MATH 587/CSCE 557 - SUMMARY OF CLASS, 2/6/07

Class Feb 1 was canceled because of bad weather. People taking the class should make sure they're getting their VIP email, since it is no longer automatically forwarded. Class announcements are made to the VIP mailboxes.

After announcing that the 1st midterm will be held in class on Feb 13 (it will cover the material on the first three homeworks) and that class on Feb 8 will focus on review for the test, I went over the solutions to the 2nd homework. Then I read from the Code Book by Simon Singh where he suggests that Babbage's solution to cracking the Viginere cryptosystem was kept quiet for 9 years so that Britain would have an edge in the Crimean War. I also read from the same book about how the ADFGX code was used by the Germans in World War I and cracked.

We discussed the different kinds of attacks some more. For shift ciphers, one plaintext is enough to find the key for a chosen plaintext attack or a known plaintext attack (since if input x leads to output y, then the key is y-x mod 26). For ciphertext only attacks frequency analysis can be used, which is discussed below.

For affine ciphers, with a chosen plaintext attack, if the encryption function is $e_k(x) = ax + b$, input 0 leads to output b and input 1 leads to output $a + b$, from which a can now be found. So two plaintext letters suffice. For a known plaintext attack, if input x_1 leads to output y_1 and input x_2 leads to output y_2, then $y_1 = ax_1 + b$ and $y_2 = ax_2 + b$. Subtracting, $y_1 - y_2 = a(x_1 - x_2)$. We can solve this for a if $x_1 - x_2$ is invertible mod 26 (otherwise there may be a solution but it won't be unique). The chances of this happening are heuristically $12/26 = 6/13$.

For a ciphertext only attack, we use frequency analysis. Here's an example.

JFFGJFDMGFSJHYQHTAGHQGAFDCCFP

The most frequent letter is F (6 times) and G (4 times), so we guess that F decrypts to E and G decrypts to T, i.e. $d_k(5) = 4, d_k(6) = 19$. If $d_k(x) = cx + d$, then $5c + d = 4, 6c + d = 19$. Subtracting, $c = 15$. Plugging back in, $75 + d = 4$, so $d = -71 = 7(\pmod{26})$. Applying $d_k(x) = 15x + 7$ to the entire message yields

MEETMEAFTERMIDNIGHTINTHEALLEY

which makes sense when spaces are inserted. For a shift cipher you'd guess which letter is E and then treat it like a known plaintext attack. If it doesn't work first time, try another letter.

For general substitution ciphers, for chosen plaintext attacks it takes 25 inputs since you need to know what A,B,...,Y map to before you know the key for sure (Z goes to whatever is left over). For known plaintext attacks, the question amounts to how long the text needs to be before 25 different letters of the alphabet arise (of course you can probably guess most of it from fewer than 25). For chosen plaintext attacks, frequency analysis together with guesswork will usually do it. (See next time.)

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX