ELECS425F
Computer and Network Security

Outline - Classical cryptology

- What is cryptology?
  - Cryptography.
  - Cryptanalysis.
- Communication model.
- Secret-key cryptography.
- Early ciphers:
  - Caesar.
  - Monoalphabetic.
  - Polyalphabetic.
- Statistical cryptanalysis.

Cryptology

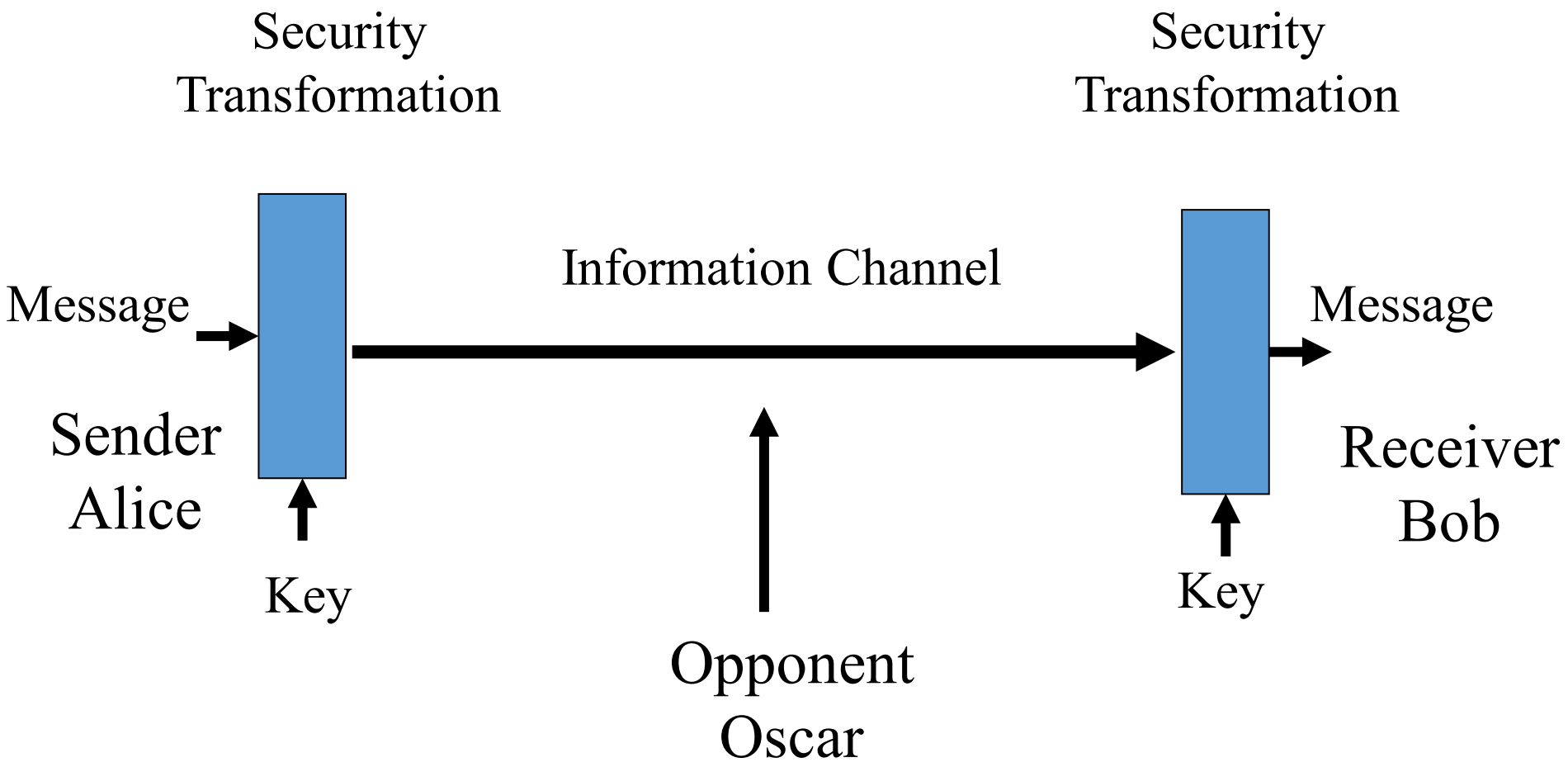
- From the Greek words:
  - *kryptos* meaning “hidden”
  - *logos* meaning “word”

  Cryptology is the art/science of secure communication. It splits into…

- **Cryptography**: Designing algorithms to ensure security: i.e. confidentiality, integrity and authenticity.

- **Cryptanalysis**: Analysing security algorithms with the aim of breaching security.

The basic secrecy channel



The basic secrecy channel

- The channel can be a *communication channel* or a *storage channel*.
- Sender (A for Alice) wants to send a message X to the Receiver (B for Bob), through this channel, such that the opponent/enemy/intruder O (O for Oscar) cannot access X.
- Alice applies a transformation, known as **encryption**, to X, referred to as the **plaintext**, to produce a garbled message Y, referred to as the **ciphertext** (or cryptogram).
- Bob applies another transformation, known as **decryption**, to Y to obtain the plaintext again.

Key dependence

- The transformations are not fixed, they are **key** dependent. The **key K** controls the transformation and is known only by Alice and Bob. The key is secret.
- Encryption and decryption are sometimes referred to as enciphering and deciphering, respectively.

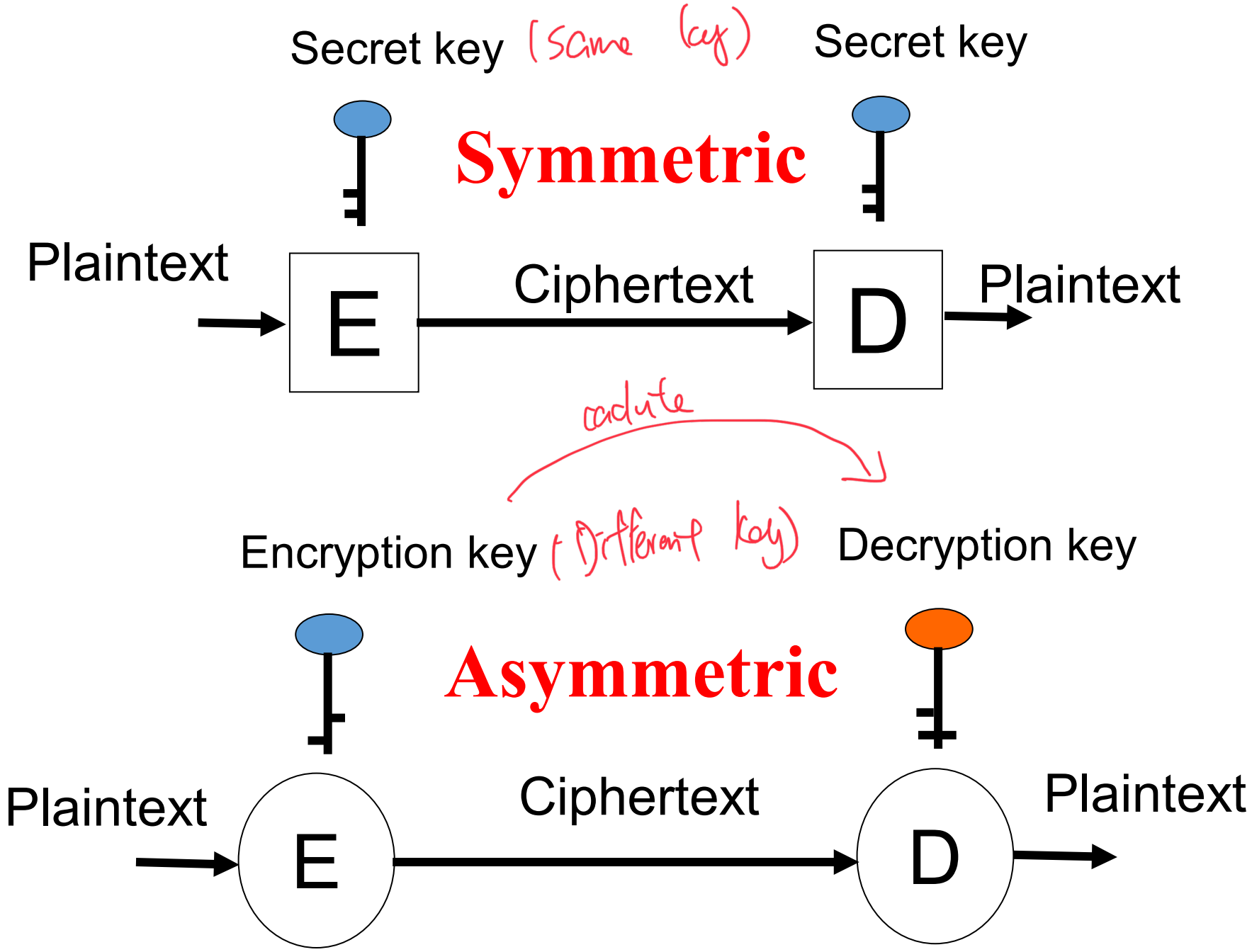
- Note that if a transformation **does not** depend on a key, it is referred to as **encoding**, with the inverse transformation being referred to as **decoding**.
  - Morse code.
  - ASCII code.

Secret-key cryptography

- In secret key cryptography the participants all have only secret key information.
- Basically this means there is no role, other than as an attacker, for anybody who doesn’t hold a secret key of some sort.
- We shall later look at public-key cryptography, where persons without secret-keys can play a role (other than an attacker) in the cryptosystem.

Symmetric and asymmetric encryption

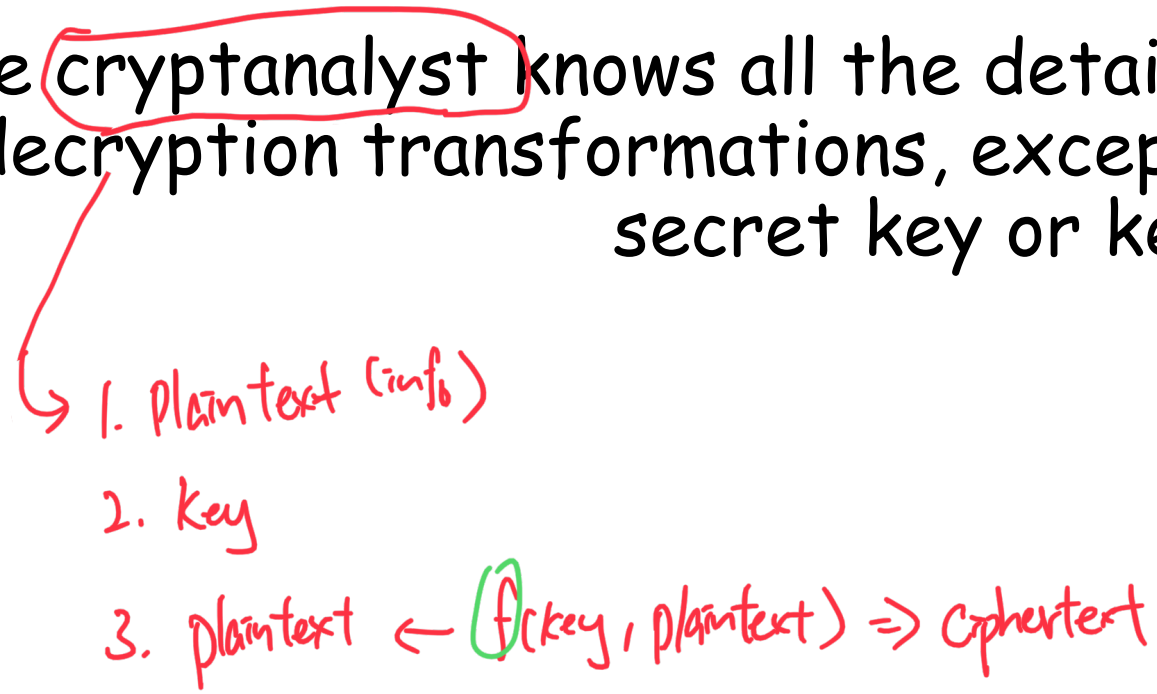
- In classical cryptography the encryption and decryption keys are the same, this is an example of a symmetric encryption scheme.
- In asymmetric encryption the encryption and decryption keys are different, this is the case in public-key encryption.


Secret key
(same key)
Secret key
Symmetric
Plaintext
E
Ciphertext
D
Plaintext
conducte
Encryption key
(Different key)
Decryption key
Asymmetric
Plaintext
E
Ciphertext
D
Plaintext

Kirchoff's Law

- This is the main assumption of cryptology.

The cryptanalyst knows all the details of the encryption and decryption transformations, except for the value of the secret key or keys.



Some possible attacks

- Oscar is trying to decrypt a particular ciphertext, and possibly (although it is harder in general) to figure out the key.
- *Ciphertext only*: Oscar knows Y. Alice and Bob don't want Oscar to figure out what either X or K is.
- *Known plaintext*: Oscar knows some X-Y pairs. Alice and Bob don't want Oscar to figure out what K is, or the correspondence between other X-Y pairs.
- *Chosen plaintext*: Oscar is allowed to choose some plaintexts (X's), and receives the corresponding ciphertexts (Y's).
- *Chosen ciphertext*: Oscar is allowed to choose some ciphertexts (Y's), and receives the corresponding plaintexts (X's).
- Some combination of these.

Easier

Epochs in cryptology

- Non-scientific cryptography: from antiquity until 1949.

  Cryptology was more an “art” than a science.

- Scientific cryptography starts with Shannon's paper :

  Communication theory of secrecy systems (1949).

  This was based on Shannon’s 1948 paper in which he had founded information theory.

- Cryptologic research really took off in 1976 with the paper *New directions in cryptography*, by Diffie & Hellman.

  They showed that secret communication is possible without a shared key.

Early ciphers

- Studying some early ciphers is useful because the empirical principles developed through their use are applied in the design and analysis of modern ciphers.
- Caesar cipher: (Julius Caesar 2000 years ago).
  - Every letter is replaced by the letter “three to the right” in the alphabet, where this operation is cyclic. A→D, B→E, … X→A, Y→B, Z→C
  - For example: CABBAGE → FDEEDJH
  - The generalised Caesar (or shift) cipher allowed the 3 to be replaced by an value between 1 and 25 inclusive.

Monoalphabetic ciphers

One letter is replaced by another fixed letter

- Also known as simple substitution ciphers, each letter of the plaintext alphabet is replaced with an element of the ciphertext alphabet.
- The substitution alphabet is the *key.*
- Consider that the plaintext and ciphertext alphabets are both the English alphabet

plaintext →	a	b	c	d	e	…	x	y	z
ciphertext →	F	G	N	T	A	…	K	P	L

Substitution table is key

a		b	a	d		d	a	y
F		G	F	T		T	F	P

- Example: Consider the plaintext and ciphertext alphabets to be the set of binary strings of length 3.

Key

↳ K=

000	001	010	011	100	101	110	111
101	111	000	110	010	100	001	011

plaintext

ciphertext

- Plaintext: 100 101 111
- Ciphertext: 010 100 011

Unique decryption: One-to-one.

- In order for decryption to be unique, we need one-to-one mappings. Consider…

a	b	c	d	e	…	x	y	z
F	G	N	G	A	…	K	P	L

a		b	a	d		d	a	y
F		G	F	G		G	F	P

- Decryption: a bad day, a dad day, a bab day, a dab day, a bad bay, a dad bay, a bab bay, a dab bay.

The same

Security: Monoalphabetic ciphers

- To decipher a ciphertext we need to know the substitution alphabet (key), or at least the subset of the key corresponding to those symbols which appear in the ciphertext.
- One can use an *exhaustive key search*, to find the full key. This means use each key to decipher the ciphertext and accept the one that produces a meaningful plaintext.
- For an alphabet of size N, the number of possible keys is N!.
- The key is N elements long.

$$N! \approx \sqrt{2\pi N}\left(\frac{N}{e}\right)^N$$

- For the English alphabet there are 26!≈4*10^(26) keys.

Weak keys

- Not every substitution alphabet is suitable.
- For example, a possible substitution which doesn't hide the message very well at all is:

a	b	c	d	e	…	x	y	z
X	B	C	D	E	…	A	Y	Z

Same replacement

c	o	m	p	u	t	e	r
C	O	M	P	U	T	E	R

- In most ciphers there are some *weak keys*.

Properties of keys

- Keys must be easy to remember, a long random string is difficult to remember.
- The key set must be large enough so that *exhaustive key search* is not easy.
- To reduce the number of keys we may restrict ourselves to an indexed subset of all possible substitutions and use the index to identify the substitution that is used.
- In this case the cipher algorithm is the collection of substitutions, and the key is the index.

- Additive and multiplicative ciphers are examples of such indexed constructions.

Additive ciphers

- Also known as *translation ciphers,* the substitution alphabet is obtained by shifting the plaintext alphabet by a fixed value. The amount of the shift is the *key*.
- For example with the key of 3 we have

a	b	c	d	e	…	x	y	z
D	E	F	G	H	…	A	B	C

which is the substitution alphabet for the Caesar cipher.

Modular addition for additive ciphers

- Additive ciphers can be described by modular addition.

a	b	c	d	e	…	x	y	z
0	1	2	3	4	…	23	24	25

- Plaintext character X, ciphertext character Y, shift (*key*) Z. Each of X,Y,Z is an element of the set {0,1,2,3,…,25} and they are related by $Y = X \oplus Z$, where $\oplus$ denotes addition modulo 26. There are 26 keys, so bits are needed to represent the key.
- This is a key space, hence *exhaustive key search* is a feasible attack against additive ciphers.

(only 26 keys) ⇓ Z : only 26 possible value

⇓ only need to try 26 time to decrpty

$$Y = X \oplus Z \ , \{0, 1, 2 \cdots , 24, 25\}$$

Example: Z is the key, Z=12

⇓

plaintext = ZOO = 25, 14, 14

Ciphertext = $\{25+12, 14+12, 14+12\}$ % 26

⇓

11, 0, 0

Multiplicative Ciphers

- The plaintext alphabet will again be taken as the set {0,1,…,25}.
- The ciphertext alphabet can be determined using modular multiplication. We multiply each plaintext alphabet by a constant value, which is the *key* for this cipher.
- Using similar notation to that for additive ciphers we write

$$Y = X \otimes Z$$ ← key

where $\otimes$ represents multiplication modulo 26.

Multiplicative Ciphers

- Not all possible numbers for the key Z will result in a one-to-one mapping. The set of keys is therefore a subset of {0,1,…,25}.
- Consider Z=2

  $1 \otimes 2 = 2$ b → c

  $14 \otimes 2 = 2$ o → c

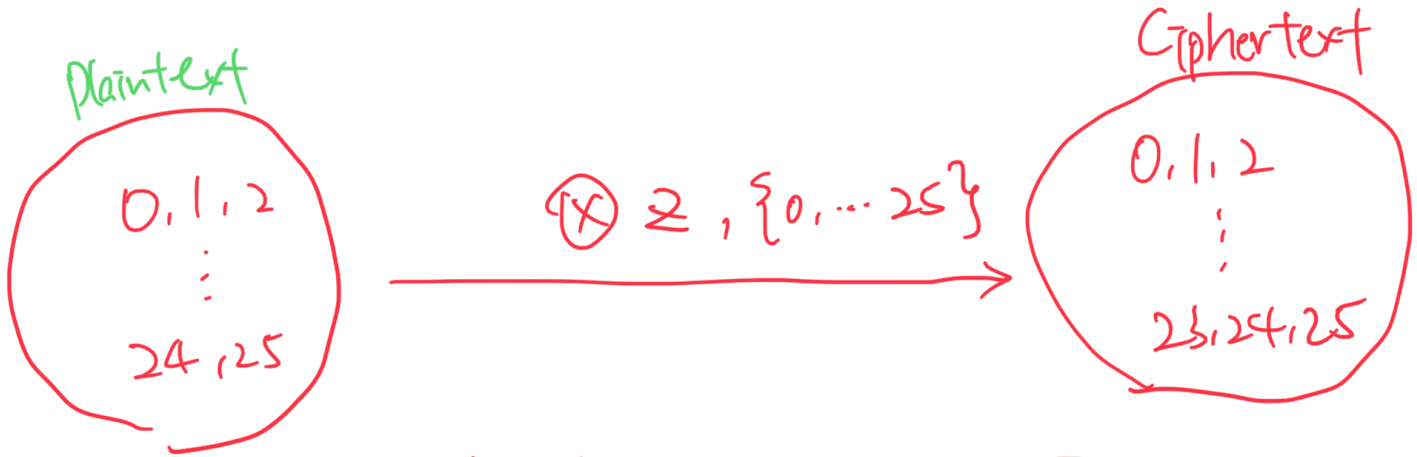
  } Decryption problem

- For all even numbers, and 13, the mapping not one-to-one. Hence the number of keys is 12, we need only 4 bits to represent the key.
- Again exhaustive key search can easily break the cipher (find the secret key).

↓

at most try 12-time

Not all possible numbers for key $z \Rightarrow$ one-to-one mapping



* not one to one mapping

⇓

decryption

have problem

Affine ciphers

- To increase the number of keys we can combine additive and multiplicative ciphers to obtain *affine* ciphers.

$$Y = \alpha \otimes X \oplus Z$$ (Add + multiple)

where $X, Y, Z \in \{0, 1, \ldots 25\}$

and $\alpha \in \{1,3,5,7,9,11,15,17,19,21,23,25\}$

- Number of keys:
  - 12*26=312.

  Still too small!

Key phrase based ciphers

- The next thing we can try and do is use a phrase as the key (index). We can also specify a starting letter for the translation alphabet.

- This increases the size of the key space but makes the key (index) easy to remember.

key

- Phrase: bubble bath
- Starting letter: e

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

- **This cipher is significantly more resistant to exhaustive key search, but …**

Statistical cryptanalysis

- **… it is insecure against statistical cryptanalysis.**

- Statistical properties of the plaintext language can be used to cancel many keys in one step and enable the cryptanalyst to find the key without trying all of them.

- Statistical analysis relies on there being a relationship between the statistical properties of the plaintext and the statistical properties of the ciphertext, since we assume the attacker has only the ciphertext.

Statistics of the English language

- The letters can be grouped according to the frequency with which they occur.

I	e
II	t a o i n s h r
III	d l
IV	c u m w f g y p b
V	v k j x q z

- The frequency of pairs of consecutive letters (bigrams) and triples of consecutive letters (trigrams) are important clues to cryptanalysts, as are spaces between words.
- Frequent bigrams:
  th, he, in, er, an, re, ed, } Frequent in 2 words
  on, es, st, en, at, to

- Frequent trigrams:
  the, ing, and, her, ere, ent } Frequent in 3 words
  tha, nth, was, eth, for, dth

- It is important to realise that frequency counts only provide clues to the actual key used. The distribution will differ from sample to sample.

- What is the plaintext associated with the following ciphertext?
- According to Kirchoff's Law the encryption algorithm is known to the attacker; it is **key phrase substitution**.

YKHLBA JCZ SVIJ JZB TZVHI JCZ VHJ DR IZXKHLBA VSS
RDHEI DR YVJV LBXSKYLBA YLALJVS IFZZXC CVI
LEFHDNZY EVBLRDSY JCZ FHLEVHT HZVIDB RDH JCLI
CVI WZZB JCZ VYNZBJ DR ELXHDZSZXJHDBLXI JCZ
XDEFSZQLJT DR JCZ RKBXJLDBI JCVJ XVB BDP WZ
FZHRDHEZY WT JCZ EVXCLBZ CVI HLIZB
YHVEVJLXVSST VI V HZIKSJ DR JCLI HZXZBJ
YZNZSDFEZBJ LB JZXCBDSDAT EVBT DR JCZ XLFCZH
ITIJZEI JCVJ PZHZ DBXZ XDBILYZHZY IZXKHZ VHZ BDP
WHZVMVWSZ.

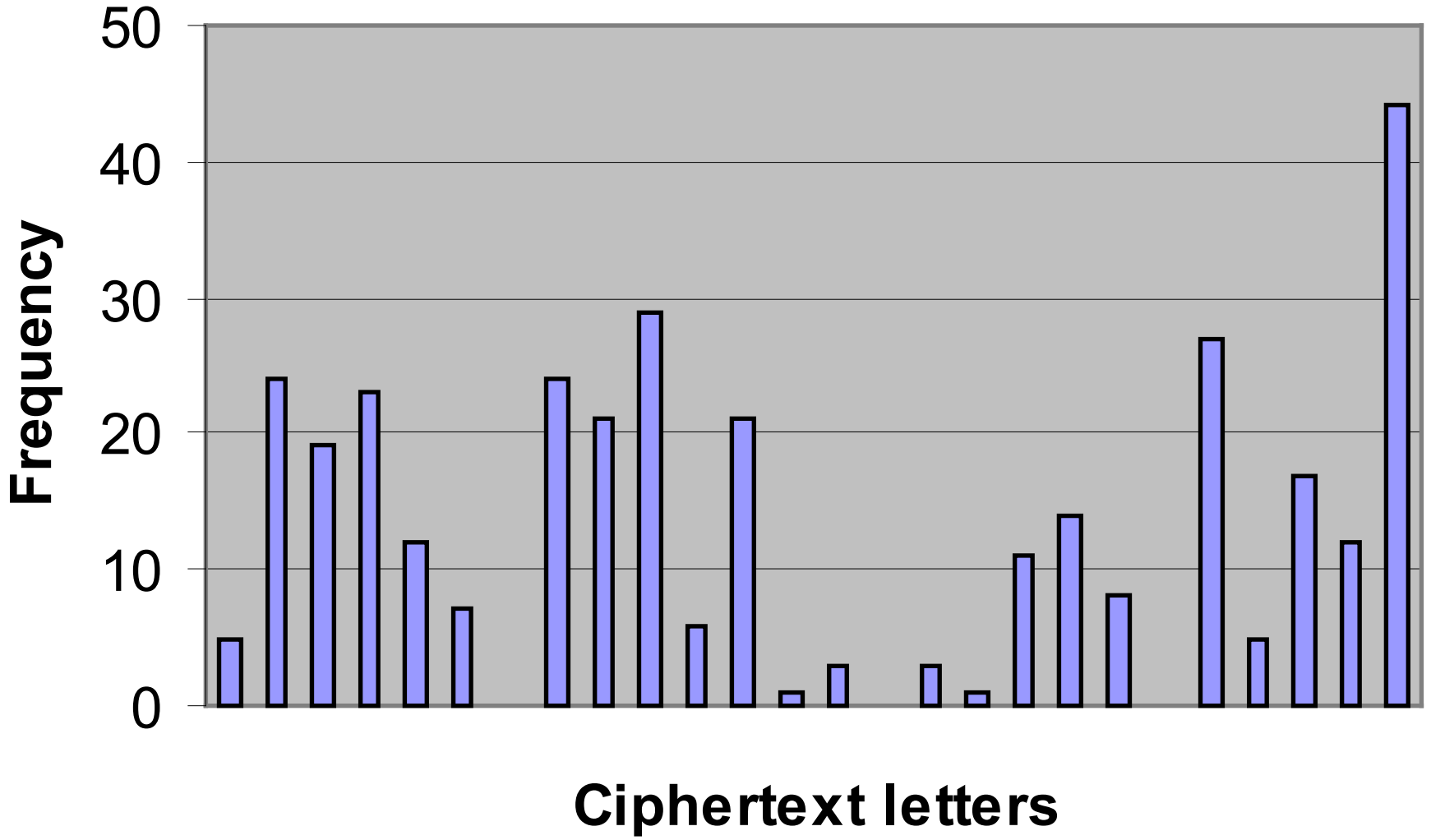
Breaking the cipher

- There are 337 letters, with frequencies:

A	B	C	D	E	F	G	H	I
5	24	19	23	12	7	0	24	21
J	K	L	M	N	O	P	Q	R
29	6	21	1	3	0	3	1	11
S	T	U	V	W	X	Y	Z	
14	8	0	27	5	17	12	**44** → e	

- This suggests we should first try **Z** being the encryption of **e**.

Ciphertext distribution



- There are 8 **JCZ** in the ciphertext, this is almost certainly **the** in the plaintext.
- The single letters will generally be **i** or **a**. In this case there is a single letter word **V** in the ciphertext.
- The word **JZB** in the ciphertext can be identified by looking at the word **teB** and noting that **B** occurs in the second frequency group for this ciphertext.
  - Some of those letters (t a o i n s h r) have already been identified.

- After a few of these kind of steps we can build up a preliminary mapping such as:

a	b	c	d	e	f	g	h	i
V				Z			C	L
j	k	l	m	n	o	p	q	r
				B	D			H
s	t	u	v	w	x	y	z	
I	J							

- Most of the time we would probably be safe to guess **f** as the starting position. With that assumption we can fill b,c,d → W,X,Y!

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

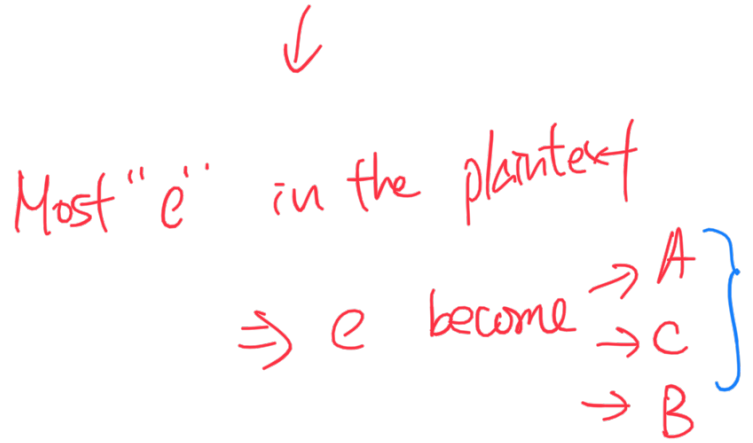
YKHLBA JCZ SVIJ JZB TZVHI JCZ VHJ DR IZXKHLBA VSS RDHEI DR YVJV LBXSKYLBA YLALJVS IFZZXC CVI LEFHDNZY EVBLRDSY JCZ FHLEVHT HZVIDB RDH JCLI CVI WZZB JCZ VYNZBJ DR ELXHDZSZXJHDBLXI JCZ XDEFSZQLJT DR JCZ RKBXJLDBI JCVJ XVB BDP WZ FZHRDHEZY WT JCZ EVXCLBZ CVI HLIZB YHVEVJLXVSST VI V HZIKSJ DR JCLI HZXZBJ YZNZSDFEZBJ LB JZXCBDSDAT EVBT DR JCZ XLFCZH ITIJZEI JCVJ PZHZ DBXZ XDBILYZHZY IZXKHZ VHZ BDP WHZVMVWSZ.

during the last ten years the art of securing all forms of data including digital speech has Improved manifold the primary reason for this has been the advent of microelectronics the complexity of the functions that can now be performed by the machine has risen dramatically as a result of this recent development in technology many of the cipher systems that were once considered secure are now breakable.

What does this tell us?

- A cipher system should not allow statistical properties of the plaintext language to pass to the ciphertext.

- The ciphertext generated by a “good” cipher system should be statistically indistinguishable from random text.



Solutions

- Polyalphabetic ciphers.
  - Vigenere ciphers.
- Statistical analysis of polyalphabetic ciphers.
  - Kasiski method
  - index of coincidence

Polyalphabetic ciphers

- A ciphertext character represents more than one plaintext character. This must be done in a way that allows the plaintext to be recovered.
  - For example, if **B** represents **n** and **t** we need to know when to decipher it as **n**, and when as **t**.
- A polyalphabetic cipher uses a sequence of substitution alphabets. If this sequence repeats after *p* characters, we say it has *period p*.

Monoalphabetic vs Polyalphabetic cipher

Monoalphabetic : 1 plaintext letter → one ciphertext letter

polyalphabetic : 1 plaintext letter → more than one ciphertext letter

- Questions: one-to-one mapping?
- Polyalphabetic is not one-to-one mapping. yes

°. ? The decryption will have more than one possible results? It is NOT good?

It depends on the cipher system. There are some methods to decrypt to a unique result.

The Vigenère cipher

- This uses 26 substitution alphabets, and a *key word* or *key phrase*.
- The substitution alphabets are cyclically related.

a	b	c	d	e	…	x	y	z
0	1	2	3	4	…	23	24	25

Key =

X	w	o	l	l	o	n	g	o	n	g
	22	14	11	11	14	13	6	14	13	6
Z	c	a	f	e	c	a	f	e	c	a
	2	0	5	4	2	0	5	4	2	0
Y	24	14	16	15	16	13	11	18	15	6
	Y	O	Q	P	Q	N	L	S	P	G

★ As the key is multiple value ⇒ Cipher text of the same word ⇒ may not the same ies.

$Y = X \oplus Z$, remember addition modulo 26.

l + f ⇒ Q , l + e ⇒ P } l → Q, P

Analysing Vigenère ciphers

one plaintext letter -> more than one ciphertext letter
The distribution of letters are changed. It is not followed the English language statistical properties

- We analyse these ciphers by
  1. Finding the period. ① Cafe [key] Period = 4
  2. Breaking the ciphertext into components each obtained from a single substitution alphabet. Break into 4 groups (componant)
  3. Solving each component using techniques discussed for monoalphabetic ciphers (statistical analysis)

↓

In each group

Component 1 : All add "c"
2 : All add "a"
3 : Add add "f"
4 : Add add "e"

key point

Find the period

Finding the period: The Kasiski method

- We observe that two identical plaintexts will be encrypted to the same ciphertext if their occurrence is *m* positions apart, where *m=0 mod p*, i.e. when *m* is a multiple of the period p.

X	w	o	l	l	…	…	w	o	l	l
Z	c	a	f	e	…	…	c	a	f	e
y	Y	O	Q	P	…	…	Y	O	Q	P

Finding the period: The Kasiski method

- We search the ciphertext for repeated segments, and measure the distances between such repeated segments. It is *likely*, but not guaranteed, that these distances will be a multiple of the period.

Finding the period: The index of coincidence.

- Consider a string of length *n*, where each element is a letter from the English alphabet $\boldsymbol{X}=x_1x_2\ldots x_n$

  $\lambda \in \{A,B,C,\ldots,Z\}$, and f_λ *frequency of* λ

$$IC(x) = \frac{\sum_{\lambda=A}^{Z} f_\lambda (f_\lambda - 1)}{n(n-1)}$$

ref (language) $\approx$

Random IC and English IC

- If X were a random string over the English alphabet we would expect the probability of each letter occurring to be the same, so that

  IC ≈ 1/26 ≈ 0.038

- If X is an English language text then we would expect, for p(λ) the probability of a particular letter being λ,

$$IC(x) \approx \sum_{\lambda=A}^{Z} p(\lambda)^2 \approx 0.065$$

- The two values 0.065 and 0.038 are sufficiently far apart that we will often be able to determine the correct keyword length.

English language letter probabilities

A	B	C	D	E	F	G	H	I
0.082	0.015	0.028	0.043	0.127	0.022	0.020	0.061	0.070
J	K	L	M	N	O	P	Q	R
0.002	0.008	0.040	0.024	0.067	0.075	0.019	0.001	0.060
S	T	U	V	W	X	Y	Z	
0.063	0.091	0.028	0.010	0.023	0.001	0.020	0.001	

IC for monoalphabetic ciphers

- For monoalphabetic ciphers the index of coincidence is the same for the plaintext as for the ciphertext.
- The IC for the ciphertext of a English text plaintext encrypted under a monoalphabetic cipher, will therefore be approximately 0.065, the same as English text.
- We can use this test to find the period.

Example

- Suppose we know the following ciphertext was generated using the Vigenère cipher. We need to find the period.

ooon yho cshexjlg nz xfksledcl ky luw h dltqupduw jjhrolww
jshehj dwns df llwpslpchlodf plzdv yohrh gm audwwyg uyg
vvqnzuss zyghd wfirustg qp djl hbp psqcl augmsmdlguof
pgmjontrf jshehj dwnslf zihj zaav u qxds fuyjw vt jcrxlgmtrfhz
xtvupdftqwz whnomkwhr vgts ftnw soq aksyaunb zlofek
jlzuehv wfiqhkzwiyv loon luw bbcbxw ac mfqq ds uch ssgi
ekw solrhka ihohjnfuoxsas wpqllf qtwz avy hlvlgn cdfns iq
sjvullpk hbx hllowh zxj usqwb xvfgpg ...

To find the period we guess and check the IC of the components. Lets try *p*=2.

oo yo sejg z fsec k lw dtudw jrlw sej ws f lplcld pzv or g adwg y vqzs zgd
frsg p j hp sc agsdgo pmotf sej wsf ij av qd fyw t cxgtfz tudtw wnmwr gs
tw o asan zoe jzev fqkwy lo lw bbw c fq s c sg ew ork iojfoss plf tz v hvg
cfs q julk b hlw zj sw xfp uzpw t ck b dabp oh ew flwh zwhl l cqj rqfb
fvfcvop vqeg glfb ew ilawd zcr ls zaz nwqd g v aqwl sr tdund qpug sl g
yaopq wvv c s shg ybclc z fk yosr sr y ...

on h chxl n xkldl y u h lqpu jhow jhh dn d lwsphof ld yhh m uwy ug vnus
yh wiut q dl b pql ummluf gjnr jhh dnl zh za u xs uj v jrlmrh xvpfqz hokh
vt fn sq kyub lfk luh wihziv on u bcx a mq d uh si k slha hhnuxa wql qw
ay lln dn i svlp hx loh x uqb vgg vfj v uw hx flwv pb k nywz jwuco v zh h
ylpa qladbn hbulu jqpa k ogqay wyok o mfh mluy w ay kzwo u vrvcd
zcql nd p cwtno shh a nh jch lydph i l ersxh u u ...

Component 1: IC = .047082
Component 2: IC = .050946

p=2	.047028	.050946					
p=3	.050229	.054603	.056575				
p=4	.047511	.051445	.047226	.051843			
p=5	.045771	.045228	.048290	.045118	.043533		
p=6	.070324	.068399	.065509	.066987	.063925	.070450	≈ 0.065
p=7	.042861	.045826	.045078	.046161	.047380	.047236	.047230

- Therefore *p=6* seems most likely.
- The keyword is ***should***.

Transposition Ciphers

- The previous ciphers are all **substitution** ciphers
- now consider classical **transposition** or **permutation** ciphers
- these hide the message by rearranging the letter order without altering the actual letters used

Transposition Cipher

- Write letters of message out in rows over a specified number of columns
- Then reorder the columns according to some key before reading off the rows

```
Key: 4312567
Plaintext:
attackpostponeduntiltwoam
```

4	3	1	2	5	6	7
A	T	T	A	C	K	P
O	S	T	P	O	N	E
D	U	N	T	I	L	T
W	O	A	M	X	Y	Z

```
Ciphertext: TTNAAPTMTSUOAODWCOIXKNLYPETZ
```

Summary - Classical cryptology

- What is cryptology?
  - Cryptography.
  - Cryptanalysis.
- Communication model.
- Secret-key cryptography.
- Early ciphers:
  - Caesar.
  - Monoalphabetic.
  - Polyalphabetic.
- Statistical cryptanalysis.