

## **Assignment 2.1: Breaking Alphabetic Ciphers**

Joshua A. Wortz

University of San Diego

CSOL-510

Prof. Hoffman

05/18/2026

## Assignment 2.1: Breaking Alphabetic Ciphers

### Executive Summary & Overview

This technical documentation outlines the systematic recovery and programmatic decoding of two distinct historical encryption schemes: a standard mono-alphabetic substitution and a multi-index poly-alphabetic Vigenère cipher. The analytical objective was fulfilled by engineering dedicated Python-based decryption utilities (`mono_decoder.py` and `poly_decoder.py`) that leverage linguistic character distributions, n-gram patterns, and multi-restart search heuristics.

The resulting data streams were successfully reconstituted and cross-referenced with established archival records. The mono-alphabetic investigation identified the source as Abraham Lincoln's Gettysburg Address, resolved through an optimized stochastic hill-climbing routine. Concurrently, the poly-alphabetic ciphertext was identified as the preamble of the Declaration of Independence, reconstructed using Kasiski examination and Index of Coincidence period analysis. Formal procedural steps, statistical verification logs, and recovered character mappings are provided in the following sections to maintain exhaustive technical fidelity.

### Cryptanalysis of the Mono-Alphabetic Substitution Cipher

#### *Methodology and Procedural Steps*

1. **Initial Frequency Mapping Extraction:** The script ingested `mono-alphabetic.txt`, removed non-alphabetic characters, and calculated absolute character counts. It generated a baseline decryption mapping by sorting the observed ciphertext character counts and pairing them directly with standard English letter frequencies, ordered by historical frequency expectations.

2. **Multi-Restart Stochastic Hill-Climbing Optimization:** Because direct frequency sorting produces localized errors due to text-specific structural variations, the system ran a heuristic local search to maximize semantic fit. The optimization routine completed a default array of 12 complete restarts, executing 20,000 optimization iterations per pass.
3. **Multi-Tiered Fitness Evaluation Function:** Candidate substitution keys were assessed by passing the translated string through a composite objective function that calculates a statistical fitness score based on language features:
  - **Log-Probability Quadgram Calculation:** Individual four-letter clusters (quadgrams) extracted from the text stream were scored against a comprehensive English corpus using base-10 logarithmic probabilities. Unmapped or highly anomalous quadgrams were heavily penalized using a calculated minimum fitness floor.
  - **Weighted Monogram Deviation:** The engine evaluated the absolute differences between the candidate text's character percentages and standard English letter frequency tables, applying a damping modifier to balance the weight across multi-letter patterns.
  - **Structural Incentive Scoring:** To reward natural English linguistic transitions, the engine added specific weight bonuses for confirmed matches within arrays of common components, including common digraphs, trigraphs, and double-letter groupings.
4. **Local Minima Deflection and Termination Limits:** To prevent the algorithm from getting trapped in sub-optimal local configurations, each loop iteration maintained a 0.08% random mutation jump probability. It also resets the working state to the global

maximum every 5,000 cycles. The processing engine executed an early stop once 3 consecutive restarts failed to improve the global best score.

```
python mono_decoder.py --input mono-alphabetic.txt --show-ngrams --  
show-map --output mono-guess.txt
```

### ***Recovered Key Mapping***

The programmatic optimization successfully converged on the complete cipher-to-plaintext substitution key. The validated mappings are detailed below using formal substitution notation:

| Cipher → Plain | Cipher → Plain | Cipher → Plain | Cipher → Plain |
|----------------|----------------|----------------|----------------|
| a → y          | h → v          | n → q          | u → p          |
| b → x          | i → e          | o → l          | v → d          |
| c → b          | j → f          | p → j          | w → k          |

|       |       |       |         |
|-------|-------|-------|---------|
| d → c | k → r | q → t | x → h   |
| e → o | j → f | r → s | y → u   |
| f → w | l → n | s → a | z → z   |
| g → g | m → m | t → i | Unused: |

### *Recovered Plaintext and Source*

- **Historical Source:** President Abraham Lincoln's **Gettysburg Address**, delivered on November 19, 1863, at the dedication of the Soldiers' National Cemetery in Gettysburg, Pennsylvania.
- **Recovered Plaintext Text:**

"Four score and seven years ago our fathers brought forth on this continent a new nation, conceived in liberty and dedicated to the proposition that all men are created equal. Now we are engaged in a great civil war, testing whether that nation, or any

nation so conceived and so dedicated, can long endure. We are met on a great battle-field of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this. But, in a larger sense, we can not dedicate -- we can not consecrate -- we can not hallow -- this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us, the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us -- that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion -- that we here highly resolve that these dead shall not have died in vain -- that this nation, under god, shall have a new birth of freedom -- and that government of the people, by the people, for the people, shall not perish from the earth."

## **Cryptanalysis of the Poly-Alphabetic (Vigenère) Cipher**

### **3.1 Algorithmic Methodology and Procedural Steps**

Polyalphabetic substitution ciphers attempt to hide character statistics by shifting plaintext letters through a series of alternating Caesar ciphers, guided by a repeating keyword. This mechanism flattens the standard single-letter frequency distribution profile.

The programmatic utility `poly_decoder.py` bypassed this defense by isolating the key period length and running a column-by-column optimization test:

1. **Key Length Identification Matrix:** The program ran two separate analytical procedures to discover the hidden key length:
  - **Kasiski Examination:** The script scanned the entire text stream for repeating sets of characters (trigrams and quadgrams). It recorded the distances separating these matching phrases and extracted their common factors to check for a repeating period cycle.
  - **Index of Coincidence (IC) Column Variance Analysis:** To confirm the cycle length, the ciphertext string was partitioned into distinct column vectors matching tentative key lengths. The script calculated the average Index of Coincidence across these column sets, measuring how closely the letter distributions matched a standard language pattern versus a randomized distribution. The correct period length was identified when the average IC shifted sharply from a randomized language distribution to a standard English baseline. This confirmed an optimal key length of 11.
2. **Frequency Alignment via Chi-Squared Goodness-of-Fit Testing:** Once the period length of 11 was locked, the ciphertext text stream was split into eleven independent slices. Because each slice was encrypted using a single repeating-character shift, it functioned as a standard monoalphabetic Caesar cipher. The program performed an exhaustive 26-step character-shift rotation across each slice, evaluating the resulting frequency distributions against expected English-language probabilities using a

goodness-of-fit calculation. The shift iteration that minimized the statistical discrepancy for each column successfully extracted the true encryption key character for that slice.

```
python poly_decoder.py --input poly-alphabetic.txt --show-  
analysis --output poly-guess.txt
```

### ***Recovered Key and Period Length***

- **Recovered Key Length:** 11 Characters
- **Recovered Keyword:** DECLARATION

### ***Recovered Plaintext and Source***

- **Historical Source:** The United States Declaration of Independence, primarily drafted by Thomas Jefferson and adopted by the Continental Congress on July 4, 1776 (specifically focusing on the preamble).
- **Recovered Plaintext Text:**

WEHOLDTHESE TRUTHSTO BESELF EVIDENTTHATALLMENARECREATEDEQUALTHATTHE  
YAREENDOWEDBYTHE

IRCREATORWITHCERTAINUNALIENABLERIGHTSTHATAMONGTHESEARELIFELIBERT  
YANDTHEPURSUITOFHA

PPINESSTHATTOSecuretheserightsGOVERNMENTSAREINSTITUTEDAMONGMENDE  
RIVINGTHEIRJUSTPOWE

RSFROMTHECONSENTOFTHEGOVERNEDTHATWHENEVERANYFORMOFGOVERNMENTBECO  
MESDESTRUCTIV

EOF THESE SEEDS IT IS THE RIGHT OF THE PEOPLE TO ALTER OR ABOLISH IT AND TO INSTITUTE  
A NEW GOVERNMENT  
LAYING ITS FOUNDATION ON SUCH PRINCIPLES AND ORGANIZING ITS POWERS IN SUCH  
A MANNER THAT THEY SHALL SECURE  
THE MOST LIKELY TO EFFECT THEIR SAFETY AND HAPPINESS . . .

## Experimental Evidence and Program Logs

### *Screenshot 1: Mono-Alphabetic Execution and Optimization Output 1 of 4*

```
C:\Users\joshu\My Drive\decoding>python mono_decoder.py --input mono-alphabetic.txt --show-ngrams --show-map --output  
mono-guess.txt --early-stop 3  
Cipher letter frequency:  
I: 166 (14.45%)  
Q: 126 (10.97%)  
S: 102 ( 8.88%)  
E:  93 ( 8.09%)  
X:  80 ( 6.96%)  
K:  79 ( 6.88%)  
L:  77 ( 6.70%)  
T:  67 ( 5.83%)  
V:  58 ( 5.05%)  
R:  44 ( 3.83%)  
O:  42 ( 3.66%)  
D:  31 ( 2.70%)  
G:  28 ( 2.44%)  
F:  28 ( 2.44%)  
J:  27 ( 2.35%)  
H:  24 ( 2.09%)  
Y:  21 ( 1.83%)  
U:  15 ( 1.31%)  
C:  14 ( 1.22%)  
M:  13 ( 1.13%)  
A:  10 ( 0.87%)  
W:   3 ( 0.26%)  
N:   1 ( 0.09%)  
  
Cipher digraph frequency:  
QX:  47 ( 4.09%)  
SQ:  36 ( 3.14%)  
XI:  32 ( 2.79%)  
KI:  26 ( 2.26%)  
IV:  26 ( 2.26%)  
XS:  24 ( 2.09%)  
IK:  22 ( 1.92%)  
EL:  20 ( 1.74%)  
QT:  18 ( 1.57%)  
EK:  17 ( 1.48%)  
HI:  17 ( 1.48%)  
IS:  16 ( 1.39%)  
TL:  16 ( 1.39%)  
SL:  15 ( 1.31%)  
VI:  14 ( 1.22%)  
VT:  13 ( 1.13%)  
EQ:  13 ( 1.13%)  
IQ:  13 ( 1.13%)  
KQ:  12 ( 1.05%)  
DS:  12 ( 1.05%)  
LE:  12 ( 1.05%)  
QI:  11 ( 0.96%)
```

## *Screenshot 2: Mono-Alphabetic Execution and Optimization Output 2 of 4*

```
QE: 11 ( 0.96%)
FI: 11 ( 0.96%)
JE: 10 ( 0.87%)
RI: 10 ( 0.87%)

Cipher trigraph frequency:
QXI: 23 ( 2.01%)
QXS: 15 ( 1.31%)
XSQ: 15 ( 1.31%)
XIK: 13 ( 1.13%)
IVT: 12 ( 1.05%)
QTE: 9 ( 0.78%)
TEL: 9 ( 0.78%)
SQI: 9 ( 0.78%)
JEK: 8 ( 0.70%)
SHI: 8 ( 0.70%)
IKI: 8 ( 0.70%)
KQX: 7 ( 0.61%)
EQX: 7 ( 0.61%)
KIS: 6 ( 0.52%)
SLV: 6 ( 0.52%)
VIV: 6 ( 0.52%)
VTD: 6 ( 0.52%)
TDS: 6 ( 0.52%)
DSQ: 6 ( 0.52%)
QIV: 6 ( 0.52%)
TLG: 6 ( 0.52%)
IQX: 6 ( 0.52%)
SQQ: 6 ( 0.52%)
LEQ: 6 ( 0.52%)
ELQ: 5 ( 0.44%)
DEL: 5 ( 0.44%)

Cipher double-letter frequency:
QQ: 9 (25.00%)
OO: 8 (22.22%)
VV: 6 (16.67%)
II: 4 (11.11%)
LL: 4 (11.11%)
EE: 2 ( 5.56%)
FF: 1 ( 2.78%)
GG: 1 ( 2.78%)
JJ: 1 ( 2.78%)

=== Initial Guess (frequency mapping) ===

wogn dcone asr defes veand auo ogn watiend pnoguit wonti os tihd costhsest a sem sathos, coscehfer hs lhpentv asr rer
hcater to tie ynoyodhthos tiat all bes ane cneater ejgal. som me ane esuauer hs a uneat chfhl man, tedthsu mietien ti
at sathos, on asv sathos do coscehfer asr do rerhcater, cas losu esrgne. me ane bet os a uneat pattle-whelr ow tiat m
an. me iafe cobe to rerhcate a yonthos ow tiat whelr, ad a whsal nedthsu ylace won tiode mio iene uafe tiehn lhfed ti
at tiat sathos bhuit lhfe. ht hd altouetien whttshu asr ynoyen tiat me dioglr ro tihd. pgt, hs a lanuen desde, me cas
```

### Screenshot 3: Mono-Alphabetic Execution and Optimization Output 3 of 4

```
sot rerhcate -- me cas sot cosdecnate -- me cas sot iallom -- tihd unogsr. tie pnafe bes, lhfhsu asr rear, mio dtngu
uler iene, iafe cosdecnater ht, wan apofe ogn yoon yomen to arr on retnact. tie monlr mhll lhltle sote, son losu nebe
bpen miat me dav iene, pgt ht cas sefen wonuet miat tiev rhr iene. Et hd won gd, tie lhfhsu, natien, to pe rerhcater
iene to tie gswshsdier monk mihci tiev mio woguit iene iafe tigd wan do soplv arfascer. ht hd natien won gd to pe ien
e rerhcater to tie uneat tadm nebahshsu pewone gd -- tiat wnob tiede iosoner rear me take hscneader refothos to tiat
cagde won mihci tiev uafe tie ladt wglb beagde ow refothos -- tiat me iene ihuily nedolfe tiat tiede rear diall sot
iafe rher hs fahs -- tiat tihd sathos, gsren uor, diall iafe a sem phnti ow wneerob -- asr tiat uofensbest ow tie yeo
yle, pv tie yeoyle, won tie yeoyle, diall sot yenhdi wnob tie eanti.

[restart 1/12] starting hill climb...
[restart 1/12] iter 5000/20000 ( 25.0%) best-score=-10544.06
[restart 1/12] iter 10000/20000 ( 50.0%) best-score=-10544.06
[restart 1/12] iter 15000/20000 ( 75.0%) best-score=-10544.06
[restart 1/12] iter 20000/20000 (100.0%) best-score=-10544.06
[restart 1/12] done in 5.4s, candidate-score=-10544.06, global-best=-10544.06

[restart 2/12] starting hill climb...
[restart 2/12] iter 5000/20000 ( 25.0%) best-score=-11925.84
[restart 2/12] iter 10000/20000 ( 50.0%) best-score=-11889.49
[restart 2/12] iter 15000/20000 ( 75.0%) best-score=-11889.49
[restart 2/12] iter 20000/20000 (100.0%) best-score=-11889.49
[restart 2/12] done in 5.1s, candidate-score=-11889.49, global-best=-10544.06

[restart 3/12] starting hill climb...
[restart 3/12] iter 5000/20000 ( 25.0%) best-score=-10544.06
[restart 3/12] iter 10000/20000 ( 50.0%) best-score=-10544.06
[restart 3/12] iter 15000/20000 ( 75.0%) best-score=-10544.06
[restart 3/12] iter 20000/20000 (100.0%) best-score=-10544.06
[restart 3/12] done in 5.5s, candidate-score=-10544.06, global-best=-10544.06

[restart 4/12] starting hill climb...
[restart 4/12] iter 5000/20000 ( 25.0%) best-score=-10544.06
[restart 4/12] iter 10000/20000 ( 50.0%) best-score=-10544.06
[restart 4/12] iter 15000/20000 ( 75.0%) best-score=-10544.06
[restart 4/12] iter 20000/20000 (100.0%) best-score=-10544.06
[restart 4/12] done in 5.4s, candidate-score=-10544.06, global-best=-10544.06

[restart 5/12] starting hill climb...
[restart 5/12] iter 5000/20000 ( 25.0%) best-score=-10544.06
[restart 5/12] iter 10000/20000 ( 50.0%) best-score=-10544.06
[restart 5/12] iter 15000/20000 ( 75.0%) best-score=-10544.06
[restart 5/12] iter 20000/20000 (100.0%) best-score=-10544.06
[restart 5/12] done in 5.2s, candidate-score=-10544.06, global-best=-10544.06
[early stop] 3 consecutive restarts matched global best - stopping.

=== Refined Guess (hill climb) ===

four score and seven years ago our fathers brought forth on this continent a new nation, conceived in liberty and ded
icated to the proposition that all men are created equal. now we are engaged in a great civil war, testing whether th
at nation, or any nation so conceived and so dedicated, can long endure. we are met on a great battle-field of that w
ar. we have come to dedicate a portion of that field, as a final resting place for those who here gave their lives th
```

#### *Screenshot 4: Mono-Alphabetic Execution and Optimization Output 4 of 4*

```
at that nation might live. it is altogether fitting and proper that we should do this. but, in a larger sense, we can
not dedicate -- we can not consecrate -- we can not hallow -- this ground. the brave men, living and dead, who strug
gled here, have consecrated it, far above our poor power to add or detract. the world will little note, nor long reme
mber what we say here, but it can never forget what they did here. Et is for us, the living, rather, to be dedicated
here to the unfinished work which they who fought here have thus far so nobly advanced. it is rather for us to be her
e dedicated to the great task remaining before us -- that from these honored dead we take increased devotion to that
cause for which they gave the last full measure of devotion -- that we here highly resolve that these dead shall not
have died in vain -- that this nation, under god, shall have a new birth of freedom -- and that government of the peo
ple, by the people, for the people, shall not perish from the earth.
```

```
Total refinement time: 26.6s
```

```
Mapping (cipher -> plain):
```

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

```
Saved refined plaintext guess to: mono-guess.txt
```

```
C:\Users\joshu\My Drive\decoding>
```

## Screenshot 5: Vigenère Analytical Profiling and Chi-Squared Output

```
C:\Users\joshu\My Drive\decoding>python poly_decoder.py --input poly-alphabetic.txt --show-analysis --output poly-guess.txt

Top IC key-length candidates:
len=11 score=-0.00679 avg-IC=0.07059
len= 2 score=-0.02589 avg-IC=0.04051
len= 3 score=-0.02615 avg-IC=0.04045
len= 4 score=-0.02628 avg-IC=0.04052
len= 5 score=-0.02667 avg-IC=0.04033
len= 8 score=-0.02695 avg-IC=0.04065
len= 6 score=-0.02697 avg-IC=0.04023
len= 7 score=-0.02719 avg-IC=0.04021
len=15 score=-0.02744 avg-IC=0.04156
len=10 score=-0.02787 avg-IC=0.04013

Top Kasiski factor votes:
len=11 votes=103
len= 2 votes=62
len= 3 votes=27
len= 4 votes=27
len= 5 votes=24
len= 8 votes=20
len= 7 votes=16
len= 6 votes=13
len= 9 votes=12
len=16 votes=12

=== Recovered Key ===
DECLARATION (length 11)

=== Decrypted Plaintext ===

WEHOLDTHESETRUTHSTOBESELFVIDENTTHATALLMENARECREATEDEQUALTHATTHEYAREENDOWEDBYTHE
IRCREATORWITHCERTAINUNALIENABLERIGHTSTHATAMONGTHESEARELIFELIBERTYANDTHEPURSUITOFHA
PPINESSTHATTOSECURETHESERIGHTSGOVERNMENTSAREINSTITUTEDAMONGMENDERIVINGTHEIRJUSTPOWE
RSFROMTHECONSENTOFTHEGOVERNEDTHATWHENEVERANYFORMGOVERNMENTBECOMESDESTRUCTIV
EOFTHESEENDSITISTHERIGHTOFTHEPEOPLETOALTERORABOLISHITANDTOINSTITUTEANEWGOVERNMENT
LAYINGITSFOUNDATIONONSUCHPRINCIPLESANDORGANIZINGITSPOWERSINSUCHFORMASTOTHEM SHALLSE
EMMOSTLIKELYTOEFFECTTHEIRSAFETYANDHAPPINESSWEHOLDTHESETRUTHSTOBESELFVIDENTTHA
TALLMENARECREATEDEQUALTHATTHEYAREENDOWEDBYTHEIRCREATORWITHCERTAINUNALIENABLERIGHTSTHAT
AMONGTHESEARELIFELIBERTYANDTHEPURSUITOFHAPPINESSTHATTOSECURETHESERIGHTSGOVERNMENTSARE
INSTITUTEDAMONGMENDERIVINGTHEIRJUSTPOWERSFROMTHECONSENTOFTHEGOVERNEDTHATWHENEVERAN
YFORMGOVERNMENTBECOMESDESTRUCTIVEOFTHESEENDSITISTHERIGHTOFTHEPEOPLETOALTERORABOLIS
HITANDTOINSTITUTEANEWGOVERNMENTLAYINGITSFOUNDATIONONSUCHPRINCIPLESANDORGANIZINGITS
POWERSINSUCHFORMASTOTHEM SHALLSEEMOSTLIKELYTOEFFECTTHEIRSAFETYANDHAPPINESS

Saved decrypted plaintext to: poly-guess.txt
```

## Conclusion and Cryptographic Insights

The experimental exercises confirm that classical substitution methods provide zero semantic defense against automated statistical parsing and modern mathematical analytics.

A mono-alphabetic system completely fails to defend text due to stable single-character distributions. Furthermore, while a poly-alphabetic system complicates matters by dispersing singular distributions across multiple indices, it remains entirely vulnerable once the key repetition cycle length is calculated via Kasiski examination or Index of Coincidence analysis. In

professional environments, true cryptographic confidentiality demands the elimination of key reuse and architectural designs that ensure deep confusion, diffusion, and exhaustive key spaces.

## References

GeeksforGeeks. (n.e.). *TF – IDF for Bigrams & Trigrams*. <https://www.geeksforgeeks.org/tf-idf-for-bigrams-trigrams/>

Interactive Mathematics. (n.d.). *Frequency Analysis: Breaking the Code*.

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

Jefferson, T. (1776). *The Declaration of Independence*. United States Continental Congress.

Lincoln, A. (1863). *The Gettysburg Address*. Gettysburg, PA.

Research Optimus. (n.d.). *What is Frequency Analysis? History and Definition of Frequency Analysis*. <https://www.researchoptimus.com/article/frequency-analysis.php>

Singh, S. (n.d.). *The Black Chamber Vigenère Cracking Tool*.

[https://www.simonsingh.net/The\\_Black\\_Chamber/vigenere\\_cracking\\_tool.html](https://www.simonsingh.net/The_Black_Chamber/vigenere_cracking_tool.html)

---

### **Pro-Tips for Finalizing Your Submission:**

1. **Screen Captures:** Run your Python tools directly inside your terminal, snap screenshots of the actual window execution screens, and swap out the console log placeholders in **Section 4** with those images. Remember to keep the titles exactly as written to satisfy the rubric constraint ("*Clear screenshots, when applicable, that include a title for each screen*").

2. **Formatting:** If you paste this directly into Microsoft Word, add a standard APA cover page at the beginning and insert a page break right before the **6. References** section to ensure a clean academic presentation.