CO 487/687: Assignment #1 Due date: January 24 (Wednesday), 11:00 am

---

NOTE: The material for Problems #1, #2, #3 will be covered in class on January 5, 5 and 8, respectively. Assignments will be submitted via Crowdmark. You will be emailed a Crowdmark link for submitting the assignment on January 12. If you do not receive the link, please send an email to ajmeneze@uwaterloo.ca.

1. **Simple substitution cipher** (10 marks)
   Retrieve "pagexy" 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 occurring 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. **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 $n = 5$,

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

(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 challenge 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 (each different from c) from Bob.

3. **RC4 stream cipher** (5+5 marks)

(a) Suppose that after the RC4 key scheduling algorithm has been executed we have $S[1] = 2$. Prove that the first two keystream bytes are different. (This shows that the RC4 keystream bytes are in principle distinguishable from a random keystream, since two randomly selected bytes are equal with probability $\frac{1}{256}$.)

(b) Agents from the National Security Agency (NSA) managed to convince programmers at a software company to implement the slight modification of RC4 described below. (The implementation will be used to secure a communications product intended for foreign markets.)

RC4* Key Scheduling Algorithm:
Input: Secret key $K[0], K[1], \ldots, K[d-1]$.
Output: $S[0], S[1], \ldots, S[255]$.

For i from 0 to 255 do:
    $S[i] \leftarrow i$
    $\overline{K}[i] \leftarrow K[i \bmod d]$
$j \leftarrow 0$
For i from 0 to 255 do:
    $j \leftarrow (\overline{K}[i] + S[i] + j) \bmod 256$
    Swap$(S[i], S[j])$

RC4* Keystream Generator:
Input: $S[0], S[1], \ldots, S[255]$.
Output: Keystream.

$i \leftarrow 0$; $j \leftarrow 0$
While keystream bytes are required do:
    $i \leftarrow (i + 1) \bmod 256$
    $j \leftarrow (j + 1) \bmod 256$
    Swap$(S[i], S[j])$
    $t \leftarrow (S[i] + S[j]) \bmod 256$
    Output$(S[t])$

Suppose now that Alice, using her secret 128-bit (16-byte) RC4* key, sends Bob a ciphertext c that is 1 Megabyte long. Suppose that an eavesdropper Eve intercepts c and is somehow able to learn the first 256 bytes of the plaintext corresponding to c. Can Eve learn anything about the remainder of the plaintext? (Explain)

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 487/687. 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 24**. Late assignments will not be accepted except in *very* special circumstances (usually a documented illness of a serious nature). A high workload because of midterm tests and assignments in other courses will *not* qualify as a special circumstance.

Instructor and TA office hours:

Monday:	1:00 pm – 2:00 pm	Alessandra Graf (MC 5029)
	3:00 pm – 5:30 pm	Alfred Menezes (MC 5026)
Tuesday:	10:30 am – 11:30 am	Priya Soundararajan (MC 5466)
	11:30 am – 12:30 pm	Sam Jaques (QNC 4114) [no office hour on Jan 9]
	1:00 pm – 2:00 pm	Luis Ruiz-Lopez (MC 5486) [no office hour on Jan 9]
	3:00 pm – 4:00 pm	Elena Bakos Lang (MC 5474) [no office hour on Jan 9]
Thursday:	2:00 pm – 3:00 pm	Chris Leonardi (MC 5494)
Friday:	1:00 pm – 3:00 pm	Alfred Menezes (MC 5026)