MATC16 Cryptography and Coding Theory
Gábor Pete
University of Toronto Scarborough

Why the attack on Vigenère works

First recall a result we had already used in the fancy version of the frequency analysis, basically the lemma of the 5-year-old child:

Lemma 1. *Let* $\mathbf{A}_0 = (.082, .015, .028, \dots, .020, .001)$ *be the frequencies of* $(a, b, c, \dots, y, z)$ *in English, and let* $\mathbf{A}_i$ *be the same vector shifted cyclically by* i *entries to the right, for* $i = 0, 1, \dots, 25$. *E.g.,* $\mathbf{A}_1 = (.001, .082, .015, \dots, .020)$. *Then the dot product* $\mathbf{A}_i \cdot \mathbf{A}_j$ *depends only on* $|i - j|$, *and is maximized when* $i = j$, *with value* $.066$.

Consider a plaintext $x_1 x_2 x_3 \dots$, a key $k_1 k_2 k_3 \dots$ given by a keyword of length p, so that $k_i = k_{\ell p + i}$ for all i, ℓ, and the resulting ciphertext $y_1 y_2 y_3 \dots$, with $y_i = x_i + k_i \pmod{26}$.

Let's estimate the frequency of coincidences between $y_1 y_2 y_3 \dots$ and its displacement by t places, i.e., the fraction of places i with $y_i = y_{i-t}$. The two letters we see at a typical place i roughly follow the typical frequencies of English letters, except, of course, that y_i is like the typical x_i shifted by k_i, while y_{i-t} is like the also typical x_{i-t} shifted by k_{i-t}. That is, the frequency of the ciphertext letter $A = 0$ as y_i has the English frequency of $0 - k_i$, the frequency of $B = 1$ has the English frequency $1 - k_i$, and so on, i.e., the frequency vector for y_i is $\mathbf{A}_{k_i}$. Similarly, the frequency vector for y_{i-t} is $\mathbf{A}_{k_{i-t}}$.

Now, to get an agreement $y_i = y_{i-t}$, we can have two A's or two B's, and so on, so have to add up the frequencies for these 26 possible agreements. If the displacement t is large enough, say at least 3, then, for a typical i, the English plaintext letters x_i and x_{i-t} are quite independent, hence the frequency of a pair of letters as (x_i, x_{i-t}) is roughly the product of the two frequencies. So, the frequency of agreements $y_i = y_{i-t}$ is roughly a sum of 26 pairwise products, namely, $\mathbf{A}_{k_i} \cdot \mathbf{A}_{k_{i-t}}$.

By the lemma, this is largest when $k_i = k_{i-t}$. Thus, we expect the largest frequency of coincidences when t is a multiple of the period p. So, if we see much more coincidences at displacements, say $t = 4, 8, 12, \dots$, than at other values, then our best guess for the keyword length is $p = 4$.

This method is unsafe for displacements $t = 1$ and 2, because of the correlations between nearby letters in the English plaintext. However, if p is the real length, then $t = \ell p$ will give many coincidences for each $\ell = 1, 2, \dots$. Therefore, many coincidences for displacement 2, say, but only few for 4, will mean that the length is probably not 2.

Once you have the keyword length p, do frequency analysis on the ciphertext letters $y_j, y_{p+j}, y_{2p+j}, \dots$ to find the shift there, for each $j = 1, 2, \dots, p$. Combine these shifts to get the keyword.