Cryptography

The *coding theory* we met yesterday dealt with issues of encoding messages, typically in text format, into a numeric or symbolic format. This was done with an eye on ensuring:

- that the coded message was uniquely decipherable;
- that the coded message wasn't overlong;
- that the code might contain measures to combat poor transmission.

There was no aspect of encoding the message to make it undecipherable to anyone who intercepted the code. This is the type of problem studied in *cryptography*.

Breaking Encryptions

Anybody trying to make unbreakable encryption has to face two realities:

- a piece of message, and its encrypted form, are likely to be intercepted;
- there will be security issues trying to transmit the encrypting and decrypting keys to any other party.

A good encryption will still be unbreakable, even in the event of such interceptions.

A simple, though not very good encryption method, is to shuffle the alphabet, encrypting each letter of the alphabet as another letter of the alphabet. Such an encryption is called a *cipher*.

Q: How many different ciphers are there of a,b,c,d,. . .,x,y,z?

Frequency Analysis

As there are 26!, which is roughly 4×10^{26}, ciphers of the alphabet, then we cannot simply try them all to decrypt a cipher. However, as different letters of the alphabet occur more or less often than others, we can perform some frequency analysis of a cipher-encrypted message to help our decryption.

The letters a,b,c,. . .,x,y,z appear, in an English text, with the following frequencies.

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

Problem

Decrypt the following cipher encryption – the original message was written in English.

Ljs vgnpgz tgis ojzshgz ysmpp. Hg vigt shqs gqah npqigs ynggwy jn mi msy czlms qy ms rgsy apcygz sc shg yji, qiw ypcty wcti qy ms xcdgy ojzshgz qtqu. Shmy aqi lg yggi, ocz geqxnpg, mi shg nmasjzg qlcdg, thmah yhcty yjaagyymdg ncymsmciy co shg npqigs qosgz gfjqp misgzdqpy co smxg. Qiw hg wmyacdgzgw q ymxnpg zjpg shqs wgyazmlgy nzgamygpu hct shmy ynggwmir qiw ypctmir wcti caajzt.

I have performed a frequency analysis on the 305 letter passage below:

Letter	a	b	c	d	e	f	g	h	i	j	k	l	m
Occurences	12	0	18	5	1	1	42	16	18	10	0	5	22
Frequency (%)	4.0	0.0	6.0	1.7	0.3	0.3	14.0	5.3	6.0	3.3	0.0	1.7	7.3
Letter	n	o	p	q	r	s	t	u	v	w	x	y	z
Occurences	11	6	14	19	3	27	10	2	2	10	4	31	16
Frequency (%)	3.7	2.0	4.7	6.3	1.0	9.0	3.7	0.7	0.7	3.3	1.3	10.0	5.3

Fill in your answer for the decrypted cipher in the table below:

L	j	s		v	g	n	p	g	z		t	g	i	s		o	j	z	s	h	g	z		y	s	m	p	p	.				
H	g		v	i	g	t		s	h	q	s		g	q	a	h		n	p	q	i	g	s		y	n	g	g	w	y				
j	n		m	i		m	s	y		c	z	l	m	s		q	y		m	s		r	g	s	y									
a	p	c	y	g	z		s	c		s	h	g		y	j	i	,		q	i	w		y	p	c	t	y		w	c	t	i	
q	y		m	s		x	c	d	g	y		o	j	z	s	h	g	z		q	t	q	u	.										
S	h	m	y		a	q	i		l	g		y	g	g	i	,		o	c	z		g	e	q	x	n	p	g	,				
m	i		s	h	g		n	m	a	s	j	z	g		q	l	c	d	g	,		t	h	m	a	h		y	h	c	t	y	
y	j	a	a	g	y	y	m	d	g		n	c	y	m	s	m	c	i	y		c	o		s	h	g		n	p	q	i	g	s
q	o	s	g	z		g	f	j	q	p		m	i	s	g	z	d	q	p	y		c	o		s	m	x	g	.				
Q	i	w		h	g		w	m	y	a	c	d	g	z	g	w		q		y	m	x	n	p	g		z	j	p	g			
s	h	q	s		w	g	y	a	z	m	l	g	y		n	z	g	a	m	y	g	p	u		h	c	t		s	h	m	y	
y	n	g	g	w	m	i	r		q	i	w		y	p	c	t	m	i	r		w	c	t	i		c	a	a	j	z	y	.	

Coded letter	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
Original letter																										

Answer

The key is:

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

and the original text came from *1089 and All That* p.45:

"But kepler went further still. He knew that each planet speeds up in its orbit as it gets closer to the sun, and slows down as it moves further away. This can be seen, for example, in the picture above, which shows successive positions of the planet after equal intervals of time. And he discovered a simple rule that describes precisely how this speeding up and slowing down occurs."

The RSA Public Key System

The RSA Public Key System was devised by Ronald Rivest, Adi Shamir, and Leonard Adleman, in 1978.

It is the most used method of encryption in the world today, being especially common on the internet. It is essentially unbreakable.

The mathematics behind the encrytion is relatively elementary, using results from classical number theory: Fermat's Little Theorem and the Chinese Remainder Theorem. What makes it practicable today is knowledge of large prime numbers.

The main principle behind RSA, is that it is relatively easy to multiply two prime numbers together, but, comparatively, almost impossible to factorise their product back to the two constituent primes. This is the idea of a one way function.

Public Key Systems

The idea behind a public-key system is that there is a public key, known to everybody, with which messages can be *encrypted*, and a private key known only to the person or company receiving the message, with which the coded message can be *decrypted*.

Given that the encrypting process is known and public, in principle it has to be the case that the decrypting process can be determined. For the system to be effectively unbreakable, determining the decrypting process has to be computationally unfeasible. At the same time, decrypting must be easy when in possession of the private key.

The type of set-up using a public key sytem might include:

- an internet company
- a customer (with his/her credit card details)
- the customer's computer
- a telephone line connecting the computer and company

Where could such a system be broken into?

Details of the RSA

In the RSA system:

- the public key includes two numbers n and e;
- the private key is n together with a different number d.

Given a (numerical) message M it is encrypted by

$$M \to M^e \pmod{n}.$$

Similarly an encrypted message C is decrypted by the message

$$C \to C^d \pmod{n}.$$

The number n is a product of two large prime numbers p and q. If you know p and q, you can work out d from e. As n is part of the public key, then in principle it is possible to factorize n to find p and q. The point is that this is nigh impossible, when each prime will, in practice, be a 1024-bit, i.e. around 300 decimal digits long.

Details of the RSA (continued)

1. Find two 'large' primes p and q.

   We will choose $p = 11$ and $q = 5$. (Obviously these aren't large at all!)

   We define

   $$n = pq = 55 \quad \text{and} \quad k = (p-1)(q-1) = 40.$$

2. Find a 'large' 'random' integer e which is coprime with k, (that is k and e have no common factors).

   As $k = 40 = 2^3 \times 5$ we can choose $e = 7$.

3. Take your message M and encode it as a whole number in the range $0 \leqslant M < n$. If your message is too long break the message into blocks in this range.

4. The message M is encrypted into the cryptogram C in the range $0 \leqslant C < n$ under the rule

   $$C = M^e \pmod{n}.$$

5. Compute the unique whole number d such that

$$de = 1 \pmod{k} \text{ and } d \text{ is in the range } 1 \leqslant d < k$$

Because e and k are coprime, we can make sense of $d = \frac{1}{e}$.

In our example, where $e = 7$ and $k = 40$, we can do this by trial and error, or spot that

$$7 \times 23 = 161 = 1 \pmod{k}.$$

So $d = 23$.

(For the small values involved here, calculating d by trial and error would not take long. For larger values d can still be calculated systematically using the Chinese Remainder Theorem and Euclidean algorithm.)

6. Given the encrypted message C, this can be decrypted back into the message M by taking

$$M = C^d \pmod{n}.$$

An Example

Take the previous values of

$$n = 55, \quad k = 40, \quad e = 7, \quad d = 23$$

and suppose that our message was assigned the value $M = 13$.

To encode it we need to calculate $C = M^e \pmod{n} = 13^7 \pmod{55}$.

As

$$13^2 = 169 = 3 \times 55 + 4 = 4 \pmod{55},$$

then

$$\begin{aligned} C &= 13^7 \\ &= 13^2 \times 13^2 \times 13^2 \times 13 \\ &= 4 \times 4 \times 4 \times 13 \\ &= 64 \times 13 \\ &= 9 \times 13 \\ &= 117 \\ &= 7 \pmod{55}. \end{aligned}$$

Decoding $C = 7$, we should obviously get $M = 13$.

In this example, M is given by $M = C^d \pmod{n} = 7^{23} \pmod{55}$.

We note

$$7^2 = 49 = -6 \pmod{55} \text{ and } 7^6 = (-6)^3 = -216 = 4 \pmod{55}.$$

So

$$\begin{aligned} D &= 7^{23} \\ &= 7^6 \times 7^6 \times 7^6 \times 7^2 \times 7^2 \times 7 \\ &= 4 \times 4 \times 4 \times (-6) \times (-6) \times 7 \\ &= 64 \times 6 \times 6 \times 7 \\ &= 9 \times 6 \times 6 \times 7 \\ &= 54 \times 42 \\ &= -1 \times 42 \\ &= -42 \\ &= 13 \pmod{55}. \end{aligned}$$

Problems

Set

$$p = 11, \quad q = 7, \quad e = 7.$$

1. What is n?
2. What is k?
3. Show that d is 43.
4. Encrypt $M = 3$.
5. Decrypt $C = 3$.

Answers

1. $n = pq = 11 \times 7 = 77$.
2. $k = (p-1)(q-1) = 10 \times 6 = 60$.
3. $43 \times e = 43 \times 7 = 301 = 1 \pmod{60}$ and so $d = 43$.
4. We wish to calculate $C = 3^7 \pmod{77}$. Note that $3^4 = 81 = 4 \pmod{77}$. Hence, we find $C = 31$ from
$$3^7 = 3^4 \times 3^3 = 4 \times 27 = 108 = 31 \pmod{77}.$$
5. We wish to calculate $M = 3^{43}$. Then, noting in $\bmod 77$ arithmetic
$$3^4 = 81 = 4 \text{ and } 2^6 = 64 = -13,$$
we have
$$\begin{aligned} 3^{43} &= (3^4)^{10} \times 3^3 = 4^{10} \times 27 = 2^{20} \times 27 \\ &= (2^6)^3 \times 2^2 \times 27 = -13 \times -13 \times -13 \times 4 \times 27 \\ &= 169 \times -13 \times 108 = 15 \times -13 \times 31 = -195 \times 31 \\ &= 36 \times 31 = 1116 = 346 \text{ taking away } 770 \\ &= 38 \text{ taking away } 308. \end{aligned}$$

How RSA works.

For the RSA system to work it should be the case the encrypting and decrypting of messages are *inverses* — that is each process will reverse the other. From Fermat's Little Theorem we know that

$$M^{(p-1)(q-1)} = 1^{q-1} = 1 \pmod{p} \text{ and } M^{(p-1)(q-1)} = 1^{p-1} = 1 \pmod{q}).$$

Because of another theorem (the Chinese Remainder Theorem) this means now that

$$M^k = M^{(p-1)(q-1)} = 1 \pmod{pq}.$$

Recall now that $de = 1 \pmod{k}$ which means that $de = 1 + ak$ for some whole number a. So if we take a message M, encrypt it to M^e and decrypt this to $(M^e)^d$ we'd find

$$(M^e)^d = M^{ed} = M^{1+ak} = M\left(M^k\right)^a = M \pmod{n},$$

showing that decryption reverses encryption and similarly, given an encryption C, then

$$\left(C^d\right)^e = C \pmod{n},$$

showing that encryption reverses decryption.

Further Questions

1. Show that there isn't an integer which leaves remainder 7 when divided by 12 and remainder 8 when divided by 15.
2. What is the remainder when 2^{1000} is divided by 11?
3. Without direct calculation, determine the next to last digit in 2^{1000}? (i.e. the one in the "tens" column.) [Hint: listing the last two digits of each power of 2 you should find a cycle of length 20 which first starts with the second power.]
4. Ask your teacher how you might use logarithms to calculate the first digit in 2^{1000}. You would need to know that

$$\log_{10} 2 = 0.30103\ldots \quad \text{and that} \quad 10^{0.03\ldots} = 1.0715\ldots$$