Example 20. (bonus challenge!) Eve, can you crack the following message?

$$W\,D\,G\,Q\,S\,D\,H\,S\,B\,Q\,K\,K\,F\,W\,S\,B$$

Word on the street is that Alice was using the Vigenere cipher with a key of size 2.

(Send me an email by 1/24 with the plaintext and how you found it to collect a bonus point.)

Example 21. The challenge from Example 19 was encrypted using a shift cipher. The key space has size 26, so a brute-force attack results in immediate success: we find that $k=2$ and that the plaintext is $E\,N\,J\,O\,Y\,T\,H\,E\,W\,E\,E\,K\,E\,N\,D$.

This is the worst kind of vulnerability: we successfully mounted a **ciphertext only attack**.

That is, just knowing the encrypted message, we were able to decrypt it (and discover the key that was used).

Attacks

So far, we considered the weakest kind of attack only: namely, a **ciphertext only attack**. And, even then, the historical ciphers prove to be terribly insecure.

However, we need to also worry about attacks where our enemy has additional insight.

- In a **known plaintext attack**, the enemy somehow has knowledge of a plaintext-ciphertext pair (m,c).
- In a **chosen plaintext attack**, the enemy can, herself, compute $c=E(m)$ for a chosen plaintext m ("gained some sort of access to our encryption device").
- In a **chosen ciphertext attack**, the enemy can, herself, compute $m=D(c)$ for a chosen ciphertext c ("gained some sort of access to our decryption device").

There exist many variations of these. Sometimes, the attacker can make several choices (maybe even adaptively), sometimes she only has partial information.

Example 22. Alice sends the ciphertext $B\,K\,N\,D\,K\,G\,B\,Q$ to Bob. Somehow, Eve has learned that Alice is using the Vigenere cipher and that the plaintext is $A\,L\,L\,C\,L\,E\,A\,R$. Next day, Alice sends the message $D\,N\,F\,F\,Q\,G\,E$. Crack it and figure out the key that Alice used! (What kind of attack is this?)

Solution. This is a known plaintext attack.

Since $m+k=c$ (to be interpreted characterwise, modulo 26, and with k repeated as necessary), we can find k simply as $k=c-m$.

For instance, since A (value 0!) got encrypted to B, the first letter of the key is B.

c			B	K	N	D	K	G	B	Q
m		$-$	A	L	L	C	L	E	A	R
k		$=$	B	Z	C	B	Z	C	B	Z

We conclude that the key is $k=BZC$. Now, we can decrypt any future message that Alice sends using this key. For instance, we easily decrypt $D\,N\,F\,F\,Q\,G\,E$ to $C\,O\,D\,E\,R\,E\,D$ (using $m=c-k$).

All of the historical ciphers we have seen, including the substitution cipher below, fall apart completely under a known plaintext attack.

Example 23. (substitution cipher) In a substitution cipher, the key k is some permutation of the letters $A,B,...,Z$. For instance, $k=FRA...$. Then we encrypt $A\to F$, $B\to R$, $C\to A$ and so on. How large is the key space?

Solution. Key space has size $26! \approx 10^{26.6} \approx 2^{88.4}$, so a key can be stored using 89 bits. That's actually a fairly large key space (for instance, DES has a key size of 56 bits only). Too large to go through by brute force.

However, still easy to break. Since each letter is always replaced with the same letter, this cipher is susceptible to a **frequency attack**, exploiting that certain letters (and, more generally, letter combinations!) occur much more frequently in, say, English text than others. For instance, Lewand's book on Cryptology lists the following frequencies:

E: 12.7%, T: 9.1%, A: 8.2%, O: 7.5%, I: 7%, N: 6.7%, S: 6.3%, H: 6.1%, R: 6%, D: 4.3%, L: 4%, C: 2.8%, ...

The rarest letters are Q and Z with a frequency of about 0.1% only. (The exact frequencies and precise ordering various between different sources and the body of text that the frequencies were obtained from.)

The most common letter pairs (digrams) are TH HE AN RE ER IN ON AT ND ST ES EN OF TE ED OR TI HI AS TO.

More information at: https://en.wikipedia.org/wiki/Letter_frequency

Comment. Note that the frequencies and even the ranking depends considerably on the source of text. For instance, using government telegrams, a military resource lists EN followed by RE, ER as the most frequent digrams. That same manual suggests SENORITA as a mnemonic to remember the most frequent letters.

http://www.umich.edu/~umich/fm-34-40-2/ (Field Manual 34-40-2, Department of the Army, 1990)

Example 24. It seems convenient to add the space as a 27th letter in the historic encryption schemes. Can you think of a reason against doing that?

In most texts, the space occurs more frequently and more regularly than any other letter. Adding it to the encryption schemes would make them even more susceptible to attacks.

Fermat's little theorem

Example 25. (warmup) What a terrible blunder... Explain what is wrong!

$$\text{(incorrect!)} \quad 10^9 \equiv 3^2 = 9 \equiv 2 \pmod 7$$

Solution. $10^9 = 10 \cdot 10 \cdot \ldots \cdot 10 \equiv 3 \cdot 3 \cdot \ldots \cdot 3 = 3^9$. Hence, $10^9 \equiv 3^9 \pmod 7$.

However, there is no reason, why we should be allowed to reduce the exponent by 7 (and it is incorrect).

Corrected calculation. $3^2 \equiv 2$, $3^4 \equiv 4$, $3^8 \equiv 16 \equiv 2$. Hence, $3^9 = 3^8 \cdot 3^1 \equiv 2 \cdot 3 \equiv -1 \pmod 7$.

By the way, this approach of computing powers via exponents that are $2, 4, 8, 16, 32, \ldots$ is called **binary exponentiation**. It is crucial for efficiently computing large powers.

Corrected calculation (using Fermat). $3^6 \equiv 1$ just like $3^0 = 1$. Hence, we are allowed to reduce exponents modulo 6. Hence, $3^9 \equiv 3^3 \equiv -1 \pmod 7$.

Theorem 26. (Fermat's little theorem) Let p be a prime, and suppose that $p \nmid a$. Then

$$a^{p-1} \equiv 1 \pmod p.$$

Proof. (beautiful!) Since a is invertible modulo p, the first $p-1$ multiples of a,

$$a, 2a, 3a, \ldots, (p-1)a$$

are all different modulo p. Clearly, none of them is divisible by p.

Consequently, these values must be congruent (in some order) to the values $1, 2, \ldots, p-1$ modulo p. Thus,

$$a \cdot 2a \cdot 3a \cdot \ldots \cdot (p-1)a \equiv 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (p-1) \pmod p.$$

Cancelling the common factors (allowed because p is prime!), we get $a^{p-1} \equiv 1 \pmod p$. □

Remark. The "little" in this theorem's name is to distinguish this result from Fermat's last theorem that $x^n + y^n = z^n$ has no integer solutions if $n > 2$ (only recently proved by Wiles).