9/4 Notes - Substitution and Vigenere Ciphers

Scott Swatling

September 11, 2019

The Caeser cipher and the Affine cipher both fall under a larger category of ciphers known as substitution ciphers.

Substitution Cipher

Definition

- Any encryption method that maps each individual character to either itself, or any other character, once

Properties

- 26! combinations
- Too big to brute force
- Remains seceptible to frequency analysis

Attacks

- Chosen Plaintext
  - Encrypting the entire alphabet, or a sentence containing every letter of the alphabet, will directly reveal the key.
    Examples
    * "abcdefghijklmnopqrstuvwxyz"
    * "The quick brown fox jumped over the lazy dog"
  - Known Plaintext
    * All the unique characters present in the plaintext will reveal, at least partially, the key for the cipher text
  - Ciphertext Only
    * frequency analysis may be used in conjunction with commonly seen patterns in language to reveal the key

Key Observation

- Simply resisting a brute force attack is not enough to gurantee the relative security of a cipher

Vigenére Cipher

The key will consist of a word or phrase converted into their respective alphanumeric values, which we may call a vector.
Example:

- converting "key"
  - $k \equiv 10 \pmod{26}$
  - $e \equiv 4 \pmod{26}$
  - $y \equiv 24 \pmod{26}$
  - the key is <10 4 24>

Encryption Steps

1. Convert the plaintext to numbers.
2. Bellow the converted plaintext, repeatedly copy the key vector until they both match in size.
3. Add the two lines in a top-down fashion then mod 26.
4. Convert the resulting line back into text.

Example

- The key will be "key" = <10 4 24>
- The plaintext will be "here is how it works"
- Step 1:

h	e	r	e	i	s	h	o	w	i	t	w	o	r	k	s
7	4	17	4	8	18	7	14	22	8	19	22	14	17	10	18

- Step 2:

h	e	r	e	i	s	h	o	w	i	t	w	o	r	k	s
7	4	17	4	8	18	7	14	22	8	19	22	14	17	10	18
10	4	24	10	4	24	10	4	24	10	4	24	10	4	24	10

- Step 3:

h	e	r	e	i	s	h	o	w	i	t	w	o	r	k	s
7	4	17	4	8	18	7	14	22	8	19	22	14	17	10	18
10	4	24	10	4	24	10	4	24	10	4	24	10	4	24	10
17	8	15	14	12	16	17	18	20	18	23	20	24	21	8	2

- Step 4:

h	e	r	e	i	s	h	o	w	i	t	w	o	r	k	s
7	4	17	4	8	18	7	14	22	8	19	22	14	17	10	18
10	4	24	10	4	24	10	4	24	10	4	24	10	4	24	10
17	8	15	14	12	16	17	18	20	18	23	20	24	21	8	2
R	I	P	O	M	Q	R	S	U	S	X	U	Y	V	I	C

- the plaintext "here is how it works" converts to "RIPO MQ RSU SX UYVIC" using the key "key"

Attacks

- Chosen Plaintext
  - Having a large string of a single character, preferably 'a', will reveal the key vector
    Examples
    * "aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa..."
    * "bbbbbbbbbbbbbbbbbbbbbbbbbbbbbbb..."
  - Known Plaintext
    * Simply reverse the encryption process by subtracting the plaintext from the ciphertext
  - Ciphertext Only
    * Using a guessed length N for the vector key, frequency analysis may be performed for every Nth character in the ciphertext
    * Babbage's Trick
      1. Write the ciphertext on one line
      2. copy the ciphertext on another line shifted to the right once
      3. repeat step 2 until there are, at most, as many lines as the length of the ciphertext
      4. Count the number of times a letter for any given line repeats when compared to the original ciphertext

      Example
      · We will use the previous example's cipher text "RIPOMQRSUSXUYVIC"
      · Step 1:

R	I	P	O	M	Q	R	S	U	S	X	U	Y	V	I	C

      · Step 2/3:

R	I	P	O	M	Q	R	S	U	S	X	U	Y	V	I	C
C	R	I	P	O	M	Q	R	S	U	S	X	U	Y	V	I
I	C	R	I	P	O	M	Q	R	S	U	S	X	U	Y	V
V	I	C	R	I	P	O	M	Q	R	S	U	S	X	U	Y
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

- step 4:
  Line 1 - 0 coincidence(s)
  Line 2 - 1 coincidence(s)
  Line 3 - 2 coincidence(s)
  ...

  We will see a noticeable spike in the number of coincidences
  on lines that are multiples of the length of the key