

Some classical cryptography

MATH2068 Number Theory & Cryptography
Week 3 Lecture 1

University of Sydney
NSW 2006
Australia

11th August 2008

Why do we study classical cryptography?



Our main objective in cryptography is to understand the **RSA** and **Elgamal** cryptosystems

But it makes no sense to study these modern cryptosystems in ignorance of their predecessors.

In particular, we need to understand the deficiencies of classical systems to see the advantages of RSA and Elgamal.

Alphabets



An *alphabet* is a set of symbols (of any kind).
The elements of this set are called the *letters* of the alphabet.

Most of the time we will use the alphabet
 $\{A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z\}$.

We will frequently use the numbers
 $0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25$
to represent these 26 letters.

Another important alphabet is the two-letter alphabet $\{0, 1\}$.

We also use alphabets whose “letters” are composite symbols,
such as AB, AC, \dots, ZZ , or $000, 001, \dots, 111$.

Messages



A *message* is a sequence of letters from some alphabet.
We imagine that someone wishes to convey a message to someone else, in such a way that noone but the intended recipient will be able to understand it.

This is achieved by an encryption process that replaces the original message (the *plaintext*) by another message (the *ciphertext*), which is the one that is actually sent.

The recipient must know the decryption process, whereby the plaintext is recovered from the ciphertext.

Typically the encryption and decryption processes are very similar; indeed, they may even be identical.

Encryption and Decryption



In mathematical terminology, encryption and decryption are functions from one set of messages to another set of messages. This could involve a change of alphabet.

We use the term *source alphabet* for the alphabet of the input messages, and *target alphabet* for the alphabet of the output messages.

There are just two basic kinds of encryption/decryption processes that people have devised: **transposition ciphers** and **substitution ciphers**.

Transposition ciphers



In a transposition cipher system the output message (ciphertext) is obtained by rearranging the order of the letters of the input message (plaintext) according to some rule.

So if the plaintext is $x_1 x_2 \dots x_\ell$ then the ciphertext is $x_{\pi(1)} x_{\pi(2)} \dots x_{\pi(\ell)}$, where the numbers $\pi(1), \pi(2), \dots, \pi(\ell)$ are just the numbers $1, 2, \dots, \ell$ in some other order.

Most transposition ciphers involve entering the letters of the input message into some geometrical configuration, and then extracting them in a different order to obtain the output message.

An example of a transposition cipher



One simple procedure is to write the input message along the rows of a rectangular array, and obtain the output message by reading down the columns.

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

The ciphertext, split into blocks of length 5 for convenience, would be AATCR NUNUE EDSS FYFTA ROORR LERUE ALJWN ELLOE GIEIA A.

Decryption is achieved by using the total length of the message to figure out the length of the short last row and then entering the cipher text into the columns.

Block transposition ciphers



Another system involves splitting the message into blocks of some fixed size m and rearranging the letters within each block according to some permutation of $\{1, 2, \dots, m\}$. We call this a *block transposition cipher*; the permutation is called the *key*.

Suppose the block length is six and the key is 634125.

(This means that in each block the 6th letter of the input message becomes the first letter of the output, the third of the input becomes the second, and so on.)

Plaintext: ENCRYPTIONANDDECRYPTIONSHOULDBEEASYIFYOUKNOWHOW

```
ENCRYP TIONAN DDECRY PTIONS HOULDB EEASYI FYOUKN OWHOW*
PCRENY NONTIA YECDDR SIOPTN BULHOD IASEEY NOUFYK *HOOWW
```

Ciphertext: PCRENYNONTIAYECDDRSIOPTNBULHODIASEEYNOUFYKHOOWW

Simple substitution ciphers



The encryption key for a simple substitution cipher is a one-to-one function from the source alphabet to the target alphabet.

Then if the input message is $x_1 x_2 \dots x_\ell$
then the output message is $f(x_1)f(x_2) \dots f(x_\ell)$.

A simple form of this was used by Julius Caesar.

Caesar's cipher



Represent the letters A – Z by the numbers 0 – 25.

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Caesar's cipher consists of replacing i by $i + 3$, reduced modulo 26 if $i + 3 \geq 26$.

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
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

So

CROSSINGRUBICONTOMORROW

becomes

FURVVVLQJUXELFRQWRPRUURZ

Translation ciphers



Of course, Caesar could have used any number n in place of 3.

A *translation cipher*, or *alphabetic shift*, is a substitution cipher where i is replaced by $i + n$ (reduced mod 26).

Just about the least secure system imaginable!

There are only 26 possible keys, and an enemy who intercepted the message could just try them all.

Such a method of attack is known as an *exhaustive key search*.

Of course the attacker needs to know the encryption method before it is possible to embark on an exhaustive key search. But still one should use a system where the number of possible keys is so large that an exhaustive key search is never feasible.

Attacks on simple substitution ciphers



For an arbitrary simple substitution cipher on the alphabet A – Z the key is a permutation of the 26 letters.

The number of possible keys is
 $26 \times 25 \times \dots \times 2 \times 1 = 40329146112660563584000000$.

If an attacker tried exhaustive key search, and could test each possible key in one nanosecond, it would take 12779535235 years to do them all. Not practical!

The obvious weakness of simple substitution ciphers is that one can make a good guess at which ciphertext letters represent which plaintext letters by examining the frequencies with which the various letters occur in the ciphertext.

Frequency analysis



The most common letters in English text are E ($\approx 12\%$), T ($\approx 9\%$), A ($\approx 8.5\%$), O, N, I, S, R, H (all about 6% to 7.5%), followed by D ($\approx 5\%$) and L ($\approx 4\%$).

It is probably easy to identify E, T and A. (Though you may confuse T and A.)

The next six most frequently occurring letters in the ciphertext are likely to represent O, N, I, S, R, H in some order. And there are only $6! = 720$ different orders to try.

When you try the right order there will be so many recognizable words that the rest will be easy.

Homophonic substitution ciphers



For these systems the target alphabet should have more letters than the source alphabet.

Each letter of the source alphabet is associated with a set of letters of the target alphabet.

Example: 26 letter source alphabet A – Z; 64 letter target alphabet consisting of 8 term sequences of 0's and 1's.

A $\rightarrow \{01101100, 11110000\}$
B $\rightarrow \{10101010, 00011110\}$
...
E $\rightarrow \{11011011, 00110011, 01010101\}$
...

An A in the plaintext can be encrypted as 01101100 or as 11110000, an E as 11011011, 00110011 or 01010101, etc..

This makes frequency analysis harder, but not impossible.

Playfair Square



An example of a *polygram substitution cipher*. The letters of the plaintext are grouped into blocks; then blocks are substituted by other blocks according to some rule.

The Playfair Square requires a 25 letter source alphabet. This can be achieved by identifying J's and I's. The letters are written in a 5 by 5 square. For example,

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

The letters of the plaintext are grouped into two-letter blocks. Double letter blocks are not allowed; if one occurs an X is inserted. So "Meet me tomorrow" is first written as

ME ET ME TO MO RX RO WX.

(We had to stick an X on the end to make the length even.)

Playfair Square encryption/decryption



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

Plaintext: ME ET ME TO MO RX RO WX.

Ciphertext: STQHSTIMAMXLIXNO.

If the letters in a block are not in the same row or the same column, they are replaced by the letters at the other two corners of the rectangle they determine.

Two letters in the same row are shifted one step to the right (wrapping if necessary).

Two letters in the same column are shifted one step down (wrapping if necessary).