The plaintext has always some structure. It is not just random data but, instead, it contains some *redundancy.* In our example we assume we know that the plaintexts in question are of English language. When coded into ASCII bit strings this assumption implies a huge redundancy for these bit strings; most ASCII codes never appear and some appear very frequently. For the illustrative purposes, however, we assume a simplified coding for this example: we use only capital letters from A to Z plus space in between words (no punctuation etc.) Letter A is coded into a number 0. Similarly B is coded to 1, C to 2 etc. Finally, Z is coded into 25 and space is coded into the number 26. The full coding is given below:

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**
0	1	2	3	4	5	6	7	8	9	10	11	12	13

O	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**	**space**
14	15	16	17	18	19	20	21	22	23	24	25	26

The encryption is now done as follows. A list of random numbers from the interval between 0 and 26 is created, and it serves as a mask. Each plaintext letter, or more precisely, the number corresponding to it, is added to one entry in the random number list in a modular fashion, i.e. the numbers 27 and 0 are considered to be equal. Also, for each number that exceeds 27, multiples of 27 are subtracted until the result is between 0 and 26. For instance, 14 + 21 = 35, but when 27 are subtracted, the result becomes 8.

For example, assume the plaintext is CAT while the mask is (3, 17, 12). Then the encoded plaintext is (2, 0, 19) and the ciphertext is (5, 17, 4).

This encryption provides perfect security as long as the mask is truly random, it is not known to the attacker and it is *used only once.* Indeed, any three-letter plaintext could be transformed to the same ciphertext (5, 17, 4) with a suitable mask. MOM encodes to (12, 14, 12) and the mask producing our ciphertext would be (20, 3, 19). Similarly, XYZ is a potential plaintext if only the mask happens to be (9, 20, 6).

We assume now that a short keastream has been repeatedly used to encrypt a two times longer English text. We try to determine the contents of the text.

The ciphertext is (21, 12, 22, 25, 21, 15, 6, 13, 6, 15, 9, 20, 13, 0, 1, 13). We split it into two halves:

Ciphertext 1: (21, 12, 22, 25, 21, 15, 6, 13)

Ciphertext 2: (6, 15, 9, 20, 13, 0, 1, 13)

We immediately see that the plaintext letter in positions 8 and 16 is the same. Let us start our analysis from the first two letters:

Ciphertext 1: 21 12

Ciphertext 2: 06 15

There exist various statistics about frequencies of letters in average English text. Also, statistics about frequencies of letter-pairs (*digrams)* can be found, see e.g. GAINES(1956). For instance, fifteen most common digrams cover almost 30 % of all cases. Thus, we could try to jump start the analysis by trying most frequent letter-pairs for the first two letters of the unknown plaintexts.

Most frequent pair in English is TH, encoded as 19 07. If the plaintext 1 would begin with TH, then the mask would be 02 05, and the plaintext 2 would begin with 04 10, decoded as EK. This is certainly possible in principle but not very promising, as EK is not among the top one hundred commonest digrams and a very few English words begin with this digram.

Let us try next TH for the beginning of the plaintext 2: in this case mask would be 14 08, and plaintext 1 would start with 07 04, i.e. HE. This is a more promising track, as HE is one of the most frequent digrams also.

The most probable continuation for the plaintext 2 is probably THE + space, encoded 19 07 04 26. This yields a mask 14 08 05 21, and the plaintext 1 would be encoded 07 04 17 04, corresponding to HERE. We are on the right track.

Plaintext 1 certainly continues with a space, hence the next element in the mask is 22, and the plaintext 2 continues with 18, i.e. the plaintext 2 is THE S??. The best tactics now seems to be trying of common three letter words that begin with S. Plaintext 2 = THE SEA would imply plaintext 1 = HERE TF, no good. THE SKY would imply HERE RD, not better. Finally, plaintext 2 = THE SUN implies plaintext 1 = HERE IS, and a very probable solution is found: “Here is the sun ”. The symbol in positions 8 and 16 is the empty space.