IALP 2011 – Cryptography

P. Stallinga

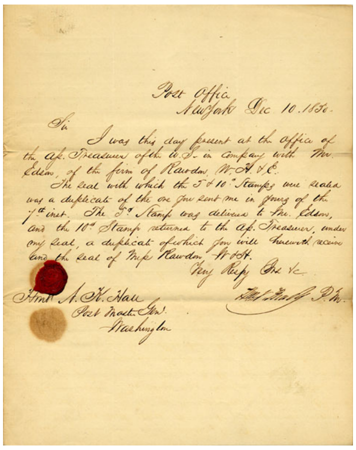
MIEET 1º ano

Whereas the bar code (of the lecture Barcodes) is simply a way of translating the human-readable information into a machine-readable format, sometimes we want to make sure that not everybody can read the information in the message or document. Even worse, can we even trust the bearer of the message? Is he not an impersonator, trying to convince us he is somebody that he is not. For this, we can make use of passwords and encryption.

Password

The most basic problem is: Do we trust the person or entity that brings us the information? The traditional way is by means of a password. This ensures that the person is indeed the person he claims to be. The idea of a password is very ancient. Sentries would challenge those wishing to enter an area or approaching it to supply a password or 'watchword'. Nowadays, everybody has passwords. Nearly everybody has accounts for e-mail, Facebook, Twitter, linked-in, Hyves, Applestore, Farmville, etc. Everything comes with a password, to make sure that the owner is the only person to have access to it. Even our PIN code for the ATM (money machine) is a password to identify ourselves.

The next step is that a person --- a 'messenger' --- would bring a message from, for instance, one king to another king. The letter would be 'sealed' by the king, thus authenticating that the message indeed originated from said king and not from an imposter. The king would use his seal ring and impress it onto molten red wax to show that it is him writing the document. The result is a letter as shown here below as an example.



Post Office
New York Dec 10. 1850.

Sir

I was this day present at the office of the as. Treasurer of the U.S. in company with Mr. Edson, of the firm of Rawdon, W. H. & E.

The seal with which the 5 & 10c Stamps were sealed was a duplicate of the one you sent me in yours of the 7th inst. The 5c Stamp was delivered to Mr. Edson, and the 10c Stamp returned to the as. Treasurer, under my seal, a duplicate of which you will herewith receive and the seal of Mess Rawdon W & H.

Very Resp. Yrs &c

Wm. V. Brady P.M.

Hon. N. K. Hall
Post Master Genl.
Washington

(National postal museum)

Simple encryption

Still, the message itself could easily be read by others. Even if the letter was closed physically by the seal, it could be opened and resealed with a little tampering. In some cases this does not matter (“pay 100 coins to the bearer of the document”, “I would like to invite you to the wedding of my daughter”), but in other cases this could be fatal (“At dawn at the 3rd of August we will attack Castle Redgrave”). Thus, the information should also be encrypted, in order that it can be read only by the intended person.

The simplest of codes is by replacing every letter of the alphabet by another letter or symbol.

Imagine the text

```
THIS IS A TEXT THAT IS ENCODED AND CAN BE READ BY NOBODY
```

Now we will write the same text with the following translation table (which is nothing more than replacing every letter by its next-nearest neighbor in the alphabet in a cyclic way ['Z' becomes 'A' again]):

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	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

The text becomes

```
UIJT JT B UFYU UIBU JT FODPEFE BOE DBO CF SFBE CZ OPCPEZ
```

Or we can even use a different character set altogether. Instead of Latin symbols, we could use (this comes out when I switch my keyboard to the Greek character set):

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
α	β	ψ	δ	ε	φ	γ	η	ι	ξ	κ	λ	μ	ν	ο	π	;	ρ	σ	τ	θ	ω	ς	χ	υ	ζ

The text becomes

```
τηισ ισ α τεχτ τηατ ισ ψοδεδ ανδ ψαν βε ρεαδβυ νοβοδυ
```

Difficult to read, especially for people that don't know Greek. And we could even use a special symbol set that we designed ourselves (see Appendix for the famous Freemason's cypher). The problem is that it is still quite easy to decode, even without knowing the 'password' (in this case the character map). Especially so if the text is somewhat longer. As an example, a property of English texts is that different characters have different frequency of use. The most often used character in English is the 'E', closely followed by 'T', etc. The first twelve most letters used are

E T A O I N S H R D L U . . .

Apart from that, the one-letter word is probably 'a'. Other often-used words are 'the' and 'and', etc. On basis of this it is not difficult to decode any text coded by such simple algorithms. Have a go at the next text:

Exercise

1) Try to find the translation table and decode the following text (New York Times, 6 June 2011) that is encoded with an algorithm shift-right-n (for both uppercase and lowercase letters), similar to the example above that was called shift-right-1. This is extremely simple and should not take you longer than a minute or so to break, five to decipher.

```
Xnshj ymj xuwnsl tk 2010, Utwyzlfq mfx gjjs gfyyjwji gd ymj ijgy
hwnxnx ymfy gjlfs ns Lwjjhj fsi mfx xuwjfi fhwtxx rzhm tk Jzwtuj'x
ujwnumjwd.

Kfhji bnym f mjfad ijgy gzwijs fsi wnxnsl nsyjwjxy wfyjx, Utwyzlfq
jsfhyji ymwjj wtzsix tk fzxyjwnyd rjfxzwjx ymfy fsljwji btwpjwx
fsi uzyx nyx jhtstrd nsyt wjhjxxnts. Gzy ymj nsyjwjxy wfyjx gjnsl
hmfwlji ktw nyx ltajwsrjsy ijgy wjrfnsji mnlm, fsi ns Rfd 2011 ymj
ltajwsrjsy fxpji ymj Jzwtujfs Zsnts ktw f knsfshnfq gfnqtzy tk 78
gnqqnts jzwtx, tw fgtzy $114 gnqqnts.

Ymj hwnxnx mfi fqwjfid qji yt ymj wjxnlsfynts tk Uwnrj Rnsnxyjw
Otxj Xthwfyjx, fkyjw tuutxnynts ufwynjx wjojhyji mnx qfxy-inyhm
fyyjruy yt uzxm ymwtzlm f ktzwym ufhpflj tk xujsinsl hzyx fsi yfc
nshwjfxjx. Rw. Xthwfyjx xyfdji ts, qjfinsl f hfwjyfpjw ltajwsrjsy.
```

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 more advanced method is applying a randomization of the characters, instead of a simple right-shift-n. See for example the next table, where we also included the coding of the 'space', to avoid easy recognition of words

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

Our text becomes

```
CX WE WEHECJMCECXHCE WEJDFGAJAEHDAEFHDETJELJHAETNEDGTGAN
```

which is already more difficult to decipher. Note that individual words can no longer be distinguished.

Let's take it a step further. Every letter in the source text will be coded in a different way. We now use really a 'password' (instead of effectively a 'passletter' above). Or better to call it a coding **key**, or code sequence since it does not let us 'pass' anything anymore, but instead just encrypts a text. The word 'key' is adequate, because with the key we can unlock the message.

	Coding key																										
Source text	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
		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

Our text with the password 'BENFICA' becomes

```
  THIS IS A TEXT THAT IS ENCODED AND CAN BE READ BY NOBODY
  BENFICABENFICABENFICABENFICABENFICABENFICABENFICABENFICA
  -------------------------------------------------------- *
= ULVXHKSAEMYMZTAXUFABITDRSKQDFHMFVF DE EJG SINIHDYARAGWFY
```

which is already quite difficult to decipher, but still not very complicated (Did you know that about 30% of the passwords in Portugal are “o glorioso”?). Still, we leave the subject here.

Exercise

2) Code a text and distribute it to everybody:
- Chose a key sequence
- Write a small text to code
- Give your key to another group (you can best use e-mail for that)
- Put your coded text on the pile together with the coded texts of the other groups
- Try to find back the text that was meant for you

Asymmetric keys

The above ideas all make use of symmetric keys. That is, both the sender and the receiver use the same coding algorithm with the same key. This key is something that is exchanged secretly and separate from the coded text, for instance during an earlier physical meeting between the sender and the receiver.

Imagine that I want people (anybody in the world) to be able to send me messages that **only I** can read. I could give everybody in the world the password, but if everybody knows the password, then everybody can also decipher the text. That is not what I want. I want everybody to be able to cypher the text, but me the only one to be able to decipher the coded text. This can be done with asymmetric keys.

To explain how this works, imagine the following situation: Imagine that **everybody knows how to do 'additions', but nobody (but I) knows how to do 'subtractions'**. Now, to send me a number that nobody else can read, I tell you to add the number (the key) 5 to it. Imagine you want to send the number 7. What you will send me is the number 7+5 = 12 and you can even send me the key 5 you used, if you want. Since nobody but I can do subtraction, nobody but I can recover the number 7. I receive your coded number, 12, and knowing the used key 5, I can subtract and find your coded number.

The real asymmetric public-key coding systems such as PGP (pretty good privacy) and TLS (transport layer security) are not exactly working this way, but it serves as a good example how encryption can be asymmetric with the receiver having more information or knowledge than the sender. In any case, you will learn cryptology and network security along the way in this course MIEET.

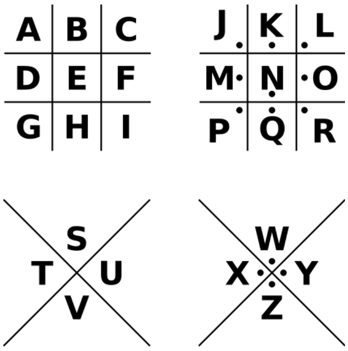
Exercise

3) Send each other encrypted e-mails using PGP
- (In case you don't have one yet): Ask for a PGP key combination at http://www.pgpi.org/
- Send your friends your **public** key
- Send somebody a message using their **public** key
- Decipher a message received by you using your **private** key

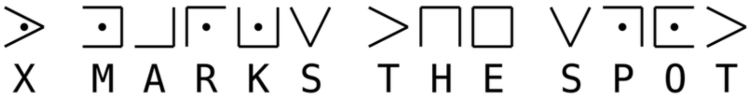
Still, remains the following to decode the following that that has kept me puzzled for ages: `“Cbhatio Uiyua Fhgjhgappat, Earthlings!”`

Appendix. Famous codes and ciphers

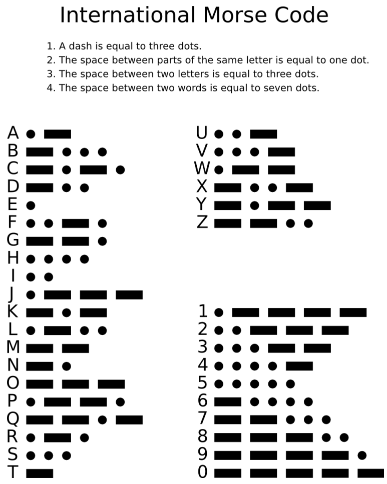
1) Pigpen cipher used by freemasons (Wikipedia), featuring in Dan Browns “The Lost Symbol” (2009), among others:



Example:



2) Morse Code (Wikipedia). Not so much a code as a translation table to make text machine processable and transmittable. (After all, it is a symmetric key that everybody knows).



International Morse Code

1. A dash is equal to three dots.
2. The space between parts of the same letter is equal to one dot.
3. The space between two letters is equal to three dots.
4. The space between two words is equal to seven dots.

A	• ▬	U	• • ▬
B	▬ • • •	V	• • • ▬
C	▬ • ▬ •	W	• ▬ ▬
D	▬ • •	X	▬ • • ▬
E	•	Y	▬ • ▬ ▬
F	• • ▬ •	Z	▬ ▬ • •
G	▬ ▬ •		
H	• • • •		
I	• •		
J	• ▬ ▬ ▬		
K	▬ • ▬	1	• ▬ ▬ ▬ ▬
L	• ▬ • •	2	• • ▬ ▬ ▬
M	▬ ▬	3	• • • ▬ ▬
N	▬ •	4	• • • • ▬
O	▬ ▬ ▬	5	• • • • •
P	• ▬ ▬ •	6	▬ • • • •
Q	▬ ▬ • ▬	7	▬ ▬ • • •
R	• ▬ •	8	▬ ▬ ▬ • •
S	• • •	9	▬ ▬ ▬ ▬ •
T	▬	0	▬ ▬ ▬ ▬ ▬

Example:

. . . — — — . . .

SOS