Cracking the Vigenère Cipher

By Joel Smith

What is the Vigenère Cipher?

- Also known as the Repeating Keyword Cipher
- Polyalphabetic
  - More secure than monoalphabetic ciphers
  - Brute force algorithms can’t decrypt it
  - Used more frequently than monoalphabetic ciphers

History

- Invented in 1553 by the Italian cryptographer Giovan Battista Bellaso
- Was known as “*le chiffre indéchiffrable*”
  - *Means “the unbreakable cipher” or “the undecipherable cipher” in French*
- *Was known to be unbreakable for three centuries*
- *Broken by Charles Babbage*

Example

- Example:
  - Plain Text: **The Quick Brown Fox Jumps Over the Lazy Dog**
  - Keyword: **Fox**
  - Cipher Text: **Yvb Vifhy Ywcts Tlc Xrrdp Tjbw hej Zxem Atu**

*Assuming **a** is **0***

Plain - T = 19
Key - F = 5
Encrypted: (19 + 5) % 26 = 24 = Y

Plain – H = 7
Key – O = 14
Encrypted: (7 + 14) % 26 = 21 = V

Creating a Program

- Step 1: Understanding ASCII
  - What is ASCII?
    - Represent: Numbers, Letters, Symbols with their respective Hex / Decimal / Binary representation.
  - Why is this necessary?
    - Very briefly:
      - **a** = 97, **b** = 98, … , **z** = 122
    - Important with:
      - IO
      - In Python: ‘a’ = 97
      - **Important!** To Normalize this we do, **97** – **97 = 0 (**First Letter in the Alphabet)
      - So, if we are pushing ‘a’ forward by ‘c’ spots, we don’t accidentally shift it by **99.**
        - **197 = ---** (Horizontal Bar) In Ascii

Creating a Program

- Step 2: File IO
  - I will cover three brief points:
    - When Encrypting & Decrypting, file formatting is preserved.
    - Files are represented as **objects**
    - A simple Statement to check if a Character is Legal
      - **return** (**ord**(char) > *64* **and ord**(char) < *91*) **or** (**ord**(char) > *96* **and ord**(char) < *123*)
        - Ord() converts a Character to Decimal in **Python**
        - Checks to make sure the character is between **a-z, A-Z**

Creating a Program

- Step 3: Methodology
  - Frequency Analysis.
    - Go through the file, gathering all letter frequencies.
      - First Checks for length 5, down to 1.
- Check out this excerpt:

LHVSY CEZZQ MPCWP UEVGH CWPWE FQHZZ WSYSV CEVGD SMUWN

LHVSY

HVSYC

VSYCE

The program goes through **every** possible pair

(Yes, I stole this excerpt from Dr. Polhill's handout)

Example Cited

Ecwoqww, Ycsu. “Jcppy Doqqe pbo Rmjhdfo.” Iwwe, 2022, qcwa.jwzdaf.llf/o2a/zp/jwyetbe/3258375/cqphRcyamye/27807694/Kwpd.

Keyword: Polhill

Polhill, John. “Break Shift and Keyword.” Bolt, 2022, bolt.bloomu.edu/d2l/le/content/3258375/viewContent/27807694/View.

Creating a Program

- Step 3: Continued

LHVSY CEZZQ MP**CWP** U**EVG**H **CWP**WE FQHZZ WSYSV C**EVG**D SMUWN
GEZLG UFZCO PPGWF PHJOE HJSXO VVESC QSSTG LFGIY OYOCS
QTHVC HJCWO CSFCT BIWEV KBVHJ SJAKU SHDSZ BVCZI TGNVG
APDNS LGGPC WPUXS CPTUO ONHQC

- A **couple** repeated words are highlighted

- In my program, they are stored as so in a Dictionary:

{'CWP' : {'InitPos': 12, 'Count': 3, 'Distance': 8}, 'EVG': {'InitPos': 16, 'Count': 2, 'Distance': 20} }

Creating a Program

- Step 4: Finding Keyword Length
  - From the Previous Example:
    - Distance between CWP = 20
    - Distance between EVG = 8
  - GCD(20, 8) = 4
  - 4 is now a candidate for the length of the keyword.
  - The candidates get tallied as the program proceeds.

  - In this case the keyword length is indeed 4

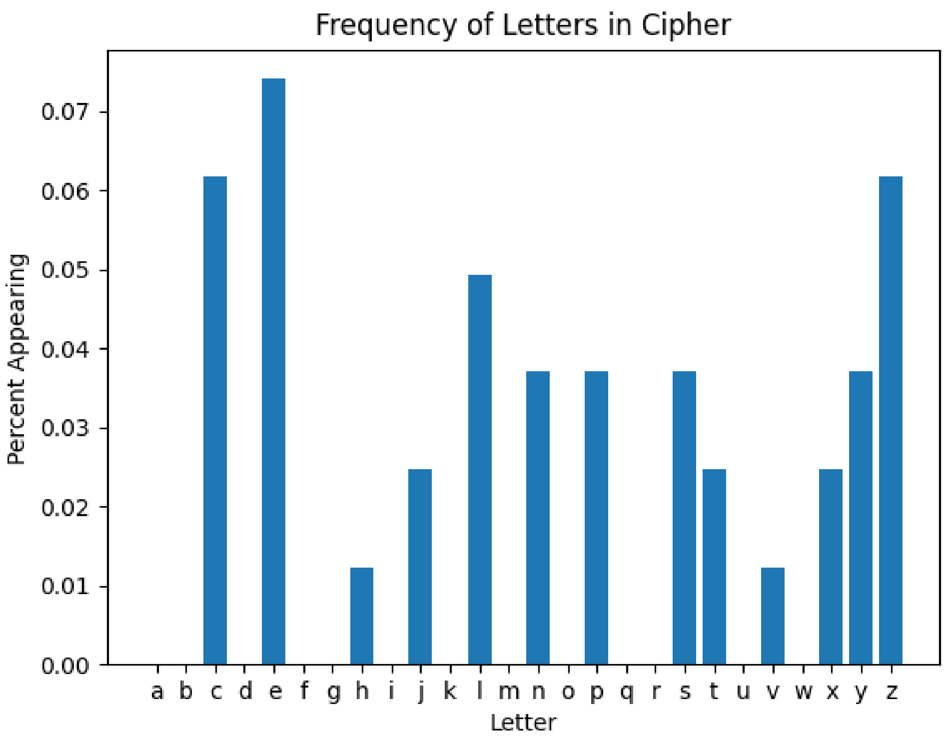
Creating a Program

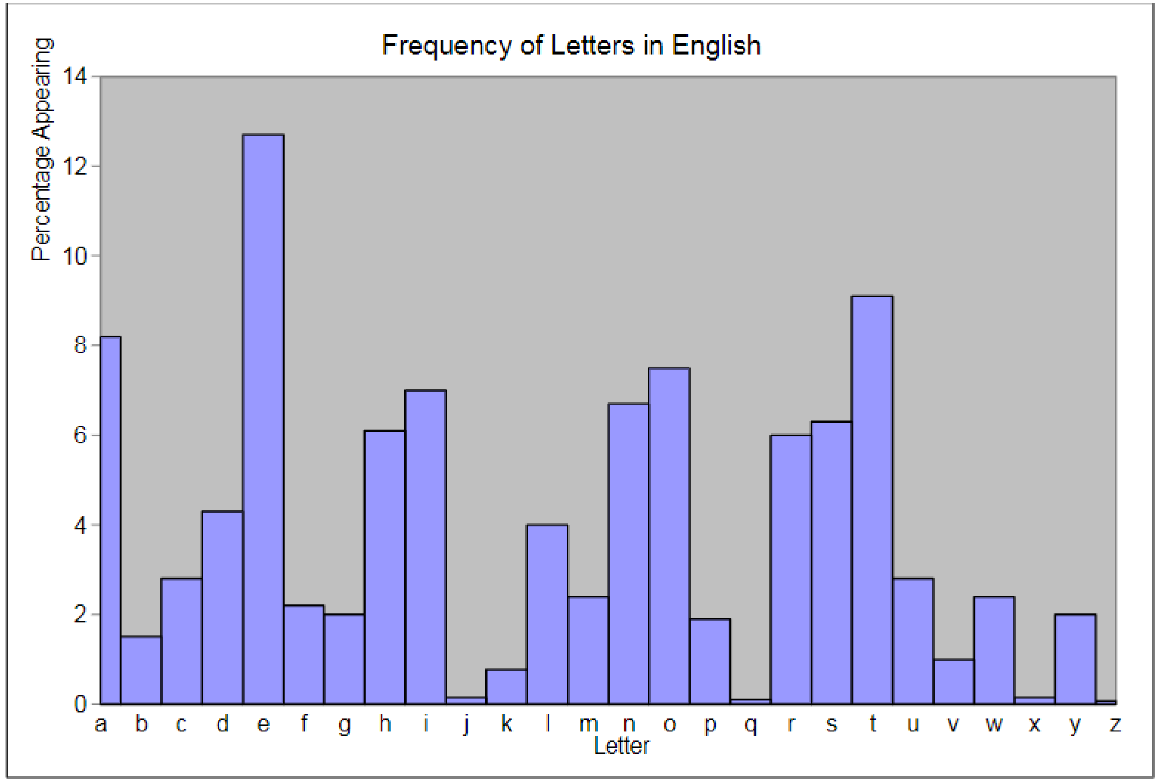
- Step 5: Compare Cipher Text Frequencies to English Letter Frequencies
  - Since we decided on a length 4 keyword
    - We can check every fourth letter

LHVS **Y**CEZ **Z**QMP **C**WPU **E**VGH **C**WPW **E**FQH **Z**ZWS **Y**SVC **E**VGD **S**MUW
NGEZ **L**GUF **Z**COP **P**GWF **P**HJO **E**HJS **X**OVV **E**SCQ **S**STG **L**FGI **Y**OYO
CSQT **H**VCH **J**CWO **C**SFC **T**BIW **E**VKB **V**HJS **J**AKU **S**HDS **Z**BVC **Z**ITG
NVGA **P**DNS **L**GGP **C**WPU **X**SCP **T**UOO **N**HQC

- Bold letters get tallied
  - The computer does this with modulo **if** (index **mod** key_len) == curr_key_pos

Frequencies





Creating a Program

- Computer takes the Dot Product of the Peaks of the graph to determine the key letter.
- After Iterating through all the positions in the keyword, the final keyword is decided.


CRACKING
THE
VIGENERE
CIPHER

Sources

Simmons, Gustavus J.. "Vigenère cipher". Encyclopedia Britannica, 14 Jul. 2021, https://www.britannica.com/topic/Vigenere-cipher. Accessed 22 April 2022.