ECE 471/571: In Class Problem #2

Cryptanalysis of the Vigenére Cipher

Electrical and Computer Engineering, University of Arizona,
Ming Li

1 Cryptanalysis of the Vigenére Cipher

Determine the plaintext given that the following ciphertext was generated using the Vigenére Cipher. Hint: first find the key vector length, and then find the key vector.

CHREEVOAHMAERATBIAXXWTNXBEEOPHBSBQMQEQERBWRVXUOAKXAOSXXWE
AHBWGJMMQMNKGRFVGXWTRZXWIAKLXFPSKAUTEMNDCMGTSXMXBTUIADNGMGPSR
ELXNJELX VRVPRTULHDNQWTWDTYGBPHXTFALJHASVBFXNGLLCHRZBWELEKMS
JIKNBHWRJGNMGJSGLXF EYPHAGNRBIEQJTAMRVLCRREMNDGLXRRIMGNSNRWCHR
QHAEYEVTAQEBBIPEEWEVKAKOEWADR EMXMTBHHCHRTKDNVRZCHRCLQOHPWQA
IIWXNRMGWOIIFKEE

1.1 Finding key vector length m

The main idea is to use trial and error in increasing length of vector to find the correct frequency. Find plaintext + key = cipher. Then arrange the cipher as a vector of size m as follows.

$$\begin{aligned} \underline{\mathbf{y_1}} &= y_1 \;\; y_{m+1} \;\; y_{2m+1}\cdots \\ \underline{\mathbf{y_2}} &= y_2 \;\; y_{m+2} \;\; y_{2m+2}\cdots \\ &. \\ &. \\ \underline{\mathbf{y_m}} &= y_m \;\; y_{2m} \;\; y_{3m}\cdots \end{aligned}$$

Every element of $\underline{\mathbf{y_i}}$ is shifted by the same key K_i of the vector key $\underline{K}$. That means that each $\underline{\mathbf{y_i}}$ is the result of encryption with a monoalphabetic cryptosystem.
Index of Coincidence of $\underline{\mathbf{x}}$: Let $\underline{\mathbf{x}} = (x_1, ..., x_n)$ be a string of alphabetic characters. Let the number of times A, B, C,...,Z appearing in $\underline{\mathbf{x}}$ be denoted as $f_0, f_1, .., f_{25}$. The Index of coincidence $I_c(\underline{\mathbf{x}})$ is the probability that two random elements of $\underline{\mathbf{x}}$ are identical. We can write the index of coincidence as:

$$I_c(\underline{\mathbf{x}}) = \sum_{i=0}^{25} \frac{\binom{f_i}{2}}{\binom{n}{2}} \tag{1}$$

Ideally, $I_c(\underline{\mathbf{x}}) = \sum_{i=0}^{25} p_i^2$ where p_i is the empirical probability (Table 1) for alphabet with index i.

So if we have the correct key vector length m, the elements of $\underline{\mathbf{y_1}} = y_1 \;\; y_{m+1} \;\; y_{2m+1}\cdots$; will have $I_c(\underline{\mathbf{y_1}}) \approx \sum_{i=0}^{25} p_i^2 \approx 0.065$. Else it will be close to $\sum_{i=0}^{25} \left(\frac{1}{26}\right)^2 = \frac{1}{26} = 0.038$, which is the index of coincidence for random character distribution.

Kasiski Test: The idea behind the Kasiski test is that it is quite infrequent to find pairs of identical segments of ciphertext of length at least three, unless these segments are the result of the encryption of the same plaintext. In that case, the distance δ of occurrence of the identical segment must be a multiple of m. That is: $\delta \equiv 0 \pmod 0$.

To find the period of the Vigenére cipher using the Kasiski test we execute the following steps:

1. Search ciphertext for pairs of identical segments with length at least 3.
2. Record distances between the *starting* positions of the segments.
3. Take Greatest Common Divisor (*gcd*) of these distances as the key vector length m.

1.2 Finding the key vector $\underline{\mathbf{K}}$

Look at $\underline{\mathbf{y_j}} = y_j \;\; y_{m+j} \;\; ...y_{\left(\frac{n}{m}-1\right)m+j}$. Let $\frac{n}{m} = n'$. Let $\hat{f}_0, \hat{f}_1, .., \hat{f}_{25}$ be the number of times (frequency) $A, B, ..., Z$ appear in $\underline{\mathbf{y_j}}$. Then the probability of alphabet corresponding to index i is:

$$\hat{p_i} = \frac{\hat{f}_i}{n'} = \frac{\hat{f}_i m}{n}. \tag{2}$$

Each $\underline{\mathbf{y_j}}$ has been associated with fixed key K_j. i.e. $\hat{f}_i = f_{i+K_j}$ where i, $i + K_j$ correspond to the indices of alphabets. Let $0 \leq g \leq 25$ and define M_g to be:

$$M_g = \sum_{i=0}^{25} \frac{p_i f_{i+g}}{n'}. \tag{3}$$

Based on this we can try different shifts $g \in \{0, ..., 25\}$ and check if:

$$M_g = \sum_{i=0}^{25} \frac{p_i f_{i+g}}{n'} \approx \sum_{i=1}^{25} p_i^2 = 0.065. \tag{4}$$

where p_i is the empirical probability (Table 1) for alphabet with index i. If the above holds, then we know that $g = K_j$. The resulting shift satisfying eq. (4), is the value for key K_j. Similarly we can find values for the remaining keys to finally get the key vector $\underline{\mathbf{K}}$.

1.3 Appendix

Table 1. Probabilities of occurence of the 26 letters in English plaintext

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

Helpful tool: http://www.cryptoclub.org/ (go to Cipher Tools → Cracking tools → Crack Vigenére)