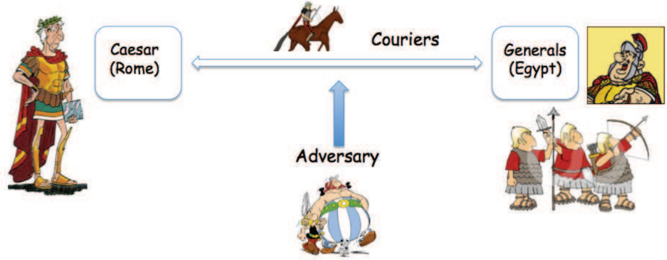
Classical Ciphers: Shift Cipher

Çetin Kaya Koç

http://cs.ucsb.edu/~koc
koc@cs.ucsb.edu



Caesar (Rome)

Couriers

Generals (Egypt)

Adversary

Shift Cipher

- Input/output: $\{a, b, \ldots, z\}$ with encoding $\{0, 1, \ldots, 25\}$

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14

p	q	r	s	t	u	v	w	x	y	z
15	16	17	18	19	20	21	22	23	24	25

- Encryption function: $E_k(x) = x + k \pmod{26}$
  Decryption function: $D_k(y) = y - k \pmod{26}$
- The encryption or decryption key: $k \in \{0, 1, 2, \ldots, 25\}$
- Key space size: 26 (or 25, if you do not count $k = 0$)
- Caesar cipher: Shift cipher with a constant encryption key $k = 3$

Shift Cipher

- For $k = 15$, `hello` is encrypted as `wtaad` since
  $E_{15}(\texttt{h}) = E_{15}(7) = 7 + 15 = 22 \pmod{26} \rightarrow \texttt{w}$
  $E_{15}(\texttt{e}) = E_{15}(4) = 4 + 15 = 19 \pmod{26} \rightarrow \texttt{t}$
  $E_{15}(\texttt{l}) = E_{15}(11) = 11 + 15 = 26 = 0 \pmod{26} \rightarrow \texttt{a}$
  $E_{15}(\texttt{o}) = E_{15}(14) = 14 + 15 = 29 = 3 \pmod{26} \rightarrow \texttt{d}$
- For $k = 12$, `eqqw` is decrypted as `seek` since
  $D_{12}(\texttt{e}) = D_{12}(4) = 4 - 12 = -8 = 18 \pmod{26} \rightarrow \texttt{s}$
  $D_{12}(\texttt{q}) = D_{12}(16) = 16 - 12 = 4 \pmod{26} \rightarrow \texttt{e}$
  $D_{12}(\texttt{w}) = D_{12}(22) = 22 - 12 = 10 \pmod{26} \rightarrow \texttt{k}$

Cryptanalysis of Shift Cipher

- Ciphertext only (CO)
  - Exhaustive key search: a paragraph of ciphertext (in order to avoid ambiguity)
  - Frequency analysis: a paragraph of ciphertext (in order to get statistically reliable frequency count)
- Known plaintext (KP): a single plaintext/ciphertext pair
- Chosen plaintext (CP): a single plaintext/ciphertext pair
- Chosen text (CT): a single plaintext/ciphertext pair

Exhaustive Key Search

- Given an encrypted text: `vnnc vn jc ljsn jc oxda yv`
- Decrypt the text with all possible keys:

  $\xrightarrow{k=1}$ `ummb um ib kirm ib nwcz xu`

  $\xrightarrow{k=2}$ `tlla tl ha jhql ha mvby wt`

  ...

  $\xrightarrow{k=8}$ `nffu nf bu dbkf bu gpvs qn`

  $\xrightarrow{k=9}$ `meet me at caje at four pm`

- A short encrypted text may have several “meaningful” decryptions:

  `vnnc` $\xrightarrow{k=9}$ `meet`

  `vnnc` $\xrightarrow{k=25}$ `wood`

- For a sufficiently long encrypted text, there will not be ambiguity

Frequency Analysis

- The most frequently occurring ciphertext is the encryption of the most frequently occurring plaintext
- In English: that would be the letter e, followed up by letters t and a

Letter frequencies (percentages) in English

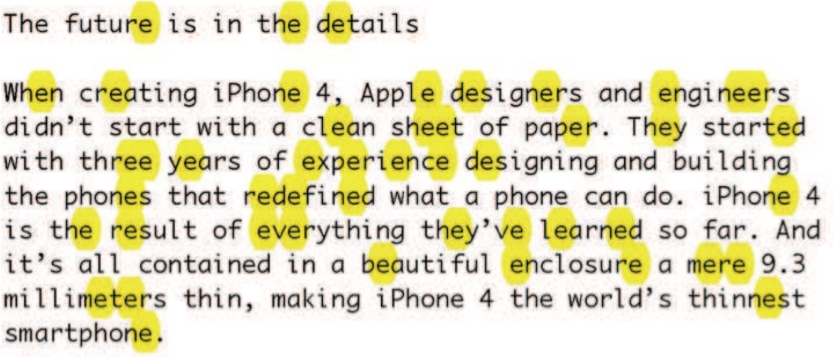
a	b	c	d	**e**	f	g	h	i	j
8.2	1.5	2.8	4.3	**12.7**	2.2	2.0	6.1	7.0	0.2

k	l	m	n	o	p	q	r	s	**t**
0.8	4.0	2.4	6.7	7.5	1.9	0.1	6.0	6.3	**9.1**

u	v	w	x	y	z
2.8	1.0	2.3	0.1	2.0	0.1

- Compute the ciphertext letter frequencies, and find the most frequently occurring the letter: this must be the ciptertext for the letter e

Occurrences of Letter e

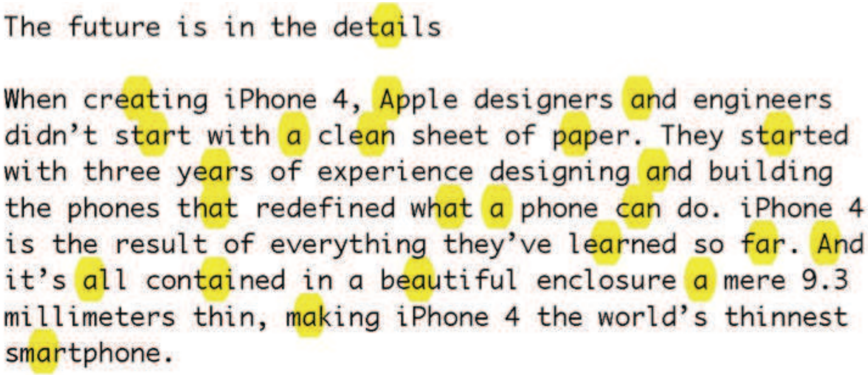


The future is in the details

When creating iPhone 4, Apple designers and engineers didn’t start with a clean sheet of paper. They started with three years of experience designing and building the phones that redefined what a phone can do. iPhone 4 is the result of everything they’ve learned so far. And it’s all contained in a beautiful enclosure a mere 9.3 millimeters thin, making iPhone 4 the world’s thinnest smartphone.

Frequency: $\frac{54}{435} \approx 12.4\%$

Occurrences of Letter a



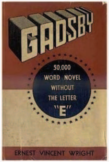
The future is in the details

When creating iPhone 4, Apple designers and engineers didn't start with a clean sheet of paper. They started with three years of experience designing and building the phones that redefined what a phone can do. iPhone 4 is the result of everything they've learned so far. And it's all contained in a beautiful enclosure a mere 9.3 millimeters thin, making iPhone 4 the world's thinnest smartphone.

Frequency: $\frac{23}{435} \approx 5.3\%$

Unusual Texts

The novel "Gadsby" by E. V. Wright is written as a lipogram*; it has 50,000 words in it without a single occurrence of the letter e



"A Void", translated from the original French "La Disparition" (The Disappearance), is a 300-page lipogrammatic novel, written in 1969 by Georges Perec, entirely without using the letter e (except for the author's name)

Cover of the English translation of *La Disparition*

However, the probability of occurrence for such texts is very low

∗ A lipogram (leipográmmatos: leaving out a letter) is a kind of constrained writing or word game consisting of writing paragraphs in which a particular letter or group of letters is avoided

Frequency Analysis

- Given the short ciphertext: `tbxqebo fp dobxq ebob`
- Frequency analysis finds the most frequently occurring letter as `b`
- The letter `b` (most probably) is the ciphertext for the letter `e`

$$\begin{aligned} E_k(\mathtt{e}) &= \mathtt{b} \\ E_k(4) &= 4 + k = 1 \pmod{26} \\ k &= 1 - 4 \;=\; -3 \;=\; 23 \pmod{26} \end{aligned}$$

- Indeed, if we decrypt the encrypted text using the key $k = 23$, we obtain:

  $\mathtt{tbxqebo\ fp\ dobxq\ ebob} \xrightarrow{k=23} \mathtt{weather\ is\ great\ here}$

Known Plaintext Scenario

- Given a (any) single plaintext/ciphertext pair (x, y), we have

$$\begin{aligned} E_k(x) &= x + k = y \pmod{26} \\ k &= y - x \pmod{26} \end{aligned}$$

- Consider the encrypted message:
  `zrrg zr ng bhe frperg ybpngvba`
  and the plaintext/ciphertext pair: $\mathtt{m} \rightarrow \mathtt{z}$

$$\begin{aligned} E_k(\mathtt{m}) &= \mathtt{z} \\ E_k(12) &= 12 + k = 25 \pmod{26} \\ k &= 25 - 12 = 13 \pmod{26} \end{aligned}$$

We find the key as $k = 13$; indeed this key decrypts the message
$\xrightarrow{k=13}$ `meet me at our secret location`

Chosen Plaintext Scenario

- Since the cryptanalyst gets to choose the plaintext, and obtains the ciphertext, she/he can select the pair (x, y) such that $x = \mathtt{a}$

$$\begin{aligned} E_k(\mathtt{a}) &= 0 + k = y \pmod{26} \\ k &= y \pmod{26} \end{aligned}$$

- In other words, the key is equal to the encoding of the letter that is the ciphertext for `a`
- Using the previous encrypted text:
  `zrrg zr ng bhe frperg ybpngvba`
- We ask and obtain the ciphertext for `a`, which is given as `n`
- Since the encoding of `n` is 13, we obtain the key as $k = 13$

Chosen Ciphertext Scenario

- Similarly, if we can choose the ciphertext y in the pair (x, y), and obtain the plaintext x, all we have to do is to solve for the linear congruence
$$E_k(x) = x + k = y \pmod{26}$$
to obtain the key as
$$k = y - x \pmod{26}$$
- In fact the difficulty of obtaining the key for all three scenarios: KP, CP, CT is about the same: Obtain a single plaintext/ciphertext and solve for the key in the above linear congruence
- Therefore, we conclude that the shift cipher is very weak

Cryptanalysis of the Shift Cipher

- The number of keys is very small: 26 (or 25)
- Ciphertext only attack succeeds by performing 26 (or 25) decryptions of a not-so-short encrypted message (in order to avoid ambiguity)
- Known plaintext attack succeeds if we obtain a single pair (x, y) of plaintext and ciphertext; we solve for the linear congruence:

$$k = y - x \pmod{26}$$

- Similarly, the chosen text attack succeeds if we obtain a single pair (x, y) of plaintext and ciphertext; we use the above equation to obtain the key