Poly-alphabetic Substitution Ciphers
Breno de Medeiros

Poly-alphabetic ciphers

- A cipher combining several substitution ciphers is called poly-alphabetic.
- Examples of such ciphers are the Vigenère cipher, and ciphers implemented by rotor machines, such as Enigma.

Vigenère Cipher

- The Vigenère Cipher implements a combination of shift ciphers.
- A simple shift cipher (by $i$ positions) substitutes an alphabet character by the character $i$ positions later. For instance, Caesar's cipher:

- $\text{abcdefghijklmnopqrstuvwxyz}$
- $\text{defghijklmnopqrstuvwxyzabc}$
**Keyword-based shift**

- Say the keyword (or phrase) is “ONE RING,” and we wish to encrypt “The meaning of things”
- **THE MEANING OF THINGS**
- **ONE RING ONE RING ONE RING ONE RING ONER**
- **1 4 13 04 17 08 13 06 14 13 04 17 08 13 06 14 13 04 17**
- **07 20 08 03 12 13 19 22 00 10 05 13 06 13 22 00 10 09**
- **H I U D M N T O A K F N G N W A K J**

**Cryptanalizing Vigenère**

- To cryptanalyze Vigenère one first finds the length $d$ of the keyword.
- This can be done using either the **index of coincidence** or the **Kasiski** methods (explained in the following).
- Then separate the ciphertext in $d$ chunks and solve each as a mono-alphabetic shift substitution cipher.

**Kasiski method**

- English and other natural languages have character sequences which are far more likely than coincidence would indicate. So much so, that in a long ciphertext, repeated occurrences of character sequences are likely to be created from keyword-aligned character sequences in the plaintext:
- **THE MEANING OF THINGS**
- **WINSOME WINSOME WINSOME**
- **PPRESMREVTSFSLEVTK**
Applying Kasiski

- For English, the Kasiski method works better with sequences of four characters.
- When such repetitions occur, likely they are keyword-aligned, so the distance reveals a multiple of keyword length.
- Find several occurrences, recover the keyword length by computing the minimum common divisor.

Index of Coincidence

- The index of coincidence IC of a string is the probability that a random pair of positions in the string contains the same character:
  - Let \( n \) be the length of the string.
  - Let \( i \) be an English character that occurs \( c \) times in the string.
  - There are \( \binom{c}{2} \) pairs of positions where \( i \) occurs, out of all \( \binom{n}{2} \) pairs of characters in the string.
  - The probability of coincidence with character \( i \) in the string is \( \frac{\binom{c}{2}}{\binom{n}{2}} \)

Computing IC

- Compute the counts of each character in the string. Let \( c(i) \) be the count of character \( i \).
  - The length of the string is \( n = \sum c(i) \)
- Compute

\[
IC = \frac{\sum c(i)(c(i)-1)}{n(n-1)}
\]
How IC can be used

• If the character frequency of the string is similar to English, then $IC \approx 0.066$
• The IC of a random string is $\approx 0.038$
• $d =$ number of simple substitutions

<table>
<thead>
<tr>
<th>$d$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC</td>
<td>0.66</td>
<td>0.52</td>
<td>0.473</td>
<td>0.445</td>
<td>0.436</td>
<td>0.427</td>
<td>0.38</td>
</tr>
</tbody>
</table>

What makes a Good Cipher

Breno de Medeiros

Security through obscurity

• **Security through obscurity** is the use of secrecy to achieve security. A security system relying on obscurity may have vulnerabilities, under the expectation that they are hard to find if the design of the system is unknown.
  – Practical example of security through obscurity: Leaving a spare key to your house under the doormat.
• Good ciphers need not to, and should never rely on security through obscurity. First formulated by Kerckhoffs (19th century), this principle is summarized in Shannon’s maxim “The enemy knows the system.”
Auguste Kerckhoffs’ principles

- A cipher should be secure even if everything about it (except the key) is known:
  - The system must be practically, if not mathematically, indecipherable;
  - It must not be required to be secret, and it must be able to fall into the hands of the enemy without inconvenience;
  - Its key must be communicable and returnable without the help of written notes, and changeable or modifiable at the will of the correspondents;
  - It must be applicable to telegraphic correspondence;
  - It must be portable, and its usage and function must not require the concourse of several people;
  - Finally, it is necessary, given the circumstances that command its application, that the system be easy to use, requiring neither mental strain nor the knowledge of a long series of rules to observe.

Shannon’s formulation

- Claude Shannon (information theory founder) proposed the characteristics of a good cipher (1949):
  - The amount of secrecy needed should determine the amount of labor appropriate for encryption and decryption.
  - The set of keys and the enciphering algorithm should be “free from complexity.”
  - The implementation of the process should be as simple as possible.
  - Errors in ciphering should not propagate and cause corruption of further information in the message.
  - The size of the enciphered message should be no larger than the text of the original message.