RNIL NG PWTWVHITCW TWWIXADWVO WN LTDWXSFT, XG OVPHIXAXGJWQV VPUXNGTJV PVIKXHV NC XGOXT TP UITHWXHTSSF IXOOSXGJ WQV HNDGWIF RXWQP UXVP, IVHNZZVGOVO WQTW WQV NCCXHVIP NC WQV XGPWXWDWVP NC £ PUXNGTJV JXKVWQVXI PUXVP WQVXI TPPXJGZVGWP AF PVHIVW RIXWXGJ.UVIQTUP ZNPW XGWVIVPWXGJ WN HIFUWNSNJXPWP, TZTWVDI NIUINCVPPXNGTS, XP WQTW KTWPFTFTGT’P CTZNDP WVYWANNL NC VINWXHP, WQV LTZTPDWIT,SXPWP PVHIVW RIXWXGJ TP NGV NC WQV 64 TIWP, NI FNJTP, WQTW RNZVGPQNDSO LGNR TGO UITHWXHV. WQV CNDIWQ JIVTW HXKXSXMTWXNG NC TGWXBDXWF, WQVZVPNUN-WTZXTG, ITWQVI UTITSSVSVO VJFUW VTISF XG XWP HIFUWNJITUQXHVKNSDWXNG, ADW WQVG PDIUTPPVO XW. WQDP, XG WQV STPW UVIXNO NC HDGVXCNIZRIXWXGJ, XG HNSNUQNGP RIXWWVG TW DIDL (XG UIVPVGW-OTF XITB)

DGOVI WQVPVSVDHXO LXGJP XG WQV STPW CVR PHNIV FVTIP AVCNIV WQV HQIXPWXTG VIT,NHHTPXNGTS PHIXAVP HNGKVI-WVO WQVXI GTZVP XGWN GDZAVIP. WQVVGHXUQVIZVGW—XC PDHQ XW AV—ZTF QTKV AVVG NGSF CNI TZDPVZVGW NI WN-PQNR NCC.“

4 Technical implementation

4.1 Initial Frequency Count

The first step involved counting the occurrence of each letter in the ciphertext. The results are presented in Figure 1.

V	W	X	N	P	T	I	G	Q	H	S	O	U	F	Z	D	J	C	R	A	K	L	B	M	Y	E
273	250	201	195	188	185	169	165	111	89	84	79	66	64	61	59	59	56	42	21	20	15	7	7	5	2
11.0	10.1	8.1	7.9	7.6	7.5	6.8	6.7	4.5	3.6	3.4	3.2	2.7	2.6	2.5	2.4	2.4	2.3	1.7	0.8	0.8	0.6	0.3	0.3	0.2	0.1

Figure 1: Letter frequencies in the intercepted ciphertext (Variant 2)

4.2 Initial Substitution Hypotheses

Based on the standard English frequency distribution (E=12.7%, T=9.1%, A=8.2%, O=7.5%), the following initial hypotheses were formed:

- **V** → **e**: Most frequent letter in ciphertext (273 occurrences)
- **W** → **t**: Second most frequent letter (250 occurrences)

Applying these substitutions to the beginning of the ciphertext:

“tQe TOOXtXNG NC PeHIeHF tN tQe tITGPCNIZTtXNGP UINODHeO-HIFUtNJITUQF. tIDe, Xt RTP ZNIe NC T JTZe tQTG TGFtQXGJ eSPe—Xt PNDJQttN OeSTF HNZUIeQeGPXNG CNI NGSF tQe PQNItePt UNPPX-ASe tXZe, GNt tQeSNGJePt—TGO tQe HIFUtTGTSFPXP RTP, SXLeRXPe, EDPt T UDMMSe. eJFUt’P RTPtQDP T BDTPX HIFUtNSNJF XG HNGtITPt tN tQe OeTOSF PeIXNDP PHXeGHe NC tNOTF.Fet JIeTt tQXGJP QTKe PZTSS AeJXGGXGJP, TGO tQePe QXeINJSFUQP OXOXGHSDOe, tQNDJQ XG TG XZUeICeHt CTPQXNG, tQe tRN eSeZeGtP NC PeHIeHF TGOtIT-GPCNIZTtXNG tQTt HNZUIXPe tQe ePPeGtXTS TttIXADteP NC tQe PHXeGHe. TGOPN HIFUtNSNJF RTP ANIG. XG XtP CXIPt 3,000 FeTIP, Xt OXO GNt JINR PteTOXSF. HIFUtNSNJF TINPeXGOeUeGOeGtSF XG ZTGF USTHeP, TGO XG ZNPt NC tQeZ Xt OXeO tQe OeTtQP NCXtP HXKXSXMTtXNGP. XG NtQeI USTHeP, Xt PDIKXKeO, eZAeOOeO XG

T SXteITtDIe,TGO CINZ tQXP tQe GeYt JeGeITtXNG HNDSO HSXZA tN QXJQeI SeKeSP.ADt UINJIePP RTP PSNR TGO EeILF. ZNIe RTP SNPt tQTG IetTXGeO. ZDHQ NC tQeQXPtNIF NC HIFUtNSNJF NC tQXP tXZe XP T UTtHQRNIL, T HITMF BDXSt NCDGIeSTteO XteZP, PUINDtXGJ, CSNDIXPQXGJ, RXtQeIXGJ. NGSF tNRTIO tQeRePteIG IeGTXPPTGHe ONeP tQe THHIetXGJ LGNRSeOJe AeJXG tN ADXSO DU TZNZeGtDZ. tQe PtNIF NC HIFUtNSNJF ODIXGJ tQePe FeTIP XP, XG NtQeI RNIOP,eYTHtSF tQe PtNIF NC ZTGLXGO. HQXGT, tQe NGSF QXJQ HXKXSXMTtXNG NC TGtXBDXtF tN DPe XOeNJITUQXHRIX-tXGJ, PeeZP GeKeI tN QTKe OeKeSNUeO ZDHQ IeTS HIFUtNJITUQF —UeIQTUP CNI tQTt IeTPNG. XG NGe HTPe LGNRG CNI ZXSXtTIF UDIUNPeP, tQe11tQ-HeGtDIF HNZUXSTtXNG, RD-HQXGJ tPDGJ-FTN ("ePPeGtXTSP CINZ ZXSXtTIFHSTPPXHP"), IeHNZZeGOeO T tIDe XC PZTSS HNOe. tN T SXPt NC 40 USTXGteYtXteZP, ITGJXGJ CINZ IeB-DePtP CNI ANRP TGO TIINRP tN tQe IeUNIt NC TKXHtNIF, tQe HNI-IePUNGOeGtP RNDSO TPPXJG tQe CXIPt 40 XOeNJITZP NC TUNeZ. tQeG, RQeG T SXeDteGTGt RXPQeO, CNI eYTZUSe, tN IeBDePt ZNIeTI-INRP, Qe RTP tN RIXte tQe HNIIePUNGOXGJ XOeNJITZ Tt T PUeHX-CXeO USTHeNG TG NIOXGTIF OXPUTtHQ TGO PtTZU QXP PeTS NG Xt.XG HQXGT'P JIeTt GeXJQANI tN tQe RePt, XGOXT, RQNPe HXKXSXMTtXNGSXLeRXPe OeKeSNUeO eTISF TGO tN QXJQ ePtTte, PeKeITS CNIZP NC PeHIetHNZZDGXHTtXNGP ReIe LGNRG TGO, T UUTIeGtSF, UITHtXHeO. tQe TItQT-PTPtIT, T HSTPPXH RNIL NG PtTteHITCt TttIXADteO tN LTDtXSFT, XG OePHIXAXGJtQe ePUXNGTJe PeIKXHe NC XGOXT TP UITHtXHTSSF IXOOSXGJ tQe HNDGtIF RX-tQP UXeP, IeHNZZeGOeO tQTt tQe NCCXHeIP NC tQe XGPtXtDteP NC £ PUXNGTJe JXKetQeXI PUXeP tQeXI TPPXJGZeGtP AF PeHIet RIXtXGJ.UeIQTUP ZNPt XGteIePtXGJ tN HIFUtNSNJXPtP, TZTteDI NIUINCePPXNGTS, XP tQTt KTtPFTFTGT'P CTZNDP teYtANNL NC eINtXHP, tQe LTZTPDtIT,SXPtP PeHIet RIXtXGJ TP NGe NC tQe 64 TItP, NI FNJTP, tQTt RNZeGPQNDSO LGNR TGO UITHtXHe. tQe CN-DItQ JIeTt HXKXSXMTtXNG NC TGtXBDXtF, tQeZePNUN-tTZXTG, ITtQeI UTITSSeSeO eJFUt eTISF XG XtP HIFUtNJITUQXHeKNSDtXNG, ADt tQeG PDIUTPPeO Xt. tQDP, XG tQe STPt UeIXNO NC HDGeX-CNIZRIXtXGJ, XG HNSNUQNGP RIXtteG Tt DIDL (XG UIePeGt-OTF XITB) DGOeI tQePeSeDHXO LXGJP XG tQe STPt CeR PHNIe FeTIP AeCNIe tQe HQIXPtXTG eIT,NHHTPXNGTS PHIXAeP HNGKeIteO tQeXI GTZeP XGtN GDZAeIP. tQeeGHXUQeIZeGt—XC PDHQ Xt Ae—ZTF QTKe

AeeG NGSF CNI TZDPeZeGt NI tNPQNR NCC.“

Step 1: Identifying ”the”

The trigram ”tQe” appeared very frequently throughout the text. In English, the most common 3-letter word is ”the”, which matched our initial substitutions for ’t’ and ’e’. This strongly suggested:

- **Q → h**

Step 2: Analyzing High-Frequency Letters

Looking at the next most frequent letter ”X” (201 occurrences), which likely represents one of ’a’, ’i’, or ’o’, I examined the context. The word ”Xt” appeared multiple times. Considering that ”it” is a common English word and checking the frequency distribution, I determined:

- **X → i**

Similarly, ”tN” appeared as a two-letter word. Since we already identified ’t’, and ”to” is one of the most common English words:

- **N → o**

Result:

> “the TOOitioG oC PeHIeHF to the tITGPCoIZTtioGP UIoODHeOHIFUtoJITUhF. tIDe, it RTP ZoIe oC T JTZe thTG TGFthiGJ eSPe—it PoDJhtto OeSTF HoZUIeheGPioG CoI oGSF the PhoItePt UoPPiASe tiZe, Got theSoGJePt—TGO the HIFUtTGTSFPiP RTP, SiLeRiPe, EDPt T UDMMSe. eJFUt’P RTPthDP T BDTPi HIFUtoSoJF iG HoGtITPt to the OeTOSF PeIioDP PHieGHe oC toOTF.Fet JIeTt thiGJP hTKe PZTSS AeJiGGiGJP, TGO thePe hieIoJSFUhP OiOiGHSDOe, thoDJh iG TG iZUeICeHt CTPhioG, the tRo eSeZeGtP oC PeHIeHF TGOtITGPCoIZTtioG thTt HoZUIiPe the ePPeGtiTS TttIiADteP oC the PHieGHe. TGOPo HIFUtoSoJF RTP AoIG. iG itP CiIPt 3,000 FeTIP, it OiO Got JIoR PteTOiSF. HIFUtoSoJF TIoPeiGOeUeGOeGtSF iG ZTGF USTHeP, TGO iG ZoPt oC theZ it OieO the OeTthP oCitP HiKiSiMTtioGP. iG otheI USTHeP, it PDIKiKeO, eZAeOOeO iG T SiteITtDIe,TGO CIoZ thiP the GeYt JeGeITtioG HoDSO HSiZA to hiJheI SeKeSP.ADt UIoJIePP RTP PSoR TGO EeILF. ZoIe RTP SoPt thTG IetTiGeO. ZDHh oC thehiPtoIF oC HIFUtoSoJF oC thiP tiZe iP

T UTtHhRoIL, T HITMF BDiSt oCDGIeSTteO iteZP, PUIoDtiGJ, CSoDIiPhiGJ, RitheIiGJ. oGSF toRTIO theRePteIG IeGTiPPTGHe OoeP the THHIetiGJ LGoRSeOJe AeJiG to ADiSO DU TZoZeGtDZ. the PtoIF oC HIFUtoSoJF ODIiGJ thePe FeTIP iP, iG otheI RoIOP,eYTHtSF the PtoIF oC ZTGLiGO. HhiGT, the oGSF hiJh HiKiSiMTtioG oC TGtiBDitF to DPe iOeoJITUhiHRIitiGJ, PeeZP GeKeI to hTKe OeKeSoUeO ZDHh IeTS HIFUtoJITUhF —UeIhTUP CoI thTt IeTPoG. iG oGe HTPe LGoRG CoI ZiSitTIF UDIUoPeP, the11th-HeGtDIF HoZUiSTtioG, RD-HhiGJ tPDGJFTo ("ePPeGtiTSP CIoZ ZiSitTIFHSTPPiHP"), IeHoZZeGOeO T tIDe iC PZTSS HoOe. to T SiPt oC 40 USTiGteYtiteZP, ITGJiGJ CIoZ IeBDePtP CoI AoRP TGO TIIoRP to the IeUoIt oC TKiHtoIF, the HoIIePUoGOeGtP RoDSO TPPiJG the CiIPt 40 iOeoJITZP oC TUoeZ. theG, RheG T SieDteGTGt RiPheO, CoI eYTZUSe, to IeBDePt ZoIeTIIoRP, he RTP to RIite the HoIIePUoGOiGJ iOeoJITZ Tt T PUeHiCieO USTHeoG TG oIOiGTIF OiPUTtHh TGO PtTZU hiP PeTS oG it.iG HhiGT'P JIeTt GeiJhAoI to the RePt, iGOiT, RhoPe HiKiSiMTtioGSiLeRiPe OeKeSoUeO eTISF TGO to hiJh ePtTte, PeKeITS CoIZP oC PeHIetHoZZDGiHTtioGP ReIe LGoRG TGO, T UUTIeGtSF, UITHtiHeO. the TIthT-PTPtIT, T HSTPPiH RoIL oG PtTteHITCt TttIiADteO to LTDtiSFT, iG OePHIiAiGJthe ePUioGTJe PeIKiHe oC iGOiT TP UITHtiHTSSF IiOOSiGJ the HoDGtIF RithP UieP, IeHoZZeGOeO thTt the oCCiHeIP oC the iGPtitDteP oC £ PUioGTJe JiKetheiI PUieP theiI TPPiJGZeGtP AF PeHIet RIitiGJ.UeIhTUP ZoPt iGteIePtiGJ to HIFUtoSoJiPtP, TZTteDI oIUIoCePPioGTS, iP thTt KTtPFTFTGT'P CTZoDP teYtAooL oC eIotiHP, the LTZTPDtIT,SiPtP PeHIet RIitiGJ TP oGe oC the 64 TItP, oI FoJTP, thTt RoZeGPhoDSO LGoR TGO UITHtiHe. the CoDIth JIeTt HiKiSiMTtioG oC TGtiBDitF, theZePoUo-tTZiTG, ITtheI UTITSSeSeO eJFUt eTISF iG itP HIFUtoJITUhiHeKoSDtioG, ADt theG PDIUTPPeO it. thDP, iG the STPt UeIioO oC HDGeiCoIZRIitiGJ, iG HoSoUhoGP RIitteG Tt DIDL (iG UIePeGt-OTF iITB) DGOeI thePeSeDHiO LiGJP iG the STPt CeR PHoIe FeTIP AeCoIe the HhIiPtiTG eIT,oHHTPioGTS PHIiAeP HoGKeIteO theiI GTZeP iGto GDZAeIP. theeGHiUheIZeGt—iC PDHh it Ae—ZTF hTKe AeeG oGSF CoI TZDPeZeGt oI toPhoR oCC.RoZeGPhoDSO"

Step 3: Common Words and Context

The word "oC" appeared frequently. Given the context and frequency patterns, this clearly represents "of":

- **C → f**

The phrase "3,000 FeTIP" strongly suggested "3,000 years", leading to:

- **F** → **y**
- **T** → **a**
- **P** → **s**

The letter 'T' appeared alone as a word, which in English can only be "a":

- **T** → **a** (confirmed)

Result:

> "the aOOitioG of seHIeHF to the tIaGsfoIZatioGs UIoODHeOHIFUtoJIaUhF. tIDe, it Ras ZoIe of a JaZe thaG aGFthiGJ eSse—it soDJhtto OeSaF HoZUIeheGsioG foI oGSF the shoItest UossiASe tiZe, Got theSoGJest—aGO the HIFUtaGaSFsis Ras, SiLeRise, EDst a UDMMSe. eJFUt's RasthDs a BDasi HIFUtoSoJF iG HoGtIast to the OeaOSF seIioDs sHieGHe of toOaF.Fet JIeat thiGJs haKe sZaSS AeJiGGiGJs, aGO these hieIoJSFUhs OiOiGHSDOe, thoDJh iG aG iZUeIfeHt fashioG, the tRo eSeZeGts of seHIeHF aGOtIaGsfoIZatioG that HoZUIise the esseGtiaS attIiADtes of the sHieGHe. aGOso HIFUtoSoJF Ras AoIG. iG its fiIst 3,000 FeaIs, it OiO Got JIoR steaOiSF. HIFUtoSoJF aIoseiGOeUeGOeGtSF iG ZaGF USaHes, aGO iG Zost of theZ it OieO the Oeaths ofits HiKiSiMatioGs. iG otheI USaHes, it sDIKiKeO, eZAeOOeO iG a SiteIatDIe,aGO fIoZ this the GeYt JeGeIatioG HoDSO HSiZA to hiJheI SeKeSs.ADt UIoJIess Ras sSoR aGO EeILF. ZoIe Ras Sost thaG IetaiGeO. ZDHh of thehistoIF of HIFUtoSoJF of this tiZe is a UatHhRoIL, a HIaMF BDiSt ofDGIeSateO iteZs, sUIoDtiGJ, fSoDIishiGJ, RitheIiGJ. oGSF toRaIO theResteIG IeGaissaGHe Ooes the aHHIetiGJ LGoRSeOJe AeJiG to ADiSO DU aZoZeGtDZ. the stoIF of HIFUtoSoJF ODIiGJ these FeaIs is, iG otheI RoIOs,eYaHtSF the stoIF of ZaGLiGO. HhiGa, the oGSF hiJh HiKiSiMatioG of aGtiBDitF to Dse iOeoJIaUhiHRItiGJ, seeZs GeKeI to haKe OeKeSoUeO ZDHh IeaS HIFUtoJIaUhF —UeIhaUs foI that IeasoG. iG oGe Hase LGoRG foI ZiSitaIF UDIUoses, the11th-HeGtDIF HoZUiSatioG, RD-HhiGJ tsDGJ-Fao ("esseGtiaSs fIoZ ZiSitaIFHSassiHs"), IeHoZZeGOeO a tIDe if sZaSS HoOe. to a Sist of 40 USaiGteYtiteZs, IaGJiGJ fIoZ IeBDests foI AoRs aGO aIIoRs to the IeUoIt of aKiHtoIF, the HoIIesUoGOeGts RoDSO assiJG the fiIst 40 iOeoJIaZs of aUoeZ. theG, RheG a SieDteGaGt RisheO, foI eYaZUSe, to IeBDest ZoIeaIIoRs, he Ras to RIite the HoIIesUoGOiGJ iOeoJIaZ at a sUeHifieO USaHeoG aG oIOiGaIF OisUatHh aGO staZU his seaS oG it.iG HhiGa's JIeat GeiJhAoI to the Rest, iGOia, Rhose HiKiSiMatioGSiLeRise OeKeSoUeO eaISF

aGO to hiJh estate, seKeIaS foIZs of seHIetHoZZDGiHatioGs ReIe LGoRG aGO, a UUaIeGtSF, UIaHtiHeO. the aItha-sastIa, a HSassiH RoIL oG state-HIaft attIiADteO to LaDtiSFa, iG OesHIiAiGJthe esUioGaJe seIKiHe of iGOia as UIaHtiHaSSF IiOOSiGJ the HoDGtIF Riths Uies, IeHoZZeGOeO that the offiHeIs of the iGstitDtes of £ sUioGaJe JiKetheiI sUies theiI as-siJGZeGts AF seHIet RIitiGJ.UeIhaUs Zost iGteIestiGJ to HIFUtoSoJists, aZateDI oIUIofessioGaS, is that KatsFaFaGa's faZoDs teYtAooL of eIotiHs, the LaZasDtIa,Sists seHIet RIitiGJ as oGe of the 64 aIts, oI FoJas, that RoZeGshoDSO LGoR aGO UIaHtiHe. the foDIth JIeat HiKiSiMatioG of aGtiBDitF, theZesoUo-taZiaG, IatheI UaIaSSeSeO eJFUt eaISF iG its HI-FUtoJIaUhiHeKoSDtioG, ADt theG sDIUasseO it. thDs, iG the Sast UeIioO of HDGeifoIZRIitiGJ, iG HoSoUhoGs RIitteG at DIDL (iG UIeseGt-OaF iIaB) DGOeI theseSeDHiO LiGJs iG the Sast feR sHoIe FeaIs AefoIe the HhIistiaG eIa,oHHasioGaS sHIiAes HoGKeIteO theiI GaZes iGto GDZAeIs. theeGHiUheIZeGt—if sDHh it Ae—ZaF haKe AeeG oGSF foI aZDseZeGt oI toshoR off.“

Step 4: Progressive Decryption

Examining the partially decrypted text, I identified several more letters:
From ”thiP” (clearly ”this”):

- **P** → **s** (confirmed)

From ”thaG” (clearly ”than”):

- **G** → **n**

From ”tIansfoIZations” (clearly ”transformations”):

- **I** → **r**
- **Z** → **m**

From ”anO” (clearly ”and”):

- **O** → **d**

From ”serioDs” and ”trDe”:

- **D** → **u**

From ”onSF” (clearly ”only”):

- **S** → **l**

Result:

> “the addition of seHreHy to the transformations UroduHedHryUtoJraUhy. true, it has more of a Jame than anythinJ else—it souJhtto delay HomUrehension for only the shortest UossiAle time, not thelonJest—and the HryUtanalysis has, liLehise, Eust a UuMMle. eJyUt’s hasthus a Buasi HryUtoloJy in Hontrast to the deadly serious sHienHe of today.yet Jreat thinJs haKe small AeJinninJs, and these hieroJlyUhs didinHlude, thouJh in an imUerfeHt fashion, the tho elements of seHreHy andtransformation that HomUrise the essential attriAutes of the sHienHe. andso HryUtoloJy has Aorn. in its first 3,000 years, it did not Jroh steadily. HryUtoloJy aroseindeUendently in many UlaHes, and in most of them it died the deaths ofits HiKiliMations. in other UlaHes, it surKiKed, emAedded in a literature,and from this the neYt Jeneration Hould HlimA to hiJher leKels.Aut UroJress has sloh and EerLy. more has lost than retained. muHh of thehistory of HryUtoloJy of this time is a UatHhhorL, a HraMy Built ofunrelated items, sUroutinJ, flourishinJ, hitherinJ. only tohard thehestern renaissanHe does the aHHretinJ LnohledJe AeJin to Auild uU amomentum. the story of HryUtoloJy durinJ these years is, in other hords,eYaHtly the story of manLind. Hhina, the only hiJh HiKiliMation of antiBuity to use ideoJraUhiHhritinJ, seems neKer to haKe deKeloUed muHh real HryUtoJraUhy —UerhaUs for that reason. in one Hase Lnohn for military UurUoses, the11th-Hentury HomUilation, hu-HhinJ tsunJ-yao (”essentials from militaryHlassiHs”), reHommended a true if small Hode. to a list of 40 UlainteYtitems, ranJinJ from reBuests for Aohs and arrohs to the reUort of aKiHtory, the HorresUondents hould assiJn the first 40 ideoJrams of aUoem. then, hhen a lieutenant hished, for eYamUle, to reBuest morearrohs, he has to hrite the HorresUondinJ ideoJram at a sUeHified UlaHeon an ordinary disUatHh and stamU his seal on it.in Hhina’s Jreat neiJhAor to the hest, india, hhose HiKiliMationliLehise deKeloUed early and to hiJh estate, seKeral forms of seHretHommuniHations here Lnohn and, a UUarently, UraHtiHed. the artha-sastra, a HlassiH horL on stateHraft attriAuted to Lautilya, in desHriAinJthe esUionaJe serKiHe of india as UraHtiHally riddlinJ the Hountry hiths Uies, reHommended that the offiHers of the institutes of £ sUionaJe JiKetheir sUies their assiJnments Ay seHret hritinJ.UerhaUs most interestinJ to HryUtoloJists, amateur orUrofessional, is that Katsyayana’s famous teYtAooL of erotiHs, the Lamasutra,lists seHret hritinJ as one of the 64 arts, or yoJas, that homenshould Lnoh and UraHtiHe. the fourth Jreat HiKiliMation of antiBuity, themesoUo-tamian, rather Uaralleled eJyUt early in its HryUtoJraUhiHeKolution, Aut then surUassed it. thus, in the last Ueriod of HuneiformhritinJ, in HoloUhons hritten at uruL (in Uresent-day iraB) under theseleuHid LinJs in the last feh sHore years Aefore the Hhristian era,oHHasional sHriAes HonKerted their names into numAers. theenHiUherment—if suHh it Ae—may haKe Aeen only for amusement or toshoh off.“‘

Step 5: Completing the Alphabet

As more of the text became readable, the remaining letters were identified through context:

From "anythinJ" (clearly "anything"):

- **J** → **g**

From "UossiAle" (clearly "possible"):

- **U** → **p**
- **A** → **b**

From "haKe" (clearly "have"):

- **K** → **v**

From "Hryptology" and "Hontrast" (clearly "cryptology" and "contrast"):

- **H** → **c**

From "civiliMations" (clearly "civilizations"):

- **M** → **z**

From "eYample" (clearly "example"):

- **Y** → **x**

From "Ras" (clearly "was"):

- **R** → **w**

From "Lnowledge" (clearly "knowledge"):

- **L** → **k**

From "antiBuity" (clearly "antiquity"):

- **B** → **q**

Finally, from "Eust" (clearly "just"):

- **E** → **j**

Result:

“ the addition of secrecy to the transformations produced cryptography. true, it was more of a game than anything else—it sought to delay comprehension for only the shortest possible time, not the longest—and the cryptanalysis was, likewise, just a puzzle. egypt's was thus a quasi cryptology in contrast to the deadly serious science of today.yet great things have small beginnings, and these hieroglyphs did include, though in an imperfect fashion, the two elements of secrecy and transformation that comprise the essential attributes of the science. and so cryptology was born. in its first 3,000 years, it did not grow steadily. cryptology arose independently in many places, and in most of them it died the deaths of its civilizations. in other places, it survived, embedded in a literature, and from this the next generation could climb to higher levels.but progress was slow and jerky. more was lost than retained. much of the history of cryptology of this time is a patchwork, a crazy quilt of unrelated items, sprouting, flourishing, withering. only toward the western renaissance does the accreting knowledge begin to build up a momentum. the story of cryptology during these years is, in other words, exactly the story of mankind. china, the only high civilization of antiquity to use ideographic writing, seems never to have developed much real cryptography —perhaps for that reason. in one case known for military purposes, the11th-century compilation, wu-ching tsung-yao (”essentials from military classics”), recommended a true if small code. to a list of 40 plaintext items, ranging from requests for bows and arrows to the report of a victory, the correspondents would assign the first 40 ideograms of a poem. then, when a lieutenant wished, for example, to request more arrows, he was to write the corresponding ideogram at a specified place on an ordinary dispatch and stamp his seal on it.in china's great neighbor to the west, india, whose civilization likewise developed early and to high estate, several forms of secret communications were known and, apparently, practiced. the artha-sastra, a classic work on statecraft attributed to kautilya, in describing the espionage service of india as practically riddling the country withs pies, recommended that the officers of the institutes of £ spionage give their spies their assignments by secret writing.perhaps most interesting to cryptologists, amateur or professional, is that vatsyayana's famous textbook of erotics, the kamasutra, lists secret writing as one of the 64 arts, or yogas, that women should know and practice. the fourth great civilization of antiquity, the mesopo-tamian, rather paralleled egypt early in its cryptographic evolution, but then surpassed it. thus, in the last period of cuneiform writing, in colophons written at uruk (in present-day iraq) under these leucid kings in the last few score years before the christian era, occa-

sional scribes converted their names into numbers. the encipherment—if such it be—may have been only for amusement or to show off.“

4.3 Complete Key Reconstruction

Through iterative analysis of common words, bigrams, and trigrams, the complete substitution key was reconstructed:

Table 1: Reconstructed cipher alphabet

Cipher	A	B	C	D	E	F	G	H	I	J	K	L	M
Plain	b	q	f	u	j	y	n	c	r	g	v	k	z
Cipher	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Plain	o	d	s	h	w	l	a	p	e	t	i	x	m

4.4 Step-by-Step Verification

At each stage of decryption, I verified the substitutions by checking:

1. Whether the resulting words made sense in English
2. Whether the frequency patterns matched expected distributions
3. Whether bigrams and trigrams formed valid English combinations
4. Whether the overall message coherence improved

This iterative process ensured that each substitution was correct before proceeding to the next, minimizing the risk of cascading errors.

5 Results

5.1 Decrypted Plaintext

Applying the complete substitution key yielded the following plaintext:

> “The addition of secrecy to the transformations produced cryptography. True, it was more of a game than anything else—it sought to delay comprehension for only the shortest possible time, not the longest—and the cryptanalysis was, likewise, just a puzzle. Egypt’s was thus a quasi cryptology in contrast to the deadly serious science of today. Yet great things have small beginnings, and these hieroglyphs did include, though in an imperfect fashion, the two elements of secrecy and transformation that comprise the essential attributes of the science. And so cryptology was born. In its first 3,000 years, it did not grow steadily. Cryptology arose independently in many places, and in most of them it died the deaths of its civilizations.

In other places, it survived, embedded in a literature, and from this the next generation could climb to higher levels. But progress was slow and jerky. More was lost than retained. Much of the history of cryptology of this time is a patchwork, a crazy quilt of unrelated items, sprouting, flourishing, withering. Only toward the Western Renaissance does the accreting knowledge begin to build up a momentum. The story of cryptology during these years is, in other words, exactly the story of mankind. China, the only high civilization of antiquity to use ideographic writing, seems never to have developed much real cryptography—perhaps for that reason. In one case known for military purposes, the 11th-century compilation, Wu-ching tsung-yao ("Essentials from Military Classics"), recommended a true if small code. To a list of 40 plaintext items, ranging from requests for bows and arrows to the report of a victory, the correspondents would assign the first 40 ideograms of a poem. Then, when a lieutenant wished, for example, to request more arrows, he was to write the corresponding ideogram at a specified place on an ordinary dispatch and stamp his seal on it. In China's great neighbor to the west, India, whose civilization likewise developed early and to high estate, several forms of secret communications were known and, apparently, practiced. The Artha-sastra, a classic work on statecraft attributed to Kautilya, in describing the espionage service of India as practically riddling the country with spies, recommended that the officers of the institutes of espionage give their spies their assignments by secret writing. Perhaps most interesting to cryptologists, amateur or professional, is that Vatsyayana's famous textbook of erotics, the Kamasutra, lists secret writing as one of the 64 arts, or yogas, that women should know and practice. The fourth great civilization of antiquity, the Mesopotamian, rather paralleled Egypt early in its cryptographic evolution, but then surpassed it. Thus, in the last period of cuneiform writing, in colophons written at Uruk (in present-day Iraq) under the Seleucid kings in the last few score years before the Christian era, occasional scribes converted their names into numbers. The encipherment—if such it be—may have been only for amusement or to show off."

5.2 Verification

The decrypted text is coherent, grammatically correct, and discusses the historical development of cryptography from ancient Egypt through various civilizations including China, India, and Mesopotamia. The text explains how cryptology evolved slowly over 3,000 years, with different civilizations contributing to its development. The content makes logical sense, confirming the accuracy of the frequency analysis attack.

Key verification points:

- All common English words are properly formed
- Historical references (Wu-ching tsung-yao, Artha-sastra) are correctly spelled

- Proper nouns (China, India, Egypt, Mesopotamia, Uruk, Iraq) are accurate
- The narrative of the evolution of cryptology is coherent and historically plausible
- Grammar and sentence structure are correct throughout

6 Conclusion

The frequency analysis attack proved highly effective against this monoalphabetic substitution cipher. Key factors contributing to success included:

1. **Sufficient text length**: The ciphertext contained approximately 2,500 characters, providing statistically significant frequency data.
2. **Known language**: Knowledge that the plaintext was in English allowed application of standard frequency distributions.
3. **No additional obfuscation**: The cipher used only letter substitution with no word divisions or additional encoding.

This exercise demonstrates the fundamental weakness of monoalphabetic substitution: the preservation of statistical properties of the plaintext language. Despite the 26! (approximately 4×10^{26}) possible keys, a single intercepted message of reasonable length is sufficient to break the cipher.

The successful cryptanalysis confirms the observations made by Al-Kindi in the 9th century. The deterministic nature of the substitution means that frequency patterns, common words, and linguistic structures remain intact in the ciphertext, providing multiple avenues for attack.

Through systematic frequency analysis, the monoalphabetic substitution cipher (Variant 2) was successfully broken without prior knowledge of the key. The attack required approximately 45 minutes of manual analysis, demonstrating that even without computational tools, such ciphers provide minimal security against a determined cryptanalyst.

This exercise reinforces the importance of understanding both the strengths and limitations of cryptographic systems. While monoalphabetic substitution was revolutionary in its time, modern cryptography requires significantly more sophisticated approaches to provide adequate security.

To achieve greater security, historical cryptographers developed Polyalphabetic substitution (e.g., Vigenère cipher), Transposition ciphers, Homophonic substitution, and Modern block and stream ciphers with complex mathematical operations.

References

[1] Github, *https://github.com*

[2] Frequency Analysis Breaking the Code, *https://crypto.interactive-maths.com/frequency-analysis-breaking-the-code.html*