11 Breaking the Vigenère Cipher

Jacques Mock Schindler

24.09.2025

The crucial weakness of the Vigenère cipher was discovered independently by two men in the midst of the 19th century. One of them was Charles Babbage, a British polymath. He discovered the weakness in order to win a bet. The other was Friedrich Wilhelm Kasiski, who was a retired Major of the Prussian army. Who, after retirement, spent his time writing. The method of breaking the Vigenère cipher is named after Prussian Major Kasiski.

The Kasiski Method - Summary

The **Kasiski method** is a technique for breaking the Vigenère cipher by exploiting repeated patterns in the ciphertext. Here's how it works:

Key Principle

When the same plaintext sequence is encrypted with the same portion of the key, it produces identical ciphertext sequences. By finding these repetitions, we can determine the key length.

Step-by-Step Process

1. **Find Repeated Sequences**
    - Scan the ciphertext for identical sequences of 3+ characters
    - Record the positions where these sequences occur
2. **Calculate Distances**
    - Measure the distance between repeated sequences
    - The key length must be a divisor of these distances
3. **Determine Key Length**
    - Find the Greatest Common Divisor (GCD) of all distances
    - Or look for the most frequent common factor

- This gives you the probable key length

4. **Frequency Analysis**

   - Divide the ciphertext into groups based on the key length
   - Each group was encrypted with the same key character
   - Apply frequency analysis to each group separately
   - Match letter frequencies to expected language patterns

Example

If “THE” appears multiple times in plaintext at positions where the key repeats, the corresponding ciphertext sequences will be identical. If these sequences are 10 positions apart, the key length is likely a divisor of 10 (1, 2, 5 or 10).

For a contrived example, if the plaintext is “the codes in the word and the message”, and the key is “crypt”, then we’d get the following ciphertext.

```
\NormalTok{p: t h e c o d e s i n t h e w o r d a n d t h e m}
\NormalTok{k: c r y p t c r y p t c r y p t c r y p t c r y p}
\NormalTok{c: V Y C r h f v q x g V Y C l h t u y c w V Y C b}
```

Note that the ciphertext pattern VYC occurs three times in the ciphertext, and each time there is a distance of 10 letters between the beginning of one VYC and the next. This happens because the same pattern of plaintext, “the”, lines up with the same part of the key, “cry”, each time, resulting in the same ciphertext. The repetition of the keyword is the liability here. The ciphertext duplicates are all 10 letters apart. Babbage and Kasiski reasoned that this implies the length of the key is a factor of 10. The factors of 10 are 10, 5, and 2. One could argue that a key of length 2 is too short to provide much security, so that a key of length 5 or 10 is more reasonable. A key of length 10 is unlikely to have as many repetitions in such a short piece of ciphertext, so a key of length 5 is where the cryptanalyst will begin their work.

A key of length 5 means that the 1st, 6th, 11th, 16th, etc., letters are all enciphered using the same key letter and hence the same alphabet from the Vigenère table. Similarly, the 2nd, 7th, 12th, 17th, etc. letters are all enciphered with the next key alphabet. So if we break up the cryptogram into 5 groups of letters we then have 5 monoalphabetic substitution ciphertexts. We can then do a frequency analysis of each group and solve each group separately. And in a standard shifted alphabet as in the normal Vigenère table, if we can find a single cipher alphabet letter we then have the entire alphabet.[1]

[1] Dooley, John F. History of Cryptography and Cryptanalysis: Codes, Ciphers, and Their Algorithms. History of Computing. Cham: Springer International Publishing, 2024. https://doi.org/10.1007/978-3-031-67485-3. p. 76.

Why It Works

The Vigenère cipher's weakness lies in key repetition. Once the key length is known, the cipher becomes multiple simple Caesar ciphers that can be broken using frequency analysis.