Unicity distance.

pl.-text — the data encryption standard, 25 char.

⇓ S subs. on A

cipher-text

frequency analysis c.-text → pl.-text.

unique

c.-text → pl.-text$_1$

→ pl.-text$_2$?

→ ...

intuitively depends on size of c.-text

---

unicity distance: shortest cipher-text with unique decryption.

m pl.-text of length n characters

S secret subs. on A Latin alphabet

$c = S(m)$

all possible substitutions on A:

$S_1 = S, S_2, \dots, S_t$, $t = 26!$

all possible decryptions of c

$m = S^{-1}(c), m_2 = S_2^{-1}(c), \dots, m_t = S_t^{-1}(c)$

model $m_2, \dots, m_t$ are random n-char. strings.

p = probability that a random n-ch. string is a sens. English text.

$$= \frac{\#\ n\text{-ch. English text}}{\#\ n\text{-char. strings}} =$$

$$\simeq \frac{2^{n \cdot H_E}}{26^n}$$

H_E binary entropy of English texts per character.

Shannon $1 \le H_E \le 1.5$.

Let's take $H_E = 1.5$

$$\Rightarrow p = \frac{2^{1.5 \cdot n}}{26^n}$$

Consider incorrect decryptions

$m_2, \dots, m_t$

random variable m_i English text

$$\gamma_i = \begin{cases} 1 & \\ 0 & \text{otherwise} \end{cases}$$

$\Rightarrow$ # English texts among $m_2, \dots, m_t$ is $\gamma_2 + \gamma_3 + \dots + \gamma_t$.

expectation of # English texts in $m_2 \dots m_t$ is

$$E(\gamma_2 + \dots + \gamma_t) = E\gamma_2 + \dots + E\gamma_t = (t-1)p$$

$$E\gamma_i = 1 \cdot p + 0 \cdot (1-p) = p$$

We want to find n s.t.

$$(t-1)p < 1.$$

$$\Leftrightarrow (26! - 1) \cdot \frac{2^{n \cdot 1.5}}{26^n} < 1$$

solve this inequality in n

$$n > \frac{\log_2(26! - 1)}{\log_2 26 - \underbrace{H_E}_{\approx 1.5}} \approx 27.$$

It means cipher-text produced with subs. cipher, of length > 27 $\Rightarrow$ likely unique decryption.

unicity distance for subs. cipher 27 characters.

example. $n = 25$. likely unique decryption.

Homophonic Ciphers.

why substitution cipher is easy to

break.

nowadays : block ciphers on complecated rounds

in the past : relevant to a modern concept of probabilistic encryption.

A Latin alphabet

$A^2 = \{aa, ab, ac, \dots, zz\}$
all digrams over A.

$|A^2| = 26^2 = 676$

partition A^2 into 26 subsets :

$$A^2 = H_a \cup H_b \cup \dots \cup H_z \qquad (*)$$

$|H_a|, |H_b|, \dots, |H_z|$ fixed.

1) key space is all partitions $(*)$

$$\binom{26^2}{|H_a|} \cdot \binom{26^2 - |H_a|}{|H_b|} \cdots = \frac{26^2!}{|H_a|!\,|H_b|! \cdots |H_z|!}$$

huge number.

↑ binomial coeff.
$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

2) Encryption

$m \in A$

$m = m_1 m_2 \dots m_N$, $m_i \in A$

each m_i is encrypted by a digram

$c_{i1} c_{i2} \in H_{m_i}$

↑ random digram

cipher-text is

$c_{11} c_{12}\ c_{21} c_{22} \dots c_{N1} c_{N2}$

3) Decryption

$\underline{c_{i1} c_{i2}} \in H_\alpha \Rightarrow m_i = \alpha$

How to provide cipher-text uniformity.

$|H_\alpha|$

q_α frequency of ch. $\alpha \in A$ in long English texts

take $|H_\alpha|$ s.t. $|H_\alpha| \approx q_\alpha \cdot 26^2$

$|H_a| \approx q_a \cdot 26^2 \approx 0.0805 \cdot 26^2 \approx 54$

$|H_b| \approx q_b \cdot 26^2 \approx 0.0162 \cdot 26^2 \approx 11$

$\dots$

$|H_z| \approx q_z \cdot 26^2 \approx 0.0009 \cdot 26^2 \approx 1$

and $|H_a| + |H_b| + \dots + |H_z| = 26^2$

let $x_1 x_2$ be any digram, $x_1 x_2 \in H_\alpha$

$$P_2(c_{i1}c_{i2} = x_1x_2) =$$

$$= P_2(c_{i1}c_{i2} = x_1x_2,\ m_i = \alpha) =$$

this event implies

$$= P_2(m_i = \alpha) \cdot P_2\left(c_{i1}c_{i2} = x_1x_2,\ m_i = \alpha \,/\, m_i = \alpha\right)$$

$$= P_2(m_i = \alpha) \cdot P_2\left(c_{i1}c_{i2} = x_1x_2 \,/\, m_i = \alpha\right)$$

$$= q_\alpha \cdot \frac{1}{|H_\alpha|} \approx \frac{1}{26^2}$$

⇒ all digrams which encrypt pl.-text characters have the same probability.

⇒ cipher-text characters are uniformly distributed. (exercise)

⇒ given cipher-text no way to find pl-text.

Two drawbacks.

1) decryption is not handy.
2) weak against known-pl. attack.

Polyalphabetic Ciphers.

block cipher with block size $t > 1$.

key space . t subst. on A Latin

$(P_1 P_2 \dots P_t)$

encryption $m = m_1 m_2 \dots m_t \mid m_{t+1} m_{t+2} \dots m_{2t} \dots$

$c = P_1(m_1) P_2(m_2) \dots P_t(m_t) \mid P_1(m_{t+1}) P_2(m_{t+2}) \dots P_t(m_{2t}) \mid \dots$

decryption with decryption key

$(P_1^{-1} P_2^{-1} \dots P_t^{-1})$

Why polyalphabetic !

$$\begin{array}{cccc} m_1 & m_2 & \dots & m_t \\ \uparrow & \uparrow & & \uparrow \\ A_1 & A_2 & & A_t \end{array}$$

P_i substitution on A_i

$A_1 = A_2 = \dots = A_t = A$ Vigenère cipher

$P_1 P_2 \dots P_t$ are shifts (rotations) on A

$|\text{key-space}| = 26^t$ simple Vigenère cipher

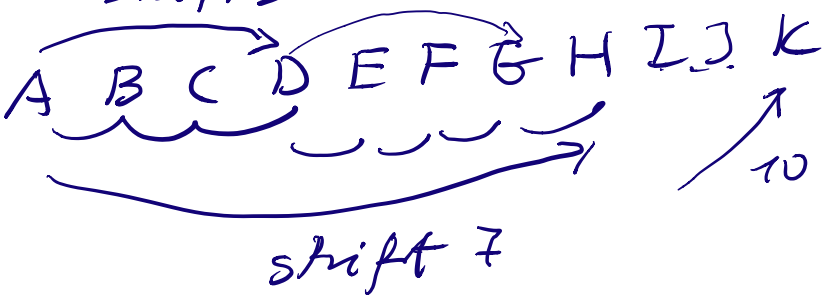
$P_1 P_2 \dots P_t$ are random subst. on A

full Vigenère cipher.

$|\text{key-space}| = (26!)^t$

Example. A = Latin, $t = 3$

encryption key $e = (p_1 p_2 p_3) = (3, 7, 10)$



THE - DAT - AEN - CRY - PTI - ONS - TAN -
WOO - gkd - dex - tyi - sas - zuc - wRx -
- DAR - D
ghb - g

$A \longrightarrow R, d$

$T \rightarrow d, a, w$

$\Rightarrow$ cipher-text is more uniform
uniformity increasing for large t.

On Tuesday study Cryptanalysis