Lecture 11: Cryptology

Cesar Cypher

In this worksheet we crack the Caesar cypher using statistical analysis.

Letter	Percentage	Letter	Percentage
E	11.16	M	3.01
A	8.50	H	3.00
R	7.58	G	2.47
I	7.54	B	2.07
O	7.16	F	1.81
T	6.95	Y	1.78
N	6.65	W	1.29
S	5.74	K	1.10
L	5.49	V	1.01
C	4.54	X	0.29
U	3.63	Z	0.27
D	3.38	J	0.20
P	3.17	Q	0.20

The frequency of letters is relevant for designing keyboards. The Qwerty keyboard for example has ESER and OI in prominent places.
The 'top twelve' letters help with about 80 percent of the text. You can remember the first 8 with the memnonic

"A SIN TO ERR".

An other thing to look for: The **top pairs** which appear are

TH HE AN RE ER IN ON AT ND ST ES EN OF TE ED OR TI HI AS TO

The most frequent **double letters** are

"LL EE SS OO TT FF RR NN PP CC"

Example

We aim to decrypt the following text:

xf uif qfpqmf pg uif vojufe tubuft,

jo psefs up gpsn b npsf qfsgfdu vojpo,

ftubcmjti kvtujdf, jotvsf epnftujd usborvjmjuz,

qspwjef gps uif dpnnpo efgfodf,

qspnpuf uif hfofsbm xfmgbsf,

boe tfdvsf uif cmfttjoht pg mjcfsuz

up pvstfmwft boe pvs qptufsjuz,

ep psebjo boe ftubcmjti uijt dpotujuvujpo gps uif vojufe tubuft pg bnfsjdb

Decoding:
Count the number of letters which occur. Since we have not much time, the 8 most frequent letters are listed in this text. Can you figure out the text?

f	appears	39 times
u	appears	29 times
p	appears	25 times
t	appears	20 times
s	appears	20 times
j	appears	20 times
o	appears	17 times
b	appears	14 times

The Vigenère Cipher

We learn how to encrypt messages using the Vigenère Cipher. This encryption was used for a long time and should be seen as an important marker in the development of substitution ciphers:

Julius Caesar	-70
Ahmad al-Qalqashandi	1400
Leon Battista Alberti	1467
Johannes Trithemius	1508
Blaise de Viginère	1586
Charles Babbage	1854
Friedrich Kasiski	1863
Arthur Scherbius	1920

Blaise de Vigenère

Assume we have a secret key like "ENIGMA". Given a text like "HARVARD IS COOL", we encrypt it using the following table: for the first letter, we use the line starting with E, for the second letter, we use the line starting with N etc.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Now it is your turn

ENIGMAE NI GMAE
HARVARD IS COOL

RSA Encryption

We want to understand the basic mechanism for RSA encryption.

Ron Rivest, Adi Shamir and Len Adleman.

An **RSA public key** is a pair (n, a) where n is an integer with secret factorization $n = pq$ and where $a < (p-1)(q-1)$ is such that there exists b with $ab = 1 \bmod (p-1)(q-1)$. Ana publishes this pair. If Bob wants to send a secrete message to Ana, he transmits to Ana the message

$$y = x^a \bmod n \ .$$

Ana can read the email by computing

$$y^b \bmod n \ .$$

Why does it work? We use the **Fermat's little theorem** which tells that $x^{p-1} - 1$ is is divisible by p and $x^{q-1} - 1$ is divisible by q. But this assumes p, q to be prime. For $n = pq$, we have $x^{(p-1)(q-1)} - 1$ divisible by pq.

Problem 1) Take $p = 3$ and $q = 5$ Verify that $2^{(p-1)(q-1)} - 1$ is divisible by $n = pq$.

Because $y^b = x^{ab} = x^{(p-1)(q-1)} = x \bmod n$, Ana gets the message, Bob has sent. But Eve has no chance to read it, because the only thing Eve can see is y and (n, a). It is believed that without the factorization of n, the message can not be read.

Let $(55, 13)$ be the public key of Ana. Assume Ana has the message $x = 4$ to submit. She computes $y = 4^{13} \bmod 55 = 9$. Because $55 = 11 * 5 = pq$, Ana knows $(p-1)(q-1) = 40$ and can obtain $b = 37$. With $x = 9^{37} \bmod 55$ she gets back 4.

Problem 2) Assume $(n, a) = (15, 2)$ is the public key of Ana. You are Bob. Send the message $x = 7$ to Ana.

Problem 3) You are now Ana and have received the message from Bob. Decipher it.