Substitution Cipher

Monday, August 31, 2020 11:35 AM

Take a look at what affine is really doing…
it's mapping each input letter to an output letter, using a pattern.

f(x) = (3x + 5) mod 26

A (0) ---> (5) F
B (1) ---> (8) I
C (2) ---> (11) L
D (3) ---> (14) O
etc.

So, using all of the possible affine cipher keys, we could create 312 different look up charts, where for each input letter, we have a correpsonding output letter.

What if, instead of constraining ourselves to these 312 charts, we allow ourselves to use ANY possible chart???

The key problem with Affine cipher is that its keyspace is small (312 for 26 letter), so someone can just try all of the keys…

For any reasonable security, a minimum requirement (but not the only one), is for there to be too many keys to try them all in a reasonable amount of time.

How many possible charts are there?

when mapping input A ---> 26 choices (A is a valid mapping…)
when mapping input B ---> 25 choices (can't choose what A's mapped to
when mapping input C ---> 24 choices.
Logic continues for the next 23 mappings.

Total # of possible charts =

26 x 25 x 24 … x 3 x 2 x 1 = 26!

This is a big number I think it's more than 10^{18}.

It is not possible to try all of the keys.

Even if we did 1 key per second….it would take us…

86,400 sec/day

10^{18}/86,400 = bigger than 10^{13} days, which I am pretty sure is a lot (more than 10 billion years…) obviously not possible.

Code Book by Simon Singh I am summarizing chapter 1.

Queen Mary of Scots…was in jail…

Hid their communication (steganography)

Message were hidden in beer barrels…runner in castle paid off…

Messages were also encrypted!

The person who was paid off, took double payment and also forwarded the messages to someone on Queen Elizabeth's staff (Francis Walsingham - double check this) who knew about breaking secret codes.

Queen Mary's Code - it had letter substitutions, like I've mentioned, but it was more complicated…it had close to 50 symbols.

It had about 20 symbols for common words like "the"…

null characters…these are symbols that translate to nothing…(when decrypting you just ignore these)

it also had a dowbleth - a symbol that means that the next letter is a double letter…

Ultimately, Walsingham was able to break the code and was thus reading all of the communication between Mary and her loyalists.

It was this evidence that essentially got Mary executed!

How did Walsingham back in the 1500s without computers or any electronics try so many billions of keys to break the code?

Answer: He didn't try all the keys!!! But still manages to figure out what the secret mapping was anyway….

Substitution ciphers have been broken since before the year 1000 AD.

The written description of how to break substitutions was first describe in around 980 AD by someone name Al Kindi (his name is a lot longer, this is an abbreviation…)

Some words more frequent than other words
Some letters more frequent than other letters
Letter frequencies are unchanged as a set in substitution (so the frequencies of all the letters are preserved in the encryption operation…)
Common digrams and trigrams retain their "mold"
Structure of vowels and consonants also retain their mold. (So in a string of 10 letters, we expect there to be some vowels…)

All of this information together reduces the chance that some mappings are likely to be correct…

So basically, we break substitution with A LOT of trial and error, utilizing frequency analysis as well as patterns of the English (or whatever language the plaintext is in) language.

Cryptool demo here…

Sample Code - Fall 2019 Homework 2 Question 1

Might guess That S maps to N since it appears twice as the second letter in the most common trigrams and these two trigrams are AND and ING. Then we have OAJ and OAY, one of these must be THE and the other one is probably THI (more likely than THO) so we can guess…

OAJ THI
OAY THE
PSD AND
JSG ING

From here, enough of the message is readable that we can start cherry picking other mappings…

End of Lecture Note about mod definition

$a \equiv b \pmod{n}$ meaning is that
$n \mid (a - b)$

$b \pmod{n}$ - it's impossible to plug into this definition, so you can't have a mod on just one side, because, mod is a relationship between two things.

$x \equiv ay + b \pmod{26}$ "26 divides the difference of a and ay+b"
$(x - b) \equiv ay \pmod{26}$ "26 divides the different of x-b and ay"
$a^{-1}(x - b) \equiv y \pmod{26}$ "26 divides the difference of $a^{-1}(x - b)$ and y"

You can't take the mod of one thing in mathematics. (In programming you can…)