CIS 3362 - 9/1/23

Substitution, even Queen Mary version was broken via freq analysis + similar techniques.

We want to thwart freq analysis!

possible 2 As in plain ⇒ diff cipher

2 ciphertext same came from 2 diff plain.

		6	14	10	13						
plain	:	G	O	K	N	I	G	H	T	S	
		20	2	5	20						
key	:	U	C	F	U	C	F	U	C	F	
		26	16	15	33						
		% 26			% 26						
		0	16		7						
		A	Q	P	H	...					

Diff positions are shifted by different values.

Relatively secure from 1500s → <u>1800s</u>.

```
void encrypt(char* plain, char* key) {
  int klen = strlen(key);
  int len = strlen(plain);

  for (int i=0; i< len; i++) {
      int val = (plain[i] - 'a' + key[i % klen] - 'a') % 26;

      printf("%c", 'a'+val);

  }
}
```

How to break

Step 1 : figure out keyword length

Kasiski Test: find repeated trigrams or longer in ciphertext.

$k_5 k_6 k_7$ $k_5 k_6 k_7$ $k_5 k_6 k_7$

71   197   350

x_1 x_2 x_3

More likely same word lined up exactly at same key ~~152/153~~ position

$\text{len}(\text{key}) \mid (x_2 - x_1)$

is divisible by

$\text{len}(\text{key}) \mid (x_3 - x_2)$

> $a \mid b$ - "b is divisible by a"
> There exists an int c such that $b = ac$.

Candidates keyword length are divisors of

$$\gcd((x_2 - x_1), (x_3 - x_2), (x_4 - x_3) \dots)$$

1^{st} gap $= 197 - 71 = 126$

2^{nd} gap $= 350 - 197 = 153$

$\gcd(153, 126)$:

$153 = 1 \times 126 + 27$

$126 = 4 \times 27 + 18$

$27 = 1 \times 18 + \boxed{9}$

$18 = 2 \times 9$

Other way to determine keyword length

Index of Coincidence

Problem: bag of candy

20 m&ms
30 Snickers
5 reeces
45 Skittles

randomly select 2 bags of candy (w/o replacement)
What is the probability I get 2 of the same bag?

$$\frac{20}{100} \times \frac{19}{99} + \frac{30}{100} \times \frac{29}{\cancel{100}\ 99} + \frac{5 \times 4}{100 \quad 99} + \frac{45}{100} \times \frac{44}{99}$$

frequencies item are $f_1, f_2, \ldots f_k$

$$\sum_{i=1}^{k} f_i = n \qquad (n \text{ total items})$$

$$IC(Set) = \sum_{i=1}^{k} \frac{f_i(f_i - 1)}{n(n-1)}$$

$$\frac{19}{5 \times 99} + \frac{87}{10 \times 99} + \frac{1}{5 \times 99} + \frac{9}{\cancel{20}_{10}} \times \frac{\cancel{44}^{22}}{99}$$

$$\frac{38 + 87 + 2 + 198}{10 \times 99}$$

$$\frac{325}{10 \times 99} = \frac{65}{2 \times 99} = \boxed{\frac{65}{198}}$$

$$\begin{array}{r} 22 \\ 9 \\ \hline 198 \\ 89 \\ \hline 287 \\ 38 \\ \hline 325 \end{array}$$

If letters all had same freq, then

$$IOC(\cancel{\text{English}}\ \text{rnd letter}) \cong \frac{1}{26} \sim .037\text{ish}$$

but because freq are diff

$$IOC(\text{English}) \sim .0676$$

98 1 1

$$\frac{98 \times 97}{100 \times 99} + 0$$

To determine keyword length
pretend key word was length = 2, 3, 4, 5, ...
for each guess, k, put the ciphertext letters in k bins.

$k = 5$

bin 1: $c_1, c_6, c_{11}, c_{16}, \ldots$] IOC(bin 1)

bin 2: $c_2, c_7, c_{12}, c_{17}, \ldots$] IOC(bin 2)

bin 3: $c_3, c_8, c_{13}, c_{18}, \ldots$] IOC(bin 3)

bin 4: $c_4, c_9, c_{14}, c_{19}, \ldots$] IOC(bin 4)

bin 5: $c_5, c_{10}, c_{15}, c_{20}$] IOC(bin 5)

If guess is correct each bin will have a high IOC.
If not, they'll not be consistently high.

What to do when I know the length of the keyword?

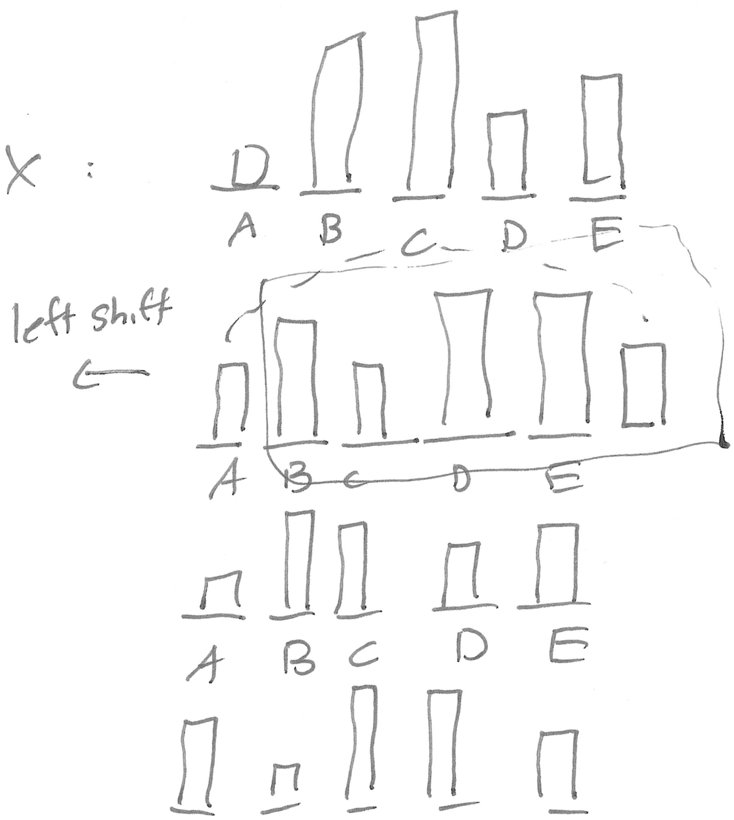
5 letter language X :

try each shift of the bin

bin 1

bin 2

bin 3



See which bar graph looks "most like" English.

Mutual Index of Coincidence Test

2 bags of candy   5 m+m   20 Snick   30 Skitt   15 Rec
                     15 m+m   30 Snick   15 Skitt   5 Rec

What's the probability if I grab 1 bag of candy from each set that I get the same bag?

$$MIC = \overbrace{\frac{5}{70} \times \frac{15}{65}}^{\text{2 mrms}} + \overbrace{\frac{20}{70} \times \frac{30}{65}}^{\text{2 snick}} + \overbrace{\frac{30}{70} \times \frac{15}{65}}^{\text{2 skittles}} + \overbrace{\frac{15}{70} \times \frac{5}{65}}^{\text{2 reeces}}$$

$$= \frac{25(1\times 3 + 4\times 6 + 6\times 3 + 3\times 1)}{25(14\times 13)}$$

$$= \frac{48}{14\times 13} = \frac{24}{7\times 13} = \boxed{\frac{24}{91}}$$