5 Unicity Distance

We now apply our findings on the redundancy to the exhaustion of the key space. We don't deal with the expenses but only consider the feasibility. We follow the simplified approach of Hellman.

Assumptions

1. All meaningful texts of length r have the same probability. [Otherwise we get more complicated formulas. For natural languages this assumption is clearly false when r is small. However for large r we might hope that it follows from the usual stochastic conditions.]
2. The densitiy $\rho(M)$ of the language M exists. [Otherwise we could derive only a bound.]
3. All keys $k \in K$ have the same probability and they are $h = \#K$ in number.
4. All encryption functions f_k for $k \in K$ respect the lengths of the texts, or in other words $f(M_r) \subseteq \Sigma^r$.

Now let $c \in \Sigma^r$ be a ciphertext. In general—if all encryption functions f_k are different—it fits h possible plaintexts of length r in Σ^r. By far not all of them are meaningful but only

$$h \cdot \frac{t_r}{n^r} \approx \frac{h \cdot 2^{r\rho(M)}}{2^{r \cdot \log_2 n}} = h \cdot 2^{-r\delta(M)}.$$

We expect a unique solution in M_r if

$$h \cdot 2^{-r\delta(M)} \leq 1, \quad \log_2 h - r\delta(M) \leq 0, \quad r \geq \frac{\log_2 h}{\delta(M)},$$

at least if all encryption functions f_k are different; otherwise we should replace $\log_2 h$ with $d = d(F)$, the effective key length of the cipher F.

This motivates the following definition:

Definition 3. For a cipher F with effective key length $d(F)$ defined on a language M of redundancy $\delta(M)$ we call

$$\mathrm{UD}(F) := \frac{d(F)}{\delta(M)}$$

the **unicity distance**.

Examples

We always assume the alphabet $\Sigma = \{\mathtt{A}, \ldots, \mathtt{Z}\}$ with $n = 26$, and the language $M =$ "English".

1. For the shift cipher we have $d = \log_2 26$, UD $\approx 4.7/3.2 \approx 1.5$, not about 5 as suspected in the introductory example. This deviation might be due to the many inexact steps in the derivation. In particular for small r the approximation $t_r \approx 2^{r\rho(M)}$ is very inexact.

2. For the monoalphabetic substitution we have $d \approx 88.4$, UD $\approx 88.4/3.2 \approx 27.6$. This result is in good concordance with empirical observations on the solvability of monoalphabetic cryptograms.

3. For the TRITHEMIUS-BELLASO cipher with period l we have $d \approx 4.7 \cdot l$, UD $\approx 1.5 \cdot l$.

4. For PORTA's disk cipher we have $d \approx 88.4 + 4.7 \cdot l$, UD $\approx 27.6 + 1.5 \cdot l$.

5. For the general polyalphabetic substitution with period l and independent alphabets $d \approx 122 \cdot l$, UD $\approx 38 \cdot l$.

6. For the One-Time Pad over the group $G = \Sigma$ we have $M = K = C = \Sigma^*$, hence $\#K = \infty$. However it makes sense to interpret $d_r/\delta_r = r \cdot \log_2 n/0 = \infty$ as unicity distance.