9.5 Die Eindeutigkeitsdistanz

Diese Erkenntnisse über die Redundanz werden nun auf eine vollständige Schlüsselsuche angewendet – der Aufwand wird dabei nicht berücksichtigt, nur die Durchführbarkeit. Die Herleitung folgt dem vereinfachten Ansatz von HELLMAN.

Annahmen

1. Alle sinnvollen Texte der Länge r sind gleich wahrscheinlich. [Sonst werden die Formeln komplizierter; für natürliche Sprachen folgt diese Annahme für genügend große r aus den üblichen stochastischen Annahmen.]

2. Die Dichte $\rho(M)$ der Sprache M existiert. [Sonst kann man nur eine Schranke herleiten.]

3. Alle Schlüssel $k \in K$ sind gleichwahrscheinlich; es gebe $h = \#K$ Stück.

4. Alle Verschlüsselungsfunktionen f_k für $k \in K$ sind längentreu, d. h., $f(M_r) \subseteq \Sigma^r$.

Sei nun ein Geheimtext $c \in \Sigma^r$ gegeben. Dazu gibt es (im allgemeinen – falls alle Verschlüsselungsfunktionen f_k verschieden sind) h mögliche Klartexte der Länge r in Σ^r. Darunter sind längst nicht alle sinnvoll, sondern nur etwa

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

Eindeutige Dechiffrierbarkeit in M_r kann man erwarten, wenn

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

falls alle Verschlüsselungsfunktionen f_k verschieden sind; sonst muss man ${}^2\log h$ durch $d = d(F)$, die effektive Schlüssellänge der Chiffre F ersetzen.

Daher ist die folgende Definition motiviert:

Definition 3. Für eine Chiffre F mit effektiver Schlüssellänge $d(F)$ auf einer Sprache M mit Redundanz $\delta(M)$ heißt

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

die **Eindeutigkeitsdistanz**.

Beispiele

Es wird stets das Alphabet $\Sigma = \{\mathtt{A}, \ldots, \mathtt{Z}\}$ mit $n = 26$ und die Sprache $M =$ „Deutsch“ angenommen.

1. Bei der Verschiebechiffre ist $d = {}^2\log 26$, ED $\approx 4.7/3.3 \approx 1.4$, nicht ungefähr 4, wie im Eingangsbeispiel vermutet. Diese Diskrepanz ist auf die vielen Ungenauigkeiten in der Herleitung zurückzuführen; insbesondere ist die Näherung $t_r \approx 2^{r\rho(M)}$ für kleine r natürlich besonders ungenau.

2. Bei der monoalphabetischen Chiffre ist $d \approx 88.4$, ED $\approx 88.4/3.3 \approx 26.8$. Dieser Wert stimmt mit empirischen Erfahrungen über die Lösbarkeit monoalphabetischer Kryptogramme recht gut überein.

3. Bei der Trithemius-Belaso-Chiffre mit Periode l ist $d \approx 4.7 \cdot l$, ED $\approx 1.4 \cdot l$.

4. Bei der Drehscheiben-Chiffre nach Porta ist $d \approx 88.4 + 4.7 \cdot l$, ED $\approx 26.8 + 1.4 \cdot l$.

5. Bei der allgemeinen polyalphabetischen Substitution der Periode l mit unabhängigen Alphabeten ist $d \approx 122 \cdot l$, ED $\approx 37 \cdot l$.

6. Bei der Vernam-Chiffre mit der Gruppe $G = \Sigma$ ist $M = K = C = \Sigma^*$, also $\#K = \infty$. Es ist aber sinnvoll, $d_r/\delta_r = r \cdot {}^2\log n/0 = \infty$ als Eindeutigkeitsdistanz anzusehen.