Cryptography I

ALFE APRTGHAL KAP BQQQ ADCQTPEMY
KSANG A KGA-ETAM MFLK FQRUQY

Math Circle

January 14, 2018

1. For as long as humans have existed, humans have tried to hide information from other humans. Today we are going to talk about various ways of hiding information. Before we go further, let's define a few things.

   - A **Scheme** is the general name for a plan or algorithm to encrypt & decode a message.
   - A **Key** is a number which you use together with a scheme to encrypt & decode a message. You can think of the key as being like the settings of the scheme.
   - **ciphertext** is what you get when you apply a scheme with a key to a secret message.

   In order for a scheme to be useful whatsoever, it should be:

   - Easy to **encrypt** a secret message into ciphertext with a key.
   - Easy to **decode** the ciphertext into a the secret message with a key.
   - Hard (or impossible) to **decipher** the ciphertext, that is to decode it with the scheme, but without the right key.

   For each of the following schemes, identify the Key in each case, and figure out how well they obey the above three properties.

   (a) Shave someone's head, tattoo a message on it, and then let their hair grow back. To decode the message, shave that person's head.

(b) For each letter in the message, flip a coin; on heads, delete that letter.

(c) Write in a different alphabet

(d) change each letter to a different one in a pre-determined manner.

2. Now, let's move onto some more mathematical schemes

   A Caesar cipher is a scheme which goes through each letter in the message and replaces it with one of the letter's neighbors in a regular way. Encrypt the message by shifting every letter forward by a prescribed amount.

   (a) for the Caesar cipher, what is the key?

   (b) Does Caesar cipher obey those above three properties?

(c) Suppose that you associated each letter with a number 0 - 25, and the message was a string of characters $a_1a_2a_3a_4, \ldots a_N$, what is a formula for b_i, the i'th letter in the ciphertext, if the key is n?

(d) Encrypt the message "MATH" with a Caesar cipher with 4 as the key.

(e) The message SKKZ NKXK was encrypted with a Caesar cipher. Decipher it!

(f) Your friend wants to send you a message encrypted with a Caesar cipher. They are worried about the small number of ciphers, and so they choose 100 different keys, and do a whole bunch of encryption. First they encrypt the entire message with the first key, and then they encrypt the entire message with the second key, then the third, etc. They figure that they now created a cipher which is 100 times more secure. What do you think?

(g) Your other friend wants to send you a secret message that is 100 characters long. Your other friend also chooses 100 keys, but she encrypts the first letter with the first key, the second with the second key, etc. What do you think of this scheme?

3. At this point you can probably guess that the Caesar cipher is pretty much useless for 'real' encryption. Here is an improved version.

   An affine cipher is an encryption scheme in which the key is a pair of numbers $a, b \in 0, 1, \ldots, 25$. The affine cipher with key (a, b) sends the number n to $an + b \bmod 26$.

   (a) What conditions do you need on the keys so that the cipher is easy to decrypt?

   (b) Consider an alphabet with p letters, with p prime. How many decodable affine ciphers are there for that alphabet? What if the alphabet has p^2 letters? What if it has pq for different primes pq? How many such ciphers does out language have?

4. This is an improvement, but let's study one of the key problems with the affine cipher, as well as every so called substitution cipher (ciphers which encode text one letter at a time).

   Not every letter is used at the same frequency in English. This is pretty obvious to anyone with even a basic knowledge of the language, but it means that you can do frequency analysis to break many substitution ciphers.

(a) Decode the message 'Yqb gvoty ebotzu yz ybkk Wzu yqb rzkzo zg yqb tfx lbyt h wzkkho' by making educated guesses about the frequency of letters. This was encrypted using an affine cipher. Can you determine the key?

(b) How many different substitution ciphers are there?

(c) Suppose that you had a super powerful computer that could try and decode cipher text. This computer could try 1 trillion substitutions per second. How many years would it take (on average) for this computer to break one message? Hint, $26! \approx 10^26$, and there are $\approx 3.2 \cdot 10^7$ seconds in a year.

(d) Can you decrypt this substitution enciphered message? "KRFC CB VFBK GBK X UBC CGYAY AMRTA? NZ ERCGYT KRA R ITXFVYT RFI R EXYFI"

(e) **What about his one? Hint, I've removed the spaces to make this one harder. "BCZOT HTLHB OCVSY HBTVP LHNCV PONHB FVAOF NFSTH HSOBC ZOTLO AAOGH TUOFS CHCAB CZOTL QFVPO NCXL"**

5. A one-time pad cipher is an encryption scheme that has a key which is as long as the encoded message. You add each element in the key to the letters in the message pairwise, and then compute the result mod26.

   (a) If I want to send a message of length N, how many possible keys are there? How many possible ciphertexts are there?

(b) How can one decode a one-time cipher?

(c) If the key has to be as long as a the message that you want to send, why would anyone use a one-time cipher? Isn't it useless?

(d) Ir can be very impractical to carry around a one-time pad; it has to be quite long to be effective. Here is an alternative. Pick a number $a_1 \in 1, \ldots, 25$, and let $a_k = (a_{k-1}^2 \text{mod} 26)$. Shift the kth letter of the text by a_k. Now instead of having to keep a huge document, you can just remember a single value. What do you think about this scheme?

(e) **Prove that the one-time cipher is the only really totally secure encryption method. That is, that without the key it's impossible to do anything with ciphertext**