Solutions to Problem Set 1

Problem 1: Cracking the Hill cipher

Suppose we are told that the plaintext

`breathtaking`

yields the ciphertext

`RUPOTENTOIFV`

where the Hill cipher is used, but the dimension m is not specified. Determine the encryption matrix. (See lecture notes, week 2, for details on the Hill cipher. Note that letters of the alphabet are encoded by the integers 0…25, and all arithmetic is performed modulo 26.)

Solution: (Thanks for Jianye Lu to let me use some of his results)

(a) This question gives you a sense of how to use known plaintext attack, our decipher requires $\lfloor 12/m \rfloor \geq m$, which means m can only be 1, 2 or 3. Otherwise, we can't solve the linear equation. You should check $m = 2$ to see whether the key is consistent. We show how to get the solution for case $m = 3$.

i. The message is divided into vectors m_i of 3 letters each: $m_1 = \{1, 17, 4\}$, $m_2 = \{0, 19, 7\}$, $m_3 = \{19, 0, 10\}$ and $m_4 = \{8, 13, 6\}$;

ii. We want to know that the three vectors m_1, m_2, and m_3 are linearly independent. We have to check that no linear combination of two of them yields the third. The easiest way to do that is to go ahead and attempt Gaussian Elimination. If the vectors are not linearly independent, we will get stuck trying to find a suitable pivot element.

But here, Gaussian elimination is not well defined when working over a ring such as $\mathbf{Z}_{26}$ that is not a field. In particular, the even numbers and 13 do not have inverses, so we can't use them as pivot elements. It can happen that no candidate pivot elements are relatively prime to 26, even though the matrix does have an inverse. Thus, Gaussian Elimination can fail even when the vectors are linearly independent and the matrix inverse does exist. One fix for this is to reduce the matrix mod 2 and mod 13 to get two new matrices. Use Gaussian elimination on each to find their inverses, then combine them together using the Chinese Remainder Theorem to get the inverse mod 26.

Fortunately, in our case, we have three linearly independent vectors $\{m_1, m_2, m_3\}$, and Gaussian Elimination, when modified to search for a pivot element that is relatively prime to 26, does succeed, so the problem can be solved without resort to Chinese remaindering.

iii. Solve $K_{3\times3}$ with $\{m_1, m_2, m_3, c_1, c_2, c_3\}$ via Gaussian Elimination (remember to implement all operations modulo 26). Or, we may use **matinvert** provided for matrix inverting modulo 26. In either way, we have

$$K_{3\times3} = \begin{bmatrix} | & | & | \\ c_1 & c_2 & c_3 \\ | & | & | \end{bmatrix} \cdot \begin{bmatrix} | & | & | \\ m_1 & m_2 & m_3 \\ | & | & | \end{bmatrix}^{-1} = \begin{bmatrix} 3 & 4 & 6 \\ 21 & 15 & 14 \\ 20 & 23 & 5 \end{bmatrix}$$

iv. Verify $K_{3\times3}$ with all plaintext and cipher pairs.

(b) One can easily assert that there is no encryption matrix for $m = 2$ or $m = 1$ by applying the above process.

(c) Some people forget to do (or mention) the verification of the key and case for $m = 2$ and $m = 1$.

Problem 2: Decrypting a substitution cipher

The file "ciphertext" in the `http://zoo.cs.yale.edu/classes/cs467/2005s/course/assignments/ps1/` subdirectory contains encrypted text using a substitution cipher . The set of valid characters is ASCII characters $32 \ldots 126$. Characters outside of this range (e.g., newline) are left unchanged. Decipher the message. Briefly describe the method that you used. (You will probably want to write some code to help you [1].)

Solution: If you are familiar with Sherlock Holmes's story called *The Adventure of the Dancing Men*,[2] we are essentially doing the same thing as there. Most people did a great job here. So, here is the outline of the cracking:

(a) We used the dictionary attack to crack this problem. We first need to assume it is written in English then sort all single characters by its frequency.

(b) The most frequently used character must be white space, so we know the separation of the article.

(c) Sort all the characters in words with lengths 2, 3, 4, 5 by their frequency.

(d) Now, we can compare this with the most frequently used English characters table online or even the frequency sorting of an English article you find.

(e) Try to guess the first few character mapping. Note that, the word "the" is very useful when you recognize it.

(f) Apply the partial map you get to the text to see something more interesting.

(g) Repeat the last two steps until most texts are clear.

(h) When you see the URL, you know the answer. With the help of computer and Internet, seems you can be a better Holmes than the real one.

[1] For your reference, the key generation program and the enciphering program are in perm_gen.cc and subciph.cc, resp.

[2] See `http://www.bakerstreet221b.de/canon/danc.htm` for this cryptographic story online. Thanks for Jianye Lu to point out the link.

Here is the final plaintext:

```
The encryption algorithm used in the TI DST tags is an unpublished,
proprietary cipher that uses a 40-bit key. The algorithm was designed
in the early 1990's by engineers at Texas Instruments, but is still
being deploying in current systems. By today's standards, a 40-bit key
is unacceptably short: advances in computing power have made such keys
susceptible to brute-force key guessing attacks. Therefore, the actual
security of the DST system rests with the secrecy of the proprietary
algorithm used in the tags. One of the most important principles in
cryptographic design states, however, that the security of a system
should be based only on the secrecy of the keys, never on the secrecy
of the algorithm.  (From http://rfid-analysis.org/)
```

Problem 3: Entropy, redundancy, and its use in enabling cryptanalysis

Textbook, exercise 3.11.
[Use the definition of redundancy given in the textbook rather than the slightly different version given in the notes.]

Solution: This is the easiest one, You should read through the textbook and apply the principle showed in the example to the new problem. The calculation is similar to what the author did in the textbook. The entropy for naming these four attacks can reasonably be as low as

$$4.7 + 2 = 6.7$$

There are 39 characters in the strings (including hyphens). So the average length of the four names is $39/4 = 9.75$. Therefore, the average number of bits per letter in these long names is

$$6.7/9.75 \text{ (bits per letter).}$$

The redundancy of these names is

$$\frac{6.7 - 6.7/9.75}{6.7} > 89\%.$$

Depending on how you count the character (whether to include white space and hyphen or not), you may get a slightly different solution.

Problem 4: DES

Consider a DES-like scheme where

- block length is 8;
- $f_i(x)$ is $(i \cdot x)^K \bmod 16 \quad (i = 1, \ldots, 4)$;
- number of rounds is 4;

Decrypt 10100101 using $K = 1101$.

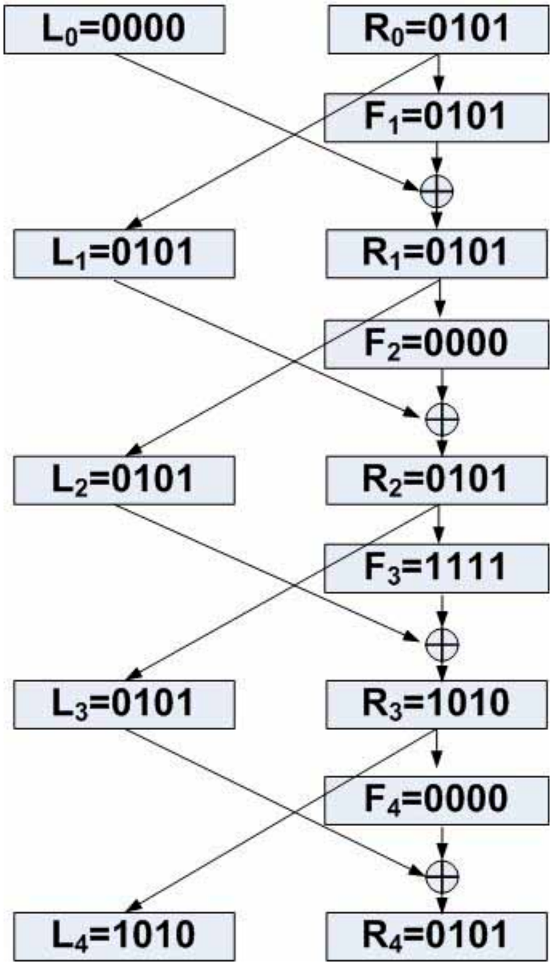

Figure 1: Simplified DES

Solution: The simplified DES is showed in Fig. 1. The encrypted string is the final result which is L4 and R4. We assume no initial permutation. Here the computation of $f_i(x)$ is $(i \cdot x)^{13} \bmod 16$

- known $L4 = 1010, R4 = 0101$
- $L3 = 0101, R3 = 1010$
- $L2 = 0101, R2 = 0101$
- $L1 = 0101, R1 = 0101$
- $L0 = 0000, R0 = 0101$