Substitution Cipher – 8/26/2024

The key issue with both shift and affine is that there aren't too many possible keys, so we could simply just try them all…

In order for any cryptosystem to be secure the **keyspace** (number of possible keys), at a minimum, has to be too many to try in a reasonable amount of time.

Affine Cipher

$f(x) = (3x + 8) \bmod 26$

$f(0) = 8$

$f(1) = 11$

$F(2) = 14$

Plaintext	Ciphertext
A	I
B	L
C	O
D	R
E	U
…	

To decrypt, have a different of the chart in reverse.

A whole chart is a key, how many possible charts could we make?

A → 26 choices
B → has to be something different, 25 choices left
C → neither what A or B maps to, so 24 choices left

The number of possible keys (charts) we could have is equal to the number of unique orderings of 26 different letters (permutations). The pattern above shows the answer is 26 x 25 x 24 x … x 1 = 26!

(This is read as “26 factorial”)

This number is pretty big! (According to IDLE, it’s about 4×10^{26})

A computer is definitely not trying all of those keys!

The substitution cipher has been broken since about 1000 AD.

In the middle east, around 980 AD, there was a person named “Al Kindi” for short, who wrote a portion of a book about how to break substitution ciphers.

Note: not all keys are equally likely, once we look at a ciphertext.

The biggest realization is that different letters in a language have different frequencies.

Al-Kindi mentioned the idea of counting the frequencies of each of the letters and then making guesses where the letters that occur more frequently in the ciphertext map to plaintext letters that occur more frequently in the target language. (So instead of having 26 possibilities for a mapping maybe we only have 2 or 3…)

In fact, there was an entire novel published that didn’t use the letter E!!! (So although letter frequencies are averages, for smaller text there won’t be exact matches.)

The substitution cipher doesn’t modify letter frequencies. The letter frequencies are invariant under the process of encryption (for substitution cipher.)

<u>What Other Information Can We Use to Break Substitution Ciphers?</u>
Common digraphs and trigraphs (patterns of 2 or 3 letters)

The information about digraphs and trigraphs along WITH letter frequency can further narrow down possibilities!!!

Patterns of consonants and vowels

Any guesses for words.

Once you get 3 or 4 letters correct, usually when you plug those in, you start seeing other words form and you end up breaking the cipher pretty quickly. In my experience, you might take 2 hours to match the first 3 or 4 letters, and then maybe another 10 minutes to finish the whole thing. So usually, you first matches are wrong, and you gotta keep on trying…

<u>Queen Mary of Scots Story</u>
Queen Elizabeth I – leader of England…but some people thought that Queen Mary of Scotland was the rightful heir to the throne…

Ultimately, Elizabeth threw Mary in jail. Mary eventually had some loyalists who were plotting to break her out and try to restore to the throne. How did Mary and those loyalists communicate?

They wrote letters that were hidden in beer barrels (outside the beer but inside the outermost casing so you couldn't see the letter easily). Just in case someone found the letter, they decided to use a secret substitution cipher code.

Steganography (hiding the fact a message is being passed)

Common words make cryptanalysis much easier, so Queen Mary's code had symbols for 20 common words. (Instead of 26 symbols, Mary's code used around 50 symbols). This code had a "backspace key" and a couple NULL characters.

Mary was executed in part because Queen Elizabeth discovered that Mary was passing message and her court cryptographer was able to break the code and read the messages.

How on earth did this happen back in the late 1500s?

Cryptographer: Sir Francis Walsingham (sp)

Mary's loyalists paid someone who worked in the jail to put the messages in the beer barrels…BUT, Walsingham and Elizabeth decide to pay this person even more money, to copy the messages and send them their way…