Lecture 3: Classical Encryption

TTM4135

Relates to Stallings Chapter 3

Spring Semester, 2025

Motivation

Apart from their intrinsic interest we study historical ciphers in order to:

- establish basic notation and terminology;
- introduce basic cryptographic operations that are still used as building blocks for modern cryptographic algorithms;
- explore the typical attacks and adversary capabilities that our cryptosystems should defend against.

Outline

Outline

Terminology

The science of *cryptology* is dual-faceted, comprising of:

- *cryptography* - the study of designing cryptosystems, and
- *cryptanalysis* - the study of breaking cryptosystems.

In practice both facets are usually studied together.

Confidentiality and authentication

- Cryptography is the science of *secret writing*. It concerns transformations of data which depend on a secret called the *key*.
- Cryptography can be used to provide *confidentiality* and to provide *authentication* (or *integrity*).
- When used for confidentiality a key is needed in order to *read* the message.
- When used for authentication a key is needed in order to *write* the message.

Cryptosystems

A cryptosystem consists of:

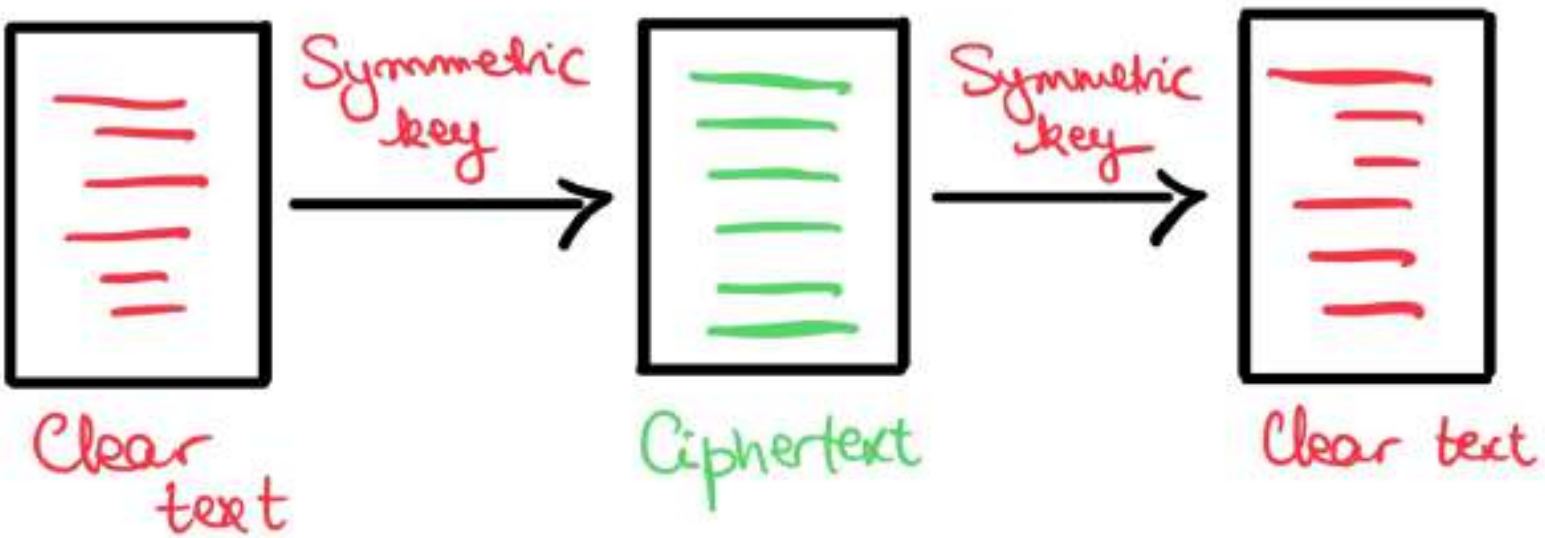
- a set of plaintexts (holding the original message);
- a set of ciphertexts (holding the encrypted message);
- a set of keys;
- a function which transforms plaintext into ciphertext (called *encryption*);
- an inverse function which transforms ciphertext back into plaintext (called *decryption*).

The encrypted message is the ciphertext.

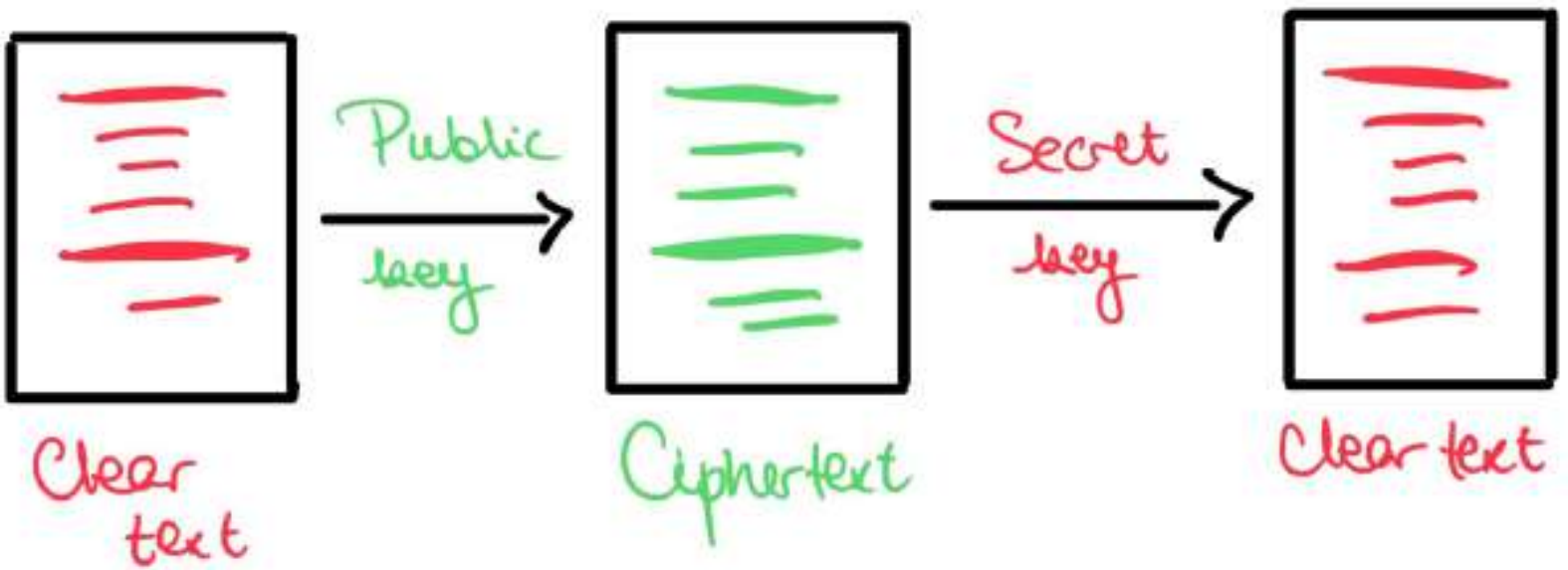
Symmetric and asymmetric cryptography

- Symmetric key cipher (also known as secret key cipher):
  - Encryption and decryption keys known only to the sender and receiver.
  - Requires a secure channel for transmission of the cryptographic key.
- Asymmetric key cipher (also known as public key cipher):
  - Each participant has a public key and a private key.
  - May allow for both encryption of messages and creation of digital signatures.
  - We study public key ciphers in a later lecture.

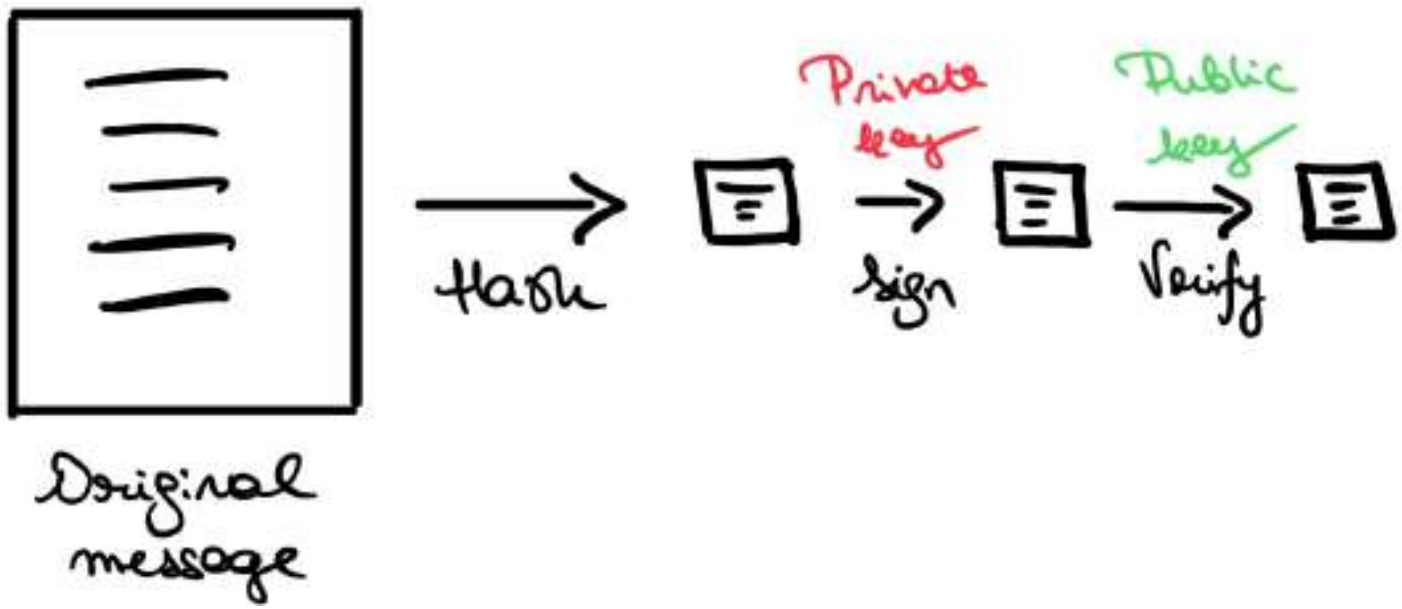
Symmetric Key encryption



Public Key encryption



Digital signatures



Notation for symmetric encryption algorithms

E	=	Encryption function
D	=	Decryption function
M	=	Message or Plaintext
C	=	Cryptogram or Ciphertext
K	=	Shared secret key

Encryption: $C = E(K, M)$

Decryption: $M = D(K, C)$

Outline

Methods of cryptanalysis

There are many methods available to an adversary who wishes to break a cryptosystem. In general we need to consider the following.

- What resources the adversary has available. This includes the computational capability of the adversary. It may also include access to various inputs and outputs of the system.
- What the adversary is aiming to achieve. This may be to retrieve the whole of the secret key or it may be as little as distinguishing two message (such as *yes* or *no*.)

Exhaustive key search

- The most basic method of attack is *exhaustive key search*, also called brute-force attack, in which the adversary tries all possible keys.
- We cannot prevent this attack so all cryptosystems must have enough keys to make exhaustive search too difficult computationally.
- Note that:
  - it may be possible for the adversary to find the key without trying exhaustive search;
  - the adversary may be able to break the cryptosystem without finding the key.

Prevention of exhaustive key search is a *minimum standard*.

Attack classification

1. **Ciphertext Only attack:** The attacker has available only the intercepted ciphertext.
2. **Known Plaintext attack:** The attacker knows a small amount of plaintext and its ciphertext equivalent.
3. **Chosen Plaintext attack:** The attacker can obtain the ciphertext equivalent of some plaintext which can be selected by the attacker; i.e. the attacker has an “inside encryptor” available.
4. **Chosen Ciphertext attack:** The attacker can obtain the plaintext equivalent of some ciphertext which can be selected by the attacker; i.e. the attacker has an “inside decryptor” available.

Which attacks should be prevented?

- A cryptosystem which can be practically attacked using only ciphertexts, is generally considered to be highly insecure.
- The modern standard is that a cryptosystem should be secure against chosen plaintext and chosen ciphertext attacks.
- History shows that chosen ciphertext attacks can often be practical to set up for an attacker.

Kerckhoffs' Principle

This principle says that an attacker has complete knowledge of the how the cryptosystem works. The decryption key is the only unknown to the attacker.

- History has shown that Kerckhoff's Principle is a reasonable assumption.
- Using a secret, non-standard algorithm can cause severe problems. This would be an example of *security through obscurity*.

Outline

Alphabets

- In our historical ciphers, we need to define the *alphabet* for the plaintext and ciphertext (they are usually the same).
- We will use the Roman alphabet: A, B, C, . . . , Z. Sometimes we include the space, sometimes we use both upper and lower case, sometimes we include punctuation.
- For some ciphers we map the alphabet to numbers and usually assume: $A = 0, B = 1, C = 2, \ldots, Z = 25$. If the space is included we map it to the number 26.
- Note that a real-world attacker needs to work out the alphabet.

Statistical attacks

- To a large extent the statistical attacks depend on using the redundancy of the plaintext. Can you read this?

```
TDY S VRY CLD
```

- In written text considerable information is available from the distribution of single letters, digrams (double letters) and trigrams (triple letters) to help in the attack.
- The exact statistics of a language will vary according to what sample is taken.

Sample statistics for English

- The following statistics give a typical distribution of English text. This particular distribution was calculated on a text passage of 143000 characters.
- In order to simplify the statistics, the text is restricted to a plaintext alphabet of 27 characters: $\{$ABCDEFGHIJKLMNOPQRSTUVWXYZ$\nabla\}$. Here ∇ represents the space character.
- The proportions shown are relative; for example the ∇ character accounts for 14.6% of all characters while 2.3% of all digrams are the E∇ digram.

Frequency (percentage) of characters and digrams

∇	14.6	A	7.0	H	2.6	V	1.3	Z	0.1
E	10.1	R	5.2	M	2.5	B	1.3	J	0.1
N	7.8	S	5.1	P	2.5	Y	0.8	Q	0.1
T	7.5	L	3.7	U	2.4	W	0.6		
I	7.1	C	3.5	G	1.7	K	0.2		
O	7.0	D	3.5	F	1.6	X	0.1		

E∇	2.3	D∇	1.7	ES	1.3	RE	1.1
∇A	2.1	TI	1.7	AT	1.3	IO	1.1
ON	1.9	AN	1.6	ND	1.3	∇I	1.1
IN	1.9	EN	1.6	N∇	1.3	ME	1.0
∇T	1.8	TH	1.6	AL	1.2	ER	0.9
S∇	1.7	NT	1.4	HE	1.2	∇O	0.9

These are typical figures but will vary with the source

Basic cipher operations

Most historical ciphers are based on a combination of two basic operations.

Transposition: the characters in the plaintext are mixed up with each other (permuted).

Substitution: each character (or set of characters) is replaced by a different character (or set of characters).

Transposition ciphers

- A transposition cipher permutes characters usually in a fixed period d and permutation f.
- We can consider the plaintext as a matrix of rows of length d
- Generally transposition ciphers can permute rows or columns and output in row or column order.

Simple transposition cipher

- The key is the pair d and f.
- Each block of d characters is re-ordered using the permutation f.
- There are $d!$ permutations of length d. (Remember that $d! = d \times (d-1) \times (d-2) \times \cdots \times 2 \times 1$.)
- When $d = 10$ there are thus $3,628,800$ keys.

Cryptanalysing a transposition cipher

- The frequency distribution of the ciphertext characters is the same as for the plaintext characters. This helps to identify a transposition cipher.
- If the period d is small then transposition ciphers can be solved by hand using the process of anagramming (restoring disarranged characters to their original position).
- We can guess the value of d and write the ciphertext in columns so that there are d columns.
- Knowledge of the plaintext language digrams and trigrams can then optimise trials.
- This process can be automated.

Simple substitution ciphers

- Each character in the plaintext alphabet is replaced by a character in the ciphertext alphabet as defined by a substitution table.
- Simple substitution ciphers are also called *monoalphabetic* substitution ciphers.
- Note that transposition ciphers permute plaintext characters while substitution ciphers permute alphabet characters.
- There are many special cases of simple substitution ciphers. We consider only two: the Caesar cipher and random simple substitution cipher.

Outline

Caesar cipher

- A cipher which moves the ith letter of an alphabet to the $(i + j)$th letter. The key is the value j.
- Instead of writing out the whole substitution table we can define encryption and decryption as follows.

$$\text{Encryption: } c_i = (a_i + j) \bmod n$$
$$\text{Decryption: } a_i = (c_i - j) \bmod n$$

where $n = 26$ or $n = 27$ (size of alphabet).

Example

If the key is $j = 1$ then CIPHER $\rightarrow$ DJQIFS

Cryptanalysis of Caesar cipher

- We only need to find where one of the most frequent characters is shifted to.
- Suppose that the ciphertext is:

  PACGHJUHHCRICGRFWRUCRICPHGLFLQH

  First count the characters. Most common characters are C and H with frequency of 5 each.

If we assume that ∇ is in the alphabet we just need to find where it is mapped to.

Trial 1: Try $\nabla \rightarrow$ H, i.e. $j = 8$.

HTVZ∇BM$\nabla\nabla$VJA. . .

not correct, since no recognisable words.

Trial 2: Try $\nabla \rightarrow$ C , i.e. j = 3

MY∇DEGREE∇OF∇DOCTOR∇OF∇MEDICINE

Outline

Random simple substitution cipher

- A cipher which assigns a random character of the alphabet to another character of the alphabet.
- Encryption and decryption are defined by the substitution table which randomly permutes the alphabet.
- If the alphabet has 26 characters, there are 26! keys which is greater than 10^{26}. This is too many keys to search even with modern computers.
- The Caesar cipher is a special case of the random simple substitution cipher.

Example

Message: THE∇EVENING∇AND∇THE∇MORNING

Substitution table (key)		
A→S	J→G	S→M
B→J	K→C	T→O
C→V	L→F	U→Q
D→I	M→K	V→D
E→N	N→B	W→P
F→Y	O→U	X→ ∇
G→W	P→H	Y→T
H→A	Q→L	Z→X
I→Z	R→R	∇ →E

Message substitution (Encryption)		
T→O	N→B	D→I
H→A	I→Z	∇→E
E→N	N→B	T→O
∇→E	G→W	H→A
E→N	∇→E	E→N
V→D	A→S	∇→E
E→N	N→B	. . .

Ciphertext: OANENDNBZBWESBIEOANEKURBZBW

Cryptanalysis of random substitution

- Use frequency analysis on the characters of the alphabet.
- Decipher the following ciphertext:

```
FJLTXCFWKOVLHKJVKCBCOTEEVLPKCKJVJSTWTJYVKJVOJSTSBPLVITWCWPVDBIT
WICKTKQLVPHYTPRBJSTQLVYTKKJSCJETSCGTUHKJPTKYLFRTPETXCBTKJFXCJTJ
STGCZHTVOCGVZJCXTJTLCJCSHWPLTPOLCWYKFOJSTCQQCLCJHKVQTLCJTKEFJSV
HJCQQLTYFCRZTETCLJSTCXVLJFATXTWJKSVHZPRTYCZYHZCJTPCJCGTLBZVEOFI
HLTKCBQTLYTWJESFYSFKZCLITFWYVWJFWHVHKVQTLCJFVWFJEVHZPQLVPHYTXVL
TJSCWYHRFYXTJTLKVOICKCBTCLKCBCZFJJZTZTKKJSCWVWTYTWJFXTQTLYHRFYX
TJTLJSTYCHKJFYKVPCFKYVWKJCWJZBLTYHQTLCJTPCWPFKWTGTLPTKJLVBTPJST
KVZTQLVPHYJJSCJPFKCQQTCLKFKJSTPFKJFZZTPECJTLWVEVWTYHRFYXTJTLVOE
CJTLQLVPHYTKXVLTJSCWYHRFYXTJTLKVOICKJSTTDQTWKTFWECJTLJSTWPVTKWV
JCXVHWJJVCYTWJFXTQTLYHRFYXTJTLJSTILTCJOCYJVLVOJSTTDQTWKTLTKFPTK
FWJSTTZTYJLFYTWTLIBJSTYVKJVOKHLGTFZZCWYTEFZZRTXFWFXHXCWPJSTITWT
LCZTDQTWKTKCPZFRFJHX ...
```

Frequency analysis of ciphertext

No.	Character	%	Frequency
1	T	15.4	110
2	J	10.2	73
3	C	8.3	59
4	K	6.7	48
5	L	6.7	48
6	V	6.3	45

- Since E and T are most frequent characters in English we can guess the E $\rightarrow$ T and T $\rightarrow$ J in the substitution table.
- We can then start looking for English words like THE or other common trigrams.

Using Cryptool

- Solving random substitution by hand can be tedious and require a lot of trial and error
- We make use of software tools such as Cryptool (see link on course website)
- These tools can automate subtasks such as frequency counts or even automate the whole process

Key

With the help of tools we can find that the key (the substitution table) is:

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

The plaintext begins: `ITREMAINSFORUSTOSAY...`

Defining polyalphabetic substitution

- Polyalphabetic substitution ciphers use multiple mappings from plaintext to ciphertext.
- The effect of the multiple alphabets is to smooth the frequency distribution so direct frequency analysis is no longer effective.
- Typical polyalphabetic ciphers are periodic substitution ciphers based on a period d.
- Given d ciphertext alphabets $C_0, C_1, \ldots, C_{d-1}$, let

$$f_i : A \rightarrow C_i$$

be a mapping from the plaintext alphabet A to the ith cipher alphabet $C_i (0 \leq i \leq d-1)$.

Encryption process

A plaintext message

$$M = m_0 \dots m_{d-1} m_d \dots m_{2d-1} \dots$$

is enciphered to

$$E(K, M) = f_0(m_0) \dots f_{d-1}(m_{d-1})\ f_0(m_d) \dots f_{d-1}(m_{2d-1}) \dots$$

For the special case $d = 1$ the cipher is monoalphabetic (a simple substitution cipher)

Random polyalphabetic substitution cipher

- Key Generation
  - Select block length d
  - Generate d random simple substitution tables
- Encryption
  - To encrypt the ith character, use the substitution table number j where $i \equiv j \pmod{d}$
- Decryption
  - Use the same substitution table as in encryption in order to reverse simple substitution

Example key for polyalphabetic substitution

Choose $d = 3$, so there are 3 ciphertext alphabets,

Key					
P:	abc	def	ghi	jkl	mno
C_1:	UWY	SX∇	TVZ	CEI	AFG
C_2:	QLM	PJO	RKN	∇XS	YUW
C_3:	MLQ	RNK	GFA	ZVT	YWU
P:	pqr	stu	vwx	yz∇	
C_1:	BDH	KNR	JOP	LMQ	
C_2:	ZVT	FGA	HDB	EIC	
C_3:	POJ	HDB	IEC	∇XS	

If P = IT∇IS∇A∇BEAUTIFUL∇DAY
then C = ZGSZFSUCLXQBNNKRSSSQ∇

Outline

Vigenère cipher

- The Vigenère cipher is a popular form of periodic substitution cipher based on *shifted* alphabets
- The key K is specified by a sequence of characters

$$K = k_0 \dots k_{d-1}$$

where $k_i (i = 0, \dots, d-1)$ gives the amount of shift in the ith alphabet, i.e.

$$f_i(p) = (p + k_i) \bmod n$$

where p is the plaintext character

- In the 19th century the Vigenère cipher was widely believed to be unbreakable

Example

M:	AT∇T	HE∇T	IME∇
K:	LOCK	LOCK	LOCK
$E(K, M)$:	LGBC	SSBC	T∇GJ

- We number the alphabet A=0, B=1 . . . Z=25, $\nabla = 26$
- In this example the first character of each 4-character group is shifted by 11, the second by 14, the third by 2 and the fourth by 10
- Shifting is computed modulo 27 so that the alphabet 'wraps around'

Cryptanalysis of Vigenère cipher

1. Identify the period length. Several different techniques for this include:
   - Kasiski method (illustrated below)
   - Autocorrelation (used in Cryptool online version)
   - Index of Coincidence (used in JCryptool and Cyberchef)
2. Attack separately d different substitution tables. Since each substitution is just a shift (Caesar cipher), this is straightforward if there are sufficient ciphertexts.

Identifying the period using autocorrelation

- Given a ciphertext C compute the correlation between C and its shift C_i for all plausible values i of the period
- Because English is non-random, there is a better correlation between two texts with the same size shift than between two texts with different size shifts
- Therefore we expect to see peaks in the value of C_i when i is a multiple of the period
- Plotting the results on a histogram can usually allows us to identify the period
- This method can be used to find the period of any periodic polyalphabetic cipher

Example Vigenère cryptanalysis

The first characters of a ciphertext are:

AUVHSGE**PELPEK**QTEDKSFNYJYATCTCCKFTSUTEFVBVVHPNMFUHBFNPV
YFVRFVUSPEEVHFNAOFLBFYJPFPTMFFMFVHBVHFJAENEGVTIGHPWSFU
HPTTMAAGVESGIHJT**PELPEK**JPTIGMPTNJPGJUAUFOXPBFUIEGTIGFJTEIQ
WFXESYIUJTIGIOVEOVIPPOGCWBKTJPGIKMIQWFXESNOOIHFOIHJTCGIXC
SBNRFCDZFEFRLZKNUGRFUTFFIOJITKNRWISAFPTTIQUHJIUYATUUSTOVP
DFFBZPOOGOGVHFIRJOAOFSUTAOIEGGAUWRFUWIKCIYESGATUODKAUG
DXKTIVHFVWPERJOETYHJEHJJAWGAMTEBFYSGCPTDFFSUKLMVHFPAUW
RFQFUJEDCSFCNEVHFGXBNTFFSUCTJQNPHHJUCMKEOVGBXEJVADJASC
CUGRPHIUUOXPIOFEFFAQCRUHRPOTIGNBVUSGOGVHFKNWGSUKGBVIP
P**WIKCIO**YGTIFPDICDPPHBPDUJESGWBUSPOEUJIOIIOJITOATVESNYHTATR
OGCSJVUBVIPPAOFHJUKFGNJPCJUIWGRFCSPPIOI**WIKCIO**AEGIUCPMGAT
WRFVONGTPUTVFYIKSTASUGMPHWPTKBPDUQFPNLPYTIGQVKCLUUCVL
FOEUJOEUBZYHJEHIGDJUEOVAOILFFTIGMPUTJPEYVRJEACNENASUGRJ

Step 1

Identify the period length d.

- Note that the sequence PELPEK and WIKCIO occur multiple times.
- The positions of some of the pairs of these strings are separated by 117 and 93 characters.
- The period is almost certain to be 1 or 3 because the only common divisors of 117 and 93 are 1 and 3.

This process is known as the Kasiski method. We can also automate the process by plotting the autocorrelation (use CrypTool).

Step 2

- Attack separately three different alphabets
- Only need to find the shift for each alphabet as in Caesar cipher
- Look for character with largest frequency and assume this is shifted from E.
- Turns out that:
  - the first has key ‘A’ (shift of 0).
  - the second has key ‘B’ (shift of 1), and
  - the third has key ‘C’ (shift of 2).

The plaintext starts:

`ATTHREEOCLOCKPRECISELYIWASATBAKERSTREET...`

Outline

Other ciphers designed for use by hand

You can find many other ciphers in the tools such as Cryptool and Cyberchef. Some examples:

- the *autokey cipher* starts off as the Vigenère cipher but once the alphabets defined by the key have been used once, uses the plaintext to define subsequent alphabets. Therefore the autokey cipher is *not* periodic.
- the *running key cipher* uses a (practically) infinite set of alphabets from a shared key. In practice the shared key can be extract from a book, when it is often called a *book cipher*.

Rotor machines

- In the early 20th century electromechanical machines were developed for encryption using *rotors* as moving alphabets.
- The most famous is the Enigma machine used by the Germans in World War II.
- Each character is encrypted using a different alphabet. The Enigma machine has a period of about 17000, so in practice it would never repeat on the same message.
- Smart's book (see the additional resources list) has a whole chapter on Enigma and how it was broken.