C&O 487: Assignment #1 Due date: January 23 (Friday)

1. **Vigenère cipher** (15 marks)
Retrieve "pagexy" from the "Assignments" 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 Vigenère cipher as described in class. The secret key word is an English word (names of cities and countries are included) having no repeated letters. 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 secret key word. Please include a *brief* (at most half page) description of the procedure you used to find the 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

2. **Hill cipher** (2+4+4 marks)
Let $n \geq 2$ be a positive integer. Let A be an invertible $n \times n$ binary matrix, and let b be a binary $1 \times n$ vector. In the Hill symmetric-key encryption scheme, the secret key is a pair (A, b). Plaintext messages m are represented as binary $1 \times n$ vectors. The encryption function is $E(m) = mA + b$. Note that all arithmetic is performed modulo 2.
For example, if $n = 5$,

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad b = [1,1,0,1,1], \quad \text{and} \quad m = [1,0,1,0,1],$$

   then $E(m) = [0,1,0,0,1]$.

   (a) Describe a decryption algorithm for the Hill cipher.

   (b) Show how an adversary can determine the secret key (A, b) using a chosen-plaintext attack.

   (c) In a *chosen-ciphertext* attack, the adversary is given a target ciphertext c. She can obtain (by asking Bob, who knows the secret key) the plaintext of any ciphertext of her choice *except for c itself*. Her task is to decrypt c.
   Show how the adversary can decrypt c by obtaining the decryptions of at most 3 ciphertexts from Bob.

3. **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.

4. **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 $K = \{0,1\}^l$ (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 $m_3 = m_1 \oplus f_2(m_2)$.
$\ldots\ldots$
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 $n = 128$, $h = 2$, $l = 256$. 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 $n = 128$, $h = 3$, $l = 128$. 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.

Please note that assignments are not weighted equally. The total marks received on assignments will be added together at the end of the course.

You are welcome to collaborate on assignments with your colleagues. 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, paper, wikipedia, a web site, solutions from previous offerings of the course, *or any other source*, please acknowledge your source. You are not allowed to solicit help from online bulletin boards, chat groups, or newsgroups.

The assignment is due at the *beginning* of class on January 23. Late assignments will not be accepted except in *very* special circumstances.
