2.5 Representations of integers in different bases

We are commonly using the **decimal system** of writing numbers:

$$1234 = 4 \cdot 10^0 + 3 \cdot 10^1 + 2 \cdot 10^2 + 1 \cdot 10^3.$$

10 is called the base, and $1, 2, 3, 4$ are the digits in base 10. To emphasize that we are using base 10, we will write $1234 = (1234)_{10}$. Likewise, we write

$$(1234)_b = 4 \cdot b^0 + 3 \cdot b^1 + 2 \cdot b^2 + 1 \cdot b^3.$$

In this example, $b > 4$, because, if b is the base, then the digits have to be in $\{0, 1, ..., b-1\}$.

Example 36. $25 = \boxed{1} \cdot 2^4 + \boxed{1} \cdot 2^3 + \boxed{0} \cdot 2^2 + \boxed{0} \cdot 2^1 + \boxed{1} \cdot 2^0$. We write $25 = (11001)_2$.

Example 37. (terrible jokes, parental guidance advised)

There is I0 types of people... those who understand binary, and those who don't.

Ok, ok, of course you knew that. How about: There are II types of people...
...those who understand Roman numerals, and those who don't.
It's not getting better:
There are I0 types of people... those who understand hexadecimal, F the rest...

Example 38. Express 49 in base 2.

Solution.

- $49 = 24 \cdot 2 + \boxed{1}$. Hence, $49 = (...1)_2$ where ... are the digits for 24.

- $24 = 12 \cdot 2 + \boxed{0}$. Hence, $49 = (...01)_2$ where ... are the digits for 12.

- $12 = 6 \cdot 2 + \boxed{0}$. Hence, $49 = (...001)_2$ where ... are the digits for 6.

- $6 = 3 \cdot 2 + \boxed{0}$. Hence, $49 = (...0001)_2$ where ... are the digits for 3.

- $3 = 1 \cdot 2 + \boxed{1}$, with $\boxed{1}$ left over. Hence, $49 = (110001)_2$.

Other bases. What is 49 in base 3? $49 = 16 \cdot 3 + \boxed{1}$, $16 = 5 \cdot 3 + \boxed{1}$, $5 = 1 \cdot 3 + \boxed{2}$, $\boxed{1}$. Hence, $49 = (1211)_3$.
What is 49 in base 7? $49 = (10)_7$.

Example 39. Bases 2, 8 and 16 (binary, octal and hexadecimal) are commonly used in computer applications.

For instance, in JavaScript or Python, 0b... means $(...)_2$, 0o... means $(...)_8$, and 0x... means $(...)_{16}$.
The digits $0, 1, ..., 15$ in hexadecimal are typically written as $0, 1, ..., 9, A, B, C, D, E, F$.
Example. FACE value in decimal? $(FACE)_{16} = 15 \cdot 16^3 + 10 \cdot 16^2 + 12 \cdot 16 + 14 = 64206$
Practical example. `chmod 664 file.tex` (change file permission)

664 are octal digits, consisting of three bits: $1 = (001)_2$ execute (x), $2 = (010)_2$ write (w), $4 = (100)_2$ read (r)
Hence, 664 means rw,rw,r. What is rwx,rx,-? 750
By the way, a fourth (leading) digit can be specified (setting the flags: setuid, setgid, and sticky).

Example 40. (another joke) Why do mathematicians confuse Halloween and Christmas?

Because 31 Oct = 25 Dec.
Get it? $(31)_8 = 1 + 3 \cdot 8 = 25$ equals $(25)_{10} = 25$.

Fun borrowed from: https://en.wikipedia.org/wiki/Mathematical_joke

2.6 Historical ciphers, cont'd

Example 41. (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. 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. (Military: EN followed by RE, ER. Used government telegrams.)

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)

2.7 Attacks

So far, we considered the weakest kind of attack only: namely, a **ciphertext only attack**. And, even then, the historical ciphers proved 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 42. 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 fall completely apart under a known plaintext attack.