Cryptanalysis of Classical Ciphers

Lorenzo Deflorian

February 14, 2026

Contents

1 Introduction

This report describes the cryptanalysis of an unknown ciphertext. The goal was to identify the encryption method, recover the encryption key (if applicable), and decrypt the message without any prior knowledge of the cipher type or key.

The analysis used classical cryptanalytic techniques including:

- Index of Coincidence (IC) analysis
- Frequency analysis
- Pattern analysis (Kasiski examination)
- Statistical tests (Friedman test, chi-squared test)

2 Cryptanalysis of Vigenère Cipher (Cipher1)

2.1 Initial Observations

The ciphertext consists of 4065 uppercase alphabetic characters with no spaces or punctuation. This format is typical of classical cipher implementations, but provides no immediate clues about the specific encryption method used.

2.2 Determining the Cipher Type

2.2.1 Index of Coincidence Analysis

The first step in cryptanalysis is to determine whether the cipher is monoalphabetic or polyalphabetic. This is done using the **Index of Coincidence (IC)**:

$$IC = \frac{\sum_{i=A}^{Z} n_i(n_i - 1)}{n(n-1)}$$

where n_i is the frequency of letter i and n is the total number of letters.

Result: The calculated IC was **0.0429**.

To interpret this value, we compare it to known benchmarks:

- English plaintext: IC $\approx$ 0.065
- Monoalphabetic substitution: IC $\approx$ 0.065 (preserves letter frequencies)
- Random text: IC $\approx$ 0.038
- Polyalphabetic cipher (Vigenère): IC $\approx$ 0.038–0.045

The low IC value of 0.0429 strongly indicates a **polyalphabetic cipher**. This rules out simple substitution ciphers, which would maintain English letter frequencies and thus have an IC close to 0.065.

However, this alone does not identify the specific polyalphabetic cipher. Possible candidates include Vigenère, Beaufort, autokey, or other polyalphabetic systems.

2.3 Key Length Determination

2.3.1 Friedman Test

To narrow down the cipher type, we estimate the key length using the Friedman test:

$$m \approx \frac{0.027n}{(n-1)IC - 0.038n + 0.065}$$

where m is the estimated key length, n is the text length, and IC is the Index of Coincidence.

Result: Estimated key length = **5.5** (rough estimate)

This suggests a relatively short key, but the Friedman test is only an approximation and can be off by a few characters. We therefore treat it as a starting point and rely on Kasiski examination and column ICs to confirm the exact length.

2.3.2 Kasiski Examination

To verify the key length estimate, we perform a Kasiski examination. This technique identifies repeated patterns (n-grams) in the ciphertext and analyzes the distances between them. If the cipher uses a repeating key (like Vigenère), these distances will often be multiples of the key length.

The analysis searched for repeated 3-grams, 4-grams, and 5-grams:

Result: Key length candidates from pattern analysis:

$$[2, 4, 8, 16, 3, 6, 12, 24, 5, 10]$$

The value **8** appears prominently in this list. The presence of multiples of 8 (16, 24) further supports this conclusion, as they would naturally occur when the same plaintext pattern aligns with the same key position multiple times.

2.3.3 Column Index of Coincidence Verification

To verify that 8 is indeed the correct key length, we divide the ciphertext into 8 columns (one for each key position) and calculate the IC for each column independently. If the key length is correct, each column should represent a monoalphabetic substitution, and thus have an IC close to 0.065.

Column	**IC**
1	0.0620
2	0.0699
3	0.0673
4	0.0615
5	0.0643
6	0.0602
7	0.0647
8	0.0660
Average	**0.0645**

Table 1: Column Index of Coincidence values for key length 8

All column IC values are close to 0.065 (typical for English text), confirming that:

1. The key length is indeed 8
2. Each column represents a monoalphabetic substitution (Caesar cipher)

3. The cipher uses a repeating key pattern

This evidence strongly suggests a **Vigenère cipher**, as it matches the characteristics:

- Polyalphabetic (low overall IC)
- Short repeating key (length 8)
- Each key position applies a simple shift (Caesar cipher)
- Columns show English-like letter frequencies when correctly aligned

2.4 Key Recovery

2.4.1 Frequency Analysis on Columns

Having identified the cipher as Vigenère with a key length of 8, we can treat each column as an independent Caesar cipher. For each column, we test all 26 possible shifts and identify which shift produces letter frequencies closest to English.

We use the chi-squared statistic to measure how well the decrypted column matches English letter frequencies:

$$\chi^2 = \sum_{i=A}^{Z} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed frequency and E_i is the expected English frequency for letter i. The shift with the minimum χ^2 value corresponds to the key character for that position.

2.4.2 Recovered Key

Position	Column Length	Recovered Shift	Key Character	Status
1	509	8	I	✓
2	508	13	N	✓
3	508	21	V	✓
4	508	4	E	✓
5	508	13	N	✓
6	508	19	T	✓
7	508	14	O	✓
8	508	17	R	✓

Table 2: Key recovery results for Vigenère cipher

Recovered Key: `INVENTOR`

2.5 Verification and Decryption

2.5.1 Decryption

Using the recovered key `INVENTOR`, we decrypt the ciphertext. The first 200 characters of the decrypted text:

```
METHODSFORTHESOLUTIONOFRUNNINGKEYCIPHERSNOANINTRODUCTIONTOMETHODSFORTHESOLUTIONOFCIPHERSNOSYNOP
TICTABLESFORTHESOLUTIONOFCIPHERSANDABIBLIOGRAPHYOFCIPHERLITERATURENOFORMULAEFORTHESOLUTIONOFGEO
METRICALTR
```

2.5.2 Verification

The decrypted text is readable English and discusses methods for the solution of running-key ciphers and related cryptanalytic literature. This confirms that:

1. The cipher type identification was correct (Vigenère)
2. The key length determination was accurate (8 characters)
3. The key recovery was successful (`INVENTOR`)
4. The decryption produces meaningful plaintext

All eight key characters were recovered correctly through frequency analysis, and the decryption produces coherent English text. This provides strong evidence that the analysis was correct.

2.6 Results Summary

- **Ciphertext length:** 4065 characters
- **Index of Coincidence:** 0.0429 (indicated polyalphabetic cipher)
- **Estimated key length (Friedman):** 5.5 (approximate)
- **Key length candidates (Kasiski):** 8 (confirmed)
- **Column IC verification:** All columns ≈ 0.065 (confirmed key length and cipher type)
- **Identified cipher:** Vigenère
- **Recovered key:** `INVENTOR`
- **Verification:** All 8 key characters recovered correctly; decryption produces readable English

3 Cryptanalysis of Substitution Cipher (Cipher2)

3.1 Initial Observations

The ciphertext consists of 4006 uppercase alphabetic characters with no spaces or punctuation. The formatting is consistent with classical substitution ciphers and provides enough text for reliable statistical analysis.

3.2 Determining the Cipher Type

3.2.1 Index of Coincidence Analysis

The Index of Coincidence (IC) for the ciphertext was computed to determine whether the cipher is monoalphabetic or polyalphabetic.

Result: IC = **0.0701**

This value is close to the English benchmark (≈ 0.065), which strongly indicates a **monoalphabetic substitution cipher**. This is notably higher than the IC observed for Cipher1 and well above the random-text baseline (≈ 0.038).

3.2.2 Frequency Analysis

We perform a simple frequency count to identify the most common ciphertext letters. The top 10 letters were:

Letter	Count
Z	602
E	406
X	300
W	281
V	256
F	254
T	253
M	246
K	235
U	197

The skewed distribution is characteristic of monoalphabetic substitution, and supports the IC-based conclusion.

3.3 Recovering the Substitution Key

Since simple frequency matching alone is insufficient to resolve the full substitution, we use a statistical scoring method based on English quadgrams. A quadgram model is built from the known English plaintext of Cipher1 (which is long enough to serve as a language model). Each candidate key is scored by the sum of log-probabilities of the decrypted quadgrams (higher is better).

To search the key space, we use a hill-climbing procedure with random restarts:

1. Start from a frequency-based key guess and several random keys.
2. At each step, swap two letters in the key to form a new candidate.
3. Decrypt with the candidate key and compute its quadgram score.

4. Accept the swap if the score improves; otherwise accept it with a small probability (simulated annealing) to escape local maxima.

5. Keep the best key found, then refine it with additional iterations on the full ciphertext.

3.3.1 Recovered Key Mapping

The recovered substitution key maps ciphertext letters to plaintext letters as:

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
UYXWTSPJBVHMRGDZKFCOLINAQE
```

3.4 Decryption and Verification

Using the recovered key, the ciphertext decrypts into readable English text. The first 200 characters are:

```
OFTENHOWEVERWHEREONECLUEISMISSINGTHEREWILLBEANOTHERPRESENTTOTAKEITSPLACEREPEATEDTRIGRAMSARELESS
LIKELYTHANREPEATEDDIGRAMSTOBEACCIDENTALANDLONGERREPEATEDSEQUENCESARESTILLLESSLIKELYTOBESOINTHEP
RESENTTABU
```

The plaintext discusses cryptanalytic reasoning about repeated n-grams and period detection. It reads as clear English and confirms that the substitution was solved correctly.

3.5 Results Summary

- **Ciphertext length:** 4006 characters
- **Index of Coincidence:** 0.0701 (indicated monoalphabetic substitution)
- **Identified cipher:** Simple substitution
- **Recovered key mapping:**
  `ABCDEFGHIJKLMNOPQRSTUVWXYZ`
  `UYXWTSPJBVHMRGDZKFCOLINAQE`
- **Verification:** Decrypted plaintext is coherent English

4 Conclusions

The ciphertexts were successfully broken using classical cryptanalytic techniques. Cipher1 was identified as a polyalphabetic cipher (Vigenère) with key length 8; the key `INVENTOR` was recovered through Kasiski examination and frequency analysis on columns. The Friedman estimate (5.5) was treated as a rough starting point and refined using Kasiski and column ICs. Cipher2 was identified as a monoalphabetic substitution cipher based on its high Index of Coincidence and was solved using frequency analysis and quadgram-scored hill-climbing.

The Index of Coincidence values provided an effective first discriminator: 0.0429 for Cipher1 (polyalphabetic) and 0.0701 for Cipher2 (monoalphabetic). These methods demonstrate that classical ciphers, while historically significant, are vulnerable to statistical cryptanalysis and should not be used for modern secure communication.