Mathematics and Cryptography

- **Math 4161 (Winter 2021) Course Outline Course description**
- This course is an introduction to mathematical cryptography, with a focus on public-key cryptography based on number theory. This course is almost entirely theorems and proofs. You will need to be able to construct a mathematical proof, including basic types of arguments that you might use (induction, proof by contradiction, contrapositive, etc.).
- **Information**
- Instructor: Jorge Mello Email: jmelloguitar@gmail.com Lectures schedule on Zoom: Tues, Thurs 10:00 - 11:20 Zoom Office Hours: Tues, Thurs 1:00 - 1:50pm
- **Grading**
- 20% Regular homework assignments 30% Mid-term test 50% Final exam
- **Access/Disability**
- Students with health-related, learning, physical, psychiatric, or sensory disabilities who require reasonable accommodations in teaching style or evaluation methods should discuss their concerns with the course instructor as soon as possible so that appropriate arrangements can be made.
- **Course textbook**
- *An Introduction to Mathematical Cryptography (Undergraduate Texts in Mathematics)*
- 2nd edition, by J. Hoffstein, J. Pipher, and J.H. Silverman

Cryptography comes from the greek Crypto (hidden) + graphy (writing) and refers to the methodology of hiding the content of messages.

Applications: Security/safety, Integrity/confidentiality, Digital signatures, Bank transactions.

For example, there is some evidence that the Roman emperor Julius Caesar used early methods of cryptography to communicate secret messages with his armed legions.

In particular, an early method that came to be known as “Caesar’s Cipher”

To illustrate, suppose that Julius Caesar received the following ciphertext from some companion during a battle, while standing up on his post.

```
j s j r d k f q q n s l g f h p g w j f p y m w t z l m n r r n s j s y q z h n z x
```

Suppose that after this, Caesar's quickly sends orders to some of his legions to move and attack the enemies at a certain point of the war map, and such decision will provide great advantages to Caesar's army for an eventual victory.

How such a cipher of random letters could bring such a decisive outcome?

In this case, if one simply shifts 5 letters up in the alphabet, then a new plaintext written at the lower row below is obtained.

```
j s j r d k f q q n s l g f h p g w j f p y m w t z l m n r r n s j s y q z h n z x
e n e m y f a l l i n g b a c k b r e a k t h r o u g h i m m i n e n t l u c i u s
```

The second line is a decrypted message, which, after broken into words and supplied with punctuation, becomes the following message to Julius Caesar

```
Enemy falling back. Breakthrough imminent. Lucius.
```

For letters like y, we walk back to the beginning of the alphabet. Thus it is helpful to see the letters on a circle rather than on a line.

To decrypt a ciphertext, find it on the inner wheel and read the corresponding plain text on the outer. To encrypt, reverse the process.

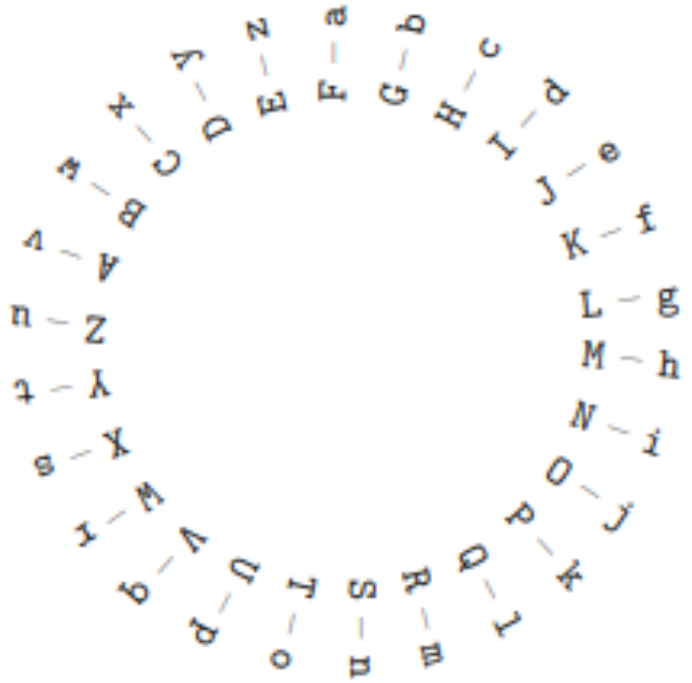

Figure 1.1: A cipher wheel with an offset of five letters

Cipherwheels of this kind have been used for centuries. A shift or translated cipher as such is a simple substitution cipher known as Caesar cipher in honour of Caesar.

Caesar’s cipher is very easy to be decrypted!!

It is enough to try all the 26 possibilities of shifts, checking which one is the
correct.

Hence, one can try to produce ciphers with substitutions more complicated than
the translations.

These are called simple substitution ciphers.

A simple substitution cipher may be viewed as a one-to-one function or rule

$$\{a,b,c,d,e,\ldots,x,y,z\} \longrightarrow \{A,B,C,D,E,\ldots,X,Y,Z\}$$

that assigns for each plaintext in the domain a ciphertext in the range.

Thus, a random permutation of the alphabet will give us a simple substitution cipher, for example, one given by the table below

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

Table 1.1: Simple substitution encryption table

With such table, one can easily encrypt a given message, namely, given

Four score and seven years ago,

We can use the referred table to find the associated encrypted ciphertext on Second line below

f	o	u	r	s	c	o	r	e	a	n	d	s	e	v	e	n	y	e	a	r	s	a	g	o
N	U	R	B	K	S	U	B	V	C	G	Q	K	V	E	V	G	Z	V	C	B	K	C	F	U

It is then customary to write the ciphertext in five-letter blocks:

NURBK SUBVC GQKVE VGZVC BKCFU

Decryption is a similar process. Suppose that we receive the message

GVVQG VYKCM CQQBV KKWGF SCVKC B

Writing now the cipher letters of the range in alphabetical order with the correspondent domain letters in the first line, we have a useful decryption table

j	r	a	x	v	g	n	p	b	z	s	t	l	f	h	q	d	u	c	m	o	e	i	k	w	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

Table 1.2: Simple substitution decryption table

The message then can be decrypted easily

G	V	V	Q	G	V	Y	K	C	M	C	Q	Q	B	V	K	K	W	G	F	S	C	V	K	C	B
n	e	e	d	n	e	w	s	a	l	a	d	d	r	e	s	s	i	n	g	c	a	e	s	a	r

We therefore obtain the real plain text decrypted message

```
Need new salad dressing. -Caesar
```

For simple substitution ciphers, we can see that there are

$$26 \cdot 25 \cdot 24 \cdots 4 \cdot 3 \cdot 2 \cdot 1 = 26! = 403291461126605635584000000.$$

possibilities of different encryption tables.

If one could check one million cipher alphabets per second, such would be able to check 60.60.24.365. 10^6, and 26! / 60.60.24.365. 10^6 is bigger than 10^{13}. Hence this person would take more than 10^{13} years to check all the possibilities, and this is bigger than the estimated age of the universe!
So, apparently, no chance of decrypting, HOWEVER…

Your opponent always uses her best strategy to defeat you, not the strategy that you want her to use. Thus the security of an encryption system depends on the best known method to break it. As new and improved methods are developed, the level of security can only get worse, never better.

In fact, one is able to decrypt simple substitution ciphers quickly using other methods and facts. For example, it is known that, in any language, certain letters of bigrams show up much more frequently than others.
In English, the letter that appear the most in texts is “e”

By decreasing frequency			
E	13.11%	M	2.54%
T	10.47%	U	2.46%
A	8.15%	G	1.99%
O	8.00%	Y	1.98%
N	7.10%	P	1.98%
R	6.83%	W	1.54%
I	6.35%	B	1.44%
S	6.10%	V	0.92%
H	5.26%	K	0.42%
D	3.79%	X	0.17%
L	3.39%	J	0.13%
F	2.92%	Q	0.12%
C	2.76%	Z	0.08%

In alphabetical order			
A	8.15%	N	7.10%
B	1.44%	O	8.00%
C	2.76%	P	1.98%
D	3.79%	Q	0.12%
E	13.11%	R	6.83%
F	2.92%	S	6.10%
G	1.99%	T	10.47%
H	5.26%	U	2.46%
I	6.35%	V	0.92%
J	0.13%	W	1.54%
K	0.42%	X	0.17%
L	3.39%	Y	1.98%
M	2.54%	Z	0.08%

Table 1.3: Frequency of letters in English text

th	he	an	re	er	in	on	at	nd	st	es	en	of	te	ed
168	132	92	91	88	86	71	68	61	53	52	51	49	46	46

(a) Most common English bigrams (frequency per 1000 words)

Let us look to the following cipher text

```
LOJUM YLJME PDYVJ QXTDV SVJNL DMTJZ WMJGG YSNDL UYLEO SKDVC
GEPJS MDIPD NEJSK DNJTJ LSKDL OSVDV DNGYN VSGLL OSCIO LGOYG
ESNEP CGYSN GUJMJ DGYNK DPPYX PJDGG SVDNT WMSWS GYLYS NGSKJ
CEPYQ GSGLD MLPYN IUSCP QOYGM JGCPL GDWWJ DMLSL OJCNY NYLYD
LJQLO DLCNL YPLOJ TPJDM NJQLO JWMSE JGGJG XTUOY EOOJO DQDMM
YBJQD LLOJV LOJTV YIOLU JPPES NGYQJ MOYVD GDNJE MSVDN EJM
```

Table 1.4: A simple substitution cipher to cryptanalyze

	J	L	D	G	Y	S	O	N	M	P	E	V	Q	C	T	W	U	K	I	X	Z	B	A	F	R	H
Freq	32	28	27	24	23	22	19	18	17	15	12	12	8	8	7	6	6	5	4	3	1	1	0	0	0	0
%	11	9	9	8	8	7	6	6	6	5	4	4	3	3	2	2	2	2	1	1	0	0	0	0	0	0

Table 1.5: Frequency table for Table 1.4—Ciphertext length: 298

th	he	an	re	er	in	on	at	nd	st	es	en	of	te	ed
168	132	92	91	88	86	71	68	61	53	52	51	49	46	46

(a) Most common English bigrams (frequency per 1000 words)

LO	OJ	GY	DN	VD	YL	DL	DM	SN	KD	LY	NG	OY	JD	SK	EP	JG	SV	JM	JQ
9	7	6 each		5 each					4 each										

(b) Most common bigrams appearing in the ciphertext in Table 1.4

Table 1.6: Bigram frequencies

LOJUM	YLJME	PDYVJ	QXTDV	SVJNL	DMTJZ	WMJGG	YSNDL	UYLEO	SKDVC
the--	-te--	----e	-----	--e-t	---e-	--e--	----t	--t-h	-----
GEPJS	MDIPD	NEJSK	DNJTJ	LSKDL	OSVDV	DNGYN	VSGLL	OSCIO	LGOYG
---e-	-----	--e--	--e-e	t---t	h----	-----	---tt	h---h	t-h--
ESNEP	CGYSN	GUJMJ	DGYNK	DPPYX	PJDGG	SVDNT	WMSWS	GYLYS	NGSKJ
-----	-----	--e-e	-----	-----	-e---	-----	-----	--t--	----e
CEPYQ	GSGLD	MLPYN	IUSCP	QOYGM	JGCPL	GDWWJ	DMLSL	OJCNY	NYLYD
-----	---t-	-t---	-----	-h---	e---t	----e	--t-t	he---	--t--
LJQLO	DLCNL	YPLOJ	TPJDM	NJQLO	JWMSE	JGGJG	XTUOY	EOOJO	DQDMM
te-th	-t--t	--the	--e--	-e-th	e----	e--e-	---h-	-hheh	-----
YBJQD	LLOJV	LOJTV	YIOLU	JPPES	NGYQJ	MOYVD	GDNJE	MSVDN	EJM
--e--	tthe-	the--	--ht-	e----	----e	-h---	---e-	-----	-e-

```
VSGLL OSCIO LGOYG,
---tt h---h t-h--.
```

We guess the word “thought”,
Then S=o, C=u, I=g

```
LOJUM YLJME PDYVJ QXTDV SVJNL DMTJZ WMJGG YSNDL UYLEO SKDVC
the-- -te-- ----e ----- o-e-t ---e- --e-- -o--t --t-h o---u
-----------------------------------------------------------
GEPJS MDIPD NEJSK DNJTJ LSKDL OSVDV DNGYN VSGLL OSCIO LGOYG
---eo --g-- --eo- --e-e to--t ho--- ----- -o-tt hough t-h--
-----------------------------------------------------------
ESNEP CGYSN GUJMJ DGYNK DPPYX PJDGG SVDNT WMSWS GYLYS NGSKJ
-o--- u--o- --e-e ----- ----- -e--- o---- --o-o --t-o --o-e
-----------------------------------------------------------
CEPYQ GSGLD MLPYN IUSCP QOYGM JGCPL GDWWJ DMLSL OJCNY NYLYD
u---- -o-t- -t--- g-ou- -h--- e-u-t ----e --tot heu-- --t--
-----------------------------------------------------------
LJQLO DLCNL YPLOJ TPJDM NJQLO JWMSE JGGJG XTUOY EOOJO DQDMM
te-th -tu-t --the --e-- -e-th e--o- e--e- ---h- -hheh -----
-----------------------------------------------------------
YBJQD LLOJV LOJTV YIOLU JPPES NGYQJ MOYVD GDNJE MSVDN EJM
--e-- tthe- the-- -ght- e---o ----e -h--- ---e- -o--- -e-
```

Just two vowels left (a and i) so we guess Y=I, this gives, N=n, for the end of itio

LOJUM	YLJME	PDYVJ	QXTDV	SVJNL	DMTJZ	WMJGG	YSNDL	UYLEO	SKDVC
the--	ite--	--i-e	-----	o-ent	---e-	--e--	ion-t	-it-h	o---u
GEPJS	MDIPD	NEJSK	DNJTJ	LSKDL	OSVDV	DNGYN	VSGLL	OSCIO	LGOYG
---eo	--g--	n-eo-	-ne-e	to--t	ho---	-n-in	-o-tt	hough	t-hi-
ESNEP	CGYSN	GUJMJ	DGYNK	DPPYX	PJDGG	SVDNT	WMSWS	GYLYS	NGSKJ
-on--	u-ion	--e-e	--in-	---i-	-e---	o--n-	--o-o	-itio	n-o-e
CEPYQ	GSGLD	MLPYN	IUSCP	QOYGM	JGCPL	GDWWJ	DMLSL	OJCNY	NYLYD
u--i-	-o-t-	-t-in	g-ou-	-hi--	e-u-t	----e	--tot	heuni	niti-
LJQLO	DLCNL	YPLOJ	TPJDM	NJQLO	JWMSE	JGGJG	XTUOY	EOOJO	DQDMM
te-th	-tunt	i-the	--e--	ne-th	e--o-	e--e-	---hi	-hheh	-----
YBJQD	LLOJV	LOJTV	YIOLU	JPPES	NGYQJ	MOYVD	GDNJE	MSVDN	EJM
i-e--	tthe-	the--	ight-	e---o	n-i-e	-hi--	--ne-	-o--n	-e-

Remember that e, t, a, o, n, r, i, s, h are the most common letter in English.
As we see below, D and G have not been assigned small letters, so
we may guess G=s from the fragment position and D must be a not r because of the part th_tunt.

	J	L	D	G	Y	S	O	N	M	P	E	V	Q	C	T	W	U	K	I	X	Z	B	A	F	R	H
	e	t	-	-	i	o	h	n	-	-	-	-	-	u	-	-	-	-	g	-	-	-	-	-	-	-
Freq	32	28	27	24	23	22	19	18	17	15	12	12	8	8	7	6	6	5	4	3	1	1	0	0	0	0

We W=p completing the word position, Z=x, M=r due to the fragment “expression”, E=c, P=l due to “conclusion”. We keep doing this until we obtain completion.

LOJUM	YLJME	PDYVJ	QXTDV	SVJNL	DMTJZ	WMJGG	YSNDL	UYLEO	SKDVC
the--	ite--	-ai-e	---a-	o-ent	a--e-	--ess	ionat	-it-h	o-a-u
GEPJS	MDIPD	NEJSK	DNJTJ	LSKDL	OSVDV	DNGYN	VSGLL	OSCIO	LGOYG
s--eo	-ag-a	n-eo-	ane-e	to-at	ho-a-	ansin	-ostt	hough	tshis
ESNEP	CGYSN	GUJMJ	DGYNK	DPPYX	PJDGG	SVDNT	WMSWS	GYLYS	NGSKJ
-on--	usion	s-e-e	asin-	a--i-	-eass	o-an-	--o-o	sitio	nso-e
CEPYQ	GSGLD	MLPYN	IUSCP	QOYGM	JGCPL	GDWWJ	DMLSL	OJCNY	NYLYD
u--i-	sosta	-t-in	g-ou-	-his-	esu-t	sa--e	a-tot	heuni	nitia
LJQLO	DLCNL	YPLOJ	TPJDM	NJQLO	JWMSE	JGGJG	XTUOY	EOOJO	DQDMM
te-th	atunt	i-the	--ea-	ne-th	e--o-	esses	---hi	-hheh	a-a--
YBJQD	LLOJV	LOJTV	YIOLU	JPPES	NGYQJ	MOYVD	GDNJE	MSVDN	EJM
i-e-a	tthe-	the--	ight-	e---o	nsi-e	-hi-a	sane-	-o-an	-e-

LOJUM YLJME PDYVJ QXTDV SVJNL DMTJZ WMJGG YSNDL UYLEO SKDVC
thewr iterc laime d--am oment ar-ex press ionat witch o-amu

GEPJS MDIPD NEJSK DNJTJ LSKDL OSVDV DNGYN VSGLL OSCIO LGOYG
scleo ragla nceo- ane-e to-at homam ansin mostt hough tshis

ESNEP CGYSN GUJMJ DGYNK DPPYX PJDGG SVDNT WMSWS GYLYS NGSKJ
concl usion swere asin- alli- leass oman- propo sitio nso-e

CEPYQ GSGLD MLPYN IUSCP QOYGM JGCPL GDWWJ DMLSL OJCNY NYLYD
uclid sosta rtlin gwoul dhisr esult sappe artot heuni nitia

LJQLO DLCNL YPLOJ TPJDM NJQLO JWMSE JGGJG XTUOY EOOJO DQDMM
tedth atunt ilthe -lear nedth eproc esses --whi chheh adarr

YBJQD LLOJV LOJTV YIOLU JPPES NGYQJ MOYVD GDNJE MSVDN EJM
i-eda tthem the-m ightw ellco nside rhima sanec roman cer

The writer claimed by a momentary expression, a twitch of a muscle or a glance of an eye, to fathom a man's inmost thoughts. His conclusions were as infallible as so many propositions of Euclid. So startling would his results appear to the uninitiated that until they learned the processes by which he had arrived at them they might well consider him as a necromancer.[7]

Next class: Number Theory related with Cryptography!