HS Projects in Classical Cryptography: TO DO LIST

by William Gasarch

Here is our TO DO list for now

IS ENGLISH PROGRAM

Here is a list of programs you should write SEPARATELY (though you should agree on the language) and test SEP but then put together.

1. Write a program that will take a text in English, remove all punctuation, remove all numbers, and break it into blocks of 5.

2. Write a program that will take a text in English and break into blocks of 5.

3. Write a program that will take a text in English convert it to a text of numbers A is 0, B is 1, etc.

   If you find a punctuation or a number then output

   BAD TEXT— THERE WAS A NUMBER

   or

   BAD TEXT— THERE WAS A PUNCTUATION

4. Write a program that will, given a text of numbers CHECK if all of the numbers are between 0 and 25.

   Output

   YES ALL OF THE NUMBERS ARE BETWEEN 0 and 25

   or

   NO THERE IS A NUMBER NOT BETWEEN 0 AND 25.

5. Write a program that will, given a text of numbers

   first check that they are all between 0 and 25, if not then output

   NO THERE IS A NUMBER NOT BETWEEN 0 AND 25.

   If they are then output

   YES ALL OF THE NUMBERS ARE BETWEEN 0 and 25

   And THEN do the following:

   Scan the text ONCE to find

how many 0's

how many 1's

etc.

Then divide these numbers by the number-of-numbers-in-the-text. Store these numbers in a 26-sized array FREQ. (If your lang allows arrays that begin at 0 then great, FREQ[i] will have the freq of i. If not then FREQ[i] stores the freq of the number i-1.)

NOTE- you could do this by scanning the text 26 times. DO NOT DO THAT. Think about how you can do it scanning just once.

NOTE- FREQ will be an array of REALS or FLOATS or whatever your lang calls them, NOT an array of natural numbers.

6. Write a program that will, given two 26-long arrays of reals computes the DOT product.

7. TEST TEST TEST these programs sep to make sure they all work. Then put them together to get the following program:

   Input a Text in English.

   Output a Freq vector of that text.

8. Run this program on a VERY LONG text in English to get good values for the freq vector. Compare your numbers to ones ones you find on the web. They should be similar.

9. Do the DOT product of the freq you get with themselves. I had said you should get around 0.068 but I really am not sure of my numbers- so please check.

10. Do the DOT product of the freq vector and the SHIFT of that vector by 1,2,3,...,25. Record the largest value. Should be about 0.035

11. Write a program that takes a text of numbers between 0 and 25 and converts to a text of English.

12. Write a Shift program to ENCRYPT, to DECRYPT, and to CRACK.

13. Work on a program to crack a linear Congruential generator.

14. Recall: Your current code goes through a text punctuation and numbers. I do not know what it does if it encounters math symbols like + and ×. This probably never came up.

    (a) Modify it to only remove punctuation. REDO all the stuff you did on SHIFT ciphers with math books. You will need to deal with LaTeX which is the usual source code for mathematics.

    (b) REDO all the stuff you did for SHIFT cipher adapted to math books. Note that you will now have 36-long freq vectors for $\{a, \ldots, z, 0, \ldots, 9\}$.