2.4. Vigenère Cipher (1523–1596)

Alphabet: $\{0, \dots, 25\}$

Key string, keyword of length k: $(s_0, \dots, s_{k-1})$

Plaintext: $a_0, \dots, a_{k-1}, a_k, \dots, a_{2k-1}, \dots$

Keystream: $s_0, \dots, s_{k-1}, s_0, \dots, s_{k-1}, \dots$

Encryption: componentwise addition mod 26

Ciphertext: $c_i = (a_i + s_{i \bmod k}) \bmod 26$.

Note: Vigenère cipher is polyalphabetic, i.e., different ciphertext char. may occur for the same plaintext char.

- Vigenère cipher with running key: use a text as key as long as the plaintext.

- Vernam cipher (1917)
  Same as Vigenère, but for each plaintext char. generate randomly a key char. (one-time pad)

2.6. Joint principles / notation

$\mathcal{X}, \mathcal{Y}$: alphabets = finite set of characters

$\mathcal{X} = \{x_1, \dots, x_m\}$, $\mathcal{Y} = \{y_1, \dots, y_d\}$

$\mathcal{X}^\ell, \mathcal{Y}^\ell$: words of length $\ell \in \mathbb{N}_0$ over $\mathcal{X}, \mathcal{Y}$

$\mathcal{M} \subseteq \bigcup_{\ell=0}^{\infty} \mathcal{X}^\ell$: set of possible plaintexts, messages

$\mathcal{C} \subseteq \bigcup_{\ell=0}^{\infty} \mathcal{Y}^\ell$: set of possible ciphertext

$M \in \mathcal{M}$ is called message or plaintext

$C \in \mathcal{C}$ is called ciphertext or cryptogram

$\mathcal{K}$: (finite) set of possible keys, the keyspace.

$K \in \mathcal{K}$ is called key.

Encryption is described by a function (encr. rule)

$$e : \mathcal{M} \times \mathcal{K} \to \mathcal{C} : (M, K) \mapsto C$$

decryption by a function

$$d : \mathcal{C} \times \mathcal{K} \to \mathcal{M} : (C, K) \mapsto M$$

Def. 2.8. A cryptosystem is a five-tuple $(\mathcal{M}, \mathcal{K}, \mathcal{C}, e, d)$ with $\mathcal{M}, \mathcal{K}, \mathcal{C}, e, d$ as above such that

$$d(e(M,K),K) = M \quad \text{for } (M,K) \in \mathcal{M} \times \mathcal{K}.$$

Cryptanalysis

General assumptions: O/E knows the cryptosystem being used.

Kerckhoff's principle: The security of a system shall not rely on the premise that the system is unknown.

Further information: language, context, statistical frequencies, etc.

Objective: determine the key

Different types of attacks:

a) Ciphertext only
b) Known plaintext (string of ciphertext and corresp. plaintext)
c) Chosen plaintext (access to the encryption machinery)
d) Chosen ciphertext (access to the decryption mach.)

b) is a minimal requirement, c) and d) are hardest, Classical systems 2.1, 2.2, 2.4, will fail.

3. Cryptanalysis of Classical System

3.1 Frequency analysis

Monoalphabetic cipher retains the frequency of characters in natural language. In English $\{E, T, A, O, I, N\}$ combine 51.75% of all frequencies.

Avoid this attack by:

non-natural language

enlarge the alphabet, eg., DES with $\mathcal{X} = \{0,1\}^{64}$

$|\mathcal{X}| = 2^{64}$

3.2. Friedman-Test

Objective: decide whether a cipher is monoalphabetic or polyalphabetic.

Alphabet: $\mathcal{Y} = \{1, \dots, m\}$

Ciphertext: $C = (C_1, \dots, C_n)$ modeled by i.i.d. r.v. $C_1, \dots, C_n$ with $P(C_i = \ell) = p_\ell$, $\ell = 1, \dots, m$

Def. 3.1.

$$I_C = I(C_1, \dots, C_n) = \frac{|\{(i,j) \mid C_i = C_j, 1 \le i < j \le n\}|}{\binom{n}{2}}$$

is called index of coincidence.

Obviously: $I_C = 1 \iff C_1 = \dots = C_n$

$I_C = 0 \iff$ all C_i are different.

Different representation of I_C:

Let $N_\ell = |\{i \mid C_i = \ell\}|$, $\ell = 1, \dots, u$

(no. of entries equal to char ℓ)

Then

$$I_C = \frac{1}{\binom{u}{2}} \sum_{\ell=1}^{u} \binom{N_\ell}{2} = \frac{2}{u(u-1)} \sum_{\ell=1}^{u} \frac{N_\ell (N_\ell - 1)}{2}$$

$$= \frac{1}{u(u-1)} \sum_{\ell=1}^{u} N_\ell (N_\ell - 1)$$

$$= \sum_{\ell=1}^{u} \frac{N_\ell}{u} \cdot \frac{N_\ell - 1}{u-1}$$

~~Hence~~ By strong law of large numbers

$$\frac{N_\ell}{u} \to q_\ell \quad \text{a.e.} \quad (u \to \infty)$$

Hence

$$I_C = \sum_{\ell=1}^{u} \underbrace{\frac{N_\ell}{u}} \underbrace{\frac{N_\ell - 1}{u-1}} \to \sum_{\ell=1}^{u} q_\ell^2 = K_C \quad (u \to \infty) \text{ a.e.}$$

Another representation of I_C:

Let $Y_{ij} = \begin{cases} 1, & C_i = C_j \\ 0, & \text{otherwise} \end{cases}$, $1 \le i < j \le u$

Then: $I_C = \frac{1}{\binom{u}{2}} \sum_{1 \le i < j \le u} Y_{ij}$

Lemma 3.3. $E(\bar{I}_c) = \sum_{\ell=1}^{u} q_\ell^2 = \kappa_c$.

Proof.

$$E(Y_{ij}) = 1 \cdot P(C_i = C_j)$$

$$= \sum_{\ell=1}^{u} P(C_i = \ell, C_j = \ell) = \sum_{\ell=1}^{u} \underbrace{P(C_i = \ell)}_{q_\ell} \underbrace{P(C_j = \ell)}_{q_\ell}$$

$$= \sum_{\ell=1}^{u} q_\ell^2 = \kappa_c$$

$$E(\bar{I}_c) = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \underbrace{E(Y_{ij})}_{\kappa_c} = \kappa_c \quad \square$$

In summary: $\bar{I}_c$ is an unbiased, strongly consistent estimator of κ_c, i.e.,

$\bar{I}_c \to \kappa_c$ a.e. $(n \to \infty)$, $E(\bar{I}_c) = \kappa_c$

By the Cauchy-Schwarz inequality

$$\underbrace{\left(\sum_{\ell=1}^{u} q_\ell\right)^2}_{=1} \le \underbrace{\left(\sum_{\ell=1}^{u} q_\ell^2\right)}_{\kappa_c} \underbrace{\left(\sum_{\ell=1}^{u} 1\right)}_{u}$$

$$\Leftrightarrow \sum_{\ell=1}^{u} q_\ell^2 \ge \frac{1}{u} \quad \text{with equality iff } q_\ell = \frac{1}{u} \ \forall \ell = 1, \dots, u$$

If $q_\ell = \frac{1}{26}$ (uniform distribution over char.)

the $K_U = \sum_{\ell=1}^{26} \frac{1}{26^2} = 0.0385$

For German language $K_G = 0.0762$

Table of K-values:

	English	French	Swedish	Russian	Arabic
K	0.066895	0.074604	0.064489	0.056074	0.075889

Application: determine I_c for a given ciphertext C

$I_c \sim 0.0762 \rightarrow$ monoalphabetic

$I_c \sim 0.0385 \rightarrow$ polyalphabetic, close to a uniform distr.