CRYPTOGRAHY

Supplies: notebook, pencil.

Curriculum (1 hour, 30 minutes)

1. On the morning of Saturday, October 15, 1586, Mary Queen of Scots was on trial for treason. Mary conspired to remove Queen Elizabeth and she sent letters to other conspirators. She had been careful to write everything in cipher. Now, her life hung on the strength of her cipher. If her accusers could break the code, she would face death.

2. Why use secret writing?
   a. To get important message securely in time of war.
   b. To communicate between spies.
   c. To have secured business communication, for example between a bank and a customer.

3. Two branches of cryptography: transposition and substitution.

   Transposition: the letters of the message are simply rearranged, e.g.

   Cat, cta, act, atc, tac, tca – 6 ways

   Consider this short sentence: I love my cat. I how many ways can the letters be rearranged?

Answer: 3628800 ways

How about this: Transfer my funds to my secret Swiss account.
Answer: 259541276001130906838415360000000 ways

Impossible to unscramble by trying all rearrangements.

4. Counting the number of rearrangements.
   The letters in "Cat" can be rearranged in 6 different ways because we can write any of the three letters first, then any of the remaining letters next, resulting in 3x2 = 6 options.

   The number 3x2x1 = 3! (read as "three factorial")

   The word "cipher" consists of 6 different letters, which can be rearranged in 6! = 720 ways

   The word "kinawa" has also 6 letters so there is 720 rearrangements but if the letters "a" are switched nothing changes, so the number of different rearrangements is 720/2 =360.

5. Transposition ciphers.
   a. "Rail Fence" transposition

T	H	I	S	I	S	M	Y	S	E	C	R	E	T	F	O	R	Y	O	U
T		I		I		M		S		C		E		F		R		O	
	H		S		S		Y		E		R		T		O		Y		U
T	I	I	M	S	C	E	F	R	O	H	S	S	Y	E	R	T	O	Y	U

Activity: pair up, create a cipher text to be deciphered by your partner

b. Columnar transposition: the text to be encrypted is arranged in columns, usually the number of columns is a factor of the number of letters in the text. Then the columns are then reordered resulting in encrypted text. To decrypt the cipher, first decide on the number of columns and then rearrange them back. Easier said than done!

   The following text has 45 letters; what are the factors of 45?

   45 = 3X3X5, so we can use 3, 5, 9, or 15 columns.

   How about a text with 60 letters? (Why 60?)

   Plain text arranged in columns of five

T	H	I	S	I	S	M	Y	N	E	W	M	E	S	S
A	G	E	F	O	R	Y	O	U	C	O	M	E	T	O
T	H	E	O	L	D	H	O	U	S	E	S	O	O	N

Cipher text 1,2,3,4,5 TO 2,4,1,5,3

H	S	T	I	I	M	N	S	E	Y	M	S	W	S	E
G	F	A	O	E	Y	U	R	C	O	M	T	O	O	E
H	O	T	L	E	H	U	D	S	O	S	O	E	N	O

To decrypt 1,2,3,4,5 to 3,1,5,2,4

Activity: using software
http://www.richkni.co.uk/php/crypta/index.php

Activity: each group creates a cipher text to be decrypted by a partner group using software. Restriction: 45 letters arranged in 5 columns.

c. Caesar Shift – used by Julius Caesar.

Easy Caesar shifts: for example

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
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

Plain text: MEET YOU AT MIDNIGHT
Cipher text: rjjy dtz fy rnisnlmy

Activity: use http://www.richkni.co.uk/php/crypta/index.php

Problem: this is very easy to decipher!

Caesar shift with a key word, e.g. Julius Caesar (easy to remember by the sender and the receiver – a very important feature)

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

Activity: code “meet you at midnight” and check that

http://www.richkni.co.uk/php/crypta/index.php

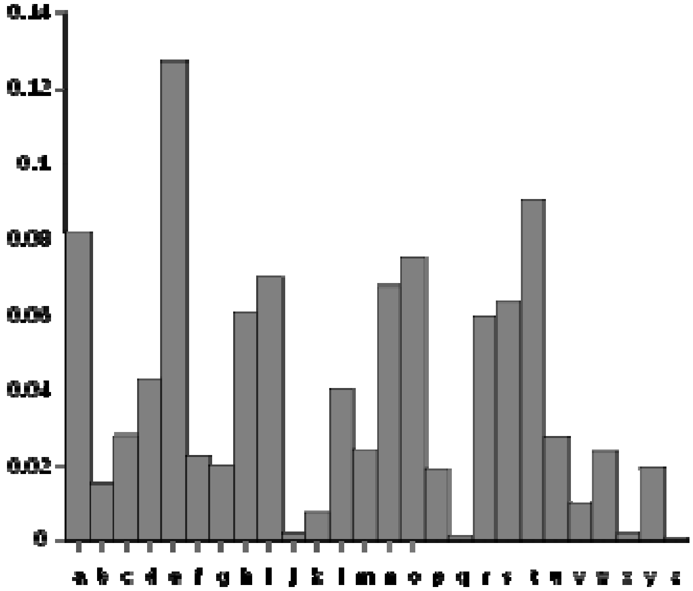
will fail to decrypt the cipher text.

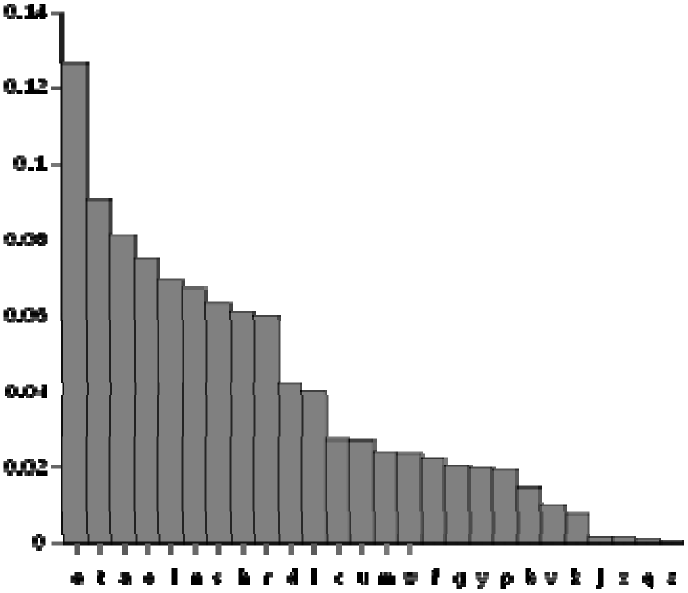
6. The Arab cryptanalysts.

In about 750 A.D. there was a golden age of Islamic civilization; an affluent society was established. Administration used encryption to protect privacy.

But also, the Arabs, in addition to employing ciphers, they destroyed them in their scholarly work.

7. Frequency analysis.




0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
e t a o i n s h r d l c u m w f g y p b v k j x q z

Letter	**Frequency**
A	8.167%
B	1.492%
C	2.782%
D	4.253%
E	12.702%
F	2.228%
G	2.015%
H	6.094%
I	6.966%
J	0.153%
K	0.772%
L	4.025%
M	2.406%
N	6.749%
O	7.507%
P	1.929%
Q	0.095%
R	5.987%
S	6.327%
T	9.056%
U	2.758%
V	0.978%
W	2.360%
X	0.150%
Y	1.974%
Z	0.074%

Relative frequencies of letters in other languages

Letter	French [5]	German [6]	Spanish [7]	Portuguese [8]	Esperanto [9]	Italian[10]	Turkish	Swedish[11]	Polish[12]
a	7.636%	6.51%	12.53%	14.63%	12.12%	11.74%	11.68%	9.3%	8.0%
b	0.901%	1.89%	1.42%	1.04%	0.98%	0.92%	2.95%	1.3%	1.3%
c	3.260%	3.06%	4.68%	3.88%	0.78%	4.5%	0.97%	1.3%	3.8%
d	3.669%	5.08%	5.86%	4.99%	3.04%	3.73%	4.87%	4.5%	3.0%
e	14.715%	17.40%	13.68%	12.57%	8.99%	11.79%	9.01%	9.9%	6.9%
f	1.066%	1.66%	0.69%	1.02%	1.03%	0.95%	0.44%	2.0%	0.1%
g	0.866%	3.01%	1.01%	1.30%	1.17%	1.64%	1.34%	3.3%	1.0%
h	0.737%	4.76%	0.70%	1.28%	0.38%	1.54%	1.14%	2.1%	1.0%
i	7.529%	7.55%	6.25%	6.18%	10.01%	11.28%	8.27%*	5.1%	7.0%
j	0.545%	0.27%	0.44%	0.40%	3.50%	0.00%	0.01%	0.7%	1.9%
k	0.049%	1.21%	0.01%	0.02%	4.16%	0.00%	4.71%	3.2%	2.7%
l	5.456%	3.44%	4.97%	2.78%	6.14%	6.51%	5.75%	5.2%	3.1%
m	2.968%	2.53%	3.15%	4.74%	2.99%	2.51%	3.74%	3.5%	2.4%
n	7.095%	9.78%	6.71%	5.05%	7.96%	6.88%	7.23%	8.8%	4.7%
o	5.378%	2.51%	8.68%	10.73%	8.78%	9.83%	2.45%	4.1%	7.1%
p	3.021%	0.79%	2.51%	2.52%	2.74%	3.05%	0.79%	1.7%	2.4%
q	1.362%	0.02%	0.88%	1.20%	0.00%	0.51%	0	0.007%	-
r	6.553%	7.00%	6.87%	6.53%	5.91%	6.37%	6.95%	8.3%	3.5%
s	7.948%	7.27%	7.98%	7.81%	6.09%	4.98%	2.95%	6.3%	3.8%
t	7.244%	6.15%	4.63%	4.74%	5.27%	5.62%	3.09%	8.7%	2.4%
u	6.311%	4.35%	3.93%	4.63%	3.18%	3.01%	3.43%	1.8%	1.8%
v	1.628%	0.67%	0.90%	1.67%	1.90%	2.10%	0.98%	2.4%	-
w	0.114%	1.89%	0.02%	0.01%	0.00%	0.00%	0	0.03%	3.6%
x	0.387%	0.03%	0.22%	0.21%	0.00%	0.00%	0	0.1%	-
y	0.308%	0.04%	0.90%	0.01%	0.00%	0.00%	3.37%	0.6%	3.2%
z	0.136%	1.13%	0.52%	0.47%	0.50%	0.49%	1.50%	0.02%	5.1%
à	0.486%	0	0	see **a**	0	see **a**	0	0.0%	0
å	0	0	0	0	0	0	0	1.6%	0
ä	0	-	0	0	0	0	0	2.1%	0
ą	0	-	0	0	0	0	0	0	see **a**
œ	0.018%	0	0	0	0	0	0	0	0

ç	0.085%	0	0	see **c**	0	0	1.26%	0	0
ĉ	0	0	0	0	0.66%	0	0	0	0
ć	0	-	0	0	0	0	0	0	see **c**
è	0.271%	0	0	0	0	see **e**	0	0.0%	0
é	1.904%	0	0	see **e**	0	see **e**	0	0.0%	0
ê	0.225%	0	0	see **e**	0	0	0	0	0
ë	0.001%	0	0	0	0	0	0	0	0
ę	0	-	0	0	0	0	0	0	see **e**
ĝ	0	0	0	0	0.69%	0	0	0	0
ğ	0	0	0	0	0	0	1.13%	0	0
ĥ	0	0	0	0	0.02%	0	0	0	0
î	0.045%	0	0	0	0	0	0	0	0
ì	0	0	0	0	0	see **i**	0	0	0
ï	0.005%	0	0	0	0	0	0	0	0
ı	0	0	0	0	0	0	5.20%*	0	0
ĵ	0	0	0	0	0.12%	0	0	0	0
ł	0	-	0	0	0	0	0	0	see **l**
ñ	0	0	0.31%	0	0	0	0	0	0
ń	0	-	0	0	0	0	0	0	see **n**
ò	0	0	0	0	0	see **o**	0	0	0
ö	0	-	0	0	0	0	0.87%	1.5%	0
ó	0	-	0	see **o**	0	0	0	0	see **o**
ŝ	0	0	0	0	0.38%	0	0	0	0
ş	0	0	0	0	0	0	1.94%	0	0
ś	0	-	0	0	0	0	0	0	see **s**
ß	0	0.31%	0	0	0	0	0	0	0
ù	0.058%	0	0	0	0	see **u**	0	0	0
ŭ	0	0	0	0	0.52%	0	0	0	0
ü	0	-	0	0	0	0	1.99%	0	0
ź	0	-	0	0	0	0	0	0	see **z**
ż	0	-	0	0	0	0	0	0	0.7%

To start deciphering the encryption it is useful to get a frequency count of all the letters. The most frequent letter may represent the most common letter in English E followed by T, A, O and I whereas the least frequent are Q, Z and X. Common percentages in standard English are:

a	b	c	d	e	f	g	h	i	j	k	l	m
8.2	1.5	2.8	4.3	12.7	2.2	2.0	6.1	7.0	0.2	0.8	4.0	2.4
n	o	p	q	r	s	t	u	v	w	x	y	z
6.7	7.5	1.9	0.1	6.0	6.3	9.1	2.8	1.0	2.4	0.2	2.0	0.1

and ranked in order:

e	t	a	o	i	n	s	h	r	d	l	u	c
12.7	9.1	8.2	7.5	7.0	6.7	6.3	6.1	6.0	4.3	4.0	2.8	2.8
m	w	f	y	g	p	b	v	k	x	j	q	z
2.4	2.4	2.2	2.0	2.0	1.9	1.5	1.0	0.8	0.2	0.2	0.1	0.1

Common pairs are consonants TH and vowels EA. Others are OF, TO, IN, IT, IS, BE, AS, AT, SO, WE, HE, BY, OR, ON, DO, IF, ME, MY, UP. Common pairs of repeated letters are SS, EE, TT, FF, LL, MM and OO. Common triplets of text are THE, EST, FOR, AND, HIS, ENT or THA.

If the results show that E followed by T are the most common letters then the ciphertext may be a transposition cipher rather than a substitution. If one of the characters has a 20% then the language may be German since it has a very high percentage of E. Italian has 3 letters with a frequency greater than 10% and 9 characters are less than 1%.

8. Cryptoanalysis of a ciphertext.

PCQ VMJYPD LBYK LYSO KBXBJXWXV BXV ZCJPO EYPD KBXBJYUXJ LBJOO KCPK. CP LBO LBCMKXPV XPV IYJKL PYDBL, QBOP KBO BXV OPVOV LBO LXRO CI SX'XJMI, KBO JCKO XPV EYKKOV LBO DJCMPV ZOICJO BYS, KXUYPD: "DJOXL EYPD, ICJ X LBCMKXPV XPV CPO PYDBLK Y BXNO ZOOP JOACMPLYPD LC UCM LBO IXZROK CI FXKL XDOK XPV LBO RODOPVK CI XPAYOPL EYPDK, SXU Y SXEO KC ZCRV XK LC AJXNO X IXNCMJ CI UCMJ SXGOKLU?

OFYRCDMO, LXROK IJCS LBO LBCMKXPV XPV CPO PYDBLK

With special symbols removed:

PCQ VMJYPD LBYK LYSO KBXBJXWXV BXV ZCJPO EYPD KBXBJYUXJ LBJOO KCPK CP LBO LBCMKXPV XPV IYJKL PYDBL QBOP KBO BXV OPVOV LBO LXRO CI SXXJMI KBO JCKO XPV EYKKOV LBO DJCMPV ZOICJO BYS KXUYPD DJOXL EYPD ICJ X LBCMKXPV XPV CPO PYDBLK Y BXNO ZOOP JOACMPLYPD LC UCM LBO IXZROK CI FXKL XDOK XPV LBO RODOPVK CI XPAYOPL EYPDK SXU Y SXEO KC ZCRV XK LC AJXNO X IXNCMJ CI UCMJ SXGOKLU OFYRCDMO, LXROK IJCS LBO LBCMKXPV XPV CPO PYDBLK

```
339 chars

a : 3 ... 0.9 %
b : 25 ... 7.4 %
c : 27 ... 8.0 %
d : 14 ... 4.1 %
e : 5 ... 1.5 %
f : 2 ... 0.6 %
g : 1 ... 0.3 %
h : 0 ... 0.0 %
i : 11 ... 3.2 %
j : 18 ... 5.3 %
k : 26 ... 7.7 %
l : 25 ... 7.4 %
m : 11 ... 3.2 %
n : 3 ... 0.9 %
o : 38 ... 11.2 %
p : 31 ... 9.1 %
q : 2 ... 0.6 %
r : 6 ... 1.8 %
s : 7 ... 2.1 %
t : 0 ... 0.0 %
u : 6 ... 1.8 %
v : 18 ... 5.3 %
w : 1 ... 0.3 %
x : 34 ... 10.0 %
y : 19 ... 5.6 %
z : 5 ... 1.5 %
```

Letter frequencies

```
o : 38
x : 34
p : 31
c : 27
k : 26
b : 25
l : 25
y : 19
j : 18
v : 18
d : 14
m : 11
i : 11
s : 7
r : 6
```

```
u : 6
e : 5
z : 5
n : 3
a : 3
q : 2
f : 2
g : 1
w : 1
t : 0
h : 0
```

2 letter sequences

```
pv => 11
lb => 11
bo => 9
xp => 9
cm => 8
op => 7
ok => 7
yp => 6
pd => 6
kx => 5
bx => 5
kc => 5
kl => 5
ro => 4
ci => 4
jc => 4
sx => 4
cp => 4
kb => 4
ey => 4
jo => 4
vx => 4
```

3 letter sequences

```
xpv => 8
lbo => 6
ypd => 6
kxp => 4
```

4 letter sequences

```
kxpv => 4
```

9. Steps in decrypting ciphertext:

   a. o, x, p are most frequent letters. Perhaps o is E, x is T, p is A?
   b. Perhaps, but then x alone would be T, not existent in English.
   c. So o = e,t or a, x = e,t or a, p = e,t or a is reasonable.
   d. Since the only letters in English that can be on its own are “a” “I”, and x is on its own, perhaps x =a.
   e. Ok, assume x =a, now what are the most frequent English three letter words?
   f. Compare with the analysis

xpv => 8
lbo => 6
ypd => 6
kxp => 4

x = a, so if xpv = and, p = n and v = d
o = e, so if lbo = the, l = t and b = h.

By the way, the only other stand alone letter is Y, so perhaps this is I?

   g. Now the ciphertext looks like this:

nCQ dMJinD thiK tiSe KhahJaWad had ZCJne EinD KhahJiUaJ thJee KCnK. Cn the thCMKand and IiJKt niDth, Qhen Khe had ended the taRe CI Sa’aJMI, KheJCKe and EiKKed the DJCMnd ZeICJe hiS, KaUinD: “DJeat EinD, ICJ a thCMKand and Cne niDhtK i haNe Zeen JeACMntinD tC UCM the IaZReK CI FaKt aDeK and the ReDendK CI anAient EinDK, SaU i SaEe KC ZCRd aK tC AJaNe a IaNCMJ CI UCMJ SaGeKtU?
eFiRCDMe, taReK IJCS the thCMKand and Cne niDhtK

h. Surely K = s; “thik” = this, what could be “thJee”?
What about “Cn”?

With a little more effort we can brake the cipher:

> Now during this time Shahrazad had borne King Shahriyar three sons. On the thousand and first night, when she had ended the tale of Ma’aruf, she rose and kissed the ground before him, saying: “Great King, for a thousand and one nights I have been recounting to you the fables of past ages and the legends of ancient kings, May I make so bold as to crave a favour of your majesty?
>
> Epilogue, Tales from the Thousand and One Nights

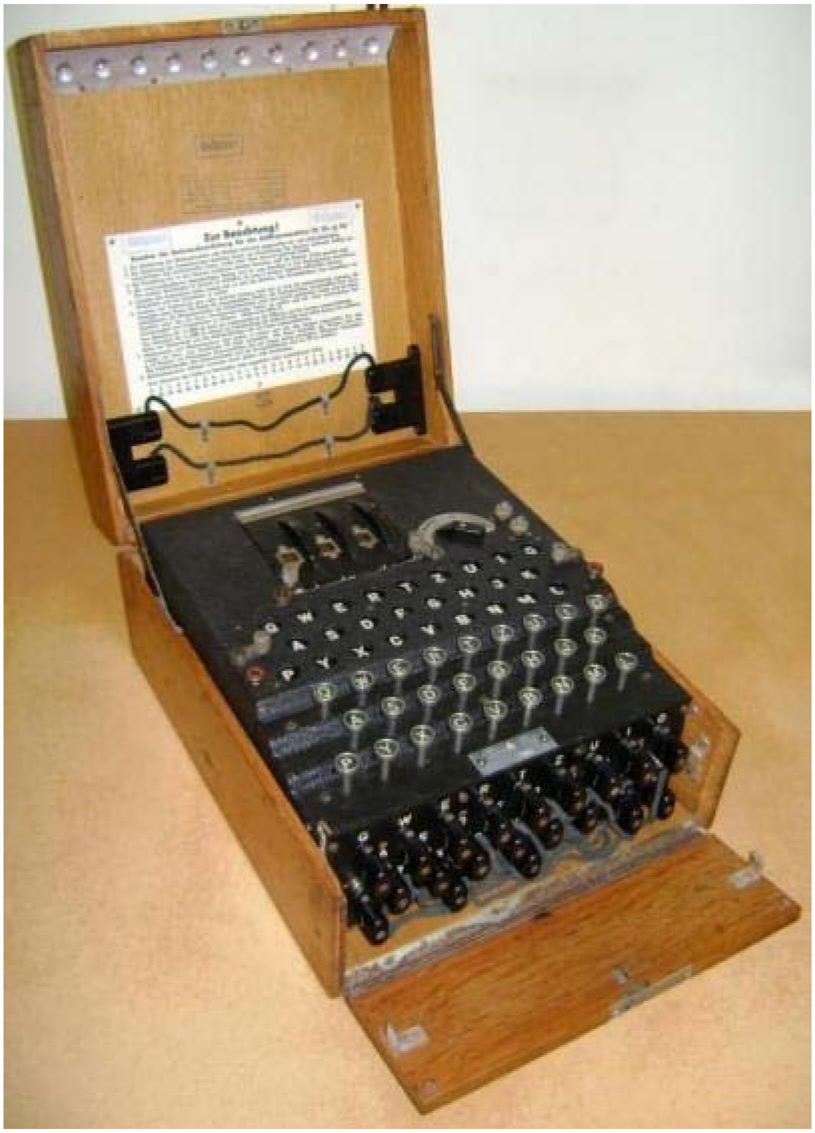
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
X	Z	A	V	O	I	D	B	Y	G	E	R	S	P	C	F	H	J	K	L	M	N	Q	T	U	W

We are dealing with Julius Caesar shift with the keyphrase inserted not at the beginning: A VOID BY GEORGES PEREC, which reduced to AVOIDBYGERSPC.

10. Now analyze this ciphertext:

BT JPX RMLX PCUV AMLX ICVJP IBTWXVR CI M LMT’R PMTN, MTN YVCJX CDXV MWMBTRJ JPX AMTNGXRJBAH UQCT JPX QGMRJXV CI JPX YMGG CI JPX HBTW’R QMGMAX; MTN JPX HBTW RMY JPX QMVJ CI JPX PMTN JPMJ YVCJX. JPXT JPX HBTW’R ACUTJXTMTAX YMR APMTWXN, MTN PBR JPCUWPJR JVCUFGXN PBL, RC JPMJ JPX SCBTJR CI PBR GCBTR YXVX GCCRXN, MTN PBR HTXXR RLCJX CTX MWMBTRJ MTCJPXV. JPX HBTW AVBXN MGCUN JC FVBTW BT JPX MRJVCGCWXVR, JPX APMGNXMTR, MTN JPX RCCJPRMEXVR. MTN JPX HBTW RQMHX, MTN RMBN JC JPX YBRX LXT CI FMFEGCT, YPCRCXDXV RPMGG VXMN JPBR YVBJBTW, MTN RPCY LX JPX BTJXVQVXJMJBCT JPXVXCI, RPMGG FX AGCJPXN YBJP RAMVGXJ, MTN PMDX M APMBT CI WCGN MFCUJ PBR TXAH, MTN RPMGG FX JPX JPBVN VUGXV BT JPX HBTWNCL. JPXT AMLX BT MGG JPX HBTW’R YBRX LXT; FUJ JPXE ACUGN TCJ VXMN JPX YVBJBTW, TCV LMHX HTCYT JC JPX HBTW JPX BTJXVQVXJMJBCT JPXVXCI. JPXT YMR HBTW FXGRPMOOMV WVXMJGE JVCUFGXN, MTN PBR ACUTJXTMTAX YMR APMTWXN BT PBL, MTN PBR GCVNR YXVX MRJCTBRPXN. TCY JPX KUXXT, FE VXMRCT CI JPX YCVNR CI JPX HBTW MTN PBR GCVNR, AMLX BTJC JPX FMTKUXJ PCURX; MTN JPX KUXXT RQMHX MTN RMBN, C HBTW, GBDX ICVXDXV; GXJ TCJ JPE JPCUWPJR JVCUFGX JPXX, TCV GXJ JPE ACUTJXTMTAX FX APMTWXN; JPXVX BR M LMT BT JPE HBTWNCL, BT YPCL BR JPX RQBVBJ CI JPX PCGE WCNR; MTN BT JPX NMER CI JPE IMJPXV GBWPJ MTN UTNXVRJMTNBTW MTN YBRNCL, GBHX JPX YBRNCL CI JPX WCNR, YMR ICUTN BT PBL; YPCL JPX HBTW TXFUAPMNTXOOMV JPE IMJPXV, JPX HBTW, B RME, JPE IMJPXV, LMNX LMRJXV CI JPX LMWBABMTR, MRJVCGCWXVR, APMGNXMTR, MTN RCCJPRMEXVR; ICVMRLUAP MR MT XZAXGGXTJ RQBVBJ, MTN HTCYGXNWX, MTN UTNXVRJMTNBTW, BTJXVQVXJBTW CI NVXMLR, MTN RPCYBTW CI PMVN RXTJXTAXR, MTN NBRRCGDBTW CI NCUFJR, YXVX ICUTN BT JPX RMLX NMTBXG, YPCL JPX HBTW TMLXN FXGJXRPMOOV; TCY GXJ NMTBXG FX AMGGXN, MTN PX YBGG RPCY JPX BTJXVQVXJMJBCT. JPX IBVRJ ACNXYCVN BR CJPXGGC

11. Enigma



Enigma was a German made coding machine used during WWII.

Breaking the code ▲

The general idea was that this military Enigma, unlike the commercial types, would be impossible to break. No one even tried to break it. However, in 1932, Poland's Biuro Szyfrow (Cipher Bureau) initiated attempts to analyze and break the Enigma messages. Although the chief of this Bureau received copies of codebooks sold by the German spy Hans-Thilo Schmidt, he did not give them to his codebreakers. He thought that keeping this information from them might stimulate their efforts. Marian Rejewski, Henryk Zygalski and Jerzy Rozicki were convinced that mathematics could solve the problem and succeeded in breaking the Enigma messages. They also developed an electro-mechanical machine, called the Bomba, to speed up the codebreaking process. Two major security flaws in the German Enigma procedures were the global groundsetting and the twice encodes message-key, a procedure to exclude errors. These flaws opened the door to cryptanalysis. In 1939 the Bureau was no longer able to break the codes due to increased sophistication in the design, new procedures and lack of funds for the code breakers. When Germany invaded Poland, the Polish Biuro Szyfrow passed its secret knowledge and several replica Enigma machines to the baffled French and British intelligence. The work of the Biuro Szyfrow was vital, not only because their pioneering work itself, but also because it convinced other cipher bureaus that it was possible to break Enigma.

Bletchley Park ▲

The Government Code and Cipher School (GC&CS) at Bletchley Park initially broke Enigma by hand. In August 1940 they started using their own Bombes, designed by Alan Turing and Gordon Welchman. It was also a rotary electro-mechanical device but it worked on an entirely different principle as Rejewski's Bomba. The Turing Bombe searched for the enigma settings for a given piece of plain and cipher text. When an Enigma message was intercepted, codebreakers had to search for cribs. These cribs were presumed pieces of plain text within the encrypted message. This could be "An Der Oberbefehlshaber", "An Gruppe", "Es Lebe Den Fuhrer" or any other standardized piece of text. Once a crib was located (there were some techniques for that) the associations between the letters of the ciphertext and their plain version were entered in the Bombe. The Bombe, which contains a large number of drums, each replicating the rotors of the Enigma, ran through all possible settings to find the key settings that belong to the given pieces of cipher and plain text. Once these settings were found all messages, encrypted with these setting, could be deciphered.

Enigma simulator:

http://russells.freeshell.org/enigma/