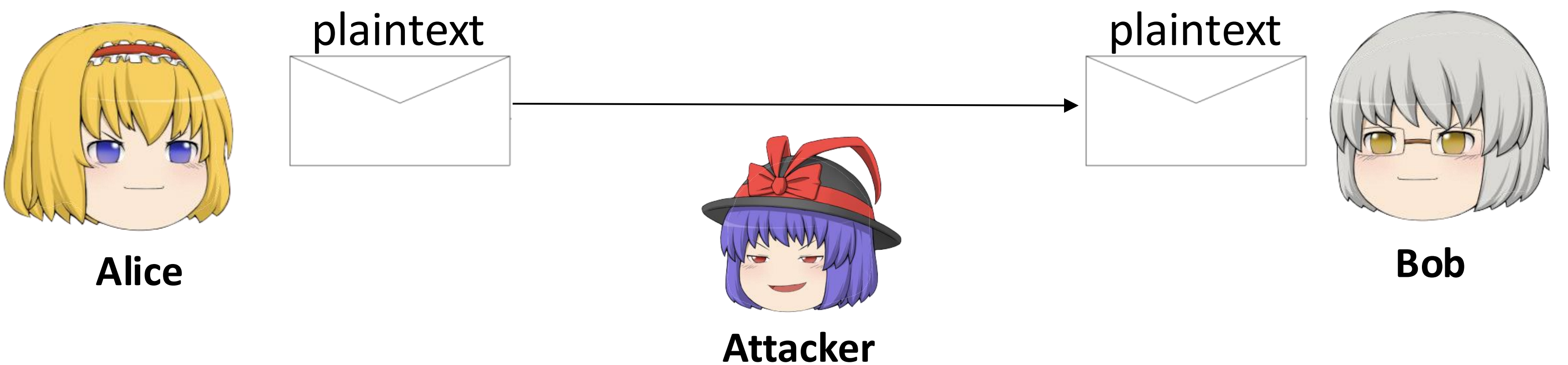
Classic Crypto

Shuai Wang

香 港 科 技 大 學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Some slides are written by Mark Stamp.

Communication could be unsafe



Cryptography

Cryptography is the science (and art?) of secret writing

A <u>cryptosystem</u> consists of:

- Key(s)
- Encryption mechanism
- Decryption mechanism

Again, How to Speak Crypto

- A *cipher* or *cryptosystem* is used to *encrypt* the *plaintext*
- The result of encryption is *ciphertext*
- We *decrypt* ciphertext to recover plaintext
- A *key* is used to configure a cryptosystem
- A *symmetric key* cryptosystem uses the same key to encrypt as to decrypt
- A *public key* cryptosystem uses a *public key* to encrypt and a *private key* to decrypt → talk about that later
- Computing plaintext from ciphertext → **hard**
- Computing plaintext from ciphertext with key → **easy**

E(plaintext,key) = ciphertext

D(ciphertext,key) = plaintext

Basic Assumption in Crypto

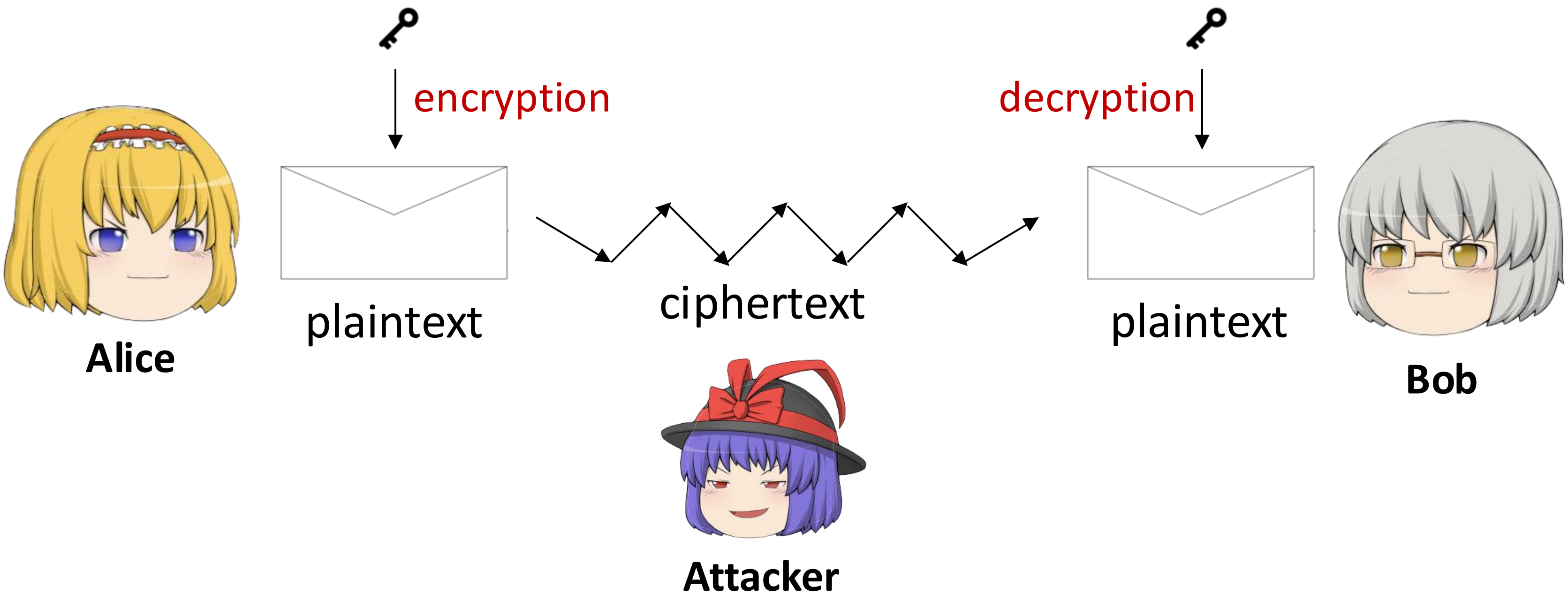
- **Kerckhoffs' Principle:**

> The key(s) of a cryptosystem should be hidden, but the mechanisms should be public.

- Why do we make such an assumption?
  - Experience has shown that secret algorithms tend to be weak when exposed
  - Secret algorithms never remain secret
  - Better to find weaknesses beforehand

Crypto as Black Box

A generic view of symmetric key crypto



What does the attacker know in this scenario?

Alice and Bob both already know some key K

Symmetric Key Cryptography

- Symmetric keys, where a single key (k) is used for **E** and **D**

$$D(E(p, k), k)) = P$$

- All (intended) receivers have access to key
- Management of keys determines who has access to encrypted data
  - But how to do that? A chicken and egg problem?
- Examples:
  - Simple substitution; codebook ← classic crypto (can be done by hand)
  - One-time pad (OTP) ← classic crypto (can be done by hand)
  - Stream ciphers ← talk later
  - Block ciphers ← talk later
- Of course, we will talk about public key crypto later.

Simple Substitution

- Plaintext: fourscoreandsevenyearsago
- Key: 3

Plaintext	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
Ciphertext	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

- Ciphertext:
  IRXUVFRUHDQGVHYHQBHDUVDJR
- Shift by 3 is “Caesar’s cipher”

Ceasar's Cipher Decryption

- Suppose we know a Caesar's cipher is being used:

Plaintext	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
Ciphertext	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

- Given ciphertext:
  VSRQJHEREVTXDUHSDQWV

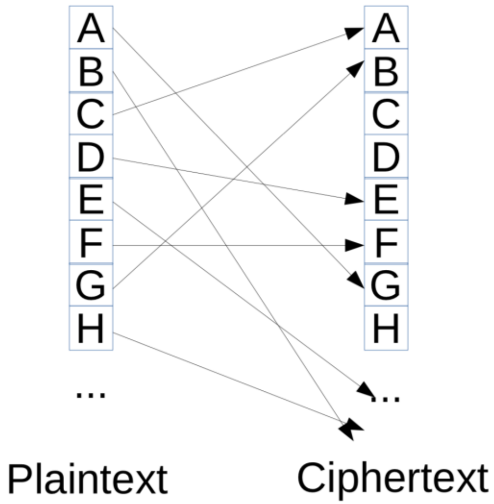
• Plaintext: spongebobsquarepants

Attack I: Try Them All

- A simple substitution (shift by n) is used
  - But the key is unknown
- Given ciphertext: CSYEVIXIVQMREXIH
- How to find the key?
- Only 25 possible keys — try them all!
- **Exhaustive key search**

Simple Substitution: General Case

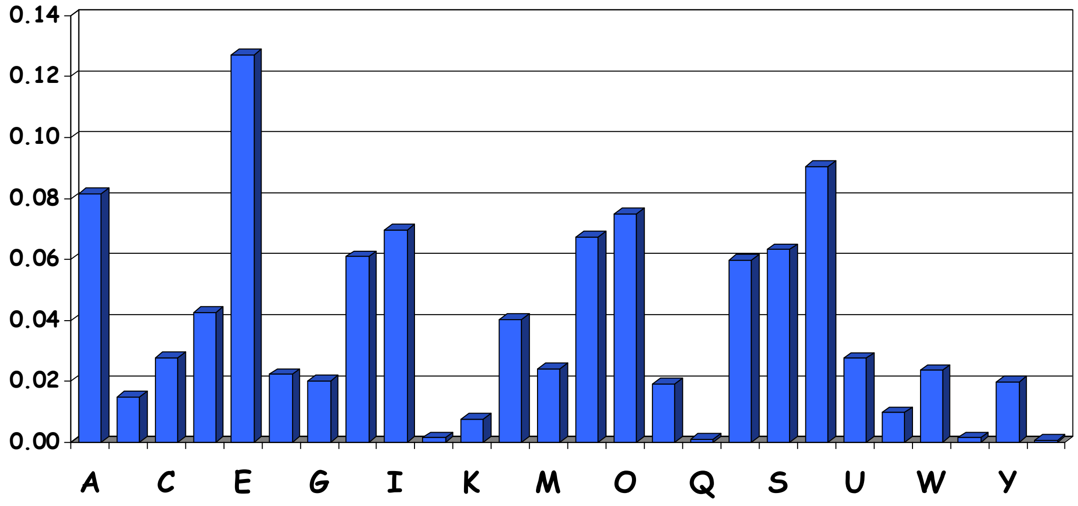
- any **permutation** of letters?



- How many variations?
  - $26! = 403291461126605635584000000 > 2^{88}$

Crypto Attack II

- Cannot try all 2^{88} simple substitution keys
- Can we be more clever?
- English letter frequency counts…

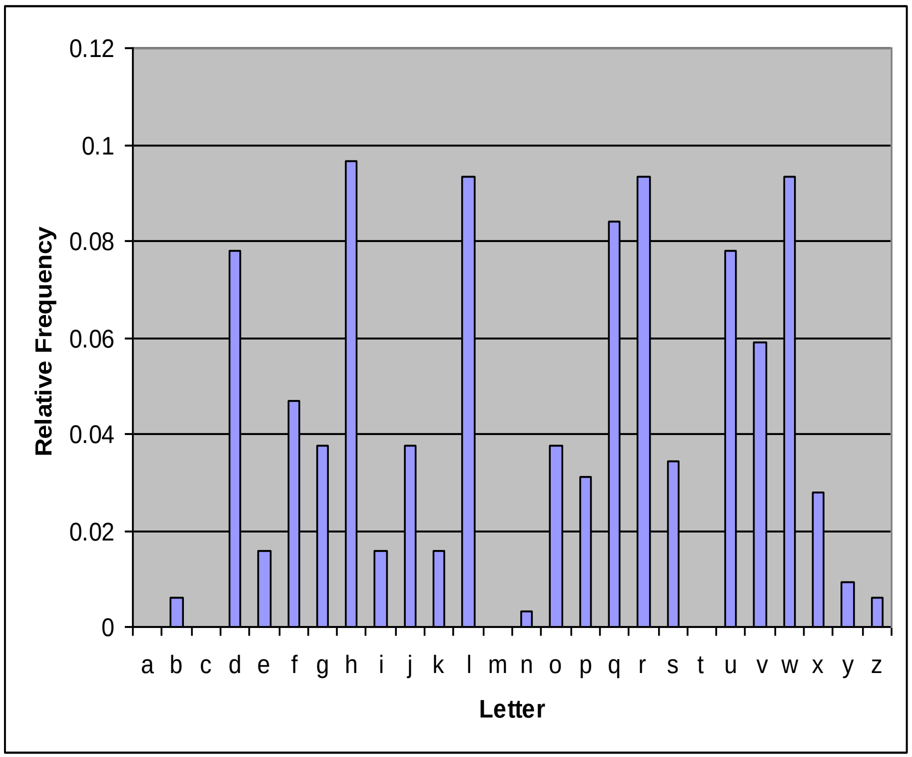


Let's see an example…

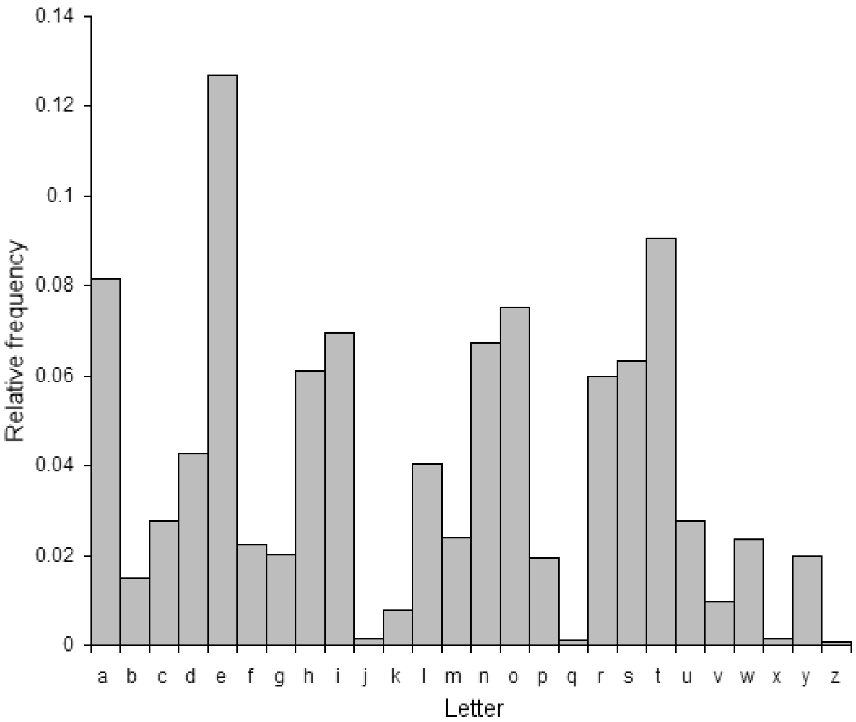
Crypto Attack II

- Vg gbbx n ybg bs oybbq, fjrng naq grnef gb trg gb jurer jr ner gbqnl, ohg jr unir whfg ortha. Gbqnl jr ortva va rnearfg gur jbex bs znxvat fher gung gur jbeyq jr yrnir bhe puvyqera vf whfg n yvggyr ovg orggre guna gur bar jr vaunovg gbqnl.

Crypto Attack II



Ciphertext distribution



English distribution

{h, l, r} → {t, a, e}?

Crypto Attack II

- Vg gbbx n ybg bs oybbq, fj**r**ng naq g**r**nef gb t**r**g gb jur**r**r j**r** ne**r** gbqnl, ohg j**r** uni**r** whfg o**r**tha. Gbqnl j**r** o**r**tva va **r**nea**r**fg gu**r** jbex bs znxvat fhe**r** gung gu**r** jbeyq j**r** y**r**ni**r** bhe puvyqe**r**a vf whfg n yvggy**r** ovg o**r**gg**r**e guna gu**r** ba**r** j**r** vaunovg gbqnl.
- It took a lot of blood, sw**e**at and t**e**ars to get to wh**e**r**e** w**e** ar**e** today, but w**e** hav**e** just b**e**gun. Today w**e** b**e**gin in **e**arn**e**st th**e** work of making sur**e** that th**e** world w**e** l**e**av**e** our childr**e**n is just a littl**e** bit b**e**tt**e**r than th**e** on**e** w**e** inhabit today.

'r' appears very frequently so likely is one of the top frequency letters (i.e., e).

Crypto Attack II

- Vg gbbx n ybg bs oybbq, fjrng naq grnef gb trg gb jurer jr ner gbqnl, ohg jr unir whfg ortha. Gbqnl jr ortva va rnearfg gur jbex bs znxvat fher gung **gur** jbeyq jr yrnir bhe puvyqera vf whfg n yvggyr ovg orggre guna **gur** bar jr vaunovg gbqnl.

- It took a lot of blood, sweat and tears to get to where we are today, but we have just begun. Today we begin in earnest the work of making sure that **the** world we leave our children is just a little bit better than **the** one we inhabit today.

Repeat this process, picking out more letters, then common words, e.g., '**the**'

Again, Principle

- Cryptosystem is **secure** if best know attack is to try all keys
  - i.e., exhaustive key search
- Cryptosystem is **insecure** if ***any*** shortcut attack is known

- **Secure**
- **No exhaustive search is feasible!**

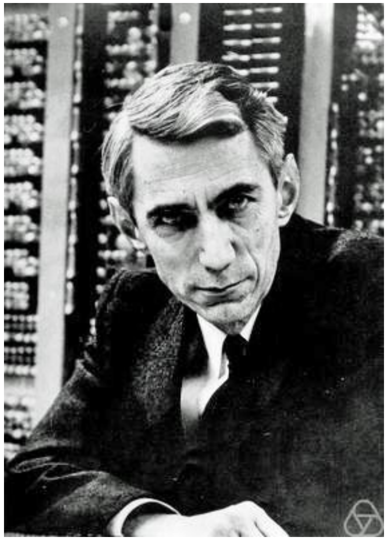
Then, is there an unbreakable cipher?

- Yes!
  - Claude Shannon proved it → One-Time Pad (OTP)!
    - The father of of Information Theory

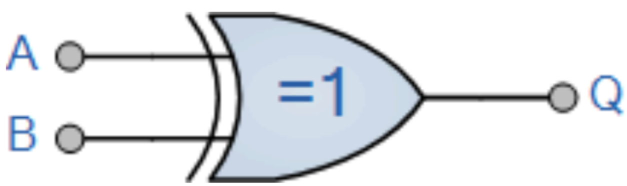
A Mathematical Theory of Communication 1948

↓

The Mathematical Theory of Communication 1949



Exclusive OR

<table>
<tr><th>Symbol</th><th colspan="3">Truth Table</th></tr>
<tr><td rowspan="5">

2-input Ex-OR Gate</td><td>B</td><td>A</td><td>Q</td></tr>
<tr><td>0</td><td>0</td><td>0</td></tr>
<tr><td>0</td><td>1</td><td>1</td></tr>
<tr><td>1</td><td>0</td><td>1</td></tr>
<tr><td>1</td><td>1</td><td>0</td></tr>
<tr><td>Boolean Expression $Q = A \oplus B$</td><td colspan="3">A **OR** B but NOT **BOTH** gives Q</td></tr>
</table>

You XOR the same thing twice, you get the same thing!

One-Time Pad: Encryption

e=000 h=001 i=010 k=011 l=100 r=101 s=110 t=111

Encryption: Plaintext ⊕ Key = Ciphertext

Salute!

	h	e	i	l	h	i	t	l	e	r
Plaintext:	001	000	010	100	001	010	111	100	000	101
Key:	111	101	110	101	111	100	000	101	110	000
Ciphertext:	110	101	100	001	110	110	111	001	110	101
	s	r	l	h	s	s	t	h	s	r

One-Time Pad: Decryption

e=000 h=001 i=010 k=011 l=100 r=101 s=110 t=111

Decryption: Ciphertext ⊕ Key = Plaintext

	s	r	l	h	s	s	t	h	s	r
Ciphertext:	110	101	100	001	110	110	111	001	110	101
Key:	111	101	110	101	111	100	000	101	110	000
Plaintext:	001	000	010	100	001	010	111	100	000	101
	h	e	i	l	h	i	t	l	e	r

One-Time Pad

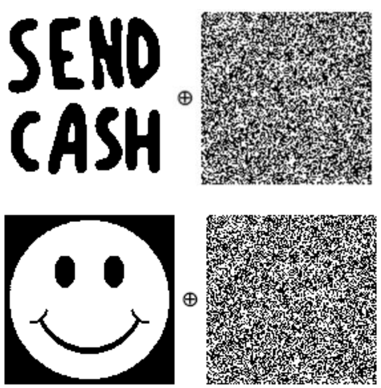
Claim the following “**key**” was used:

	s	r	l	h	s	s	t	h	s	r
Ciphertext:	110	101	100	001	110	110	111	001	110	101
“**key**”:	101	111	000	101	111	100	000	101	110	000
“Plaintext”:	011	010	100	100	001	010	111	100	000	101
	k	i	l	l	h	i	t	l	e	r

e=000 h=001 i=010 k=011 l=100 r=101 s=110 t=111

One-Time Pad

- **Provably** secure
  - Ciphertext gives **no** useful info about plaintext
    - Assume value of each bit in **k** is equally likely.
  - All ciphertexts are ***equally likely***
- BUT, only when be used correctly
  - Pad must be random, used only once



Shannon, Claude (1949). “Communication Theory of Secrecy Systems” Bell System Technical Journal 28 (4): 656–715.

One-Time Pad

- **Provably** secure
  - Ciphertext gives **no** useful info about plaintext
    - Assume value of each bit in **k** is equally likely.
  - All ciphertexts are ***equally likely***
- BUT, only when be used correctly
  - Pad must be random, used only once
  - Pad (key) is same size as message
- So, why not distribute pad instead of msg?

Shannon, Claude (1949). "Communication Theory of Secrecy Systems" Bell System Technical Journal 28 (4): 656–715.

Real-World One-Time Pad

- Project *VENOVA*
  - Russia spies encrypted messages from U.S. to Moscow in 30’s, 40’s, and 50’s
  - Thousands of messages
- Spy carried one-time pad into U.S.
- Spy used pad to encrypt secret messages
- Repeats within the “one-time” pads made cryptanalysis possible

Codebook Cipher

- Literally, a book filled with “codewords”
- Zimmerman Telegram encrypted via codebook

Februar	13605
fest	13732
finanzielle	13850
folgender	13918
Frieden	17142
Friedenschluss	17149
:	:

- Modern block ciphers are codebooks!
- More about this later…

Codebook Cipher: Additive

- Codebooks also (usually) use **additive**
- Additive — book of "random" numbers
  - Encrypt message with codebook
  - Then choose position in additive book
  - Add in additives to get ciphertext
  - Send ciphertext and additive position ← not secret
  - Recipient subtracts additives before decrypting
- Why use an additive sequence?
  - Same word encodes into different representations.