

Cryptanalysis

Cryptanalysis is the discipline of deciphering a ciphertext without having access to the keytext (see cryptosystem), usually by recovering more or less directly the plaintext or even the keytext used, in cases favorable for the attacker by reconstructing the whole cryptosystem used. This being the worst case possible for the attacked side, an acceptable level of security should rest completely in the key (see Kerckhoffs' and Shannon's maximes). "A systematic and exact reconstruction of the encryption method and the key used" (Hans Rohrbach, 1946) is mandatory if correctness of a cryptanalytic break is to be proved, e.g. when a cryptanalyst is witness for the prosecution.

1) Terminology

Cryptanalysis can be *passive*, which is the classical case of intercepting the message without giving any hint that this was done, or *active*, which consists of altering the message or retransmitting it at a later time, or even of inserting own messages (some of these actions may be detected by the recipient).

A *compromise* is the loss (or partial loss) of secrecy of the key by its exposure due to cryptographic faults. We shall describe various kinds of key compromises.

A *plaintext-ciphertext compromise* is caused by a transmission of a message in ciphertext followed (e.g. because the transmission was garbled) by transmission of the same message in plaintext. If information on the encryption method is known or can be guessed, this results in exposure of the key. This attack may be successful for a plaintext of several hundred characters.

A *plaintext-plaintext compromise* is a transmission of two *isologs* i.e. two different plaintexts, encrypted with the same keytext. If the encryption method is such that the encryption steps form a group (see key group and pure cryptosystem), then a 'difference' $p_1 - p_2$ of two plaintexts p_1, p_2 and a 'difference' $c_1 - c_2$ of two ciphertexts c_1, c_2 may be defined and the role of the keytext is cancelled out: $c_1 - c_2 = p_1 - p_2$. Thus, under suitable guesses on the plaintext language involved, e.g. on probable words and phrases, a 'zig-zag' method (see below), decomposing $c_1 - c_2$, gives the plaintexts and then also the keytext. This compromise is not uncommon in case of a shortage of keying material. It is even systemic if a periodic key is used.

A *ciphertext-ciphertext compromise* is a transmission of two *isomorphs*, i.e. the same plaintext, encrypted with two different keytexts. Exchanging the role of plaintext and keytext, this case is reduced to and can be treated as a plaintext-plaintext compromise. This compromise is even systematic in message sets, where the same message is sent in different encryption to many places, such as it is common in public key cryptosystems.

One speaks of a *brute force attack* or *exhaustive key search* if all possible keytexts are tried out to decrypt a ciphertext (knowing or guessing the cryptosystem used). At present, with the still growing speed of supercomputers, every ten years the number of trial and error steps that are feasible is increased by a factor of roughly 2^5 .

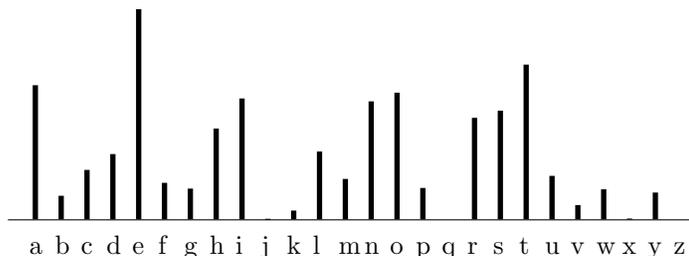
Further commonly used terminology will be given now. In a *ciphertext-only attack*, only one or more ciphertexts under the same keytext are known. In a *known-plaintext attack* one knows one or more matching pairs of plaintext-ciphertext. Frequently, this attack is carried out with rather short fragments of the plaintext (e.g. probable words and phrases). In a *chosen-plaintext attack* one can choose plaintexts and obtain the corresponding ciphertexts. Sometimes this can be done with the proviso that the plaintexts may be chosen in a way that depends on the previous encryption outcomes. How to foist the plaintext on the adversary is not a cryptographers problem, but one of cunning and is to be executed by the secret services. Finally, in a *chosen-ciphertext attack* one has the possibility to choose different ciphertexts to be decrypted, with the cryptanalyst having access to the decrypted plaintext. An example may

be the investigation of a tamperproof decryption box, with the hope of finding the key.

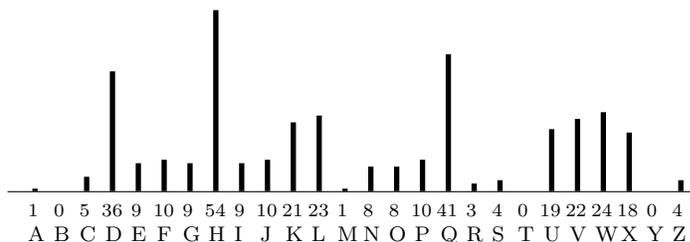
2) Statistical approaches to classical cryptosystems

We shall now discuss some statistical methods that can be used by the cryptanalyst.

Frequency matching is a cryptanalytic method for breaking monoalphabetic (Cæsar type) encryptions. One determines the frequency of the characters in a ciphertext and compares them with the frequency of the characters in a language known or assumed to be used for the plaintext. To give an example: the frequency profile of the English language looks like



If a ciphertext of 349 characters has the following distribution:



it is easy to guess a Cæsar encryption that counts down three letters in the standard alphabet: $a \doteq D$, $b \doteq E$, $c \doteq F$, ..., $z \doteq C$. More difficult is the situation if a mixed alphabet is to be expected. Then the first step is to group the letters in cliques: the most frequent ones, the very rare ones, and some in between

{etaoin} {srh} {ld} {cumfpgwyb} {vk} {xjqz} ,

and to refine the decryption within these cliques by trial and error.

Depth is a notion used in connection with the cryptanalysis of polyalphabetic encryptions. It means the arrangement of a number of ciphertexts supposedly encrypted with the same keytext—for periodic polyalphabetic encryption broken down according to the assumed period.

Example: a depth of five lines:

```
TCCVL  ES KPT  XMPVW  HYMVG  XBORV  CWARF
VLLBV  CKWFP  EHECF  CGNZE  KKKVI  HDDI D
MYYRD  MJ WMC  UI GLO  KMXLR  EWHXM  TJ HAS
BKQTZ  BZ WKW  ZXGZO  VTBAT  KWMGM  RJ KLP
MYYVH  BWJ DX  CPCZO  HVTSI  VMEBS  OHRAU .
```

The lines of a depth are isologs: they are encrypted with the same key text and represent a plaintext-plaintext compromise.

By forming differences of the elements in selected columns, a reduction of depth to a monoalphabetic (Cæsar type) encryption is accomplished. This makes it possible to derive the keytext

TRUTH ISSOP RECI O US THA TI TNE EDS AB

which decrypts the depth (by means of the Vigenère table) to

```
a l i c e  w a s b e  g i n n i  n g t o g  e t v e r  y t i r e
c u r i o  u s e r a  n d c u r  i o u s e  r c r i e  d a l i c
t h e y w  e r e i n  d e e d a  q u e e r  l o o k i  n g p a r
i t w a s  t h e w h  i t e r a  b b i t t  r o t t i  n g s l o
t h e c a  t e r p i  l l a r a  n d a l i  c e l o o  k e d a t .
```

Forming a depth is possible as soon as the value of the period of the periodic polyalphabetic encryption has been found, for instance by the Kasiski method below. Forming a depth is not possible, if the key is non-periodic. But even for periodic polyalphabetic encryptions, forming a depth of sufficiently many elements (usually more than six) is not possible if the keytext is short enough.

When the alphabets used in a polyalphabetic periodic substitution are a mixed alphabet and a shifted version of it, *symmetry of position* is the property that for any pair of letters their distance is the same in all rows of the encryption table. For a known period, it may allow, after forming a depth, the complete reconstruction of the substitution (Auguste Kerckhoffs, 1883).

Kasiski's method: If in a periodic polyalphabetic encryption the same plaintext sequence of characters happens to be encrypted with the same sequence of key characters, the same ciphertext sequence of characters will occur. Thus, in order to determine the period of a periodic polyalphabetic encryption, the distance between two “parallels” in the ciphertext (pairs, triples, quadruples etc. of characters) is to be determined; the distance of genuine parallels will be a multiple of the period. The greatest common divisor of these distances is certainly a period—it may, however, not be the smallest period. Moreover, the period analysis may be disturbed by faked parallels. Kasiski developed in 1863 this fundamental test for key periodicity and shattered the widespread belief that periodic polyalphabetic encryption is unbreakable.

The *Kappa test* is based on the relative frequency $\kappa(T, T')$ of pairs of text segments $T = (t_1, t_2, t_3, \dots, t_M)$, $T' = (t'_1, t'_2, t'_3, \dots, t'_M)$ of equal length, $M \geq 1$, with the same characters at the same positions (that is why this method is also called the *index of coincidence*, often abbreviated to I.C., William F. Friedman 1925). The value of Kappa is rather typical for natural languages, since the expected value of $\kappa(T, T')$ is $\sum_{i=1}^N p_i^2$, where p_i is the probability of occurrence of the i -th character of the vocabulary to which T and T' belong. For sufficiently long texts, it is statistically roughly equal to $1/15 = 6.67\%$ for the English language and $1/12.5 = 8\%$ for the French language and the German language. Most importantly, it remains invariant if the two texts are polyalphabetically encrypted with the same keytext. If, however, they are encrypted with different keytexts or with the same key sequence, but with different starting positions, the character coincidence is rather random and the value of Kappa is statistically close to $1/N$, where N is the size of the vocabulary. The Kappa test applied to a ciphertext C and a cyclically shifted versions $C^{(u)}$ of the ciphertext, where u denotes the number of shifts, yields the value $\kappa(C, C^{(u)})$. If the keytext is periodic with period d , then for $u = d$ and for all multiples of d , a value significantly higher than $1/N$ will occur, while in all other cases a value close to $1/N$ will be found. This is the use of the Kappa examination for finding the period; it turned out to be a sharper instrument than the Kasiski method.

The Kappa test may also be used for adjusting two ciphertexts C, C' which are presumably encrypted with the same keytext, but with different starting positions (called *superimposition*). By calculating $\kappa(C^{(u)}, C')$, a shift d , determined as a value of u , for which $\kappa(C^{(u)}, C')$ is high, brings the two ciphertexts $C^{(d)}$ and C' ‘in phase’ i.e. produces two isologs. In this way, a depth of n texts can be formed by superimposition from a ciphertext-ciphertext compromise of n ciphertexts.

The *De Viaris attack* is a cryptanalytic method invented by Gaëtan Henri Léon de Viaris, 1893 to defeat a polyalphabetic cryptosystem proposed by Étienne Bazeries, in which the alphabets did not form a Latin square. (A *Latin square* for a vocabulary of N characters is a N -by- N matrix over this alphabet such that each character occurs just once in every line and in every column.)

Pattern finding is a cryptanalytic method that can be applied to monoalphabetic encryptions. It makes use of patterns of repeated symbols. For example, the pattern 1211234322 with “signature” 4+3+2+1 (4 two’s, 3 ones, 2 three’s

and 1 four) most likely allows in English nothing but *peppertree*, the pattern 1213143152 with the signature 4+2+2+1+1 nothing but *initiation* (Andree 1982, based on Merriam-Webster's Dictionary).

Non-coincidence exhaustion: some cryptosystems show peculiarities: genuine selfreciprocal permutations never encrypt a letter by itself. Porta encryptions even always encrypt a letter from the first half of the alphabet by a letter from the second half of the alphabet and vice versa. Such properties may serve to exclude many positions of a probable word (a *probable word* is a word or phrase that can be expected to be present in a message according to the context; it may form the basis for a known-plaintext attack).

Zig-zag exhaustion: for encryptions with a key group (see key), the difference of two plaintexts is invariant under encryption: it equals the difference of the corresponding ciphertexts. Thus in case of a plaintext-plaintext compromise (with a depth of two), the difference of the ciphertexts can be decomposed into two plaintexts by starting with probable words or phrases in one of the plaintexts and determining the corresponding plaintext fragment in the other plaintext, and vice versa. This may lead in a zig-zag way ("cross-ruff") to complete decryption.

Theoretically, the decomposition is unique provided the sum of the relative redundancies of the two texts is at least 100%. For the English language, the redundancy (see information theory) is about 3.5 [bit/char] or 74.5% of the value $4.7 \approx \log_2 26$ [bit/char].

Multiple anagramming is one of the very few general methods for dealing with transposition ciphers, requiring nothing more than two plaintexts of the same length that have been encrypted with the same encryption step (so the encrypting transposition steps have been repeated at least once). Such a plaintext-plaintext compromise suggests a parallel to Kerkhoffs' method of superimposition. The method is based on the simple fact that equal encryption steps perform the same permutation of the plaintext letters. The ciphertexts are therefore written one below the other and the columns thus formed are kept together.

Friedrich L. Bauer

References

Bauer, F.L., *Decrypted Secrets; Methods and Maxims of Cryptology*, Springer Verlag, Berlin, etc., 1997.