CO 487/687: Assignment #1 Due date: January 23 (Monday)

NOTE: The material for Problems #1, #2, #3 will be covered in class on January 6, 9 and 11, respectively. Assignments will be submitted via Crowdmark; you will be given instructions for using Crowdmark on January 13.

1. **Simple substitution cipher** (10 marks)
   Retrieve "page*xy*" from the "Assignment #1" section on the course web site, where *xy* are the last two digits of your student ID number. This page contains ciphertext that was generated using a simple substitution cipher.

   The secret key (a permutation of the English alphabet) was generated from a *key letter* and a *key word*. The key word is an English word (names of cities and countries are allowed) having no repeated letters. The secret key is then derived by writing the key word beneath the key letter in the alphabet, and then writing the remaining letters of the alphabet in cyclic order after the key word. For example, if the key letter is **g** and the key word is **SLOPE**, then the secret key is:

   a b c d e f **g** h i j k l m n o p q r s t u v w x y z
   U V W X Y Z **S L O P E** A B C D F G H I J K M N Q R T

   Using this secret key, the plaintext `cat` is encrypted to `WUJ`.

   All punctuation and spaces were removed from the plaintext, which was then blocked off into groups of 5 letters prior to encryption. Your task is to recover the *key letter*, the *key word*, and the name of the *book* from the which the plaintext was taken.
   You can solve this problem by hand, by writing a computer program, or by using any free software you can find on the internet.
   Please submit the key letter, key word, name of book, and a *brief* (at most half a page) description of the procedure you used to find the key letter and key word.

   The following are the letters of the English alphabet, grouped by letters whose frequencies are approximately equal. The letters in each group are listed in order of decreasing frequency.

   Group 1: e
   Group 2: t a o i n s h r
   Group 3: d l
   Group 4: c u m w f g y p b
   Group 5: v k j x q z

   It may also help to know that the most commonly occuring digrams in the English language, in decreasing order of frequency, are:

   th he in er re on an en at es ed te or ti st

2. **Weakness in RC4** (4+4+2 marks)

   (a) Suppose that after the RC4 key scheduling algorithm has been executed we have $S[1] \neq 2$ and $S[2] = 0$. Prove that the second keystream byte of RC4 is 0.

   (b) (*This exercise shows that the second keystrem byte of RC4 is biased towards 0.*) Estimate the probability that the second keystream byte of RC4 is 0. (The probability is assessed over all secret keys.) State any assumptions you may make.

   (c) Let m be a message consisting of at least two bytes. Suppose that Alice uses RC4 to encrypt m for 1024 different users. (For each user, m is encrypted with the secret key that Alice shares with that user.) Show that an eavesdropper who captures the resulting 1024 ciphertexts has a good chance of recovering the second byte of the plaintext.

3. **Feistel ciphers** (3+3+4 marks)
Recall that Feistel ciphers are a class of block ciphers with parameters n (half the block length), h (the number of rounds), and l (the key size). Then $M = \{0,1\}^{2n}$ (the plaintext space), $C = \{0,1\}^{2n}$ (the ciphertext space), and $K = \{0,1\}^{\ell}$ (the key space). A key scheduling algorithm determines subkeys $k_1, k_2, \ldots, k_h$ from a key k. Each subkey k_i determines a function $f_i : \{0,1\}^n \to \{0,1\}^n$. Encryption takes h rounds:

Plaintext is $m = (m_0, m_1)$, where $m_0, m_1 \in \{0,1\}^n$.
Round 1: $(m_0, m_1) \to (m_1, m_2)$, where $m_2 = m_0 \oplus f_1(m_1)$.
Round 2: $(m_1, m_2) \to (m_2, m_3)$, where $m_3 = m_1 \oplus f_2(m_2)$.
.....
Round h: $(m_{h-1}, m_h) \to (m_h, m_{h+1})$, where $m_{h+1} = m_{h-1} \oplus f_h(m_h)$.
The ciphertext is $c = (m_h, m_{h+1})$.

(a) Using notation similar to the description of encryption provided above, give an algorithm for the decryption process.

(b) Consider a Feistel cipher with parameters $n = 128$, $h = 2$, $\ell = 256$. Given $k \in \{0,1\}^{256}$, the key scheduling algorithm simply sets k_1 to be the leftmost 128 bits of k, and k_2 to be the rightmost 128 bits of k. Finally, f_i is defined by $f_i(x) = x \oplus k_i$.
Show that this block cipher is totally insecure—that is, given a single plaintext-ciphertext pair (m, c), the secret key k can be easily recovered.

(c) Consider a Feistel cipher with parameters $n = 128$, $h = 3$, $\ell = 128$. Given $k \in \{0,1\}^{128}$, the key scheduling algorithm sets $k_1 = k$, $k_2 = k'$ (where k' denotes the right cyclic shift of k), and $k_3 = k''$ (where k'' denotes the right cyclic shift of k'). [For example, if $k = 001011$, then $k' = 100101$ and $k'' = 110010$.] Finally, f_i is defined by $f_i(x) = x \oplus k_i$.
Show that this block cipher is totally insecure—that is, given a single plaintext-ciphertext pair (m, c), the secret key k can be easily determined to be one of two possible values.

You should make an effort to solve all the problems on your own. You are also welcome to collaborate on assignments with other students presently enrolled in CO 487. However, *solutions must be written up by yourself.* If you do collaborate, please *acknowledge your collaborators* in the write-up for each problem. *If you obtain a solution with help from a book, research paper, a web site, or elsewhere, please acknowledge your source.* You are *not* permitted to solicit help from online bulletin boards, chat groups, newsgroups, or solutions from previous offerings of the course.

The assignment should be submitted via Crowdmark before 11:00 am on January 23. Late assignments will not be accepted except in *very* special circumstances (usually a documented illness of a serious nature). High workloads because of midterms and assignments in other courses will *not* qualify as a special circumstance.

Office hours:
Alfred Menezes (MC 5026): Monday 3:00-5:00 pm, Friday 1:00-3:00pm
Ted Eaton (MC 5481): Tuesday 10:30-11:30 am
Gabriel Gauthier-Shalom (MC 5113): Thursday 3:00-4:00 pm
Philip Lafrance (MC 5497): Tuesday 12:30-1:30 pm
Christopher Leonardi (MC 5494): Tuesday 2:00-3:00 pm
Cathy Wang (MC 6313): Thursday 2:00-3:00 pm