International Baccalaureate

Extended essay cover

SAMPLE A

Diploma Programme subject in which this extended essay is registered: MATHEMATICS

(For an extended essay in the area of languages, state the language and whether it is group 1 or group 2.)

Title of the extended essay: WHICH ARE THE MOST USEFUL MODERN CRYPTOGRAPHIC METHODS?

Candidate's declaration

If this declaration is not signed by the candidate the extended essay will not be assessed.

The extended essay I am submitting is my own work (apart from guidance allowed by the International Baccalaureate).

I have acknowledged each use of the words, graphics or ideas of another person, whether written, oral or visual.

I am aware that the word limit for all extended essays is 4000 words and that examiners are not required to read beyond this limit.

This is the final version of my extended essay.

Candidate's signature: _____ Date: 25/02/2009

IB Cardiff use only: A: [illegible] B: ✓

Supervisor's report

The supervisor must complete the report below and then give the final version of the extended essay, with this cover attached, to the Diploma Programme coordinator. The supervisor must sign this report; otherwise the extended essay will not be assessed and may be returned to the school.

Name of supervisor (CAPITAL letters) _____

Comments

Please comment, as appropriate, on the candidate's performance, the context in which the candidate undertook the research for the extended essay, any difficulties encountered and how these were overcome (see page 13 of the extended essay guide). The concluding interview (viva voce) may provide useful information. These comments can help the examiner award a level for criterion K (holistic judgment). Do not comment on any adverse personal circumstances that may have affected the candidate. If the amount of time spent with the candidate was zero, you must explain this, in particular how it was then possible to authenticate the essay as the candidate's own work. You may attach an additional sheet if there is insufficient space here.

This was a very well researched essay and was borne out of an interest cultivated when went to a lecture on the Enigma code in School. We spent a day in the University of West of England University Library and then he set about carrying out all of the calculations on his own. I talked this through with him explaining to me how he did it. He needed to research several topics beyond the Higher IB course and clearly understood them fully. The concluding interview certainly showed an excellent insight into some of the complex techniques dealt with in the essay.

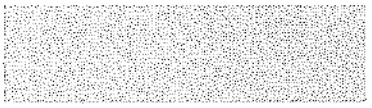
The essay was well written and easy to follow with good presentation throughout using correct and appropriate terminology. He planned the essay carefully and stated how he did this. All of his examples were original and not taken from any books or the internet. is a very good Higher student and has demonstrated this throughout the essay.

I have read the final version of the extended essay that will be submitted to the examiner.

To the best of my knowledge, the extended essay is the authentic work of the candidate.

I spent [8] hours with the candidate discussing the progress of the extended essay.

Supervisor's signature: _____ Date: 26.2.09



A very good essay showing excellent knowledge and understanding which would have benefited from a more focused (and less ambitious) research question.
- Its strengths are: well-worked out examples chosen by the candidate, personal contribution, good knowledge.
- Drawbacks are: - In describing LCA and ECA the candidate gradually loses the balance and the essay becomes a sequence of recipes / hard to follow if the general method is not presented.
- The evaluation and comparison is much weaker and does not follow clearly from the examples, partly this is due to the broad research question.

"Which are the most useful modern cryptographic methods?"

Extended Essay in Maths

IB Diploma 2007-09

Abstract

My essay looks at some of the many cryptographic methods currently available, and determines which of these is the most useful to most people. The essay begins by introducing cryptography, and its significance to modern society. The essay then starts to look at some methods, quickly increasing in complexity, going through three different ciphers where one letter/symbol represents another in the cipher. For these methods, I encrypted a message using the system, and then showed how it could be broken using cryptanalysis. I next looked at the Vigenère cipher and its vulnerability to cryptanalysis, and the perfect security provided by one-time pad. I then moved to the main point of the essay, which is the choice between RSA and elliptic curve calculus. I encoded and decoded a PIN number using each, and explained what makes them so effective as methods of cryptography. Finally in my evaluation, I looked at some statistics for RSA and ECC, and concluded that for a couple of reasons, ECC was the most useful cryptographic method. I suggested that this was likely to change before too long due to advances in technology and mathematics.
Many of the methods I have looked at have been explained by numerous textbooks, and I used some of these to help me to understand how and why they worked. I have then used some of these methods to present my essay, but using my own numbers and doing all of the calculations myself.

Table of Contents

Extended Essay: Topic Area of Mathematics

Introduction

My research question is:

Which are the most useful modern cryptographic methods?

In answering this question, I will investigate different methods of cryptography that might be used in the present day – these will range from a simple shift cipher to the RSA algorithm and ECC. I will encrypt a message using each method, and then attempt to decrypt it using both the correct way of decryption, and also as a third party wanting to find out what the information is without the knowledge of the sender.

By the analysis of the ease of breaking the encryption, I will be able to judge which are the most secure and least vulnerable to outside decryption, but this is not the only measure of how useful the cryptographic method is. Other factors will include the difficulty of use, the extent to which they have a problem with key distribution, their practicality and, for some methods, the bit size required to gain satisfactory security.

The messages encrypted by each method will differ in length and complexity. This will be explained throughout, but briefly this is because some need a certain length of message in order to be effective, and others can only be done simply using small messages.

The issue of cryptographic method is a key choice for millions of internet users. Many important things are done over the internet, particularly given the surge in internet banking, and users want to be sure that their details are secure from any "eavesdroppers". They need this to be simple enough to implement (at least for a computer) but secure enough to hold against cryptanalysis. I am seeking the method best suited to fit this purpose.

Definitions

Cryptography – The process or skill of communicating in or deciphering secret writings or ciphers.[1]

Plaintext – the intelligible original message of a cryptogram.[2]

Ciphertext – the encoded version of a message or other text; cryptogram.

Cryptanalysis – The analysis and deciphering of cryptographic writings or systems

[1] http://dictionary.reference.com/browse/cryptography, http://dictionary.reference.com/browse/cryptanalysis

[2] http://dictionary.reference.com/browse/plaintext, http://dictionary.reference.com/browse/ciphertext

Simple Ciphers

Shift Cipher

I will begin with a simple shift cipher. This is the easiest way of encrypting a message, which entails only the transposition of the whole alphabet. Since there are only 26 positions that this can take (25 not including the initial position) it is also the easiest to break, as it is very simple to move the alphabet along until the encoded passage makes perfect sense.

Monoalphabetic Cipher

Monoalphabetic ciphers involve the random pairing of letters in the ciphertext alphabet to the original plaintext alphabet – known as monoalphabetic substitution[3]. This is more complex than the shift cipher as the arrangement of the alphabet is no longer the same, so the pattern is gone. There are a total of 403,291,461,126,605,635,584,000,000 ($26! \approx 4.033 \times 10^{27}$) permutations of the alphabet, and trying all of these permutations is completely unfeasible, since at a rate of one rearrangement per second this would take around a billion times the lifetime of the universe to try them all! However this is fairly easy to break using the letter frequencies of the language the plaintext is written in.

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

Here is an example of an arrangement that can be created using this substitution method. I have quickly mixed the alphabet at random, and will now encrypt and decrypt a paragraph using letter frequencies.

DSNS CU QG SOQTYVS BX QG QNNQGESTSGJ JDQJ ZQG AS ZNSQJSW MUCGE JDCU UMAUJCJMJCBG TSJDBW. C DQIS HMCZFVL TCOSW JDS QVYDQASJ QJ NQGWBT, QGW KCVV GBK SGZNLYJ QGW WSZNLYJ Q YQNQENQYD MUCGE VSJJSN XNSHMSGZCSU.

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
Normal Frequency[4]	8	2	3	4	13	2	2	6	7	0	1	4	2	7	8	2	0	6	6	9	3	1	2	0	2	0
Ciphertext Frequency	2	3	6	5	2	1	8	1	1	9	1	2	4	6	1	0	11	0	12	3	4	4	4	1	4	4

The high frequency of the letter S suggests that this is the letter E, and this makes sense in the text, so this will be replaced. The letters J and Q are likely to be A and T, and by looking through the text it becomes clear that J is T and Q is A, since the word AT (QJ) is a real word but TA would not make sense.

[3] http://www.bbc.co.uk/dna/h2g2/alabaster/A583878
[4] http://www.simonsingh.net/The_Black_Chamber/frequencyanalysis.html

DENE CU AG EOATYVE BX AG ANNAGEETEGT TDAT ZAG AE ZNEATEW MUCGE TDCU UMAUTCTMTCBG TETDBW. C DAIE HMCZFVL TCOEW TDS AVYDAAET AT NAGWBT, AGW KCVV GBK EGZNLYT AGW WEZNLYT A YANAENAYD MUCGE VETTEN XNEHMEGZCEU.

Other letters that can now be easily deduced are D in TDAT (H), N in HENE (R), C on its own (I), I in HAIE (V), G and W in AGW (N and D), X in DIXXERENT (F), and so on, until the whole alphabet is known. This has shown that the characteristics of the ciphertext betray the plaintext, and allow simple decoding.

Anecdotal evidence? Discussion?

Homophonic Substitution

Homophonic substitution is a method deliberately intended to make it more difficult to frequency analyse the ciphertexts. Each letter can be replaced by one of a number of symbols, the number dependent upon how common the letter is in the alphabet. For example, the letter D has 4% frequency in the English language, so this would have 4% of the symbols to choose from when enciphering.

While this method makes frequency analysis more difficult, it is still not impossible, since the letter D still has only so many symbols, and none of these symbols can also mean another letter. Therefore while securer, the method still has the same flaw as before.

too superficial and vague to be of interest in the context of this essay

Vigenère

The Vigenère cipher is far more complicated and harder to break, since the frequency of letters is very unlikely to have a graph similar to that of the language of the plaintext. It involves the use of a keyword, which is repeated along the length of the message. A Vigenère square can be used, in which the rows and columns of the square are used to locate the letter needed for the ciphertext; however it can be done without the use of a table too. If each letter is given its number in the alphabet (e.g. B = 2, T = 20), and the keyword is also numerated in this way, the sum of these two numbers modulo 26 gives the number that the letter in the ciphertext will take.

I will encrypt the following message: **The Vigenère cipher is strong.**
Using the keyword: **Maths.**

Plaintext	T	H	E	V	I	G	E	N	E	R	E	C	I	P	H	E	R	I	S	S	T	R	O	N	G
Keyword	M	A	T	H	S	M	A	T	H	S	M	A	T	H	S	M	A	T	H	S	M	A	T	H	S

So, to encrypt the first letter of this sentence, I need to give each letter its number.

$$T = 20, M = 13.$$
$$20 + 13 \pmod{26} = 7 = G.$$

The second letter is encrypted the same way, as follows:

$$H = 8, A = 1$$
$$8 + 1 \pmod{26} = 9 = I.$$

By continuing this method through the rest of the text, the complete ciphertext is:

GIY DBTFHMKR DCXARS CA LGSIVZ

As you can see, the same letter in the plaintext can come up as different letters in the ciphertext. The letter E, for example, appears 5 times in the plaintext, and is encrypted to Y, F, M, R and R. This makes it impossible to use frequency analysis, since the most common letter in English is spread across different letters in the ciphertext.

However, the letter E was encrypted twice to R, so there is a weakness to the Vigenère cipher. It is hard to cryptanalyse this short text, but with more text, it is possible to decode.

GIY QGGFLVTGJIVTY CUKVNMUCKRBNM IEPAZTZNY QL N DBIEYFHOBAH
NEH LFUZ VHSLQVHMOU IEJGIKVMS IBZFX IM FUOLXAUM IZRE MQQGFYV
MB OCVXGFYV BG MYIWF UI I JHBFQYVDUBBBO NPTG JM EBQFFG
KRDIOGVAYL UL UBM PBSFLL YFULBAH OVBIFLABGJYA

At first glance, this looks completely unbreakable. Using the letter frequencies, like we would have done for the previous ciphers, no information can be gained about the plaintext, as seen below. There are no clear peaks, and no amount of letter replacing would help anyway, since the same plaintext letter codes to different letters in the ciphertext.

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
Normal Frequency	8	2	3	4	13	2	2	6	7	0	1	4	2	7	8	2	0	6	6	9	3	1	2	0	2	0
Ciphertext Frequency	4	8	2	2	3	7	6	4	7	3	2	6	6	3	4	2	4	2	2	2	6	6	1	2	6	3

There is, however, a method of cracking this code. The first and most important step is to find the length of the keyword, which is done by the *Kasiski* test.[5] This involves analysing the ciphertext for any repeated bigrams and trigrams (sequences of letters present in many words e.g. -ing, -ed), and then examining the distance between them. This works because English has a large repetition of bigrams and trigrams, and longer pieces of text are likely to match these up every so often. I will now analyse my ciphertext to find some of these repetitions (marked in red).

Repeated Sequence	
TGJ	130 letters apart
BAH	125 letters apart
GFY	10 letters apart
FYV	10 letters apart

Now, to find the length of the keyword, I need to find a factor common to all of these distances. Some of these distances may have occurred by chance, while others will reveal information about the keyword.

$$\gcd(130, 125) = 5$$
$$\gcd(130, 10) = 10$$
$$\gcd(125, 10) = 5$$

My best guess for the keyword is 5. Of course, I can go back if this seems wrong and change it to 10, but it is more usual to have a keyword of between 4-8 letters in length. Longer keywords have more security, but it is much easier to use a smaller word. It is also normally a real word for ease of remembrance. These two things help to simplify the cryptanalysis.

Now that I (probably) know the length of the keyword, the rest of the cryptanalysis is as before. I will take letters at the interval prescribed by the keyword, and then compare the letter frequencies of the ciphertext with those of normal English. The displacement of each set of frequencies is likely to indicate which letter it is in the keyword.

For each method I have so far considered, knowledge of the plaintext language is key to decipherment. If the coding is split into words, it can be useful to look for common one- or two- letter words; if not, then common bigrams, trigrams and double letters can be used to make letter identification quicker and easier. A table of these for some major European languages is included in the appendices.

One time pad

One time pad involves the use of a keyword of the same length as the message. Often this will be just a jumble of letters. This method of cryptography is perfectly secure, since there is no way of gaining the plaintext from the ciphertext by textual analysis, and they bear no relationship to one another. A determined cryptanalyst could try every possible rearrangement of letters to reach the plaintext, but this could take years if the message was not short. The only problem with this method is that both parties need copies of the keyword, and so implementing this method securely is extremely difficult, due to the necessity of transferring keys between parties.

The previous 7 pages could have been omitted since they have no significant impact on the subject of this EE (which is already too ambitious!)

[5] p57, Cryptography: An Introduction by Nigel Smart

RSA

In 1977, a brand new technique of cryptography was invented, and this revolutionised encryption. RSA was the first public-key encryption technique, and was so strong and useful that it is still the most widely used technique today. Public key encryption is one where anyone can encode a message by using the public key, but only the intended recipient should be able to open it, using their private key.

Since the complexity of this method is so much more complex than the previous methods I have looked at, encrypting a message is not really feasible. A computer would probably break a message into manageable chunks, and calculate them in an instant – doing it by hand means that I would have to do very small groups. I have instead decided to encrypt a PIN number, like on a credit card – this is a realistic place where encryption is widely used, and will show the encryption method well. A PIN number is 4 digits – I will encode two separate groups of two. I have chosen the number 1316 to encode. I will also use much smaller numbers than a computer would use; this is to make the maths more manageable and to avoid over complicating things.

First I need two prime numbers, p and q. I will choose 19 and 37. The product of these two numbers, 703, is the public key, which I can tell people sending me messages.

I must also choose another number e, which must be relatively prime to (p-1) (q-1). (p-1)(q-1) = 648, so e can be 5. This is also part of the public key, so I tell people that too.

I now have enough information to encrypt a message. The formula for encryption[6] is:

$$C = M^e \pmod{pq},$$

so for my first message, 13^5 (mod 703) will give me my encrypted message. $13^5 = 371293$, and 371293 (mod 703) = 109, so the encrypted message is 109. Encryption of the second part gives 1048576 (mod 703) which is 403, so the pin number 1316 has encrypted to 109 403, which bears no relation to the plaintext.

how does it work? why? what is the mathematics behind this?

To decode the message, I need another number, d. This should be such that $de = 1 \pmod{(p-1)(q-1)}$, or for this system, $5d = 1 \pmod{648}$. The lowest number for which this is the case is 389, since $5 \times 389 = 3 \times 648 + 1$.

The equation for decryption[6] is:

$$M = C^d \pmod{pq},$$

So for the first message, 109^{389} (mod 703) should give me my starting number. This is much easier said than done! The first part of the calculation gives the number:

no explanation here of why it works so this merely a recipe with very limited mathematical content)

[6] http://mathcircle.berkeley.edu/BMC3/rsa/node4.html

3 62166018 68314968 68728587 96325447 33745850 10575665 78369239 38910910
72799895 97392806 02686796 77944065 04893439 09579687 95606943 02138730
66867000 55937777 30524493 94047930 02226282 88766615 18565645 19845968
71384847 25469239 32213804 58752833 22322469 82732896 00226161 90875771
49387257 44940391 68079617 18194021 80914735 95903928 81340739 35855358
05267076 50247516 11445852 23148703 66342841 77666088 17500267 67224117
61020965 41220767 00920219 58528068 13971539 19404741 54908597 81378036
10075172 37086240 02620374 82580294 62090199 17992314 06521935 73367710
58811812 81100923 35807851 46504318 47756149 73329116 50433217 51977619
71120504 06468038 21368485 21607678 09971535 64602940 97186785 85589529
22519943 64574187 91632685 74169436 19196240 32684097 99021268 00767850
77916218 44379614 42364327 58985528 59022524 74869854 01224129 79621560
77677824 54168399 46928189

The reader is lost in the details...

Unfortunately, this number alone is too large for me to find the modulo of, so I will use a trick of the modulo function to find the result. I can calculate all the partial results in that modulo, and by repeated squaring of 109, I can get all the exponents that are powers of 2.

$109^{1} \pmod{703} = 109$
$109^{2} \pmod{703} = 633$
$109^{4} \pmod{703} = 682$
$109^{8} \pmod{703} = 441$
$109^{16} \pmod{703} = 453$
$109^{32} \pmod{703} = 636$
$109^{64} \pmod{703} = 271$
$109^{128} \pmod{703} = 329$

If I now take powers that add up to 389 (128 x 3, 4 and 1) and multiply the results from above together, I will then be able to find the modulus of that to get the original number back.

$329^{3} \times 682 \times 109 = 2647272001682.\ 2647272001682 \pmod{703} = 13.$

For the second part, I will use exactly the same method.

$$M = 403^{389} \pmod{703}$$

$403^{1} \pmod{703} = 403$
$403^{2} \pmod{703} = 16$
$403^{4} \pmod{703} = 256$
$403^{8} \pmod{703} = 157$
$403^{16} \pmod{703} = 44$
$403^{32} \pmod{703} = 530$
$403^{64} \pmod{703} = 403$

These are obviously the decoded message.

$403^{6} \times 256 \times 403 = 441952187970181987072 \pmod{703} = 16.$

The strength of RSA lies in the difficulty of factoring a number, because this is a far harder problem than it looks. For example, if I am given a number such as 82817953, there are only so many ways that I can factorise it, none of which are easy.

Of course, the simplest way of doing this is to check every prime number to see which are factors. For this number, this is a fairly simple problem, and a computer can factor it in a short amount of time, but to do it by hand is far more painstaking.

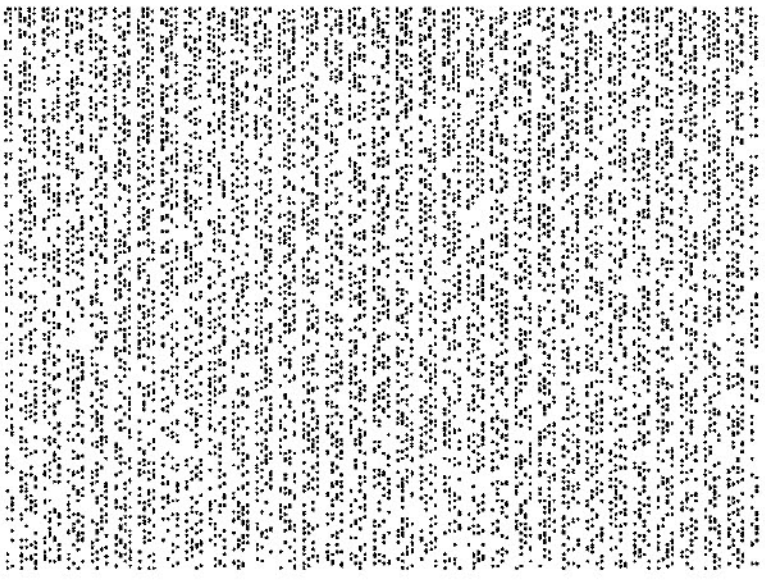
To factorise this, I would work through all the prime numbers up to √82817953 until I find a factor – the first of these would be 8887. Once you have one factor, the problem becomes easier; it turns out that this number is a semiprime (the product of two prime numbers, so that it has only four factors including 1 and itself), the hardest of all to factorise (the other factor of this number is 9319).

There are several algorithms designed to speed up the process, and a computer can check thousands of prime numbers per second. However, there are infinitely many prime numbers (the largest known, as of September 2008, is $2^{43,112,609} - 1$ (this number is 12,978,189 digits long))[7], and since the numbers used in RSA are so large, this is computationally infeasible. The current record for integer factorisation is held by team at the German Federal Agency for Information Technology Security (BSI), who factorised the 193 digit number[8]

310 7418240490 0437213507 5003588856 7930037346 0228427275 4572016194 8823206440 5180815045 5634682967 1723286782 4379162728 3803341547 1073108501 9195485290 0733772482 2783525742 3864540146 9173660247 7652346609.

The factorization of this number was accomplished using a prime factorization algorithm known as the general number field sieve. The process took several months using 80 computers. The two 97-digit factors found using this sieve are

1634733 6458092538 4844313388 3865090859 8417836700 3309231218 1110852389 3331001045 0815121211 8167511579
x
1900871 2816648221 1312685157 3935413975 4718967899 6851549366 6638539088 0271038021 0449895719 1261465571



The distribution of all the prime numbers in the range of 1 to 76,800[9], from left to right and top to bottom, where each pixel represents a number. Black pixels mean that number is prime and white means it is not prime.

[7] http://www.mersenne.org/
[8] http://mathworld.wolfram.com/news/2005-11-08/rsa-640/
[9] http://upload.wikimedia.org/wikipedia/en/d/d5/PrimeNumbersSmall.png

Elliptic Curve Cryptography (ECC)

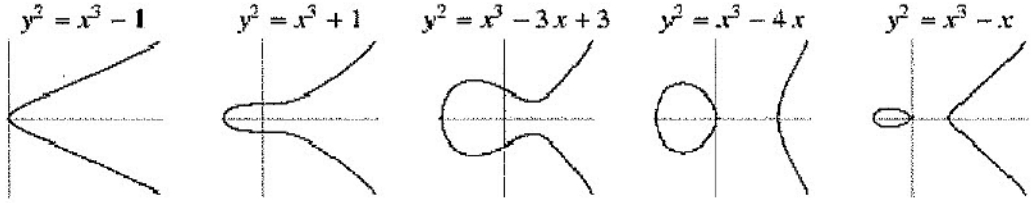
ECC is another public key cryptography technique. It relies on the use of elliptic curves.

An elliptic curve is a curve with the formula:

$$y^2 = x^3 + ax + b$$

where a and b have no repeated factors.

Examples of elliptic curves are shown below[10]:



An elliptic curve is suitable for use for cryptography only if it meets the following requirement:

- $4a^3 + 27b^2 \neq 0$[11]

The curve is also used in conjunction with the modulo function, so that the formula now reads:

$$y^2 = x^3 + ax + b \pmod p \text{ where } p > 3$$

So, I will now find an elliptic curve, and then show how it can be used to encrypt a message.

$$y^2 = x^3 + 2x + 5 \pmod 7$$

If this curve is to be suitable for my purposes, I will have to check it fits the requirement:

$$4 \times 2^3 + 27 \times 5^2 \pmod 7 = 32 + 675 \pmod 7 = 707 \pmod 7 = 0$$

This curve is unsuitable.

I will try a different curve instead, since the last one did not work.

$$y^2 = x^3 + 2x + 3 \pmod 7$$

$$4 \times 2^3 + 27 \times 3^2 \pmod 7 = 32 + 243 \pmod 7 = 275 \pmod 7 = 2.$$

This shall be my elliptic curve.

The next stage is find the points on the curve $y^2 = x^3 + 2x + 3$. I will do this using a table.

The quadratic residues (mod 7) are listed as follows:

$0^2 \pmod 7 = 0$	$1^2 \pmod 7 = 1$	$2^2 \pmod 7 = 4$	$3^2 \pmod 7 = 2$
$4^2 \pmod 7 = 2$	$5^2 \pmod 7 = 4$	$6^2 \pmod 7 = 1$	

Numbers which share this residue have integer points on the curve. So when x = 2, $x^3 + 2x + 3 = 1 \pmod 7$. This is the same as when $y^2 = 1$ or 36, so when x is 2, y = 1 and/or 6.

[10] http://mathworld.wolfram.com/EllipticCurve.html
[11] Basic Methods of Cryptography, p152

x	$z = x^3 + 2x + 3 \pmod 7$	Quadratic Residue	y	(x, y)
0	3	no		
1	6	no		
2	1	yes	1, 6	(2, 1), (2, 6)
3	1	yes	1, 6	(3, 1), (3, 6)
4	5	no		
5	5	no		
6	0	yes	0	(6, 0)

I have found nine points on the curve. The points can be generated in order, choosing an arbitrary point to begin with. This point will be called point P (x_1, y_1), and point Q is another point on the elliptic curve (x_2, y_2). I need to use some formulae to find these points:

$$x_3 = \sigma^2 - x_1 - x_2 \pmod p$$

$$y_3 = \sigma (x_1 - x_3) - y_1 \pmod p$$

$$\sigma = \begin{cases} (y_2 - y_1)/(x_2 - x_1), & \text{if } P \neq Q \\ (3x_1^2 + a)/(2y_1), & \text{if } P = Q \end{cases}$$

I will begin with the point, (2, 1). I will now find 2P, another point on the same curve.

$$2P = (2, 1) + (2, 1)$$

I will use the second formula for calculating σ, since P = Q in this case:

$$\sigma = (3 \times 2^2 + 2)/(2 \times 1) = 14/2 = 0 \pmod 7$$

$$x_3 = 0^2 - 2 - 2 = 3 \pmod 7$$
$$y_3 = 0\,(2 - 3) - 1 = 6 \pmod 7$$

I have found the point (3, 6) which I know is another point on the elliptic curve. Next I will calculate 3P, by adding P and 2P together.

$$3P = (2, 1) + (3, 6)$$

$$\sigma = (6 - 1)/(3 - 2) = 5/1 = 5 \pmod 7$$

$$x_3 = 5^2 - 2 - 3 = 6 \pmod 7$$
$$y_3 = 5\,(2 - 6) - 1 = -21 = 0 \pmod 7$$

I can then go on to find 4P and 5P, which result in (3, 1) and (2, 6) respectively. Further values of αP can provide nonsense results like dividing by ∞. These points are a rearrangement of the points found in the table above.

These results form the basis of ECC. The difficulty is in finding the value of α given αP. In my example this will be easy to calculate, but with large numbers it is totally impractical. There are two common systems of encoding by ECC; I will show them both in turn.

The El-Gamal system

Let F be an elliptic curve, and P be a point of curve F. Also let $Q = \alpha P$, where α is a secret key. P and Q are the public key. M is the message to be encrypted, in the form $M = (u_1, u_2)$. M must be a point on curve F. k is a random number. The encipherment is defined as:

$$C = e(M, k) = (v_1, v_2)$$

where: $v_1 = kP$ and $v_2 = M + kQ$

I will choose k = 2 as my random number. P will be (2, 1) and $\alpha = 4$. Q is hence 4P, which is (as calculated above) (3, 1).

$v_1 = kP = 2 \times (2, 1) =$ (as shown earlier) (3, 6)

$$kQ = 8P = 2\,(3, 1)$$

$$\sigma = (3 \times 3^2 + 3)/2 = 30/2 = 1 \pmod 7$$

$$x_3 = 1^2 - 3 - 3 = 2 \pmod 7$$
$$y_3 = 1\,(3 - 2) - 1 = 1 - 1 = 0 \pmod 7$$

I will choose my message to be (3, 6).

$$v_2 = (2, 0) + (3, 6)$$

$$\sigma = (6 - 0)/(3 - 2) = 6/1 = 6 \pmod 7$$

$$x_3 = 6^2 - 2 - 3 = 3 \pmod 7$$
$$y_3 = 6\,(2 - 3) - 0 = -6 = 1 \pmod 7$$

So $v_1 = (3, 6)$ and $v_2 = (3, 1)$.

This is the encrypted "message." The formula for decryption is:

$$d\,(C, \alpha) = v_2 - \alpha\, v_1.$$

So in my case, this reads:

$$d\,(C, 4) = (3, 1) - 4\,(3, 6)$$

But this is easy to solve, since 4 (3, 6) = 2 (3, 6) + 2 (3, 6), and I already know that 2 (3, 6) = (3, 1), since (3, 6) is 2P. But for kQ I had to double (3, 1), and I know that this equals (3, 6), so:

$$d\,(C, 4) = (3, 1) - (3, 6) = (3, 6),$$ which was my message.

(The subtraction was done by saying 4P – 2P = 2P.)

The main disadvantage of this system is that the message must be a point on the curve, which severely limits the amount of messages you are able to send.

The Menezes-Vanstone system

Let F be an elliptic curve, and P be a point of curve F. Also let $Q = \alpha P$, where α is a secret key. P and Q are the public key. M is the message to be encrypted, in the form $M = (u_1, u_2)$. M must be a point on curve F. k is a random number. The encipherment is defined as:

$$C = e(M, k) = (y_0, y_1, y_2)$$

where: $y_0 = kP$, $(c_1, c_2) = kQ$, $y_1 = c_1u_1 \pmod p$ and $y_2 = c_2u_2 \pmod 7$

I will again choose $k = 2$ as my random number. P will be (2, 1) and $\alpha = 4$. Q is (3, 1). However, I will this time choose co-ordinates not on the curve. Therefore M will be (2, 2).

$$y_0 = kP = (3, 6)$$
$$(c_1, c_2) = kQ = 2\,(3, 1) = (3, 6)$$
$$y_1 = c_1u_1 = 2 \times 3 \pmod 7 = 6$$
$$y_2 = c_2u_2 = 6 \times 2 \pmod 7 = 5$$

$$C = ((3, 6), 6, 5)$$

Deciphering is done by using the formula:

$$d(C, \alpha) = (y_1c_1^{-1} \pmod p, y_2c_2^{-1} \pmod 7)$$

$$= (6/3 \pmod 7, 5/6 \pmod 7) = (2, 2) \text{ (which was the original message).}$$

This method is much more effective than the El-Gamal version, since any message can be encrypted. Both have the huge advantage that they are incredibly difficult to break , since the difficulty of finding α given αP is so difficult with large numbers.

All these are downloaded recipes with very little mathematical activity involved (only what is needed to apply the recipe) and no attempt at understanding why it works

Evaluation

I have looked at many different methods of cryptography in this essay, but these vary in strength and usefulness. I have studied both private and public key encryption techniques, and methods workable by hand or only by computer. But which are the most useful methods?

The simple ciphers are useless if security is important, since they stand up to little scrutiny and can be broken in seconds by a computer using letter frequencies. Vigenère is much harder to break, but has the flaw of the repeated bigrams and trigrams, which allow analysis and thus the revealment of the keyword length; after this, analysis is easy. This can be used for small messages, since breaking them tends to require a fair amount of text in order to frequency analyse, but they are unsuitable for use for messages requiring substantial security.

One time pad is the only technique I have studied which is perfectly secure. If implemented correctly, it is completely unbreakable; however, as with many great encryption devices such as the Enigma machine, it is very difficult to implement them perfectly, and thus security can be compromised. It is normal to have a record of all the pads over a space of time, and the need to have this causes vulnerability to hacking or theft. During the cold war the hot line between the Kremlin and the White House was reportedly secured with a one-time pad[12], but they are only necessary when perfect security is imperative, due to the difficulties of implementation.

RSA and ECC are both public key encryption techniques. While both are very tough to break due to the complexity of the mathematics in the encryption, neither has perfect security, since while the private keys are onerous to find, with enough time and force answers can be found, allowing the decipherment of every message encrypted with those numbers. ECC is securer than RSA, as it can use any integer as α while RSA is restricted to prime numbers. Much of the choice between RSA and ECC is down to preference: the ciphertext of ECC is much longer than that of RSA, since it is double the length of the plaintext; however, ECC uses much smaller numbers in the calculations.

The telling difference between RSA and ECC is the required key size for computational security. The tables below[13] show a striking difference in the MIPS-Years (number of steps processed in one year at one million instructions per second) taken to break them using the most effective known methods.

ECC

Key Size	MIPS-Years
150	3.8×10^{10}
205	7.1×10^{18}
234	1.6×10^{28}

RSA

Key Size	MIPS-Years
512	3×10^{4}
768	2×10^{8}
1024	3×10^{11}
1280	1×10^{14}
1536	3×10^{16}
2048	3×10^{20}

[12] Cryptology, by Albrecht Beutelspacher, page 54
[13] Cryptography and Network Security: Principles and Practice, by William Stallings, page 199

in what sense?

This means that in general ECC requires far smaller keys than RSA, and so is more efficient. I believe that ECC is the better of these methods since it is simpler in terms of key length and time needed for encryption and decryption (my RSA calculations took the computer several minutes). I also believe that public key cryptography is more effective than private key cryptography, as it means that only one person needs to know the private key for messages to be sent and received. The perfect secrecy of one time pad is necessary only for messages that you want to never be broken, and I feel that the computational security of ECC with fairly long key length is sufficient for the encoding of all but the highest level communications. I therefore believe that ECC is currently the most useful cryptographic technique.

= a rather narrow base for decision.

What is it?

I believe that cryptography will continue to evolve, especially with the possibility of quantum cryptography. It is important to remember how new RSA and ECC are (late C20th), and how quickly the key size necessary for computational security grows. Therefore cryptography is a rapidly advancing field, with both many challenges but also many possibilities.

not discussed in the essay

This EE was both too ambitious (too large an area) and too modest (no attempt at justification of the methods).
This was supposed to be an essay in mathematics, yet it contains almost no mathematical activity, just downloaded recipes (admittedly somewhat complex).
A cookbook is not an essay in chemistry
The explanations of the ECC method is quite opaque
If focused on one method, properly explained and justified this could have been a good essay

Bibliography

1. http://dictionary.reference.com/browse/cryptography
http://dictionary.reference.com/browse/cryptanalysis
The American Heritage ® Dictionary of the English Language, Fourth Edition
Copyright © 2006 by Houghton Mifflin Company.
Accessed: December 15, 2008

2. http://dictionary.reference.com/browse/ciphertext
http://dictionary.reference.com/browse/plaintext
Dictionary.com Unabridged (v 1.1)
Based on the Random House Unabridged Dictionary, © Random House, Inc. 2006.
Accessed: December 15, 2008

3. http://www.bbc.co.uk/dna/h2g2/alabaster/A583878
Written 27 June, 2001
Accessed: December 15, 2008

4. http://www.simonsingh.net/The_Black_Chamber/frequencyanalysis.html
© Simon Singh
Accessed: December 15, 2008

5. Cryptography: An Introduction, by Nigel Smart
© Mcgraw-Hill College 2004
ISBN 978-0077099879

6. http://mathcircle.berkeley.edu/BMC3/rsa/node4.html
© Tom Davis on December 17, 2000
Accessed: December 15, 2008

7. http://www.mersenne.org/
© 1996-2008 Mersenne Research, Inc.
Accessed: December 15, 2008

8. http://mathworld.wolfram.com/news/2005-11-08/rsa-640/
© Eric W. Weisstein 2005
Accessed: December 15, 2008

9. http://upload.wikimedia.org/wikipedia/en/d/d5/PrimeNumbersSmall.png
Accessed: December 15, 2008

10. http://mathworld.wolfram.com/EllipticCurve.html
Weisstein, Eric W. "Elliptic Curve." From MathWorld--A Wolfram Web Resource.
© 1999-2008 Wolfram Research, Inc.
Accessed: December 15, 2008

11. Basic Methods of Cryptography, by Jan C. A. Van Der Lubbe,
translated into English by Steve Gee
© Cambridge University Press 1998
First published in Dutch (Basismethoden Cryptografie) © VSSD 1994
ISBN 978-0521555593

12. Cryptology, by Albrecht Beutelspacher
Originally published in German © Friedr. Vieweg & Sohn Verlagsgesellschaft mbH,
Braunschweig/Wiesbaden (entitled Kryptologie. 2. Auflage)
© 1994 by The Mathematical Association of America (Incorporated)
ISBN 978-0883855041

13. Cryptography and Network Security: Principles and Practice, by William Stallings
© Prentice-Hall Inc 1995
This edition (3rd) © Prentice-Hall Inc 2003
ISBN 978-0130914293

Appendix 1 – Characteristics of some major European languages

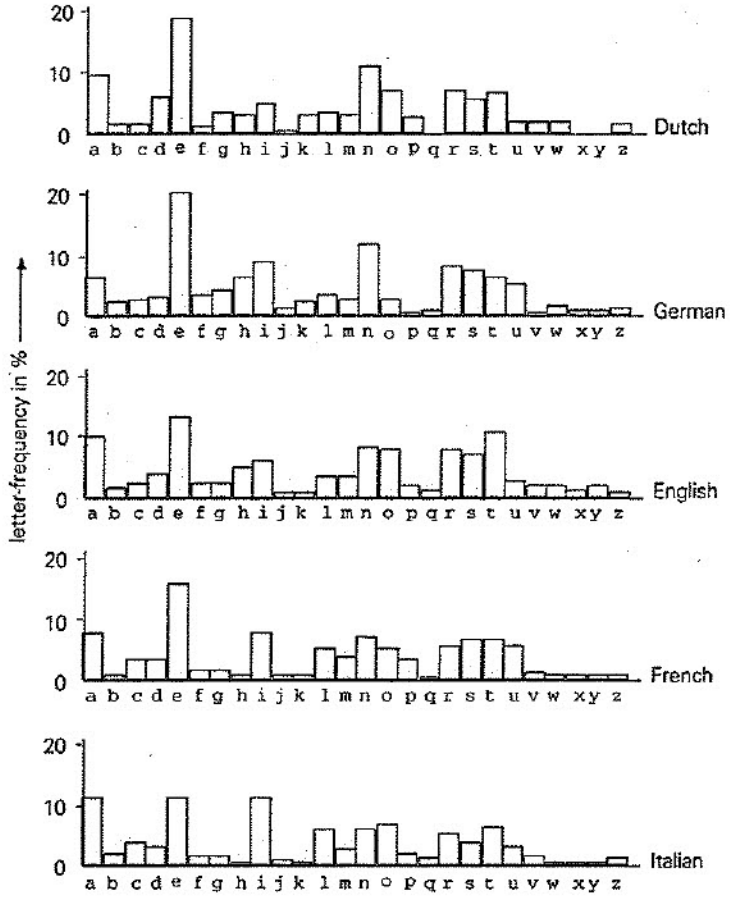
Language	English	French	German	Italian	Spanish
Commonest Letters	E T A O I N S H R D	E N A S R I U T O L	E N R I S T U D A H	E I A O R L N T S C	E A O S R I N L D C
Common Bigrams	th er on an re he in ed nd ha at en es of or nt ea ti to it st io le is ou ar as de rt ve	es en le de on ou nt re ne ed te em se er ar me an it et ie ti el ns ur ede les lle que ait eme ion	en er ch de ge ei ie in ne be el te un st di un ue se au re he it ri tz	er es on re de di ti si el en la al nt ra co ta to le li an in io ar or	es en el de la os ar ue ra re er as on st ad ai or ta co se ac ec ci ia
Common Trigrams	the and tha ent ion tio for tis nde has nce edt oft sth men	eur ell sse est dan del men des tio ese ans ter ons qui ais ous ent	ein ich den der ten cht sch che die ung gen und nen des ben rch	che ere zio del que ari ato eco edi ide esi idi ero par nte sta men	que est ara ado aqu del cio nte osa ede per ist nei res sde
One Letter Words	a i	a y o	-	e a i o	a e o u y
Two Letter Words	of, to, in, it, is, be, as, at, so, we, he, by, or, on, do, if, me, my, up, an, go, no, us, am	au, ce, ci, de, du, en, et, il, je, la, le, ma, me, ne, ni, on, ou, sa, se, si, un	ab, am, an, da, er, es, ob, so, wo, im, in, um, zu, du, ja	di, in ha, ho	en, la, de, lo, el, se
Common Doubles	ss ee tt ff ll mm oo	ss ll ee nn tt ff cc rr mm pp	ee tt ll ss dd mm nn	ll ss tt ee pp nn bb gg cc	ee ll rr aa ss cc dd nn

Source: http://www.simonsingh.net/The_Black_Chamber/hintsandtips.htm

Accessed: December 15, 2008

Appendix 2 – Letter Frequencies in some major European languages

Figure 2.1. Relative frequency of occurrence of letters for several languages.



Source: Basic Methods of Cryptography, by Jan C. A. Van Der Lubbe, translated into English by Steve Gee

First published in Dutch (Basismethoden Cryptografie) © VSSD 1994
ISBN 978-0521555593

Page 13

Assessment form (for examiner use only)

Candidate session number	0	0	

		Achievement level		
		First examiner	maximum	Second examiner
Assessment criteria	**A** research question	2	2	1
	B introduction	1	2	2
	C investigation	4	4	3
	D knowledge and understanding	3	4	3
	E reasoned argument	3	4	2
	F analysis and evaluation	2	4	2
	G use of subject language	3	4	3
	H conclusion	1	2	1
	I formal presentation	4	4	4
	J abstract	2	2	2
	K holistic judgment	3	4	2
	Total out of 36	28		25

Name of first examiner: _____ Examiner number: _____
(CAPITAL letters)

Name of second examiner: _____ Examiner number: _____
(CAPITAL letters)