IS53012B/A Computer Security

Dr Ida Pu

Room 10, 29 St James
Goldsmiths, University of London

2019-20 (since 2007)

Part I

Cryptography

Cryptology

Confidentiality is the state of being secret.

cryptography Secret writing; The process or skill of communicating in or deciphering secret writing (or ciphers), including *encryption* (and *decryption*).

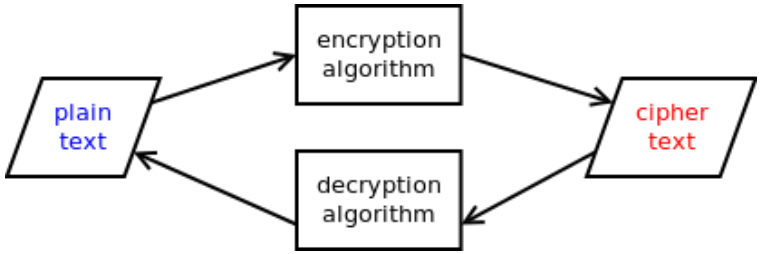
cryptanalysis The analysis and deciphering of cryptographic writings or systems

cryptology studies of cryptography related topics, i.e. cryptography or cryptanalysis

Classic cryptography means encryption, which began from as early as thousands of years ago, 2000 B.C. in Egypt.

- classic cryptography: using a pen and paper, now is often regarded as insecure or impractical
- early 20 century: Enigma rotor machine, with electronics and computing technology

Model of cryptography systems

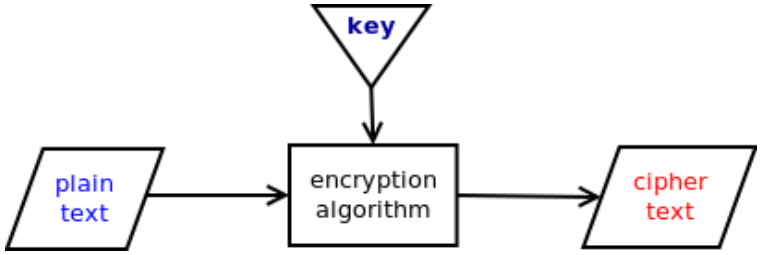


Let m be a plaintext and c the ciphertext, $\mathbf{f}$ be the encryption algorithm and $\mathbf{g}$ the decryption algorithm. We have

- $c = \mathbf{f}(m)$, with an encryption key e
- $m = \mathbf{g}(c)$, with a decryption key d.

where $\mathbf{g}(\mathbf{f}(m)) = m$.

Models of encryption

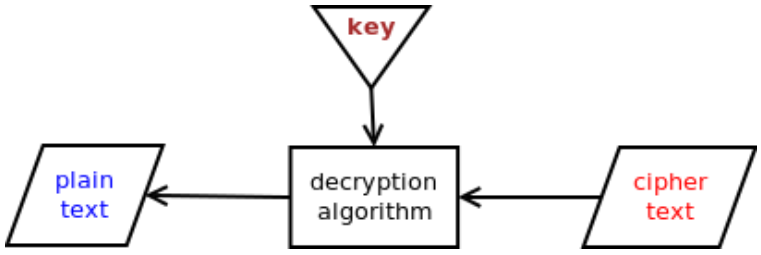


Let m be a plaintext and c the ciphertext, e be the key for encryption and d for decryption. We have

$$c = \mathbf{f}(e, m)$$

but we need a decryption algorithm $\mathbf{g}$, such that $\mathbf{g}(d, \mathbf{f}(e, m)) = m$.

Model of decryption



Let m be a plaintext and c the ciphertext, e be the key for encryption and d for decryption. We have

$$m = \mathbf{g}(d, c)$$

where $\mathbf{g}(d, \mathbf{f}(e, m)) = m$.

Concepts I

cryptology includes two disciplines: cryptography and cryptanalysis

cryptography refers to the study of concealing information using mathematical transformations. It used to mean *encryption and decryption*, the study of secret writing, a term from Greek word meaning 'hidden'. Today cryptography encompasses many aspects such as

- encryption
- authentication
- digital signatures
- one-way functions
- key generation, exchange and management

cryptoanalysis refers to the practice of revealing information hidden by cryptography using analytical and mathematical techniques.

Concepts II

plaintext the text to be encrypted

ciphertext the text that has been encrypted, or to be decrypted

encryption secret writing; a process of making the information not understandable

decryption recovering the original plaintext from a chipertext

cryptosystem often means an encryption system or *cipher*, can be described as a pair of invertible functions (f, g):

- function $\mathbf{f}$ referring to the encryption process
- function $\mathbf{g}$ to the decryption process.

Encryption

Intruder interceptor, the person to whom we would like to hide the information from and the intruder may try to

- block: affect availability
- intercept: affect confidentiality
- modify: affect integrity
- fabricate a message: affect integrity

Classic cryptosystem I

i.e. An encryption system involves

- alphabet and keyspace
- encryption and decryption algorithms
- personnel:

encrypt 'Alice'
decrypt 'Bob'
cryptanalyst 'Charles', who intercepts the crypted message
cryptographer one who studies the cryptosystems
interceptor 'Charles', the intruder

Classic cryptosystem II

A classic cryptosystem works as follows:

- Alice keeps the encryption key e secret and Bob keeps the decryption key d secret.
- Alice and Bob may know each other's keys, e.g. in symmetric key algorithms, or they may not, e.g. in public key cryptosystems

Classic encryption algorithms

Not limited to

1. Shift cipher, e.g. Caesar's cipher
2. Random substitution cipher
3. Block cipher, e.g. Transposition cipher
4. One-time pads, e.g. Vernam cipher

Shift cipher I

A technique based on letter substitution.
For convenience of discussion, we use capital letters for the plain text and lower case for encrypted message.
Example: $cipherchar(i) = plainchar(i + 4)$

i	1	2	3	4	5	6	7	8	9	..					
plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
cipher	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s
	..								24	25	26				
	P	Q	R	S	T	U	V	W	X	Y	Z				
	t	u	v	w	x	y	z	a	b	c	d				

Shift cipher II

Therefore,
'THE DOG BIT THE MAN' → xli hsk fmx xli qer
→ xlihskfmxxliqer
Decryption method: using the reverse shift:
$plainchar(i) = cipherchar(i - 4)$
When the shift is 3, it is called Caesar's cipher, named after Julia Caesar.
Shift cipher can encrypt a message in English alphabet A..Z, which can be extended to a larger alphabet, for example, including digits 0-9 and other symbols.

Shift cipher III

- Each letter in the plaintext is replaced by the letter after a shift of k places along in the alphabet (e.g. where k is between 0 and 25 inclusive, wrapping possibly around to the beginning of the alphabet)
- To decrypt the message, each letter is replaced by the letter $26 - k$ places along.

However, it is very easy to detect or decrypt this kind of ciphertexts.

Random substitution cipher

Each letter in the plaintext alphabet is replaced by a *random* letter in the ciphertext alphabet.

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

So

plain	S	E	E.	Y	O	U.	S	I	X.	I	N.	P	A	R	K
cipher	w	i	i	c	s	y	w	m	b	m	r	u	g	v	o

Transposition cipher I

A technique based on the rearrangement of the plaintext.
A long message can be broken up into shorter blocks. Each block is then encrypted and decrypted separately.

- Key: a string with a number of non-repeated letters.
- The letters of the key are each given a number according to their order in the alphabet
- Write the message continuously under the key
- The ciphertext is derived by outputting the letters of columns in the order of column numbers

Transposition cipher II

Example: 'THE QUICK BROWN FOX JUMPED OVER THE LAZY DOG ONCE'
'I AM THE' cannot be a key. Why?
But 'IAMTHE' can be a key.

Transposition cipher III

Encryption:

key:	I	A	M	T	H	E
alphabetic order:	*4*	*1*	*5*	*6*	*3*	*2*
	T	H	E	Q	U	I
	C	K	B	R	O	W
	N	F	O	X	J	U
	M	P	E	D	O	V
	E	R	T	H	E	L
	A	Z	Y	D	O	G
	O	N	C	E	_	_

Ciphertext: hkfprzniwuvlg_uojoeo_tcnmeaoeboetycqrxdhde

Transposition cipher IV

Decryption:

- Know the key
- Number the key in alphabetic order
- Compute: Number of Lines = Length(message)/Length(Key)
- Input the message by columns in alphabetic order

Transposition cipher V

Example: Number of lines $= \frac{42}{6} = 7$

key:	I	A	M	T	H	E
alphabetic order:	*4*	*1*	*5*	*6*	*3*	*2*
1		h				I
2		k				w
3		f				...
4		p				
5		r				
6		z				
7		n				

Plaintext: THE QUICK BROWN FOX JUMPED OVER THE LAZY DOG ONCE

One-time pads - Vernam cipher

Named after Gilbert Vernam from AT&T, using arithmetic notation and a *pad* of random digits (e.g. 2 digits)
Example: Encrypt message 'SECURITY'

alphabet	A	B	C	D	...			
i (0–25)	0	1	2	3	...			
plaintext:	S	E	C	U	R	I	T	Y
i	18	4	2	20	17	8	19	24
r:random number	10	11	24	65	23	98	89	80
$i + r$	28	15	26	85	40	106	108	104
mod 26	2	15	0	7	14	4	6	0
ciphertext	c	p	a	h	o	e	g	a

Decryption? Use the identical pad: 10 11 24 65 23 98 89 80

One-time pads - Vernam cipher

alphabet	A	B	C	D	...			
i (0–25)	0	1	2	3	...			
ciphertext	c	p	a	h	i	e	g	a
i	2	15	0	7	14	4	6	0
pad r	10	11	24	65	23	98	89	80
$i - r$	-8	4	-24	-58	-9	-94	-83	-80
mod 26	-8	4	-24	-6	-9	-16	-5	-2
$+26$ if < 0:	18	4	2	20	17	8	19	24
plaintext:	S	E	C	U	R	I	T	Y

The *one-time pad* can achieve *perfect secrecy*, but is not used for all encryption.

Cryptosystem models

- Symmetric key algorithms: The same key is used for both encryption and decryption. Assume that Alice and Bob know each others' key
- The characters of the plaintext message m come from a message alphabet M and the ciphertext c are characters of the ciphertext alphabet C
- Alphabets M and C may be the same but they could be different
- A cryptosystem refers to all the aspects of a particular encryption system, including
  1. information about the plaintext and ciphertext alphabets
  2. encryption and decryption algorithms
  3. method of blocking and the valid keys, etc.
- The keyspace K is the set of all possible encryption keys. We usually include the trivial keys in the key space
- The enemy in the cryptosystem, Charles, intercepts the cipher message c and tries to gain unauthorised information m.

Key space size

- Shift cipher: The key $\boxed{i}$ can only be $1, 2, \cdots,$ or 26. Hence, the key space is 26.
- Transposition cipher: A key of length n is a string $\boxed{k_1}, k_2, \cdots, k_n$. The order of each symbol in the key can be represented as $1, 2, \cdots, n$. So $2, 1, 3, 4, 5, 6$ is another order, and $5, 1, 3, 4, 2, 6$ is another. Since k_1 can be one of $1, \cdots, n$, there are n possibilities, and k_2 can have $n-1$ possibilities, $\cdots$, so the number of possible keys with different orders is $n!$.
  In our example, when $n = 6$, there are $6! = 6 \times 5 \times 4 \times 3 \times 2 = 720$
- Vernam cipher: A key of length n is a string $k_1, k_2, \cdots, k_n$. Since k_i can be an integer $\in [0, m]$, there are m possibilities for k_1, and $m-1$ for k_2, $\cdots$. Therefore, the size of the key space is $m(m-1)(m-2)\cdots(m-n+1)$, where $m \geq n$.

Charles, or cryptanalist

Charles needs to determine the plaintext without determining either key, encryption key k_e or decryption key k_d

- Having the decryption key, Charles can decrypt the ciphertext c (and other ciphertexts encrypted using the same encryption key)
- Having the encryption key, he can determine the decryption key or masquerade as Alice

Good assumption: Charles has knowledge of

- encryption and decryption algorithms
- key space K
- time (the computer power)

Potential attacks I

- ciphertext only
- full plaintext (known or guess)
- partial plaintext (chosen message, ciphertext or both)
- encryption algorithms

Charles' techniques:

Try all possible decryption keys The success of this will depend on the

- size of the keyspace K
- redundancy in the message

How long it takes to recover the key depends on

- number of keys
- how long it takes to investigate each key

Potential attacks II

Analyse the ciphertext statistically Example:

- A frequency count on the letters of the ciphertext, say in English, may indicate the substitutions used for the most commonly used letters
- Redundancy in the English language will give away the rest of the key

Differential cryptanalysis Charles may

- generate a large number of messages with similar contents
- persuade Alice to encrypt them
- analyse the corresponding ciphertexts

A 'secure' cryptosystem

Confusion difficult to predict the consequence by changing one or a number of characters in the plaintext, i.e. the corresponding relationship between a ciphertext and plaintext is difficult to guess. e.g. Substitution.

Diffusion a change in the plaintext would affect many parts of the ciphertext. e.g. Permutation.

Example

- A large alphabet M to make it hard to do statistical analysis
- A large keyspace K to make it hard to do exhaustive keyspace search
- High speed of execution (for high message throughput)
- The same algorithm for encryption and decryption (to reduce costs)

Shannon's characteristics of Good ciphers I

- The level (amount) of secrecy required should determine the amount of labour appropriate for the encryption and decryption
- The set of keys and the encryption algorithm should be free from complexity. The choice of keys and the type of plaintext should not affect too much on the performance of the cryptosystems.

Example

Good or bad?

- An algorithm works only on plaintext having an equal number of `As` or `Es`
- The sum of the values of the letters of the key must be a prime.

Shannon's characteristics of Good ciphers II

- The implementation of the process should be as simple as possible
- Errors in ciphering should not propagate nor cause corruption of further information in the message

Example

- Consider the transposition cipher: one missing letter in a column transposition during the encryption, or the decryption, would cause errors in the entire remaining process.

- The size of the ciphertext should be no larger than the plaintext. Why? The ciphertext cannot possibly carry more information than the plaintext. The extra data can only give the cryptanalyst more hints to detect the patterns.

Trustworthy encryption systems

- based on sound mathematics
- analysed by competent experts and found to be sound
- has stood the 'test of time'

Three popular encryption algorithms (in commercial world):

- Data encryption standard (DES)
- Rivest-Shamir-Adelman, name after the inventors (RSA)
- Advanced encryption standard (AES)