Locks and letters

by Andreica Amos, Balea Patricia, Căinap Maria, Cîmpean Flavia, Păcurar Flavia, Pașca Teodora, Rus Carina, Turcan Mihail

Colegiul Național "Emil Racoviță" (Cluj, Romania)

Task

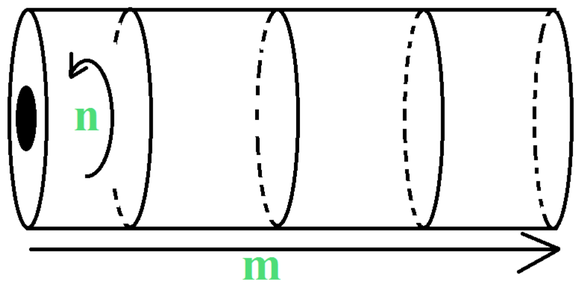
We want to create the most secure lock with cipher we possibly can. But its users are lazy and they use common English words as the opening code. How should we choose the number of barrels, the number of letters per barrel and the letters to write on each barrel in order to maximize the number of combinations and thus render a thief's task much more difficult?

Methods and Results

Supposing the cipher can be made up of any given combination of the letters of the English alphabet (26 in total), then the number of possibilities is as follows:

$$C\binom{n}{26} * n^m$$

Where *m* is the number of barrels and *n* the number of letters per barrel.



Given the further constraint -all passwords must be common English words- and assuming the thief is aware of this particular rule, the end result depends on the database we choose for our search.

For this, we devised a computer program counting the number of words of any introduced type (when given in the form of **a** for a word of 5 letters with the third letter 'a') and displaying the words themselves:

```
    /// C++ libraries ///
#include <iostream>
#include <fstream>

using namespace std;

    /// FILES ///
ifstream in ("data.in");
ofstream out ("data.out");
/*
be sure to create the two files
in the C++ project folder
*/

    /// Declaration ///
int n[19];

 int main ()
{
   char x;  //char data type allows us to
   //store charachters inside it

   int c;   //counter will be used to count
   //number of letters in word
  c=0;      //initialise counter with 0
```

```
/// COUNTING LOOP ///


while(in.get(x))    //reads the file char by char

{
   c++;             //update counter regardless
   if(x=='\n')      // if reached new line
     {
       n[c-1]++;    //increment frequency vector
       /*
       we use c-1 to compensate for the new line
       */
       c=0;         //reset counter
     }
}


 /// SHOW RESULTS ///
for(int i=0;i<19;i++)
{

  out<<" "<<i<<" :  "<<n[i]<<'\n';
}

return 0;
```

To find the proper number of barrels, letters per barrel and letters to write on these barrels, we first needed to find for what number of letters the English language has most words.

To further conduct our research, we needed a common database of reference. For this, we used a program suggested by the French researcher, which uses 92518 words and working on the same principle (found here: http://www.informathiques.fr/resources/exploredico.html)

Number of letters	Number of words
4	2593
5	5170
6	8459
7	11934
8	13057
9	12009
10	9911
	TOTAL: 63133 words with length between 4 and 10 letters

Using this, we found the most number of words to correspond to 8 letters. Theoretically speaking, this means that the most efficient lock is one with 8 barrels and 26 letters per barrel (all of the letters of the English alphabet on every barrel).

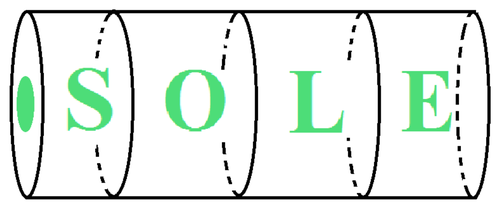
Further Research

Because a lock with 26 letters per barrel would prove to be somewhat impractical if not impossible to manufacture, we have decided to particularize the problem to a more efficient type of lock, based on the demands of the customer.

Take for example the situation of a lock with 4 barrels and 2 letters per barrel. Which letters should we choose to write on it to prevent the thief from easily breaking in?

Using the same program as previously:

TOP	Letter in first position	Number of words beginning with it
1.	s	243
2.	b	190
3.	p	183
4.	t	170
5.	c	159
6.	l	156
7.	d	155
8.	h	135
9.	m	135
10.	f	133
11.	r	132
12.	w	129
13.	g	127
14.	a	104
15.	n	63
16.	o	63
17.	e	61



Letter in second position	Number of words beginning with 'b'	Number of words beginning with 's'	Total
a	42	26	68
b	0	0	0
c	0	9	9
d	0	0	0
e	28	27	55
f	0	0	0
g	0	0	0
h	0	21	21
i	16	20	36
j	0	0	0
k	0	8	8
l	12	24	36
m	0	3	3
n	0	10	10
o	41	31	72
p	0	17	17
q	0	0	0
r	14	0	14

Letter in third position	"ba"	"sa"	Sum	"bo"	"so"	Sum	TOTAL
a	1	0	1	3	3	6	7
b	2	0	2	1	1	2	4
c	1	2	3	1	2	3	6
d	1	0	1	3	2	5	6
e	0	0	0	0	0	0	0
f	0	1	1	0	2	2	3
g	1	4	5	1	0	1	6
h	2	0	2	0	0	0	2
i	2	2	4	1	1	2	6
j	0	0	0	0	0	0	0
k	1	1	2	0	0	0	2
l	5	2	7	5	4	9	16
m	0	1	1	1	1	2	3
n	6	5	11	5	2	7	18
o	0	0	0	7	2	9	9
p	1	1	2	1	1	2	4
q	0	0	0	0	0	0	0
r	7	1	8	2	2	4	12

fourth position	"bal"	"ban"	"bol"	"bon"	"son"	"sol"	"sal"	"san"	TOTAL
a	0	0	1	0	0	0	0	0	1
b	0	0	0	0	0	0	0	0	0
c	0	0	0	0	0	0	0	0	0
d	1	1	1	1	1	1	0	1	6
e	1	1	1	1	1	1	1	1	7
f	0	0	0	0	0	0	0	0	0
g	0	1	0	1	1	0	0	1	4
h	0	0	0	0	0	0	0	0	0
i	0	1	0	0	0	0	0	0	1
j	0	0	0	0	0	0	0	0	0
k	1	1	0	1	0	0	0	1	4
l	1	0	1	0	0	0	0	0	2
m	1	0	0	0	0	0	0	0	1
n	0	0	0	0	0	0	0	0	0
o	0	0	0	0	0	1	0	0	1
p	0	0	0	0	0	0	0	0	0
q	0	0	0	0	0	0	0	0	0
r	0	0	0	0	0	0	0	0	0

For the specified criteria, we have 'b' and 's' on the first barrel, 'a' and 'o' on the second, 'l' and 'n' on the third and 'e' and 'd' on the last.

Presently, we are working on a program to conduct the process of choosing the proper letters for certain introduced parameters.