Analyzing Vigenere Cipher

Mario Enriquez

British Columbia Institute of Technology

COMP 7481, 8D

Carly Orr

April 20,2017

Contents

[1 Research 3](#_Toc480492587)

[1.1 Background 3](#_Toc480492588)

[1.2 About this project 3](#_Toc480492589)

[1.3 Objectives 5](#_Toc480492590)

[1.4 Relevance 5](#_Toc480492591)

[2 Experiment 5](#_Toc480492592)

[2.1 Algorithm implementation 5](#_Toc480492593)

[2.2 Test and Results 6](#_Toc480492594)

[2.3 Conclusions 9](#_Toc480492595)

[2.4 Future Improvements 9](#_Toc480492596)

[2.5 References 9](#_Toc480492597)

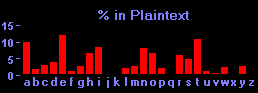
# Research

## Background

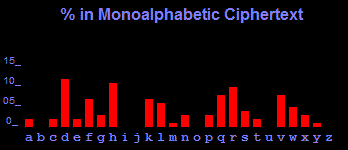
Vigenere is one of the classic ciphers, it’s a substitution stram cipher based on a method of encrypting alphabetic text by using a series of interwoven Caesar ciphers based on the letters of a keyword. It is a form of polyalphabetic substitution.

## About this project

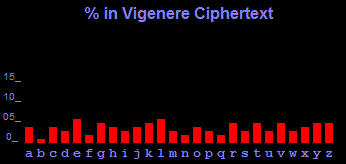
Vigenere is a strong cipher, it’s strength comes from being able to eliminate the frequencies of letters that affect the Caesar Cipher, as this example from the Black Chamber Website Shows:



Now, if you encrypt this message using a monoalphabetic substitution cipher, you can see how the frequency distribution changes:



The high frequencies have merely moved to different letters (e.g., the highest peak has moved from E to D, because E has been encrypted as D), and they can be used to crack the cipher. Now, let us encrypt the text using the Vigenere cipher and you will see why it is a better cipher:



As you can see, the frequency distribution is now much flatter. The peaks are less obvious, because each letter has been encrypted in 8 different ways, because the keyword is 8 letters long. The peak that was at E has been shared among 8 other letters. A flatter frequency distribution means a much stronger cipher.

I choose this project because:

* I wanted to try and find vulnerabilities since I thought it was secure cipher
* The original plan was to choose biggest word in document to try and get the word from it
* I discovered that there is a well-documented Vigenere weakness
* The weakness, as Charles Babbage discovered, is that repetitions in the ciphertext indicated repetitions in the plaintext and that the space between such repetitions hinted at the length of the keyword.
* If I got the factorials for the distance between sequences we would be able to discern the keyword’s length
* In addition, we could try to select the sequences of 3-characters with the most repetitions to find parts of the keyword this is based on the Frequency of Words in English as shown below

Frequency of Words in English

|  |  |
| --- | --- |
| **ENGLISH** | |
| Order Of Frequency Of Single Letters | E T A O I N S H R D L U |
| Order Of Frequency Of Digraphs | th er on an re he in ed nd ha at en es of or nt ea ti to it st io le is ou ar as de rt ve |
| Order Of Frequency Of Trigraphs | the and tha ent ion tio for nde has nce edt tis oft sth men |
| Order Of Frequency Of Most Common Doubles | ss ee tt ff ll mm oo |
| Order Of Frequency Of Initial Letters | T O A W B C D S F M R H I Y E G L N P U J K |
| Order Of Frequency Of Final Letters | E S T D N R Y F L O G H A K M P U W |
| One-Letter Words | a, I. |
| Most Frequent Two-Letter Words | of, to, in, it, is, be, as, at, so, we, he, by, or, on, do, if, me, my, up, an, go, no, us, am |
| Most Frequent Three-Letter Words | the, and, for, are, but, not, you, all, any, can, had, her, was, one, our, out, day, get, has, him, his, how, man, new, now, old, see, two, way, who, boy, did, its, let, put, say, she, too, use |
| Most Frequent Four-Letter Words | that, with, have, this, will, your, from, they, know, want, been, good, much, some, time |

## Objectives

* Learn about the key attributes of the Vigenere Cipher
* Learn about the process of generating a table for the Vigenere Cipher as well as encoding and decoding
* Discern strategies to improve on the Vigenere Cipher security
* Learn about the weaknesses of a ciphertext encoded in Vigenere and try obtaining the plaintext based on them

## Relevance

Even with modern algorithms of encryption like AES, DES, Blowfish, etc. Vigenere remains a strong secure encryption algorithm and with the correct changes we can improve on the security of the code.

It’s important to know of the weaknesses of the ciphers that are available to use to either correct them or to try and obtain the ciphertext from them.

# Experiment

## Algorithm implementation

I created an algorithm to generate a Vigenere Table, encrypt a document, and analyze the sequences that



## Test and Results

|  |  |  |  |
| --- | --- | --- | --- |
| Test | Description | Expected Results | Actual Results |
| 1 | Encrypt a Document | Document Encrypted | Encryption Successful. |
| 2 | Try obtaining the sequences from the document and het factorial and possible key | List of sequences is printed, shows distances and possible keyword | Sequences are found in the document, parts of the keyword are shown. Success. |
| 3 | Compare sequences to place in document to try and verify | We would be able to determine with more clarity the results of the previous step | Results verified. Success. |

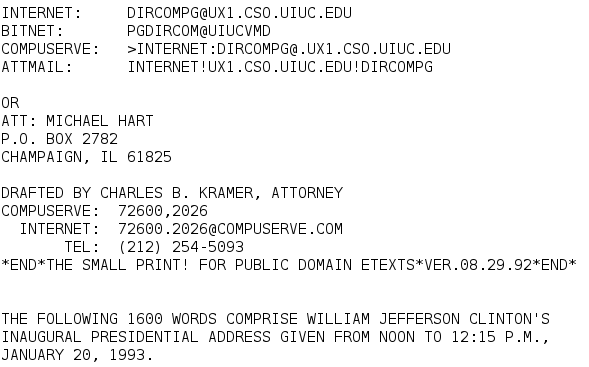


Figure 1 Document

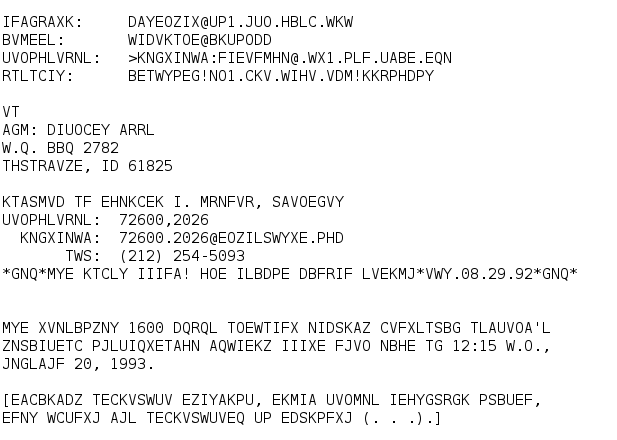


Figure 2 Document Encrypted

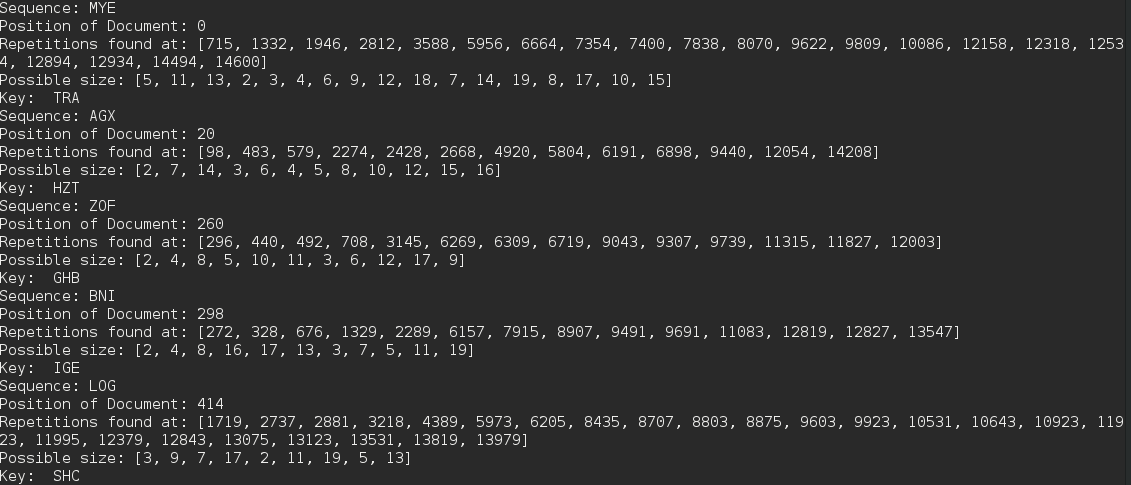


Figure 3 Results from analysis



Figure 4 Show Position, first 3 characters are "the", verifying the success of analysis

## Conclusions

There are several steps we can take to improve Vigenere security

* The longer and varied character the key has, more difficult is to find it as it leads to less repetitions
* Encrypting the ciphertext again with a different keyword can decrease chances of obtaining plaintext. Obviously, the main problem with this method is that you require more keys
* The shorter the document, it’s more difficult to find sequences since not many words are used or the key will not generate many repetitions
* Keeping the document format along with punctuations make it easier to decipher
* For different languages, the experiment would need to be adjusted only on the reverse search of the keyword. The sequence search would remain the same

## Future Improvements

* Test in other languages
* User can modify the length of the chain to search

## References

* Cryptanalysis Hints. (n.d.). Retrieved April 20, 2017, from <https://www3.nd.edu/~busiforc/handouts/cryptography/cryptography%20hints.html>
* Singh, S. (n.d.). Cracking the Vigenère Cipher. Retrieved April 20, 2017, from <http://www.simonsingh.net/The_Black_Chamber/vigenere_cracking_tool.html>
* Vigenère cipher. (2017, March 23). Retrieved April 20, 2017, from https://en.wikipedia.org/wiki/Vigen%C3%A8re\_cipher