Dpef Dsbdljoh

Encryption was one of the earliest applications of computers. Nowadays, devising and cracking a simple code is within the powers of the Basic programmer

All our communications with others are codified. Whether we use speech or written language, both are presented in such a way as to be intelligible only if the person receiving the message can interpret the code. The same is true of our conversations with computers. Most home computers rely on a dialect of BASIC in order to be accessible to people, but we know that the machine itself does not use this language to perform its functions: it must first interpret the BASIC statements into a purely numerical form that it can then use to set up the switching sequences defined in the program, and thus produce the desired results. Codes of this sort — human and programming languages — are easily accessible in our everyday lives. Anyone can learn French, German, BASIC or FORTRAN, given the effort and the will.

Data Compression

Computer users who need to store large numbers of text files are constantly searching for ways to compress the data in those files. One way to achieve this is by tokenisation. In much the same way as Sinclair's ZX series of microcomputers produce a whole BASIC reserved word at the touch of a single key, a token can be substituted for a commonly-used word or phrase.

In addition, encoding techniques are also used to compress the data even further. Compact, a Unix utility, is generally reckoned to compress text files by an average of 38 per cent, and Clip, which runs under CP/M, regularly achieves even better results. Compactor, which runs on the Commodore 64, performs the same function for BASIC programs by removing REMs, unnecessary spaces and so on

But there is another type of encoding (more accurately called 'encryption') that has the very opposite of communication as its objective: its purpose is to deny understanding to all but the small group for whom the communication is intended. Until the second half of the 20th century, the transmission of information in a form not generally intelligible was restricted to governments and a few large industrial interests. But more recently, with the ever increasing use of vulnerable public telephone lines for the exchange of information, most of it with some commercial value, the practice of encryption has become more commonplace.

Cyphers and codes range from the very simple — the addition or subtraction of a given value to every byte, perhaps, or the formatted substitution of one character for another wherever it occurs — to the immensely complex cyphers that are being worked on in the most recent advances in number theory. These cyphers contain no element of repetition whatsoever, and hence are not vulnerable to frequency analysis decoding methods.

The simplest of all meaningful encryption techniques is perhaps Caesar's Cypher (which was probably first used at the time of the Roman Empire). The decryption of Caesar's Cypher requires only the message and a knowledge of the key, so there are no bulky code books or documents to be concealed, and no sophisticated machines required. Here is a simple message encrypted in Caesar's Cypher:

FMKC AMKNSRCP AMSPQC

We can make a few assumptions about these encoded words because of the way in which the encyphered groups are spaced out (though, of course, this could be intended to create confusion!). The most obvious thing that strikes us is that the message consists of three words: the first has four letters, the second has eight and the last has six. We can also assume that the second and third words begin with the same letter, and that the first and last words end with the same letter. The common ending letter here (C) is also one of the two letters in the message with the highest frequency (the other is M). This observation is of considerable value to the cryptanalyst — at least, as long as he knows which language he is working with. In English, the letter that occurs most frequently is E, followed by T.

With a sample as small as the one we have here (a total of only 17 letters, which any statistician will tell you is an insufficiently large sample upon which to base any analysis), our results are likely to be fallible. But let's try frequency substitution anyway, and see if the results are meaningful. Let's substitute the E for the C first:

FMKe AMKNSReP AMSPQe

The message is still meaningless, but there are other clues. What about the relationship between the original letter and the one we substituted for it? C is two places in front of E in the alphabet. What happens if we put the rest of the message through the same transformation? Two places behind M (our other most commonly occurring letter) is O, so let's try adding that piece of information:

FoKe AoKNSReP AoSPQe

In the first word we now have: '(something) vowel (something) vowel', which is a valid English construction. Furthermore, the final vowel is E, which is a common occurrence in English, so