Cryptology: Lab Assignment

Cryptanalysis of the Vigenere Cipher

Usama Zafar
Group # 13

February 27, 2024

Part B: Breaking the Cipher

In this part of the lab assignment we perform cryptanalysis of the modified Vigenere Cipher and try to break different encrypted texts provided. Based on the separately submitted code we were able to break all Cipher texts provided.

Running the Code

The cryptanalysis program is written using Python programming language. The user will need an appropriate Python interpreter install as well as numpy library to run the code. Once everything is at hand the program can be simply run using the commands:

```
>> python cryptanalysis_shortkeys \
                --crypto_path={crypto_file}
                --min_key={minimum_key_size}
                --max_key={maximum_key_size}
```

```
>> python cryptanalysis_longkeys \
                --crypto_path={crypto_files_folder}
                --min_key={minimum_key_size}
                --max_key={maximum_key_size}
```

For the short keys version the program reads the crypto-text from the specified crypto_file and tries to decrypt it using various techniques while adhering the to key size limits provided. For the long key version it reads all crypto texts under a specific folder and analyzes them together with assumption that all crypto-texts are encrypted with same key.

Cryptanalysis Techniques

The implemented program uses the following procedure to break the ciphertexts:

1. Estimate the key size m of a given ciphertext using the Kasiski and Friedman tests.
2. Split the ciphertext into m chunks that are m distance apart.
3. Treat each of the chunks as a separate Caesar cipher and again use Friedman's index of coincidence along with the Swedish language's character frequency table to find the shift.
4. Once all shifts are found, decrypt the cipher text using the combined shifts as decryption key.

1. Determine Key Length

In order to determine the key length, we apply both Kasiski test as well as Friedman test. Using kasiski test we obtain matching sub strings and the distance between these sub strings gives us possible key lengths. Next we use Friedman test to compute index of coincidence and verify which of these key lengths is a good enough candidate.

2. Solve individual Ceaser Ciphers

After the key length is determined, we apply a simple frequency analysis on individual Ceaser ciphers m-distance apart in the ciphertext. This way we can determine best possible key at each individual position.

3. Frequency Table of Swedish Characters

The frequency table of swedish characters is given by Table 1. It was obtained from Wikipedia which in turn obtained it from the website: Practical Cyrptogprahy.

Character	Frequency	Character	Frequency	Character	Frequency
a	0.09383	k	0.0314	u	0.01919
b	0.01535	l	0.05275	v	0.02415
c	0.01486	m	0.03471	w	0.00142
d	0.04702	n	0.08542	x	0.00159
e	0.10149	o	0.04482	y	0.00708
f	0.02027	p	0.01839	z	0.0007
g	0.02862	q	0.0002	å	0.0134
h	0.0209	r	0.08431	ä	0.018
i	0.05817	s	0.0659	ö	0.0131
j	0.00614	t	0.07691		

Table 1: Frequency of Swedish Characters

4. Runtime

Both types of ciphertexts i.e. with short-keys and longer-key take approximately 20-30 ms to run. In the updated version of the code attacked I was able to break all ciphers.

Part C: Reasoning about Ciphers

The motivated answers to each of the questions are as follows:

1. Impact of Language

In theory, having increased number of characters (29) in Swedish language as opposed to 26 characters in English language should increases the difficulty of Vigenere Cipher especially when it comes to brute-force attacks. But in practice, with advanced techniques like frequency analysis, Kasiski test, and Friedman test this slight increase becomes irrelevant especially if you have good statistical data of the frequency of characters.

2. Impact of Key Length

The difficulty of breaking Vigenere Cipher is directly proportional to the key length, shorter keys provide more chances of repetitions and identifiable patterns especially when the text being encrypted is significantly longer than the key. Whereas longer keys to some extent mitigate this problem (although not entirely). Additionally, Vigenere Cipher with key as long as the plaintext provides perfect secrecy so naturally the argument regarding key size follows.
Other characteristics that might impact the security of a particular key include the plaintext itself and how it is distributed when the key is repeated.

3. Breaking Cipher of Unknown language

If the language of the ciphertext is unknown, it becomes very difficult to break the cipher as you cannot perform frequency analysis without knowing the target language. However there are still some tricks that can be used to decipher such a ciphertext. You can do the following steps:

1. Determine the character set from the ciphertext. Since you have access to the ciphertext you can easily determine what characters the language has (Latin derived languages are fundamentally different from say Persian or Arabic derived languages).

2. Use the kasiski tests to determine possible key lengths.

3. Using the character set derived in step 1, try to decipher the text. Here we can use the frequency table of as many major language of the category of languages that the ciphertext belongs to as we can.

4. Finally, the broken cipher would need to be verified with dictionary of most of the major languages so that the correct one is selected.

4. Modified Vigenere Ciphers

(a) Using the reverse of plaintext as key make the cipher similar to one-time pad. We observe the following about it:

1. This cipher defines a crypto system as it has a finite plaintext pool, finite cyrptotext pool, a finite key-space as well as an encryption and decryption rule.

2. Since this is similar to one-time pad it is unconditionally secure because the key is unique and as long as the message itself.

3. Given a message of length n, the computational cost of brute force attack would require $29^{\frac{n}{2}}$ attempts since the plaintext is mirror of the ciphertext. Knowing one-half automatically reveals information about the other half.

4. As the cipher is unconditionally secure, it would be near impossible (computationally infeasible) to break the cipher for longer text messages. For smaller text messages you could carry out brute force attack. The only vulnerability such a system might have is related to the key exchange and handling.

(b) Using subtraction instead of addition simply makes Vigenere cipher kind of a symmetric in terms of encryption and decryption without much impact on other properties. We note that:

1. This cipher defines a cryptosystem same as Vigenere cipher.

2. The cipher would not be more secure than the original Vigenere cipher as this cipher still has same vulnerabilities. i.e. is is susceptible to frequency analysis, Kasiski test, and Friedman's index of coincidence analysis. In summary it will be as secure as the Vigenere cipher itself.

3. Given a message of length n, the computational cost of brute force attack would still require 29^n attempts.

4. Similarly to breaking Vigenere cipher, I would use Kasiski test, and index of coincidence analysis in conjunction with frequency analysis to break it.

(c) Transforming the plain text before encryption adds confusion to the encrypted cipher text. We note that:

1. This cipher defines a cryptosystem because it involves encryption and decryption algorithms and a key.

2. Since the cipher adds another layer of confusion to the plain text I suspect it would be more secure than the original Vigenere cipher. Such a cipher would transform the plaintext to hide the statistical properties of characters making it secure against Kasiski test and frequency analysis. However this is just a conjecture, how secure it is compared to original Vigenere cipher is unclear.

3. The brute force attack on this modified cipher would still involve trying every possible key for the Vigenere cipher, but now with the additional complexity of considering the obfuscation / transformation.

4. In order to break such a cipher, cryptanalysis of this modified cipher would involve analyzing both the transformation and Vigenere cipher encryption. The frequency analysis, Kasiski test and Friedman test would still be relevant but would need to be done in different systematic way that takes the transformation into account.

(d) Appending plain text to the key in an interesting idea that makes the key management very easy and to some extend yields a one-time pad like cipher. We note the following about this cipher:

1. This cipher still defines a cryptosystem because it involves encryption and decryption algorithms and a key.

2. This modification creates a longer key and in turn increases the complexity of the cipher by appending the plaintext to the initial key making it safer than Vigenere cipher unless a part of the plaintext is known. If part of the plaintext is known along with its position, it would be possible to decryption the subsequent text in forward direction. The security of this technique heavily relies on the security properties of the initial key (including randomness, length etc).

3. The brute force attack on this modified cipher would involve guessing every possible key combination and the time will depend on the length of the key. Shorter keys can be guess quickly while longer initial keys would take more time.

4. Cryptanalysis of this modified cipher can be carried out by a brute force attack on the key from the beginning until a set of possible keys is obtained. Plugging them in and checking against a known dictionary would make it possible to break this cipher. This technique would be vulnerable to brute force attack if the key length is small.

(e) Similar to previous but appending cipher text instead of plain text is again an interesting idea but might have same vulnerabilities as the previous one. We note the following about this cipher:

1. This cipher still defines a cryptosystem because it involves encryption and decryption algorithms and a key.

2. This modification although creates a longer key and increases the complexity of the cipher by appending the plaintext to the initial key it however is more vulnerable than Vigenere cipher. It is also vulnerable to known plain text-attacks. Since a single known plaintext (or part thereof) can yield all the information regarding encrypted text.

3. The brute force attack on this modified cipher would involve guessing every possible key combination and the time will depend on the length of the initial key. In theory it would be a onetime pad hence an infinitely secure but in practice shorter initial keys can be guessed quickly. Hence a smart brute force attack might be able to crack it in reasonable time.

4. Similar to cryptanalysis of previous approach a brute force attack on the key from the beginning part of the encrypted text until a set of possible keys is obtained would be the best way to break this cipher. Plugging all possible key candidates and checking against a known dictionary would make it possible to break this cipher. Another more effective method would be to apply a rolling text attack i.e. compare the text against itself while rolling along until a perfect match (based on known dictionary) is found. This would immediately reveal the entire plaintext except for the initial part encrypted with a key. As a final step a brute force attack can be applied to the initial part of the text.

Conclusion

In this lab we first implemented Vigenere Cipher, then tried to break a bunch of ciphertexts using various cryptanalysis techniques. The hardest part for me was to understand the cryptanalysis techniques and how they can be used in practice. Implementation part was relatively simple. I was not able to break longer keys in the first submission but found a bug with my implementation and subsequently was able to break longer keys as well in this submission.