2.4. Vigenère cipher (1523–1596)

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

Key string, key word of length k: $(s_0, \ldots, s_{k-1})$

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

Key stream: $s_0, \ldots, s_{k-1}, s_0, \ldots, s_{k-1}, \ldots$

Encryption: componentwise addition mod 26.

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

Note: Vigenere cipher is polyalphabetic.

Extension:

Vernam cipher (1917) (also polyalphabetic)

Same as Vigenère, but for each plaintext use a randomly generated key stream of the same length. (one-time pad)

Disadvantage: tedious key distribution over a secure channel.

2.6. Joint principles of the above Cryptosystems

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

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

$\mathcal{X}^\ell, \mathcal{Y}^\ell$: 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 ciphertexts

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

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

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

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

Encryption: is described by a function (encryption rule)

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

decryption by a function (decryption rule)

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

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

$$e: \mathcal{M} \times \mathcal{K} \to \mathcal{C}, \quad d: \mathcal{C} \times \mathcal{K} \to \mathcal{M}$$

such that $d(e(M,K),K) = M$

for all $(M,K) \in \mathcal{M} \times \mathcal{K}$. ⊥

Cryptoanalysis (2.6.1)

General assumption: O knows the cryptosystem being used.

Kerckhoff's principle.

Further information: language, context, statistical frequency of characters, etc.

Objective: determine the key K.

Different levels of attacks:

a) Ciphertext only (a string of ciphertext only)
b) Known plaintext (ciphertext & plaintext known)
c) Chosen plaintext (access to the encryption rule)
d) Chosen ciphertext (access to the decryption rule)

b) is a minimal requirement

c) and d) are the hardest.

Classical system 2.1, 2.2, 2.4 will fail

3. Cryptanalysis of Classical Systems

Monoalphabetic ciphers retain the frequencies of characters in natural languages. In English $\{E, T, A, O, I, N\}$ combine 57.75% of all frequencies.

English. from Menezes et.al. p.247/248

Cryptography

Char.	Frequ.	Char.	Frequ.	Char.	Frequ.	Char.	Frequ.
A	8,04	H	5,49	O	7,60	V	0,99
B	1,54	I	7,26	P	2,00	W	1,92
C	3,06	J	0,16	Q	0,11	X	0,19
D	3,99	K	0,67	R	6,12	Y	1,73
E	12,51	L	4,14	S	6,54	Z	0.09
F	2,30	M	2,53	T	9,25		
G	1,96	N	7,09	U	2,71		

Frequency of single characters in English text.

Digr.	Frequ.	Digr.	Frequ.	Digr.	Frequ.	Digr.	Frequ.
AN	1,81	ER	2,13	ON	1,83	TE	1,30
AT	1,51	ES	1,36	OR	1,28	TH	3,21
ED	1,32	HE	3,05	RE	1,90	TI	1,28
EN	1,53	IN	2,30	ST	1,22		

Frequency of 15 common digrams in English text.

Char.	Frequ.	Char.	Frequ.	Char.	Frequ.	Char.	Frequ.
A	6,51	H	4,76	O	2,51	V	0,67
B	1,89	I	7,55	P	0,79	W	1,89
C	3,06	J	0,27	Q	0,02	X	0,03
D	5,08	K	1,21	R	7,00	Y	0,04
E	17,40	L	3,44	S	7,27	Z	1,13
F	1,66	M	2,53	T	6,15		
G	3,01	N	9,78	U	4,35		

Frequency of single characters in German text.

Digr.	Frequ.	Digr.	Frequ.	Digr.	Frequ.	Digr.	Frequ.
EN	3,88	TE	2,26	EI	1,88	ES	1,52
ER	3,75	DE	2,00	IE	1,79		
CH	2,75	ND	1,99	IN	1,67		

Frequency of 10 common digrams in German text.

E,T,A,O,I,N: 51,75%

-O-A- -O-N- T-E -ANT- AN- --I-T- AN- T---E-
T-E- ON A- -E -ET--NE- TO O-EN T-E -OO-.
T-E -I--T AN- -O-- AI- --E-T IN, -E-IN-IN-
-I- O- --AT -E ---T -OO- -I-E. -O-A- -A- A
--A-- -AN, -IT- -E-I-ATE -I-- --EE--ONE- AN-
-ON- E-E-A--E-. -E -A- A--A-- -EEN --O-- O- -I-
-OO-- AN- --E--E- -E--. NO- -E -EA-E- T-AT T-E
--A- -T----E AN- -ON- -AI- -A-E -I- A--EA- -E-E-T.

E,T,A,O,I,N + H,D: 61,23%

-O-A- -O-ND THE -ANT- AND -HI-T- AND T---ED
THE- ON A- HE -ET--NED TO O-EN THE DOO-.
THE -I-HT AND -O-D AI- --E-T IN, -E-INDIN-
HI- O- -HAT HE ---T -OO- -I-E. -O-A- -A- A
--A-- -AN, -ITH DE-I-ATE HI-H -HEE--ONE- AND
-ON- E-E-A-HE-. HE HAD A--A-- -EEN --O-- O- HI-
-OO-- AND D-E--ED -E--. NO- HE -EA-ED THAT THE
--A- -T----E AND -ON- HAI- -ADE HI- A--EA- -E-E-T.

E,T,A,O,I,N + H,D: 61,23% + RE

-O-A- -O-ND THE -ANT- AND -HI-T- AND T---ED
THE- ON A- HE RET--NED TO O-EN THE DOO-.
THE -I-HT AND -O-D AI- --E-T IN, RE-INDIN-
HI- O- -HAT HE ---T -OO- -I-E. -O-A- -A- A
--A-- -AN, -ITH DE-I-ATE HI-H -HEE--ONE- AND
-ON- E-E-A-HE-. HE HAD A--A-- -EEN --O-- O- HI-
-OO-- AND DRE--ED -E--. NO- HE -EARED THAT THE
--A- -T----E AND -ON- HAI- -ADE HI- A--EA- -ERE-T.

Full text *(American Short Stories 1998, p.49)*

GOPAL FOUND THE PANTS AND SHIRTS AND TUGGED
THEM ON AS HE RETURNED TO OPEN THE DOOR.
THE LIGHT AND COLD AIR SWEPT IN, REMINDING
HIM OF WHAT HE MUST LOOK LIKE. GOPAL WAS A
SMALL MAN, WITH DELICATE HIGH CHEEKBONES AND
LONG EYELASHES. HE HAD ALWAYS BEEN PROUD OF HIS
LOOKS AND DRESSED WELL. NOW HE FEARED THAT THE
GRAY STUBBLE AND LONG HAIR MADE HIM APPEAR BEREFT.

Avoid this attack by:
non-natural languages, compression, enlarged alphabets, e.g., DES with $\mathcal{X} = \{0,1\}^{64}$, $\#\mathcal{X} = 2^{64}$.

3.2. Friedman-Test

Objective: decide whether a cipher is mono- or polyalphabetic. Alphabet: $\mathcal{X} = \{1, \dots, m\}$
Ciphertext: $C = (C_1, \dots, C_n)$ modeled by i.i.d. r.v. $C_1, \dots, C_n$ with
$P(C_i = \ell) = q_\ell$, $\ell = 1, \dots, m$

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. ⌟

Obvious: $I_C = 1 \iff C_1 = \dots = C_n$
$I_C = 0 \iff$ all C_i are different

Different representation of I_C:
Let $N_\ell = \#\{i \mid C_i = \ell\}$, $\ell = 1, \dots, m$
Then $I_C = \frac{1}{n(n-1)} \sum_{\ell=1}^{m} N_\ell (N_\ell - 1)$.

By the strong law of large numbers.

$$\frac{N_\ell}{n} \to q_\ell \quad (n\to\infty) \quad \text{a.e.} \quad \forall \ell = 1,\dots,u$$

Hence:

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

Another representation of I_c:

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

Then $$I_c = \frac{1}{\binom{n}{2}} \sum_{1\le i<j\le n} Y_{ij}$$

Lemma 3.3. $E(I_c) = \sum_{\ell=1}^{u} q_\ell^2 = K_c$

Proof. $$E(Y_{ij}) = P(C_i = C_j) = \sum_{\ell=1}^{u} P(C_i = \ell, C_j = \ell)$$

$$= \sum_{\ell=1}^{u} \underbrace{\left(P(C_i = \ell)\right)^2}_{q_\ell} = \sum_{\ell=1}^{u} q_\ell^2 = K_c$$

Hence $$E(I_c) = \frac{1}{\binom{n}{2}} \sum_{1\le i<j\le n} K_c = K_c. \quad \square$$

I_c is a strongly consistent unbiased estimator of K_c.

By Cauchy-Schwarz inequality

$$\underbrace{\left(\sum_{\ell=1}^{m} q_\ell\right)^2}_{=1} \le m \sum_{\ell=1}^{m} q_\ell^2 \iff \sum_{\ell=1}^{m} q_\ell^2 \ge \frac{1}{m}$$

with equality iff $q_\ell = \frac{1}{m} \quad \forall \ell = 1, \dots, m.$

If $q_\ell = \frac{1}{26}$ (uniform distribution) then

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

For German language $\kappa_G = 0.0762$

Application: determine I_C for a given ciphertext C.

$I_C \sim 0.0762 \rightarrow$ monoalphabetic, frequencies unchanged

$I_C \sim 0.0385 \rightarrow$ polyalphabetic, uniform distribution

Table κ-values:

	English	French	Russian	Arabic	Chinese
κ	0.066895	0.074604	0.056074	0.075889	?