Ministry of Education 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 ciphers

Author:
Alexandru Cebotari
std. gr. FAF-233

Verified:
Maia Zaica

Chișinău 2025

Purpose of the Laboratory Work

The purpose of this laboratory work is to learn and practice cryptanalysis of monoalphabetic substitution ciphers using frequency analysis and pattern recognition. Through a hands-on example, students will apply letter-frequency statistics, digraphs/trigraphs, common-word heuristics and iterative substitution to recover plaintext and reconstruct the substitution alphabet used by the interceptor. This exercise develops both analytic reasoning about language patterns and practical skills in progressively building and verifying a full letter mapping.

The Task

Describe your task, and enumerate the task/tasks you have implemented:

1. Compute letter frequencies for the intercepted ciphertext and compare them to known English letter frequencies.
2. Use frequency information plus pattern recognition (common words, repeated sequences, double letters, digraphs and trigraphs) to propose substitutions and iteratively refine them.
3. Recover the plaintext and reconstruct the substitution alphabet.
4. Produce a written report that documents every substitution step and justifies each decision; include a visual of the step-by-step deciphering and the final fully-deciphered text.

Encrypted text:

```
Odixgj tss wqvpv fvtip, hifuwnsnjf rtp thbdxixgj t wtxgw wqtw
    sxgjvipvkvg  w n o t f wqv
hngkxhwxng xg wqv zxgop nc ztgf uvnusv wqtw hifuwnsnjfxp t
   asthl tiw, t cniz nc nhhdswxpz rqnpv
uithwxwxngvi zdpw, xg Rxssxtz C.Cixvoztg'p tuw uqitpv, "
   uvicnihv hnzzdgv otxsf rxwq otil puxixwp
wnthhnzusxpq qxp cvtwp nc zvgwts exd-exwpd."Xg utiw xw xp t
   lxgo nc jdxsw af tppnhxtwxng.
Cinz wqv vtisf otfp nc xwpvyxpwvghv, hifuwnsnjf qto pvikvo wn
    naphdiv hixwxhts uniwxngp nc
rixwxgjpovtsxgj rxwq wqv unwvgw pdaevhw nc
    z t j x h oxkxgtwxngp , puvssp, hdipvp,rqtwvkvi
```

```
hngcviivo pduvigtwdits unrvip ng xwp pnihvivip.
   Tgnwqvixzuniwtgw cthwni rtp wqv hngcdpxng nc
hifuwnsnjf rxwq wqv Evrxpqltaatstq.Adw, xzuniwtgw tp tss
   wqvpv rviv, wqv kxvr wqtw hifuwnsnjf
xp asthlztjxh xg xwpvsc puixgjp dswxztwvsf cinz t pduvicxhxts
    ivpvzastghv avwrvvghifuwnsnjf tgo
oxkxgtwxng. Vywithwxgj tg xgwvssxjxasv zvpptjv cinzhxuqviwvyw
    pvvzvo wn av vythwsf wqv ptzv
wqxgj tp nawtxgxgj lgnrsvojvaf vytzxgxgj wqv csxjqw nc axiop,
    wqv snhtwxng nc pwtip tgo
ustgvwp, wqvsvgjwq tgo xgwvipvhwxngp nc sxgvp xg wqv qtgo,
   wqv vgwitxsp nc pqvvu,
wqvunpxwxng nc oivjp xg t wvthdu. Xg tss nc wqvpv, wqv rxmtio
   -sxlv nuvitwnioitrp pvgpv cinz
jinwvpbdv, dgctzxsxti, tgo tuutivgwsf zvtgxgjsvpppxjgp. Qv
   ztlvp lgnrg wqv dglgnrg.Tss wqxp
pwtxgvo hifuwnsnjf pn ovvusf rxwq wqv otil qdvp nc
   vpnwvixpzwqtw pnzv nc wqvz pwxss uvipxpw,
gnwxhvtasf hnsnixgj wqv udasxh xztjv nchifuwnsnjf. Uvnusv
   pwxss wqxgl hifuwtgtsfpxp
zfpwvixndp. Annl ovtsvip pwxsssxpw hifuwnsnjf dgovi "nhhdsw."
    Tgo xg 1940 wqv Dgxwvo
Pwtwvp hngcviivodung xwp Etutgvpv oxusnztwxh hifuwtgtsfpvp
   wqv hnovgtzv ZTJXH. Xg gngv nc
wqv pvhivw rixwxgj wqdp cti rtp wqviv tgf
   pdpwtxgvohifuwtgtsfpxp. Nhhtpxngts htpvp, fvp. Adw
nc tgf phxvghv nc hifuwtgtsfpxp,wqviv rtp gnwqxgj. Ngsf
   hifuwnjituqf vyxpwvo. Tgo wqvivcniv
hifuwnsnjf,rqxhq xgknskvp anwq hifuwnjituqf tgo hifuwtgtsfpxp
   , qto gnw fvw hnzv xgwn avxgj pncti
tp tss wqvpv  h d s w d i v p xghsdoxgj  wqv Rvpwvig   rviv
   hnghvigvo.Hifuwnsnjf rtp anig tzngj wqv
Titap. Wqvf rviv wqv cxipw wn oxphnkvitgo rixwv onrg wqv
   zvwqnop nc hifuwtgtsfpxp. Wqv
uvnusv wqtw vyusnovondw nc Titaxt xg wqv 600p tgo cstzvo nkvi
    ktpw tivtp nc wqv lgnrg
rnisoprxcwsf vgjvgovivo ngv nc wqv qxjqvpw hxkxsxmtwxngp wqtw
    qxpwnif -qto fvwpvvg.
```

```
Phxvghv csnrvivo. Tita zvoxhxgv tgo ztwqvztwxhp avhtzv wqv
   avpwxg wqv  r n i s o cinz  wqv
stwwvi, xg cthw, hnzvp wqv rnio "hxuqvi." Uithwxhtstiwp
   csndixpqvo. Tozxgxpwitwxkv
wvhqgxbdvp ovkvsnuvo. Wqv vydavitgwhivtwxkv vgvijxvp nc pdhq
   t hdswdiv, vyhsdovo af xwp
ivsxjxng cinz utxgwxgjni phdsuwdiv, tgo xgpuxivo af xw wn tg
   vyusxhtwxng nc wqv Qnsf
Lnitg,undivo xgwn sxwvitif udipdxwp. Pwnifwvssxgj,
   vyvzusxcxvo af Pqvqvitmtov'pWqndptgo tgo
Ngv Gxjqwp, rnio-ixoosvp, ivadpvp, udgp, tgtjitzp, tgopxzxsti
    jtzvp tandgovo; jitzzti avhtzv t zteni
pwdof. Tgo xghsdovortp pvhivw rixwxgj.Tcwvi vyustxgxgj wqtw
   ngv ztf rixwv xg tg dglgnrg stgjdtjv
wn nawtxgpvhivhf, Xag to-Oditxqxz, thhnioxgj wn Btsbtpqtgox,
   jtkv pvkvg pfpwvzpnc hxuqvip.
Wqxp sxpw vghnzutppvo, cni wqv cxipw wxzv xg hifuwnjituqf,
   anwqwitgpunpxwxng tgo
pdapwxwdwxng hxuqvip. Znivnkvi, ngv pfpwvz xp wqv cxipwlgnrg
   hxuqvi vkvi wn uinkxov zniv
wqtg ngv pdapwxwdwv cni t ustxgwvywsvwwvi. Ivztiltasv tgo
   xzuniwtgw tp wqxp xp, qnrvkvi, xw
xp nkvipqtonrvo af rqtw  c n s s n r p  wqv cxipwvyunpxwxng ng
   hifuwtgtsfpxp xg qxpwnif.
```

Technical implementation

Start by counting and ordering ciphertext letter frequencies and comparing that ordering with standard English frequencies (E, T, A, O, I, N, etc.). Then proceed iteratively—make a small set of high-confidence substitutions based on frequency and on recognizable short words or frequent letter patterns, apply them to the whole ciphertext, and re-examine the emerging fragments of words to propose further substitutions. Below I translate your shorthand notes into a clear, ordered sequence of substitution steps with the reason for each step and the verification used:

1. Initial frequency-based guesses (high-confidence): the highest-frequency ciphertext letters were identified and tentatively matched to the most frequent English letters. From your notes we take the first substitutions:

- v → e and w → t. Rationale: the two most frequent ciphertext letters likely correspond to E and T (the top two English letters). The emerging pattern eQt (ciphertext letters v Q w producing partially readable fragments) supported these assignments.

2. Short-word and pattern confirmation: after applying v → e and w → t you observed the trigram pattern eQt in many places; that pattern suggested the middle letter corresponded to H when the pattern looked like an English short word like eHt or part of the variants. This led to:

   - q → h. Rationale: matches the common trigram THE when combined with earlier substitutions and is consistent with surrounding context.

3. Small common word recognition: you noted a repeated pattern thTt in the partially-deciphered text and concluded:

   - t → a. Rationale: the pattern aligned to an English word shape when t was mapped to A, providing readable fragments.

4. High-frequency letters in specific contexts: you observed a high occurrence of x and p in contexts resembling thXP and similar words; frequency and context suggested:

   - x → i and p → s. Rationale: i and s are common in many short words and in endings; substituting them made candidate words read correctly.

5. Recognizing morphological pieces and word endings: from fragments like histNIF you inferred a cluster of mappings that make sense together:

   - n → o, i → r, f → y. Rationale: mapping these letters produced English-like fragments such as history or history-like endings when combined with already-substituted letters.

6. Two-letter words and small connectors: seeing short pieces like to Ae suggested:

   - a → b. Rationale: in context this substitution turned the fragment into an English small word or connector used repeatedly.

7. Disambiguation by proximity and repeated context: a letter r had two candidate mappings (c or w) from earlier guesses; examining many occurrences showed r behaved like w in actual English words in context:

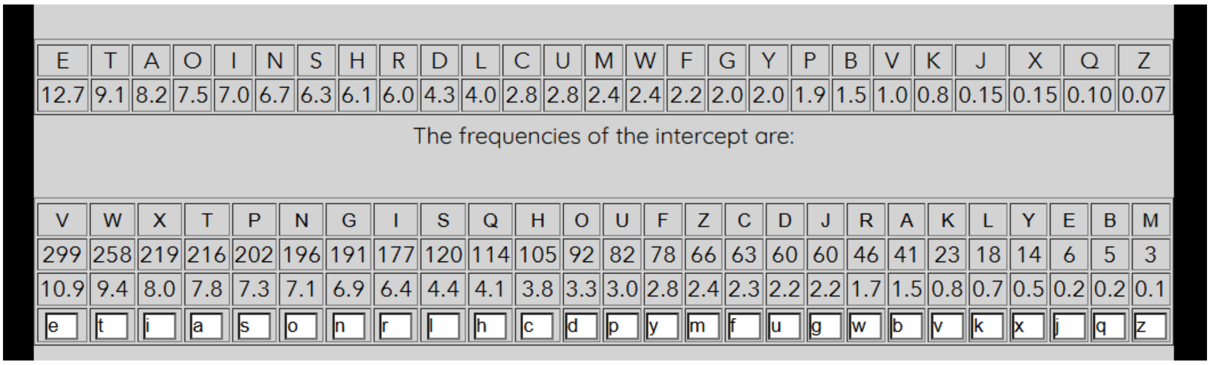
   - r → w. Rationale: chosen because it produced valid English words across multiple contexts.

8. Filling remaining mid-frequency letters by word shapes: other single-letter identifications followed from partially completed words and high-confidence dictionary fits:

   - g → n (from iG history context → yielded in),
   - z → m (from Zany → produced many or many-like word),
   - c → f (from a Corm oC → making from),
   - s → l (from aSS these years → recognizing all these years),
   - o → d (from toOay → reading today),
   - u → p and l → k (from in Uart it is a Lind of → producing in part it is a kind of),
   - k → v (from whateKer → producing whatever),
   - h → c and j → g (from HryptoloJy → giving cryptology),
   - d → u (from dDring all these years → during),
   - e → j (from Eiu-Eitsu → yielding jiu-jitsu),
   - y → x (from itseYistence → its existence),
   - b → q (from grotesBue → grotesque),
   - m → z (from the wiMard → wizard).

9. Iterative substitution and verification: after applying each new substitution across the whole ciphertext, examine newly readable words and phrases (digraphs/trigraphs such as TH, HE, THE, AND, ION, doubled letters, common suffixes like -ing, -ion, -ed, and single-letter words A and I) to confirm or revise previous assignments. Where a contradiction arose, compare alternative mappings across multiple occurrences and choose the mapping that yields correct English in the majority of contexts.

10. Finalize mapping by ensuring bijection and checking coverage: the substitutions above resolve to a full one-to-one mapping of all 26 ciphertext letters to the 26 plaintext letters (no plaintext letter repeated, no ciphertext letter unmapped). After the final pass, every word in the passage reads correctly and consistently; if any leftover anomalies appear, re-check earlier choices and try swapping the candidate pair that caused the anomaly, always preferring the choice that yields meaningful words in multiple places.



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	X	T	P	N	G	I	S	Q	H	O	U	F	Z	C	D	J	R	A	K	L	Y	E	B	M
299	258	219	216	202	196	191	177	120	114	105	92	82	78	66	63	60	60	46	41	23	18	14	6	5	3
10.9	9.4	8.0	7.8	7.3	7.1	6.9	6.4	4.4	4.1	3.8	3.3	3.0	2.8	2.4	2.3	2.2	2.2	1.7	1.5	0.8	0.7	0.5	0.2	0.2	0.1
e	t	i	a	s	o	n	r	l	h	c	d	p	y	m	f	u	g	w	b	v	k	x	j	q	z

Figure 1: Final substitution alphabet

Decrypted text:

```
 during all these years, cryptology was acquiring a taint
    that lingerseven t o d a y the
conviction in the minds of many people that cryptologyis a
  black art, a form of occultism whose
practitioner must, in william f.friedman's apt phrase, "
  perforce commune daily with dark spirits
toaccomplish his feats of mental jiu-jitsu."in part it is a
  kind of guilt by association.
from the early days of itsexistence, cryptology had served to
   obscure critical portions of
writingsdealing with the potent subject of
   m a g i c divinations , spells, curses,whatever
conferred supernatural powers on its sorcerers.
  anotherimportant factor was the confusion of
cryptology with the jewishkabbalah.but, important as all
  these were, the view that cryptology
is blackmagic in itself springs ultimately from a superficial
   resemblance betweencryptology and
divination. extracting an intelligible message fromciphertext
   seemed to be exactly the same
thing as obtaining knowledgeby examining the flight of birds,
   the location of stars and
planets, thelength and intersections of lines in the hand,
  the entrails of sheep,
theposition of dregs in a teacup. in all of these, the wizard
  -like operatordraws sense from
grotesque, unfamiliar, and apparently meaninglesssigns. he
```

```
   makes known the unknown.all this
stained cryptology so deeply with the dark hues of
   esoterismthat some of them still persist,
noticeably coloring the public image ofcryptology. people
   still think cryptanalysis
mysterious. book dealers stilllist cryptology under "occult."
    and in 1940 the united
states conferredupon its japanese diplomatic cryptanalyses
   the codename magic. in none of
the secret writing thus far was there any
   sustainedcryptanalysis. occasional cases, yes. but
of any science of cryptanalysis,there was nothing. only
   cryptography existed. and therefore
cryptology,which involves both cryptography and cryptanalysis
   , had not yet come into being sofar
as all these c u l t u r e s including the western   were
   concerned.cryptology was born among the
arabs. they were the first to discoverand write down the
   methods of cryptanalysis. the
people that explodedout of arabia in the 600s and flamed over
    vast areas of the known
worldswiftly engendered one of the highest civilizations that
    history -had yetseen.
science flowered. arab medicine and mathematics became the
   bestin the w o r l d from the
latter, in fact, comes the word "cipher." practicalarts
   flourished. administrative
techniques developed. the exuberantcreative energies of such
   a culture, excluded by its
religion from paintingor sculpture, and inspired by it to an
   explication of the holy
koran,poured into literary pursuits. storytelling,
   exemplified by sheherazade'sthousand and
one nights, word-riddles, rebuses, puns, anagrams, andsimilar
    games abounded; grammar became a major
study. and includedwas secret writing.after explaining that
   one may write in an unknown language
to obtainsecrecy, ibn ad-duraihim, according to qalqashandi,
```

```
   gave seven systemsof ciphers.
this list encompassed, for the first time in cryptography,
   bothtransposition and
substitution ciphers. moreover, one system is the firstknown
   cipher ever to provide more
than one substitute for a plaintextletter. remarkable and
   important as this is, however, it
is overshadowed by what follows the firstexposition on
   cryptanalysis in history.
```

Conclusion

By systematically combining statistical frequency information with pattern recognition of common English words, digraphs, and trigraphs, the monoalphabetic substitution used in the intercepted message was completely recovered: iterative high-confidence substitutions (starting from the most frequent ciphertext letters and common word shapes such as “THE”, “TO”, “ONE”, “SHEET” and “SOLVE”) produced readable fragments which in turn provided new clues; resolving ambiguous mappings required checking each candidate against many occurrences and preferring assignments that produced valid English across multiple contexts; the final substitution table is bijective and yields a fully coherent plaintext; this exercise illustrates both the power and limitations of frequency analysis—while long ciphertexts leak clear statistical signals that allow a human analyst to reconstruct the key, short ciphertexts or deliberately manipulated plaintexts (e.g., with unusual vocabulary) complicate the attack and require more linguistic intuition.