Attack on Cryptography

By

Mohd Zaid Waqiyuddin Mohd Zulkifli

April 2008

Table of Contents

1. Attack on Caesar Cipher.

Although historically Julius Caesar used a shift of 3 for his cipher, any ciphering based on alphabet shifting of the plaintext is called Caesar cipher. This is a very simple method of ciphering, and provides very little security. However it is still applied mostly in forum or bulletin board to post possibly offensive materials, or to provide answers to riddles where there would be no accident of unintended glimpse. For shift value 3, we have “A” encrypted to “D”, “B” encrypted to “E”, “C” encrypted to “F” and so on. “X”, “Y” and “Z” will be encrypted to “A”, “B” and “C” respectively.

Caesar cipher only has 25 possibilities of a key. A direct brute-force attack testing each key is simplest and fastest for attacking the ciphertext.

For example, suppose we intercepted a ciphertext below and we suspected it had been encrypted with Caesar Cipher.

KIMAIZKQXPMZQA MIAG

We could then start our brute-force attack.

For shift of 1, we have obtain,

CIPHER	K	I	M	A	I	Z	K	Q	X	P	M	Z	Q	A	M	I	A	G
PLAIN	J	H	L	Z	H	Y	..	..	..	..	..	..	..	..	..	..	..	..

It is already apparent that 1 is not the key and we may continue with 2 and so on. With key = 8, we finally get intelligible result.

CAESARCIPHERISEASY

Attack was successful.

2. Attack on Monoalphabetic Substitution Cipher

The previously mentioned Caesar Cipher is actually a type of monoalphabetic substitution cipher, where each character is uniquely mapped to another character. The relationship between the plaintext and the ciphertext is one-to-one.

Caesar Cipher implements the following function

$$C_i = (M_i + K) \bmod 26$$

to convert plaintext to ciphertext.

There also exist monoalphabetic ciphers which are based on different functions such as Affine and Atbash cipher. These ciphers are trivial can be attacked by applying the inverse of the underlying mathematical function.

Next, consider the situation where two parties to communicate with monoalphabetic cipher pre-determined the assignment without basing it on mathematical function.

For example,

MESSAGE	A	B	C	D	E	F	G	H	I	J	K	L	M	N
CIPHER	J	Q	C	K	G	U	M	W	D	Y	R	T	E	Z

O	P	Q	R	S	T	U	V	W	X	Y	Z
H	I	P	X	A	V	O	L	B	S	F	N

Suppose we intercept the following message, without the knowledge of the table above.

UXGPOGZCFJZJTFADADAJEJNDZMZHBBGZGGKQGVVGXCDIWGX

How can we possibly decode this?

2.1. Frequency Analysis

In the 9th century, an Arab polymath Abu Yusuf Yaqub ibn Ishaq al-Sabbah Al-Kindi in *A Manuscript on Deciphering Cryptographic Messages* describe frequency analysis as a method to defeat monoalphabetic substitution cipher.

This method makes use of the characteristic of any given stretch of written language where certain letters or combinations of letters occur with varying frequency.[1] For instance, letters A,E,I,S,T and R occur more frequently in a sufficiently long English text compared to letters J,X and Z.

LETTER	A	B	C	D	E	F	G	H	I	J	K
FREQUENCY	3	2	2	4	1	2	8	1	1	4	1

M	N	O	P	Q	T	U	V	W	X	Z
1	1	1	1	1	1	1	2	1	3	5

[1] http://en.wikipedia.org/wiki/Frequency_analysis_(cryptanalysis)

Now, we compare to the statistic of English language.

Letter	Probability	Letter	Probability
A	0.082	N	0.067
B	0.015	O	0.075
C	0.028	P	0.019
D	0.043	Q	0.001
E	0.127	R	0.060
F	0.022	S	0.063
G	0.020	T	0.091
H	0.061	U	0.028
I	0.070	V	0.010
J	0.002	W	0.023
K	0.008	X	0.001
L	0.040	Y	0.020
M	0.024	Z	0.001

We see that G is the most frequent (8 occurrences), hence it is highly likely that it represents 'E'. However trial and error is needed, as the frequency distribution of letters in a ciphered message often do not follow the above standard completely, especially for short message.

Subsequent effort with the help of the table would reveal the message to be

FREQUENCY ANALYSIS IS AMAZING NOW WE NEED BETTER CIPHER

4. Types of Attack

1) Ciphertext-only
2) Known Plaintext
3) Chosen Plaintext
4) Chosen Ciphertext
5) Side Channel Attack

<u>Ciphertext-Only</u>

All attacks described so far are examples of ciphertext-only attack where the attacker only has ciphertext. This type of attack is most common, but also most difficult because of lack of information.

Known Plaintext

Information can never make things harder. With this type of attack, the attacker possesses a string of plaintext, x and the corresponding ciphertext, y.
Consider this example of known plaintext attack with monoalphabetic substitution cipher.

The following ciphertext is intercepted and is known to contain information about a person called "ANDERSON" and a place called "MISSISSIPPI".

JZKGXAHZDAVWGZGBWGJKHUAIDGADZEDAADAADIID

Here, rather than applying frequency analysis which might not be helpful particularly on short ciphertext, we could use our information of the plaintext to devise stronger attack.

Since we know the message contains the word MISSISSIPPI, we look for a sequence of 11 letters where the 3rd, 4th, 6th and 7th letters are the same and so are the 2nd, 5th, 8th and 11th. It would not take long to notice that the sequence 'EDAADAADIID' is the ciphertext for 'MISSISSIPPI'.

Similarly for ANDERSON, we look for a sequence of 8 letters where all characters are different except for the 2nd and the 8th. A small branch of computing studies called regular expression along with appropriate software will be helpful to speed up the search, but eventually the attacker will find that 'JZKGXAHZ' represents 'ANDERSON'

With the newly gained information so far, the ciphertext has been decrypted to

ANDERSONIS<u>VW</u>ENE<u>BW</u>EADO<u>U</u>SPIESINMISSISSIPPI

Subsequent effort of cryptanalysis may eventually reveal the secret,

ANDERSON IS THE NEW HEAD OF SPIES IN MISSISSIPPI

Here, it can be seen that information of plaintext opens up new possibilities of attacking methods. This type of attack is possible with encryption of documents which are known to follow certain templates. For example, an email usually starts with 'Dear Sir' or 'Dear Madam' and ends with 'Yours Sincerely' or 'Regards'.

Chosen Plaintext Attack

This attack is different from Known Plaintext Attack in such way that the attacker can choose which plaintext is to be encrypted, and later analyse the relationship of the output ciphertext to get the key used for encryption.

For example, suppose we want to attack communication from Alice to Bob which is encrypted by monoalphabetic substitution cipher. The intercepted messages so far could not be solved using frequency analysis. And we know how helpful it is if we can get Alice to send an encrypted message to Bob which contains the word 'MISSISSIPPI'.

Here, we can send an email to Alice, "Please tell Bob that saying Mississippi will take exactly one second". Then, whether Alice sends a fresh email to Bob or simply forward our written email, we can intercept the message and obtain some information about the mapping of plaintext to ciphertext used in encryption of communication from Alice to Bob.

This type of attack is even stronger as the attacker has more control of the operation.

Chosen Ciphertext Attack

This type of attack is normally associated with the decryption process where the opponent has obtained temporary access to the decryption machinery.[2] He may then select a ciphertext string to construct the corresponding plaintext string.

5. Attack on Polyalphabetic Substitution Cipher

It was evident that monoalphabetic substitution ciphers had a lot of weaknesses, so cryptographers came up with a stronger solution, polyalphabetic cipher. Whereas monoalphabetic substitution cipher has one-to-one relationship between plaintext and ciphertext, polyalphabetic substitution cipher has one-to-many relationship.

This means the letter 'E' in plaintext may be encrypted to 'J' or 'X'. This is a useful encryption technique against frequency analysis as the letters frequencies are more obscured.

Viginere Cipher

This is a type of polyalphabetic substitution cipher. With this cipher, if the encryption key is "SECRETKEY", the first letter of the message will be encrypted with 'S' , second letter of the message with 'E' and so on, following the order of the key's character order.

---

[2] Cryptography Theory and Practice

If the encryption reaches the last character of the key, the next message's letter will be encrypted with the first character of the key again and the cycle continues.

Attacking Viginere Cipher

This ciphertext is from Cryptology Theory and Practice.

**CHREEVOAHMAERATBIAXXWTNXBEEOPHBSQMQEQERBWRVXUOAKX
AOSXXWEAHBWGJMMQMNKGRFVGXWTRZXWIAKLXFPSKAUTEMNDC
MGTSXMXBTUIADNGMGPSRELXNJELXVRVPRTULHDNQWTWDTYGBPH
XTFALJHASVBFXNGLLCHRZBWELEKMSJIKNBHWRJGNMGJSGLXFEYP
HAGNRBIEQJTAMRVLCRREMNDGLXRRIMGNSNRWCHRQHAEYEVTAQE
BBIPEEWEVKAKOEWADREMXMTBHHCHRTKDNVRZCHRCLQOHPWQAI
IWXNRMGWOIIFKEE**

The first operation is to guess the length of the key used to encrypt. This can be done by performing Kasiski test, performed with the following steps.

1) Record where similar sequences of letters occur in many places.
   Here we notice the sequence CHR occurs in five places beginning at position 1, 166, 236, 276, and 286.

2) Calculate the distance between occurrences and the first occurrence.
   a) 166 – 1 = 165
   b) 236 – 1 = 235
   c) 276 – 1 = 275
   d) 286 – 1 = 285

3) Calculate the greatest common divisor of the calculated values.
   gcd (165, 235, 275, 285)

4) The result is the likely length of the key used.

If we had guessed the key length correctly, the complexity of polyalphabetic substitution cipher is reduced to that of monoalphabetic substitution cipher. Suppose the guessed key length is 5, we may then proceed by dividing the ciphertext into a group of 5. Group 1 formed by 1^{st},6^{th} ,11^{th} … letters, Group 2 by 2^{nd}, 7^{th}, 12^{th} … letters. Frequency analysis can then be formed on these individual groups.

6. Conclusion.

Cryptography is the heart of security. While strong cryptography does not guarantee strong security, weak cryptography certainly guarantees weak security. Equally important is the protocol and management involved in implementing the cryptography.

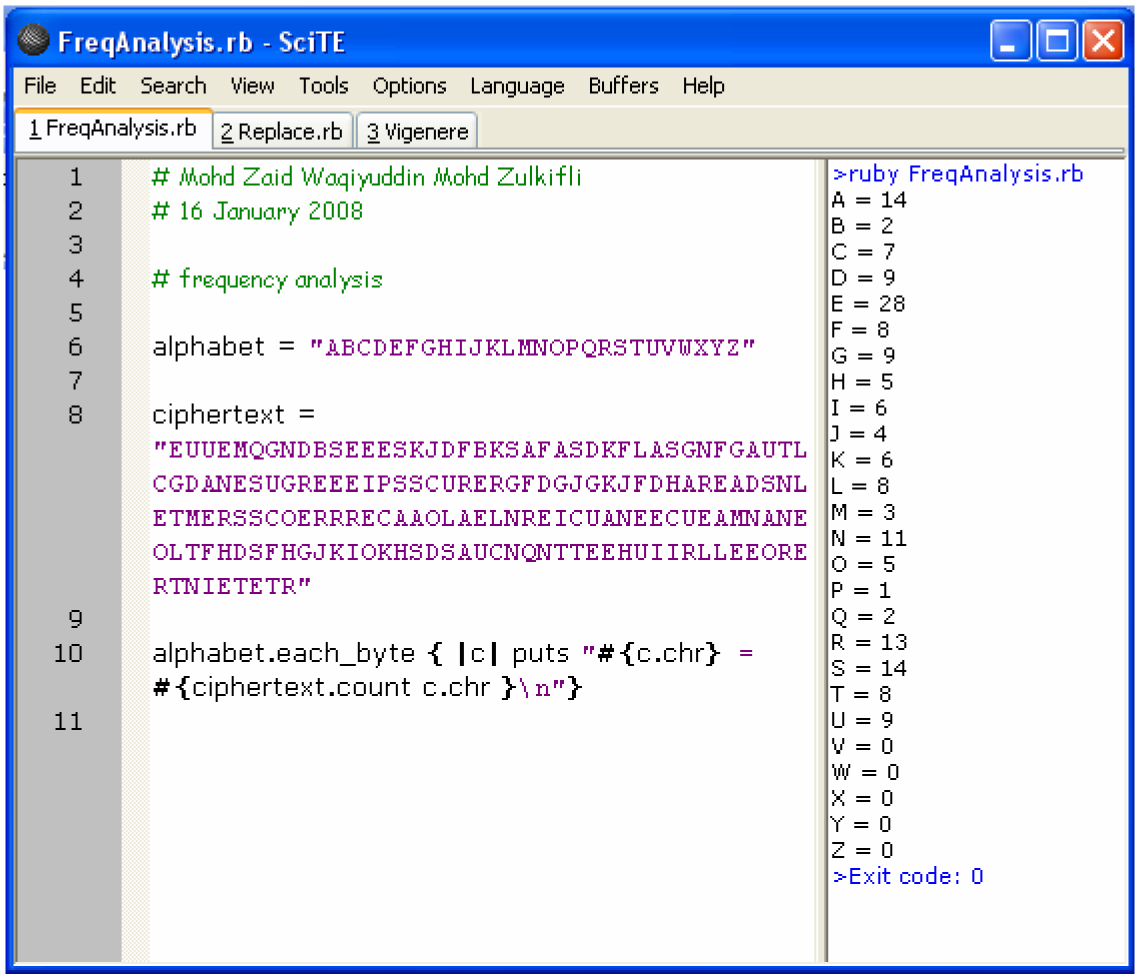
Perfect secrecy can be achieved with Vernam Cipher, as proved by Shannon in his paper. While the dream of perfect security is made possible with the latest development in quantum cryptography, the study of cryptography and cryptanalysis will still be going on, partly due to the impracticality to implement quantum cryptography in certain areas. However, the impracticality of perfect security is often not a problem, as the main concern is to make the attack to imperfect security instead to be impractical.

7. Special Chapter: Roles of Computers in Cryptanalysis.

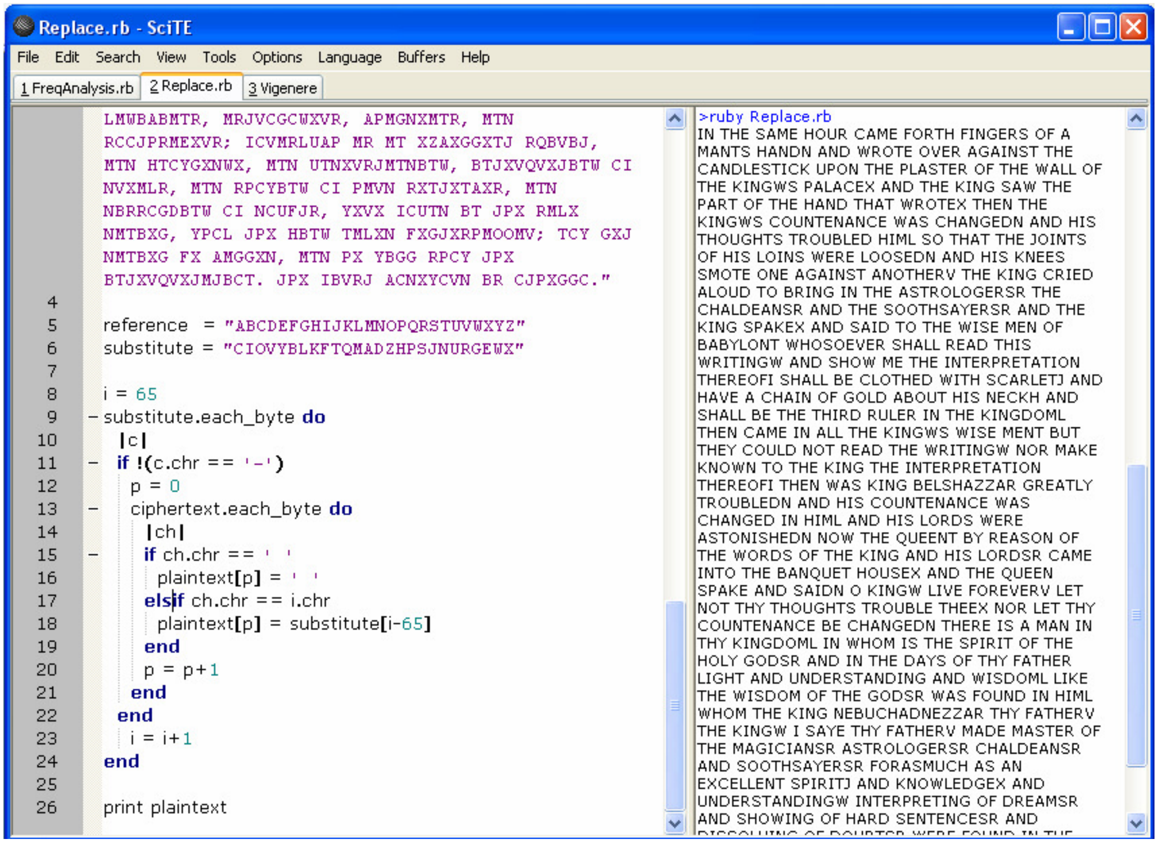
The invention of digital circuit to facilitate calculations had opened the gateway to larger possibilities of encryption algorithms. Similarly, cryptanalysis technique was made better. Brute-force attack is more feasible, causing current encryption to require large key size. For example RSA now requires 1024- bit key to be secure.

To conclude the report, here are some codes in Ruby, useful to attack classical ciphers such as Caesar Cipher and Vigenere Cipher.

<u>Frequency Analysis</u>



Replacing characters.



This Ruby program helps in substituting alphabet as noted in the part

> reference = “ABCDEFGHIJKLMNOPQRSTUVWXYZ”
> substitute = “CIOVYBLKFTQNADZHPSJNURGEWX”

which means occurrence of ‘A’ will be replaced by ‘C’, ‘B’ by ‘I’, ‘C’ by ‘O’ and so on.

Analysing Occurrences in Vigenere Cipher.

The right side shows the occurrence and the index where the repetitions happen which will help to perform Kasiski method on Vigenere Cipher.

Viginere.rb * SciTE

File Edit Search View Tools Options Language Buffers Help

1 FreqAnalysis.rb | 2 Replace.rb | 3 Viginere.rb *

```ruby
ciphertext =
"KQOWEFVJPUJUUNUKGLMEKJINMWUXFQMKJBGWRLFNFGHUDWUUMBSVLPSNCMUEKQCTESWREEKOYSSIWCTUAXYOTAPXPLWPNTCGOJBGFQ
HTDWXIZAYGFFNSXCSEYNCTSSPNTUJNYTGGWZGRWUUNEJUUQEAPYMEKQHUIDUXFPGUYTSMTFFSHNUOCZGMRUWEYTRGKMEEDCTVRECFBD
JQCUSWVBPNLGOYLSKMTEFVJJTWWMFMWPNMEMTMHRSPXFSSKFFSTNUOCZGMDOEOYEEKCPJRGPMURSKHFRSEIUEVGOYCWXIZAYGOSAANY
DOEOYJLWUNHAMEBFELXYVLWNOJNSIOFRWUCCESWKVIDGMUCGOCRUWGNMAAFFVNSIUDEKQHCEUCPFCMPVSUDGAVEMNYMAMVLFMAOYFNT
QCUAFVFJNXKLNEIWCWODCCULWRIFTWGMUSWOVMATNYBUHTCOCWFYTNMGYTQMKBBNLGFBTWOJFTWGNTEJKNEEDCLDHWTVBUVGFBIJGYY
IDGMVRDGMPLSWGJLAGOEEKJOFEKNYNOLRIVRWVUHEIWUURWGMUTJCDBNKGMBIDGMEEYGUOTDGGQEUJYOTVGGBRUJYS"

seqLength = 4
i = 0
j = 0
occurence = 0

ciphertext.each_byte do
 if (i+2 < ciphertext.length)
  j = 0
  occurence = 0

  ciphertext.each_byte do
   if ciphertext[i..i+seqLength-1] == ciphertext[j..j+seqLength-1]
    occurence = occurence + 1
   end
   j = j+1
  end

 if occurence > 1
 print i , "       " , ciphertext[i..i+seqLength-1], " = #{occurence} \n"
 end

 i = i+1
 end
end
```

```
>ruby Viginere.rb
4      EFVJ = 2
105    WXIZ = 2
106    XIZA = 2
107    IZAY = 2
108    ZAYG = 2
154    EKQH = 2
176    NUOC = 2
177    UOCZ = 2
178    OCZG = 2
179    CZGM = 2
193    EEDC = 2
224    EFVJ = 2
256    NUOC = 2
257    UOCZ = 2
258    OCZG = 2
259    CZGM = 2
263    DOEO = 2
264    OEOY = 2
295    WXIZ = 2
296    XIZA = 2
297    IZAY = 2
298    ZAYG = 2
308    DOEO = 2
309    OEOY = 2
349    IDGM = 3
374    EKQH = 2
438    FTWG = 2
440    WGMU = 2
483    FTWG = 2
493    EEDC = 2
514    IDGM = 3
560    WGMU = 2
574    IDGM = 3
>Exit code: 0
```

8. Reference

1. Cryptography, Theory and Practice (Third Edition). Douglas R. Stinson. Chapman & Hall/CRC. 2006

2. An Introduction to Cryptography (Second Edition) Richard A. Mollin. Chapman & Hall/CRC 2007

3. Information Security Principles and Practice. Mark Stamp. Wiley. 2006

4. Cryptography, Information Theory and Error Correction: A Handbook for the 21st Century. Aiden A. Bruen, Mario A. Forcinito. Wiley. 2005.

5. Applied Cryptography (Second Edition). Bruce Schneier. John Wiley & Sons, Inc. 1996