

COMP4127/COMP7850

Information Security

Classical Crypto System

Learning outcomes

After this lecture you shall be able to:

- Name some classical crypto;
- Identify different attacker Models;
- Understand the brute-force attack;

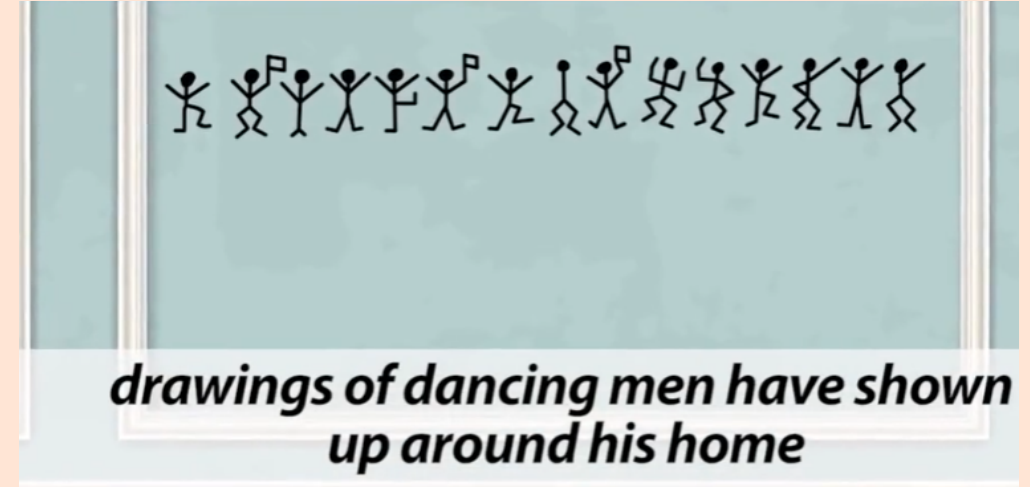
- **Cryptography** is the science of secret writing.
- It is the art of concealing a message within another message or a physical object.
- The word cryptography comes from the Greek words *kryptos*, meaning "hidden" or "secret", and *graphein*, meaning "writing".
- Cryptography was mainly about encryptions (provide data secrecy)
- Now we use this word to refer all mathematical techniques to provide data security
 - Digital signatures
 - Authentication
 - Proof of knowledge

Cryptography in History

- 1900BC, Egyptian use an alternative form of symbols, **hieroglyphs**, on the wall of a tomb.
- Hieroglyphs are different than the writing system used in Egypt at that time and only nobels and priests could read them.



- *Adventure of Dancing Men* by Arthur Conan Doyle.



- Are you able to read this?

tsitpaB ytisrevinU

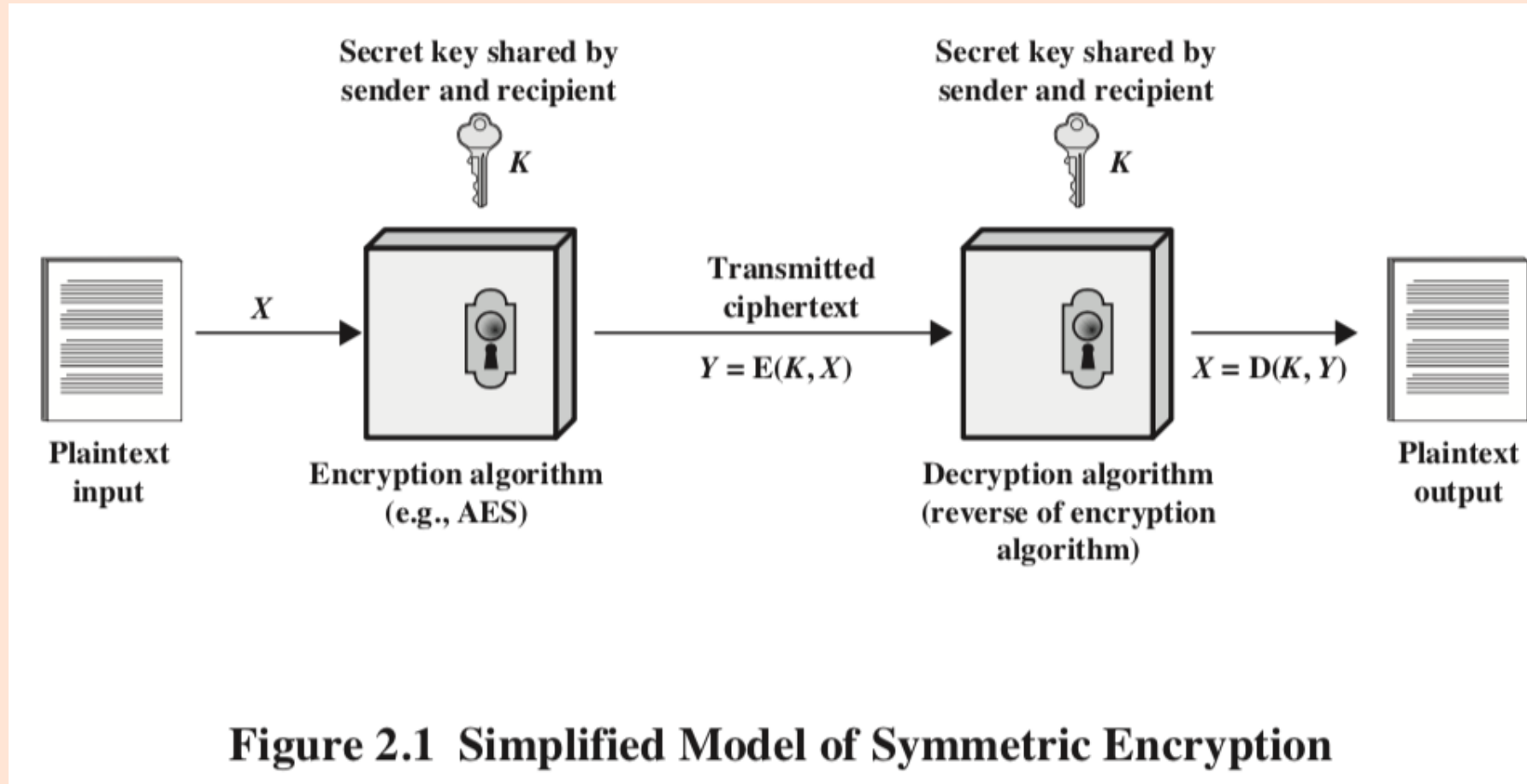
I111L111O111V111E111H111K111B111U

1D3^5 ^&3 8U|_|_3+P&00|=

- Once you know the pattern, this encryption no longer work!

Important of Keys

- Encryption is composed of a **Key** and an **Encryption algorithm**.



Security needs a key!

Cryptography in History

- 100BC Julius Caesar conveyed secret messages to his army using a special code.
- Caesar substituted each letter of the alphabet with a different letter, according to a fixed system.
- Idea: Shift the character by a fixed number of letter.
- e.g. Key = 3

A	B	C	D	...	X	Y	Z
D	E	F	G	...	A	B	C

```
plain: meet me after the toga party  
cipher: PHHW PH DIWHU WKH WRJD SDUWB
```



Caesar Cipher - mathematically speaking

- Let message m be a n characters string, K be an integer.
- $M = m_0m_1m_2 \dots m_{n-1}$
- $K \in [1, 25]$



$$C = c_0c_1c_2 \dots c_{n-1}$$
$$c_i = E_K(m_i) = m_i + K \pmod{26}$$



mod 26: divided by 26 and find the remainder. e.g.
 $40 \pmod{26} = 14$

Brute Force

There are only 25 keys to try. Thus, it is insecure against ciphertext only attacks.



How about setting K be a number larger than 25?

	PHHW	PH	DIWHU	WKH	WRJD	SDUWB
KEY						
1	oggv	og	chvgt	vjg	vqic	rctva
2	nffu	nf	bgufs	uif	uphb	qbsuz
3	meet	me	after	the	toga	party
4	ldds	ld	zesdq	sgd	snfz	ozqsx
5	kccr	kc	ydrpc	rfc	rmey	nyprw
6	jbbq	jb	xcqbo	qeb	qldx	mxoqv
7	iaap	ia	wbpan	pda	pkcw	lwnpu
8	hzzo	hz	vaozm	ocz	objv	kvmot
9	gyyn	gy	uznyl	nby	niau	julns
10	fxxm	fx	tymxk	max	mhzt	itkmr
11	ewwl	ew	sxlwj	lzw	lgys	hsjlg
12	dvvk	dv	rwkvi	kyv	kfxr	grikp
13	cuuj	cu	qvjuh	jxu	jewq	fqhjo
14	btti	bt	puitg	iwt	idvp	epgin
15	assh	as	othsf	hvs	hcuo	dofhm
16	zrrg	zr	nsgre	gur	gbtn	cnegl
17	yqqf	yq	mrfqd	ftq	fasm	bmdfk
18	xppe	xp	lgenc	esp	ezrl	alcei

Key cannot be too small

Hu Fu (虎符) in ancient China

- A Hu Fu has two half which can be combined to form a whole tiger.
- One half is kept by the King, the other half is kept by the general.
- Message of the King will be accompanied by the Hu Fu.



Hu Fu (虎符) in ancient China

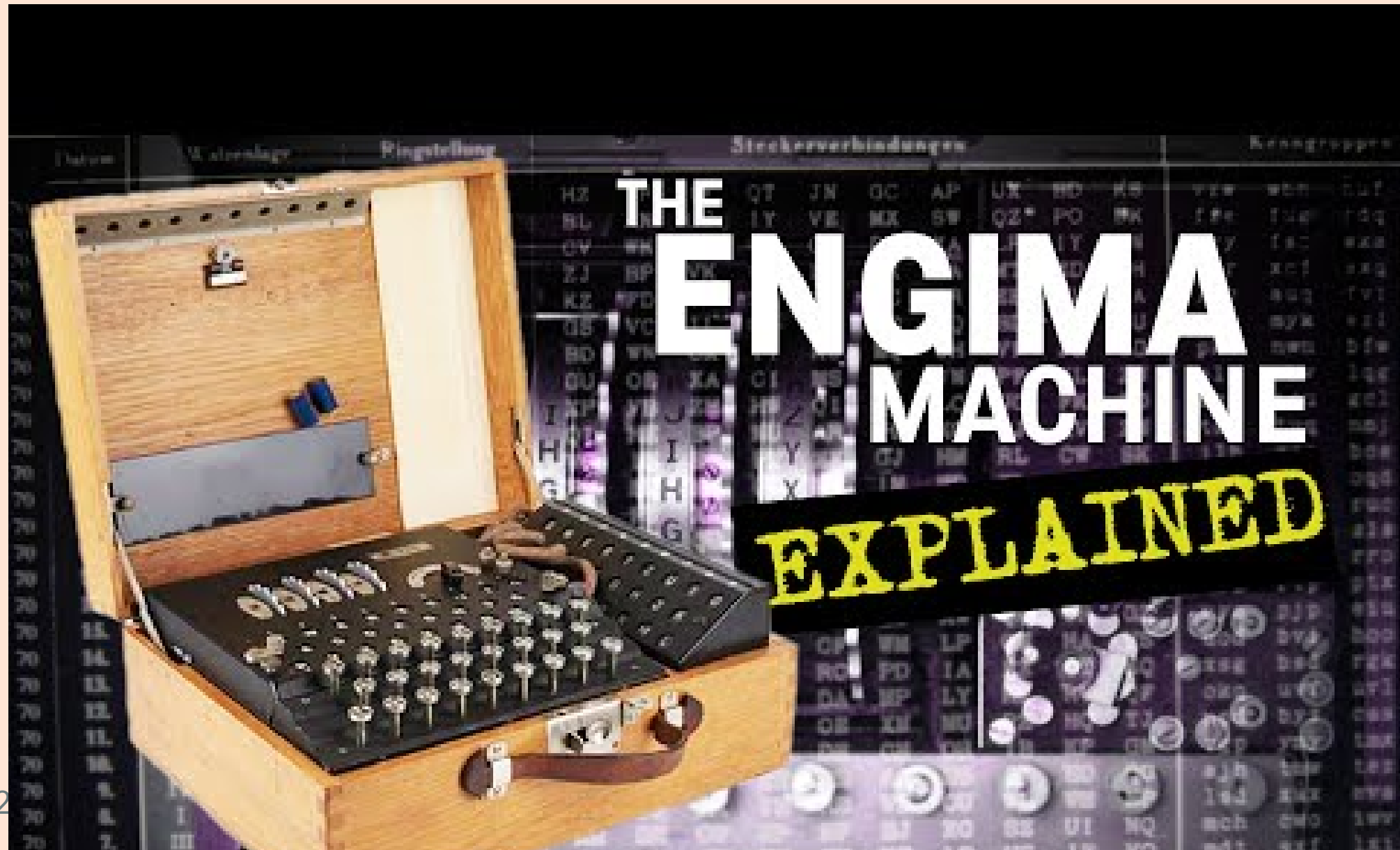
- **Zhuge Liang** (181-234) was a famous Chinese military strategist and politician during the Three Kingdoms period.
- He occupied a state from his enemy Cao Cao and took his Hu Fu.
- He forged a fake message to other Cao Cao's states to move away their army from the castles with the token Hu Fu.



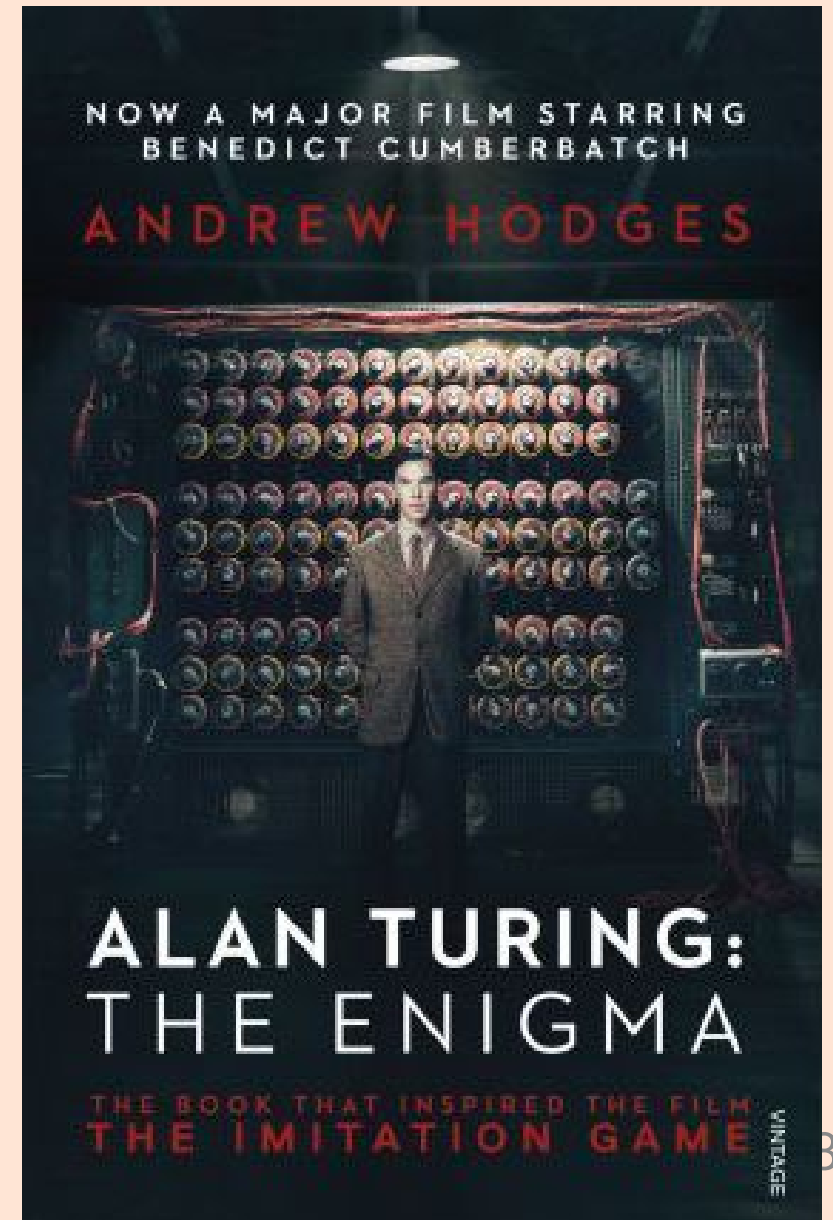
Application in WWII (Enigma)

- Enigma machine was used by German army in WWII.
- It was a cipher machine that used a series of rotating disks to encrypt and decrypt messages.
- The disks were called rotors.





Heard about the Turing Award?



Take a closer look at an encryption system

Encryption and Decryption

- An unencrypted message is called a *plaintext* or a *message*.
- An encrypted message is called a *cipher* or a *ciphertext*.
- The process that transforms a plaintext is called encryption.
- The process that inverses the encryption is called *decryption*.

We denote

- $E_K(X)$ for encrypting message X using key K .
- $D_K(C)$ for decrypting cipher C using key K

- Cryptanalysis means **attack**.
- Attack by the algorithm plus some sample plaintext-ciphertext pairs.

Common types of attacks:

- **Breaking the algorithm:** The attacker tries to exploit the weakness of an encryption algorithm.
- **Brute-force attacks:** The attacker tries every possible key on a piece of cipher-text until an intelligible translation into plaintext is obtained.
- **Non-crypto attacks:** Such as stealing the key, exploiting the weakness of implementation, guessing the random number generation, side-channel attacks, etc...

A naïve encryption

Plaintext: phone number of my crush, 9876-5432

Key: my home number, only my family know it, 2111-0000

Ciphertext: Plaintext + Key

$$C = E_K(P) = 98765432 + 21110000 = 119875432$$

💡 Why it is not good?



1. Home number isn't that secret;
2. Can't be used twice;
3. Key can be deduced if plaintext is known;
- ...

Then what is a **secure encryption**?

OK, how can we break it?

- Brute-force attack means trying every possible key until the correct one is found.
- It is a **brute-force attack** if the attacker tries every possible key on a piece of cipher-text until an *intelligible translation* into plaintext is obtained.

Intelligible translation

- The attacker can read the plaintext.



Supercomputer Fugaku has 7.6M core running 442000 TFlop/s = 4.42×10^{17} /s or 2^{58} operations per second[^].

key size	time required
56 bits $\approx 7.2 \times 10^{16}$	instant
64 bits $\approx 1.8 \times 10^{19}$	20 seconds
80 bits $\approx 1.2 \times 10^{24}$	7.9 days
128 bits $\approx 3.4 \times 10^{38}$	10^{13} years

We generally assume any computation or brute-force requiring 2^{100} is **unsolvable**.

$$C = E_K(P) = 98765432 + 21110000 = 119875432$$



Can we brute-force the encryption algorithm above?

- We know the key is 8 digits long.
- We need at most 10^8 tries to find the key.
- We can try all possible keys in a few seconds.
- However, we need to *verify* if the decrypted message is correct. Can we?

Brute Force

Brute force on this Caesar cipher is possible, because only one of the message is meaningful!

	PHHW	PH	DIWHU	WKH	WRJD	SDUWB
KEY						
1	oggv	og	chvgt	vjg	vqic	rctva
2	nffu	nf	bgufs	uif	uphb	qbsuz
3	meet	me	after	the	toga	party
4	ldds	ld	zesdq	sgd	snfz	ozqsx
5	kccr	kc	ydrpc	rfc	rmey	nyprw
6	jbbq	jb	xcqbo	qeb	qldx	mxoqv
7	iaap	ia	wbpan	pda	pkcw	lwnpu
8	hzzo	hz	vaozm	ocz	objv	kvmot
9	gyyn	gy	uznyl	nby	niau	julns
10	fxxm	fx	tymxk	max	mhzt	itkmr
11	ewwl	ew	sxlwj	lzw	lgys	hsjlg
12	dvvk	dv	rwkvi	kyv	kfxr	grikp
13	cuuj	cu	qvjuh	jxu	jewq	fqhjo
14	btti	bt	puitg	iwt	idvp	epgin
15	assh	as	othsf	hvs	hcuo	dofhm
16	zrrg	zr	nsgre	gur	gbtn	cnegl
17	yqqf	yq	mrfqd	ftq	fasm	bmdfk
18	xppe	xp	lgenc	esp	ezrl	alcei

Substitution Cipher

- Also known as monoalphabetic ciphers.
- Each character is uniquely substituted by another character. e.g.

A	B	C	D	...	X	Y	Z
V	C	Z	E	...	H	B	J

```
plaintext: BAD BOY  
ciphertext: CVE CLB
```



Substitution Cipher - in Math

Let $P_k : \{A, B, C, \dots\} \rightarrow \{A, B, C, \dots\}$ be a permutation.

Let message m be a n characters string, $M = m_0m_1m_2\dots m_{n-1}$

Substitution cipher is defined as $C = c_0c_1c_2\dots c_{n-1}$ where

$$c_i = E_K(m_i) = P_k(m_i)$$

💡 When we use the word *permutation*, it must be a one-to-one mapping. No two characters will be mapped to the same cipher.

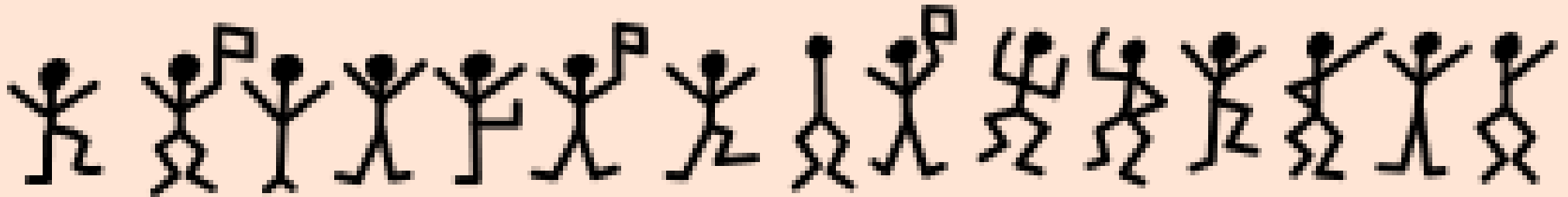
The key complexity = $26! \approx 4 \times 10^{26}$



! is read as "factorial". $5! = 5 \times 4 \times 3 \times 2 \times 1$.

- The code is much less secure if some "puzzles" are found.
- It is not secure against chosen plaintext attacks.
- It is not secure against known plaintext attacks.
- It is not secure against ciphertext only attacks, even Shylock Holmes knows how to break it.

See the *Advanture of Dancing Men* by Arthur Conan Doyle.



- If we count the frequency of each letter in a message, we can find out the most frequent letter.

Welcome to Hong Kong Baptist University.

As a global university, and one of the top research-led liberal arts institutions in Asia, we are proud of our heritage. From HKBU's humble beginnings in 1956 as a post-secondary college, our Founding Fathers instilled a liberal arts philosophy and a caring, creative, and global culture. Those principles are as important today as they were then.

Over time the University has gone from strength to strength, and become a leader in a wide range of fields. Together with our students, faculty and staff members, as well as our friends in the community, we are committed to academic excellence in teaching, research and service, and this can be witnessed through our diverse range of excellent programmes for students, as well as the knowledge breakthroughs we have achieved for the betterment of the world. In addition, our whole-person and liberal arts education ensures that we nurture a huge diversity of talent, and we are dedicated to developing future leaders who can shape the future with confidence, competence and commitment.

I hope your visit to our website will give you a better picture of our University and our outstanding students, staff, research and educational initiatives, and we look forward to seeing you on our campus.

frequency of e = 137/1073

Frequency analysis

- Count the freq of single character
- Count the freq of two consecutive characters
- Look for I, A, AM, THE
- ...

https://en.wikipedia.org/wiki/Frequency_analysis

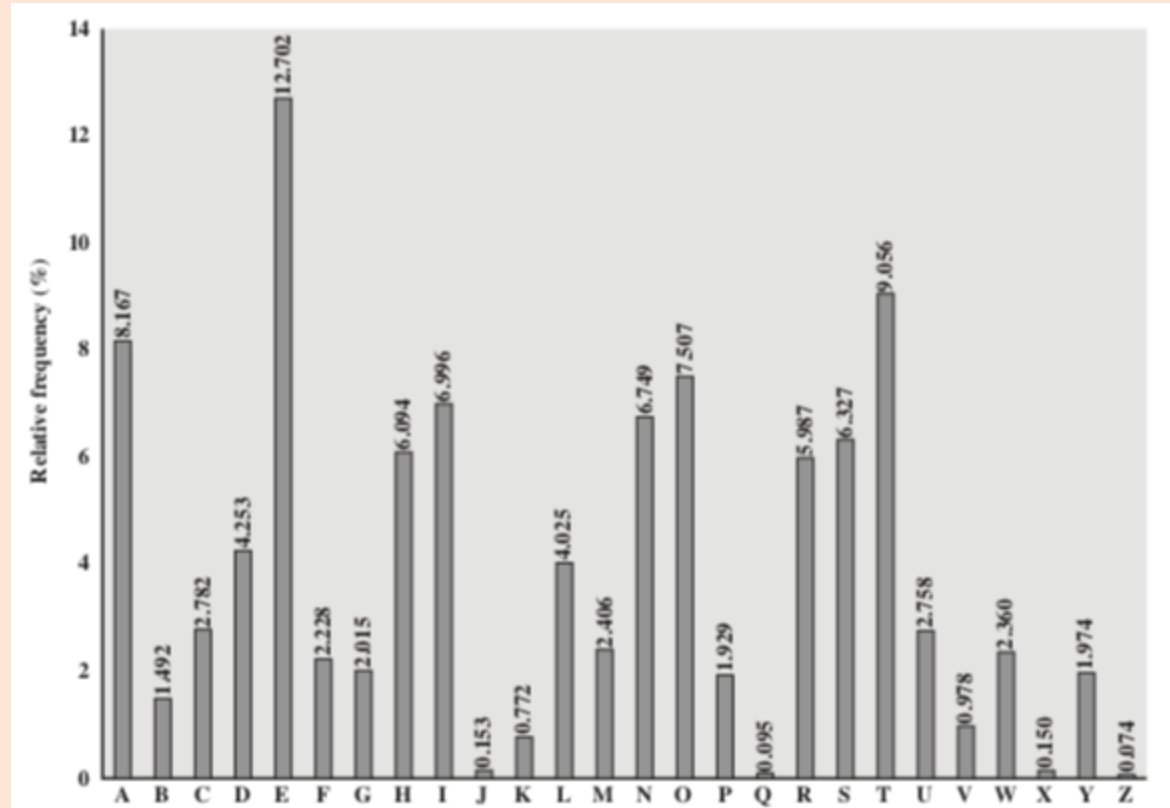


Figure 2.5 Relative Frequency of Letters in English Text

Try one...

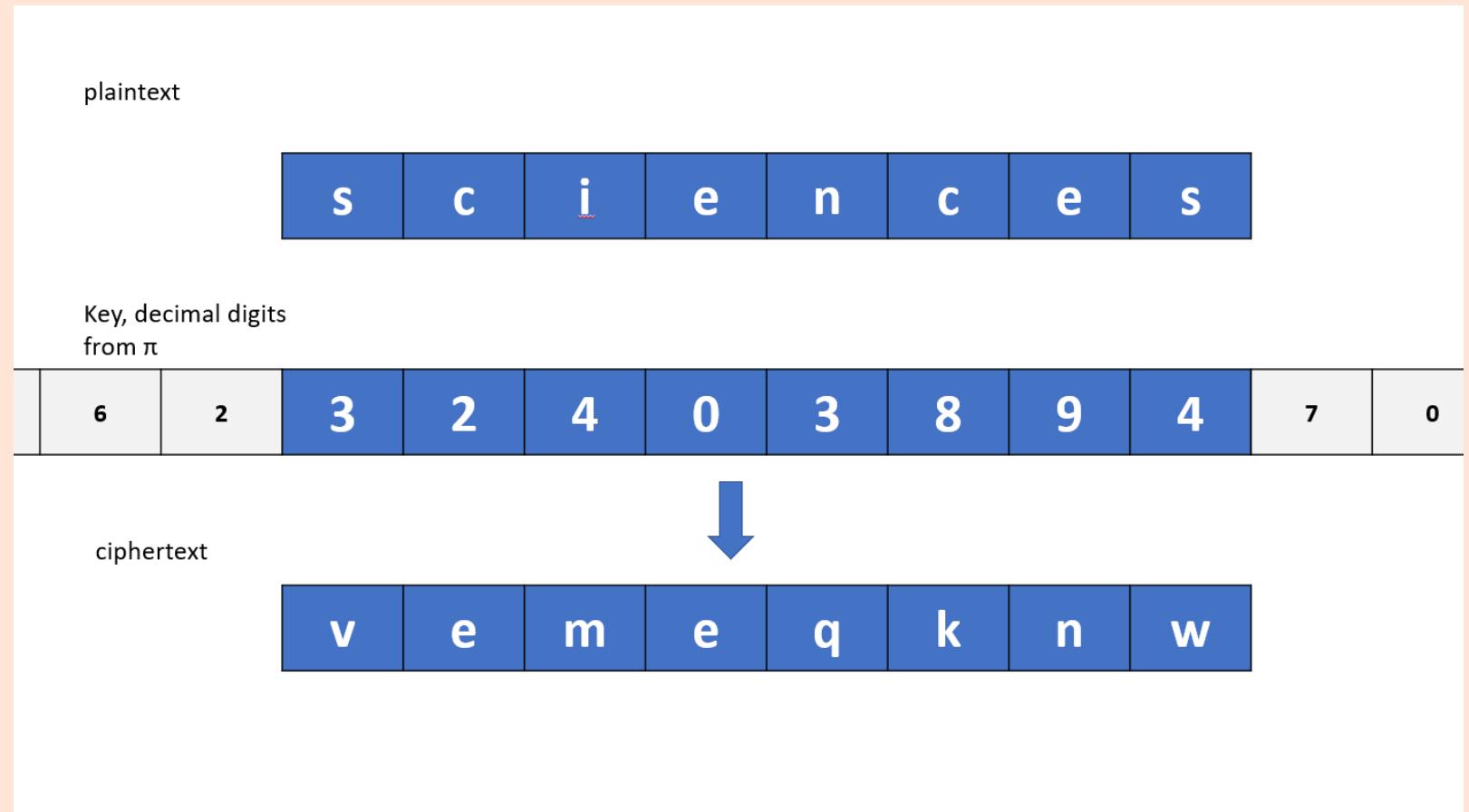


Rbo rpktigo vcrb bwucja wj kloj hcjd, km sktpqo, cq rbwr loklgo
vcgg cjqcqr kj skhcja wgkja wjd rpycja rk ltr rbcjaq cj cr.
-- Roppy Lpwrsborr

- Finding Frequency: <https://trinket.io/python3/5d07ba6b62>
- Cracking tool: <https://www.dcode.fr/monoalphabetic-substitution>
- Frequency reference 1: https://en.wikipedia.org/wiki/Letter_frequency
- Frequency reference 2: <http://www.viviancook.uk/SpellStats/DigFreqs.html>

Try another one... PiEncryptor

- A cipher is generated using this idea
- A key, starting from an unknown decimal position from π is used.



Open it from <https://comp7850.hkbu.app>

It is characterized by the following dimensions:

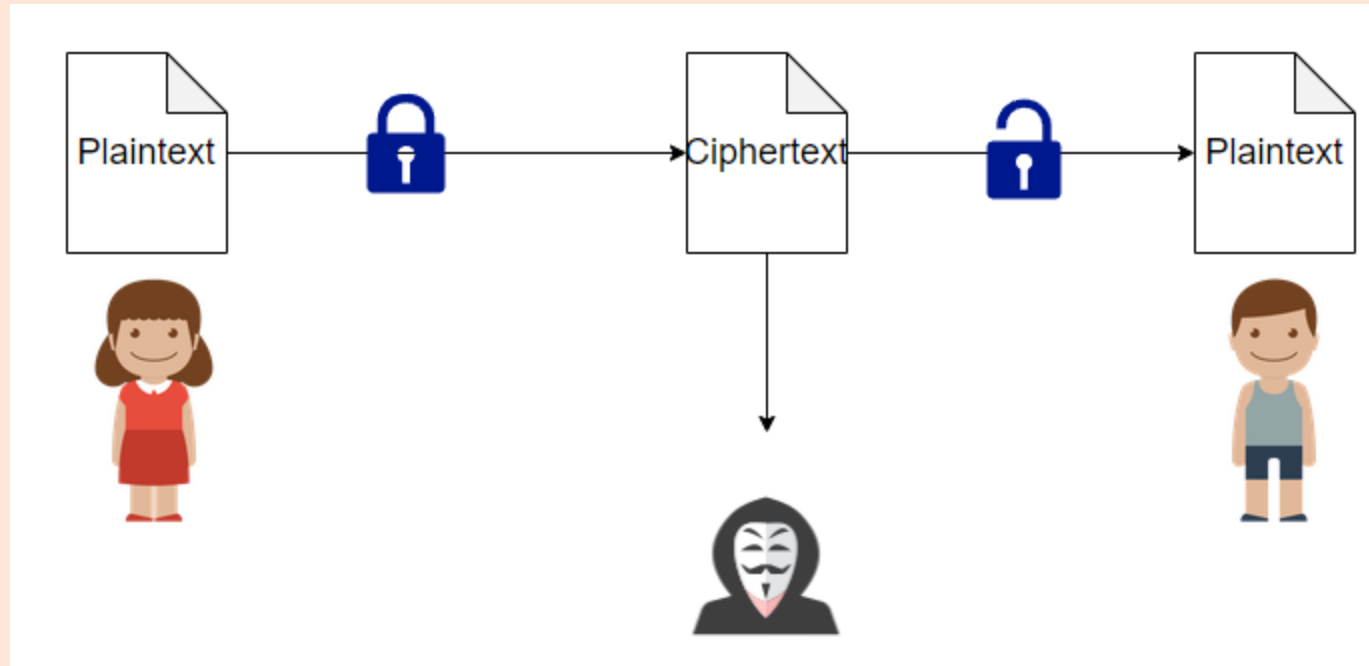
- **Attacker models:** How strong is your attacker, with what resources/capabilities.
 - e.g. Knowing only ciphertext versus Knowing both ciphertext and plaintext
- **Security Goal:** What goals does your attacker wants to achieve. Some goals can be achieved easier than the others
 - e.g. confidentially, indistinguishable, non malleable
- **Assumptions:** What is the computational limitation
 - e.g. factorization is hard problem, exhaustive search is not possible beyond 2^{60} , ...

We assume attackers with different levels of capabilities.

- **Ciphertext only attacks:** Attacker knows only the ciphertext and is needed to deduce the plaintext
- **Known plaintext attacks:** Attacker is given pairs of plaintext-cipher to investigate
- **Chosen plaintext attacks:** Same as above except the attacker can choose the plaintext on his own.
- **Chosen ciphertext attacks:** Attacker can choose the cipher and obtain the plaintext

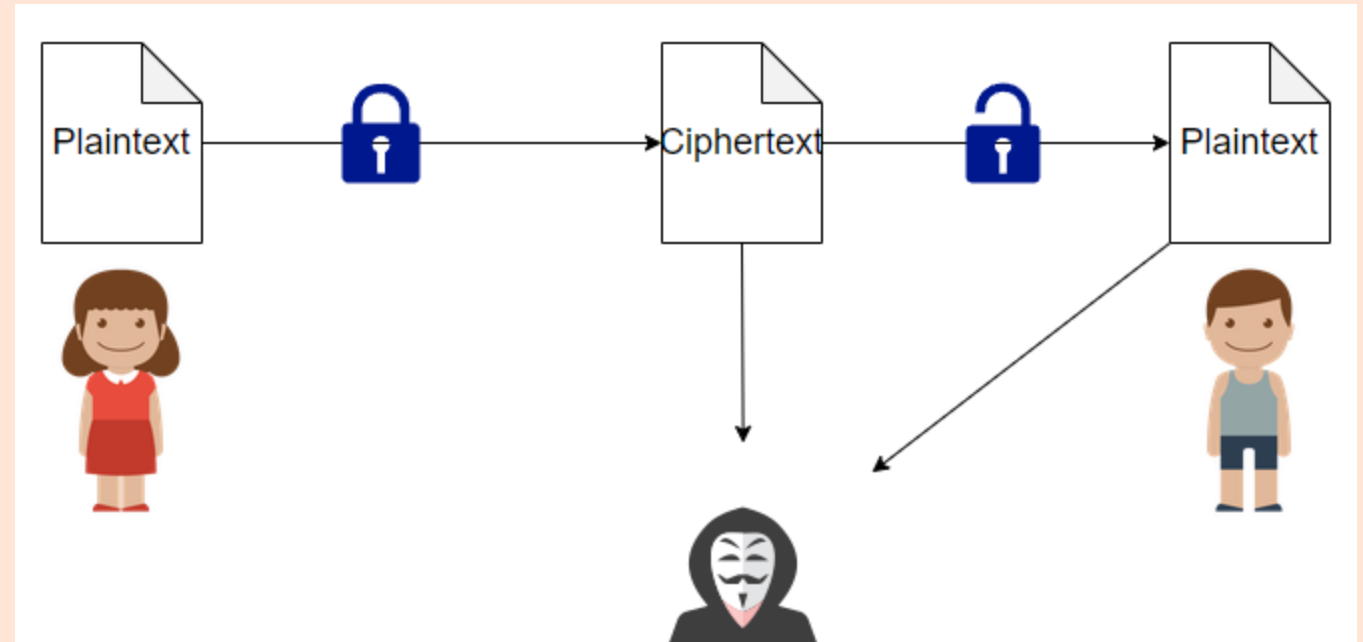
Ciphertext only attacks



- This model describes an attacker obtains the encrypted ciphertext only.
- He has no idea (or only a rough idea) of what the sender and receiver are talking.
- He needs to figure out the plaintext of a given ciphertext.



Known plaintext attacks

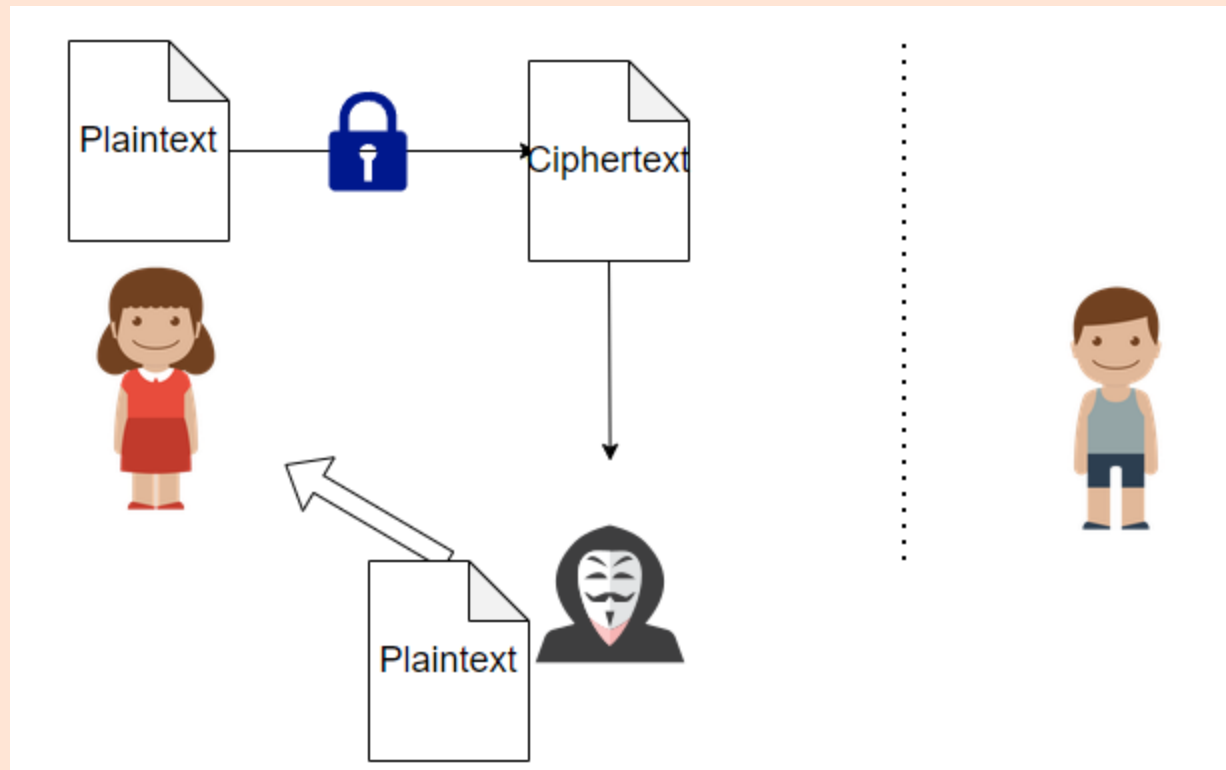
- This model describes an attacker has some ciphertext and the corresponding plaintext obtained from the **history**.
- Then he needs to figure out the plaintext of a given ciphertext.



  e.g A cracker taps the encrypted cable TV signal of a football match. The cracker knows the content of the TV signal (football match). He then try to use these to crack the TV box.

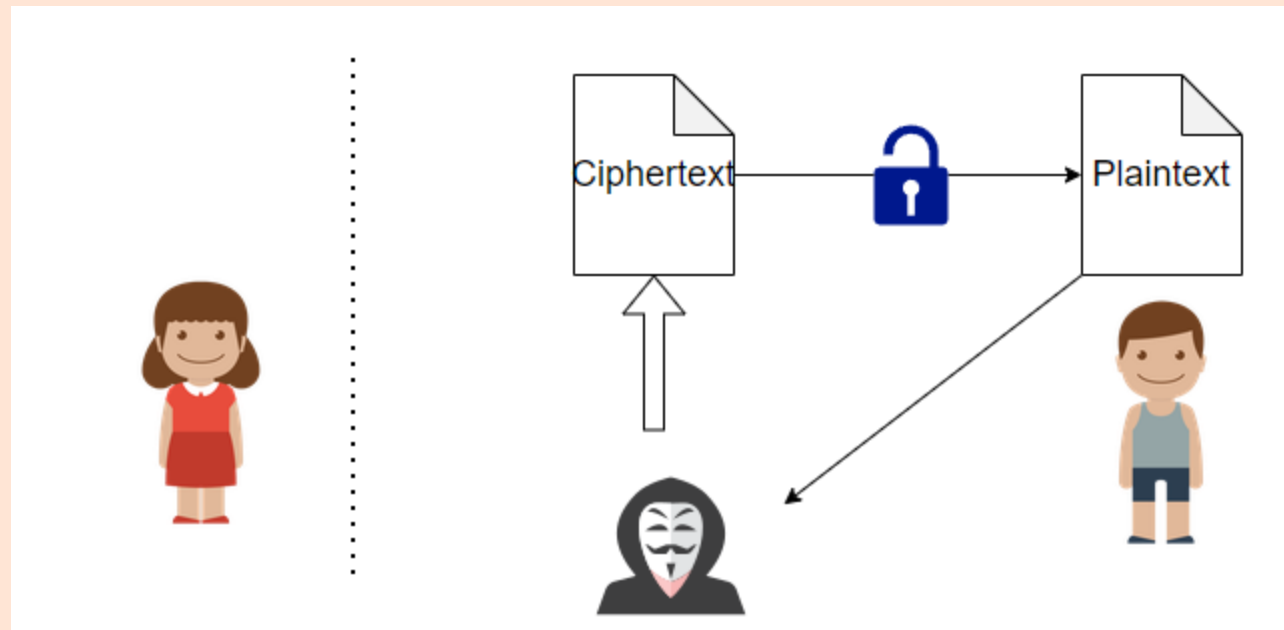
Chosen plaintext attacks

- This model allows the attacker to request the encryption of certain plaintext and observe the corresponding ciphertext.
- During the process, a blackbox / proxy will encrypt for him.
- Then he needs to figure out the plaintext of a given ciphertext.



Chosen ciphertext attacks

- This allows the attacker to request a decryption of certain ciphertext to obtain some plaintext.
- Then he needs to crack a new ciphertext with the exception that he cannot request the crack of the ciphertext he is challenged.

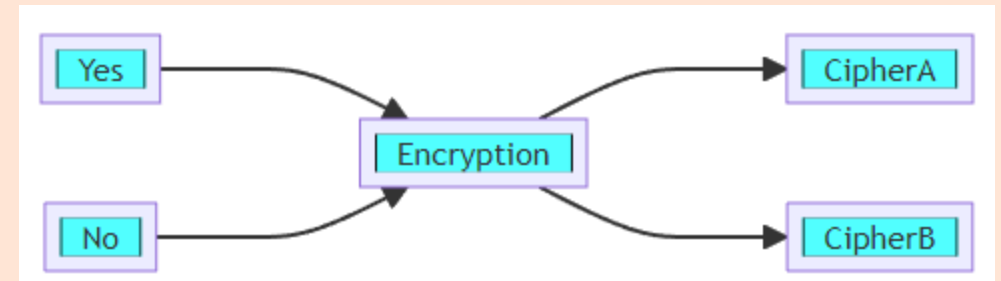


💡 Which attacker models is/are the naïve encryption insecure against?

- In encryption we assume that the attacker is trying to achieve some goals.

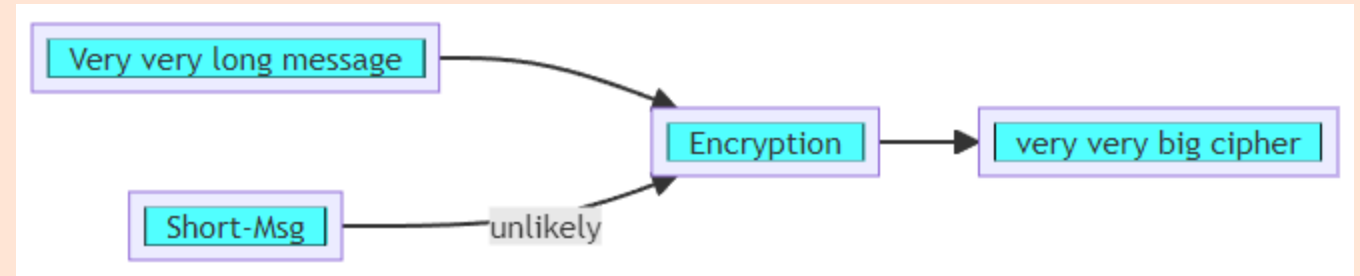
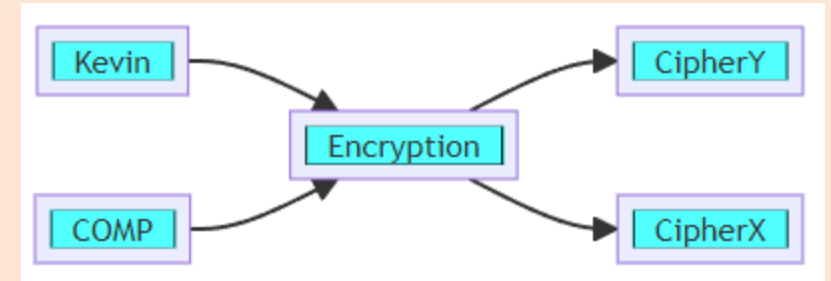
Indistinguishability (IND)

- Suppose an attacker have seen two CipherA and CipherB, he knows they are coming from the plaintext "Yes" or "No" (but don't know which one is which).
- He must not be able to tell which one is which, otherwise the encryption is insecure.



Security Goal - Indistinguishability

- Pushing the goal a little higher, for **any** two pairs of plaintext, other than just "Yes" and "No", the attacker should not be able to tell which one is which.
- To be fair, these plaintext should be about same length, otherwise the attacker can tell which one is longer.



Non-malleability

- An encryption algorithm is **malleable** if it is possible to transform a ciphertext into another ciphertext which decrypts to a related plaintext.
- For example, suppose that a bank uses a stream cipher to hide its financial information, and a user sends an encrypted message containing, say, "TRANSFER \$0000100.00 TO ACCOUNT #199."
- If an attacker can modify the message on the wire, and can guess the format of the unencrypted message, the attacker could change the amount of the transaction, or the recipient of the funds, e.g. "TRANSFER \$0100000.00 TO ACCOUNT #227".

It is always better to **over-estimate** the ability of your attackers, rather than under-estimate them.

- Computation: attacker might have super-computers / a computer cluster
- Network: attackers might have control over the network/communication channel, they can send/drop/inject/view your packet.
- Some problems are hard ($NP \neq P$): There are some problems in algorithm that has no polynomial time solutions, e.g. Travel Salesman Problem.

The sender and receiver shared a very very long bit stream, randomly generated. Each bit are independent of each other.

Encryption is done bit-wise, with Exclusive-OR $c_i = m_i \oplus k_i$.

- ⚠️ k_i should never be re-used.

Plaintext	Key	XOR
0	0	0
1	0	1
0	1	1
1	1	0



Fact: $x \oplus y \oplus y = x$

One-time Pad

Sender:

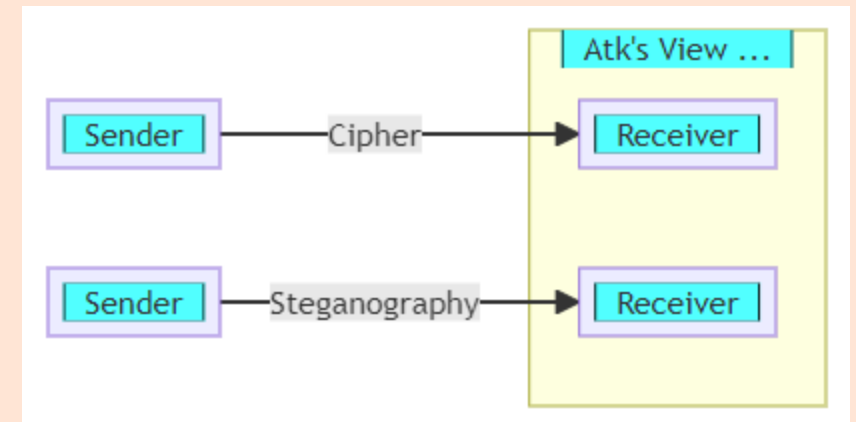
```
Plaintext: 1110 0010 1010 0101 0101  
Key:      0011 1100 1010 0000 1111  
-----  
Cipher:   1101 1110 0000 0101 1010
```

Receiver:

```
Cipher:   1101 1110 0000 0101 1010  
Key:      0011 1100 1010 0000 1111  
-----  
Plaintext: 1110 0010 1010 0101 0101
```

One-time pad is **unconditionally secure** against all kinds attacker models if key is **never reused**.

- Steganography is also known as data hiding. Instead of sending some encrypted messages, it is not letting other people to observe an encrypted message is sent.
- To get a more secure result, it is used together with data encryption.
- Adversary (government/police/wife) sees a cipher knows you are sending some secret but he/she does not know you are sending message in steganography!



Steganography

- Famous Example:
 1. Mooncake
 2. 藏頭詩
- You can/should further encrypted the message before hiding.



- Classical Crypto
- Attacker Model [in encryption]
- Brute-force attack
- Different types of security

References

[Huzaifa Sidhpurwala, A Brief History of Cryptography, Red Hat, 2019](#)