MATH 587/CSCE 557 - SUMMARY OF CLASS, 3/8/07

There is no homework today because of spring break. I offered the class the option of handing in homework after spring break since some are struggling with breaking the Vigenere cipher for homework 6. The problem seems to be an obsession with E's. For Singh's cipher challenge, for each of the cosets the most frequent letter stood for E. This is not guaranteed in general. Sometimes assigning the most frequent letter of the coset to stand for E can make the 2nd most frequent letter stand for something silly like Q or Z. Refining this idea can lead to a correct decipherment. Alternatively the method of computing the k which maximizes $S_k = p_0q_k + p_1q_{k+1} + ... + p_{25}q_{k-1}$, where $(p_0, ..., p_{25})$ stands for the typical frequencies of English letters and $(q_0, ..., q_{25})$ stand for the frequencies of A,B,...,Z in the coset, works very well.

Note that $(p_0, ..., p_{25}) = (0.082, 0.015, 0.028, 0.043, 0.127, 0.022, 0.020, 0.061, 0.070, 0.002, 0.008, 0.040, 0.024, 0.067, 0.075, 0.019, 0.001, 0.060, 0.063, 0.091, 0.028, 0.010, 0.024, 0.001, 0.020, 0.001)$

We then computed the unicity distance (or nominal unicity point) for the general substitution cryptosystem. As noted last time, it's $\log_2(|K|)/R = \log_2(26!)/3.1$, which is about 29. This means that 30 letters should be about enough to ensure unique decipherment, whereas for say 20 letters there may be several possible decipherments.

I described perfect secrecy. This means that, given the ciphertext, we are no wiser as to what the plaintext is. For instance, if we have a ciphertext of 9 letters encrypted by a Vigenere keyword of length 9, then the plaintext could be anything. The limit as keyword length $\to \infty$ is the idea behind the one-time pad.

We next discussed shift registers, which are meant to mimic one-time pads. These are stream ciphers where the plaintext is $x_1x_2...$, the key is a stream $k_1k_2...$, and the ciphertext $y_1y_2...$ is obtained by adding, so $y_i = x_i + k_i$. Usually this is addition $\pmod{26}$ (as with Vigenere) or $\pmod 2$ as with bit streams. If $k_1k_2...$ were random, then it would be a one-time pad. The goal is to generate a key stream that seems random.

A linear feedback shift register (LFSR) with n cells is represented by a sequence $x_0, x_1, x_2, ...$ with $x_n = c_0x_0 + c_1x_1 + ... + c_{n-1}x_{n-1}$, and then $x_{n+1} = c_0x_1 + c_1x_2 + ... + c_{n-1}x_n$ etc. where x_i is 0 or 1 for each i. Considering the case $c_0 = c_1 = 1, c_2 = c_3 = 0$ and $x_0 = x_1 = x_2 = x_3 = 1$, $x_{i+4} = x_i + x_{i+1}$ and we get sequence 111100010011010111000..., which is a pseudorandom bit stream periodic with period 15. I noted how using the theory of finite (Galois) fields you can get similar sequences with period $2^n - 1$ for any n. This grows fast with n so it looks like we are getting good pseudorandom bit streams this way. We will cryptanalyze these next time.