Historical ciphers I

Tanja Lange

Eindhoven University of Technology

2WF80: Introduction to Cryptology

Caesar cipher

Most famous historical cipher, here with our current alphabet.

```
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
| | |                                           |
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
```

Encryption just maps letters from the top row to the matching ones in the bottom row.

```
HELLO BOB
KHOOR ERE
```

Caesar cipher

Most famous historical cipher, here with our current alphabet.

```
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
| | |                                           |
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
```

Encryption just maps letters from the top row to the matching ones in the bottom row.

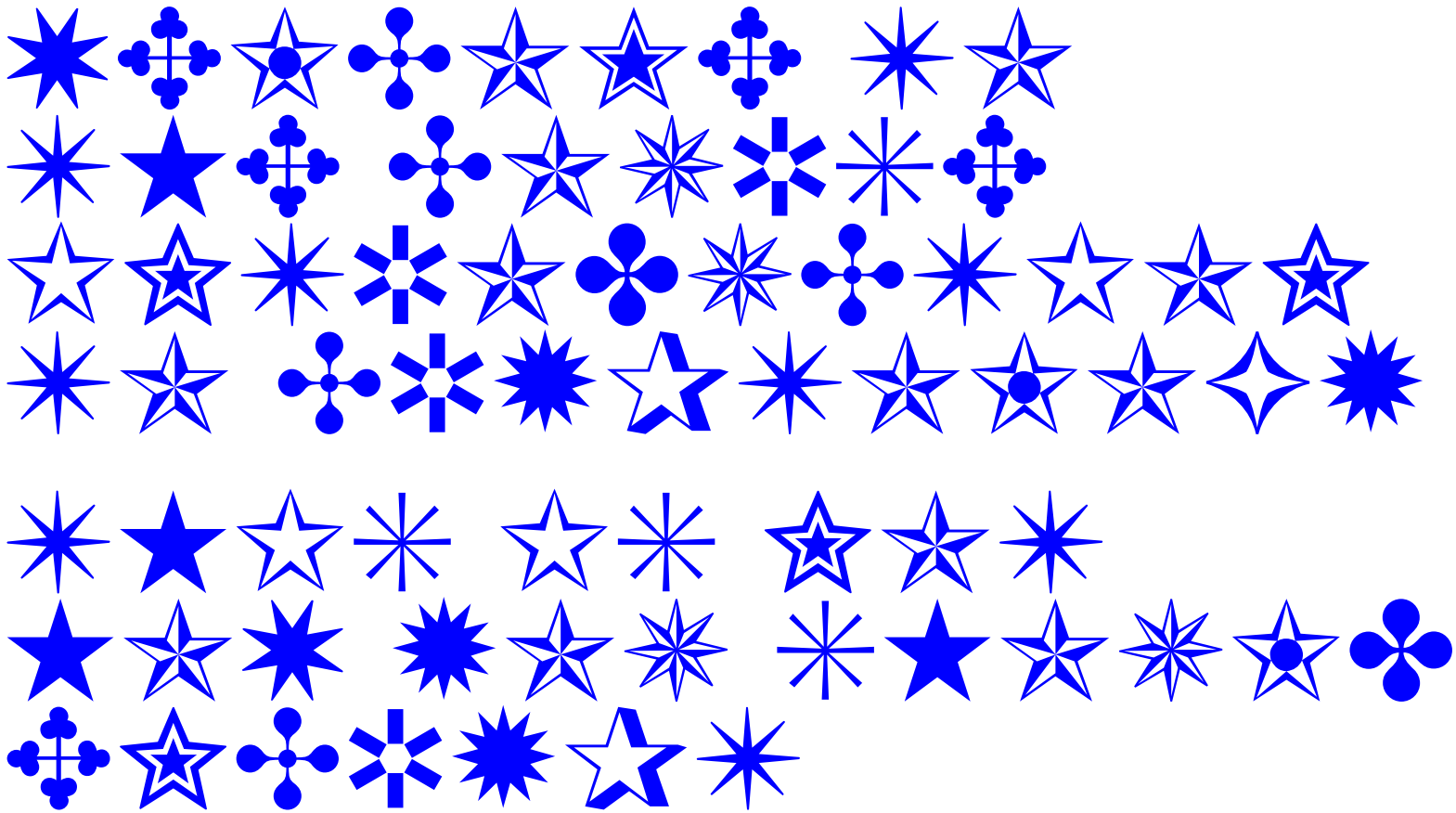
```
HELLO BOB
KHOOR ERE
```

Decryption maps from bottom row to top row.
Figure out what this decrypts to:

```
HQFUBSWLRQ ZRUNV
```

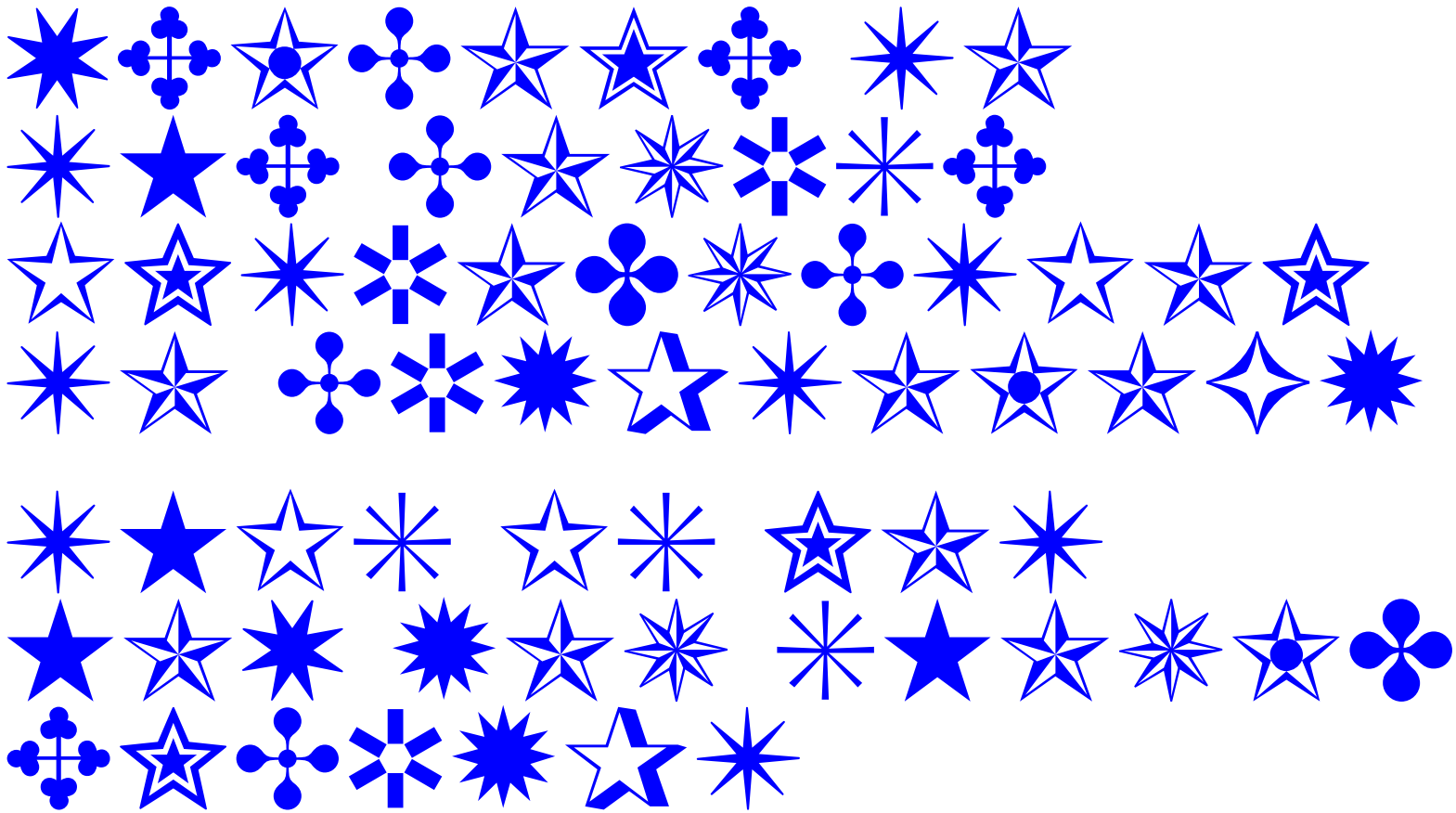
Substitution cipher

Each letter is replaced by a symbol.



Substitution cipher

Each letter is replaced by a symbol.



Hint: look for short words and repeated combinations of letters.

Where are the keys?

- For the substitution cipher the key is the knowledge of the full alphabet in symbols, e.g. A $\mapsto$ 

Where are the keys?

- For the substitution cipher the key is the knowledge of the full alphabet in symbols, e.g. A $\mapsto$ 
(This does not help you other than learning that there was no A in the sample.)

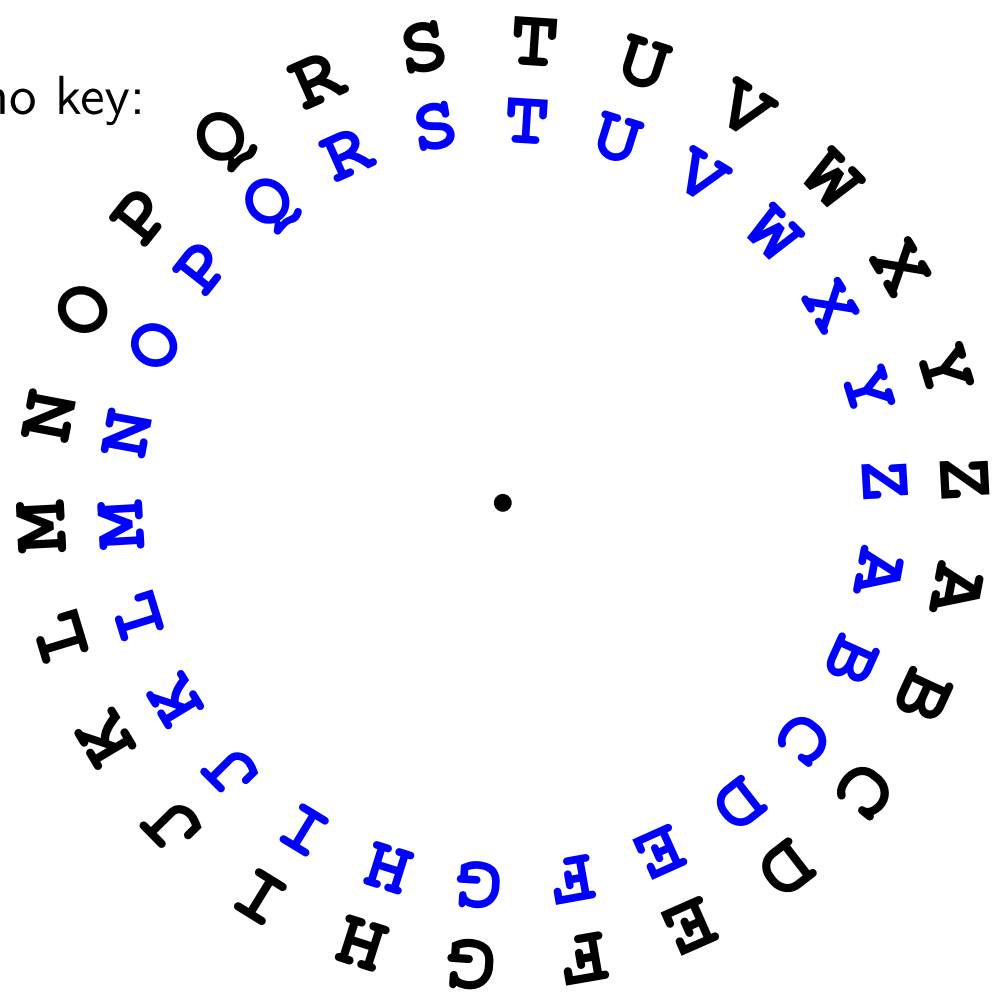
Where are the keys?

- For the substitution cipher the key is the knowledge of the full alphabet in symbols, e.g. A $\mapsto$ 
  (This does not help you other than learning that there was no A in the sample.)
- For the Caesar cipher there is no key:
  knowing the system is knowing
  everything.–
  But we can turn the
  Caesar cipher
  into a keyed cipher:

Where are the keys?

- For the substitution cipher the key is the knowledge of the full alphabet in symbols, e.g. A $\mapsto$ ✡
  (This does not help you other than learning that there was no A in the sample.)
- For the Caesar cipher there is no key: knowing the system is knowing everything.–
  But we can turn the Caesar cipher into a keyed cipher:

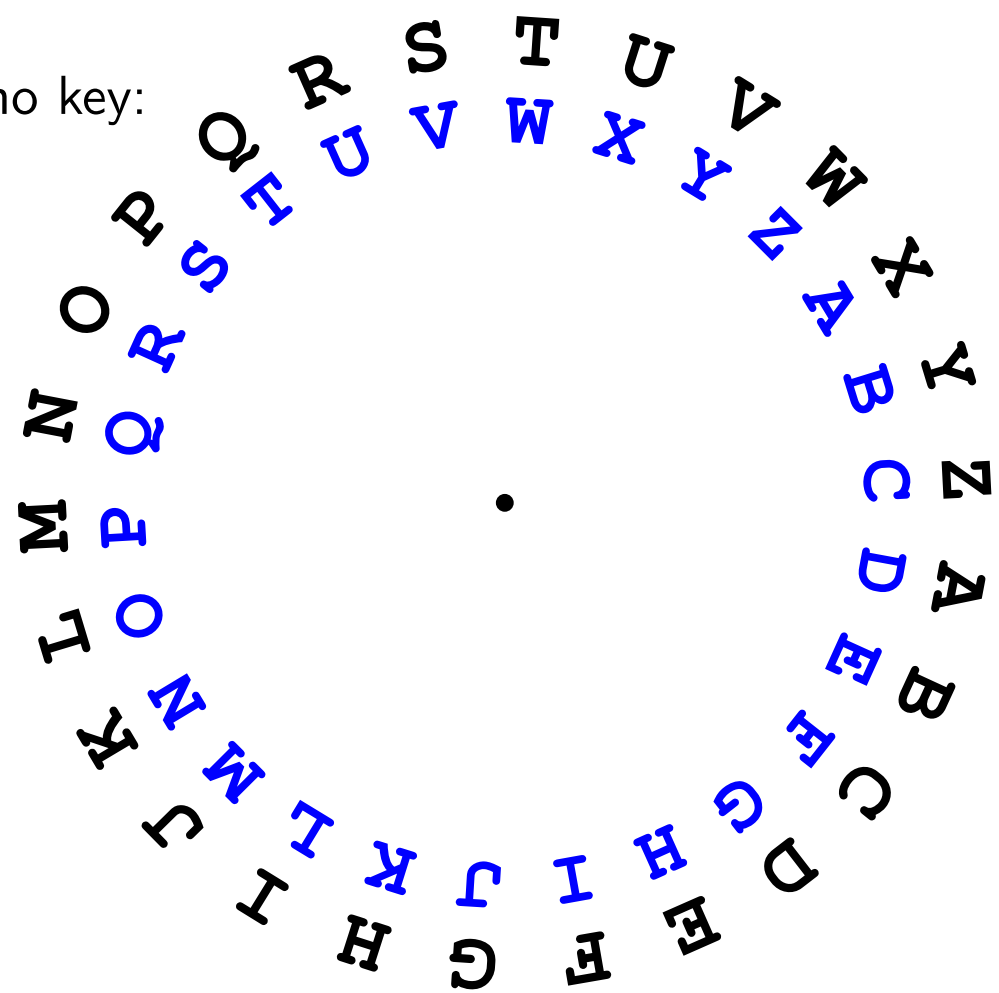
  Key is the shifting distance, e.g. 3 means A $\mapsto$ D;



Where are the keys?

- For the substitution cipher the key is the knowledge of the full alphabet in symbols, e.g. A $\mapsto$ ✡
  (This does not help you other than learning that there was no A in the sample.)
- For the Caesar cipher there is no key: knowing the system is knowing everything.–
  But we can turn the Caesar cipher into a keyed cipher:

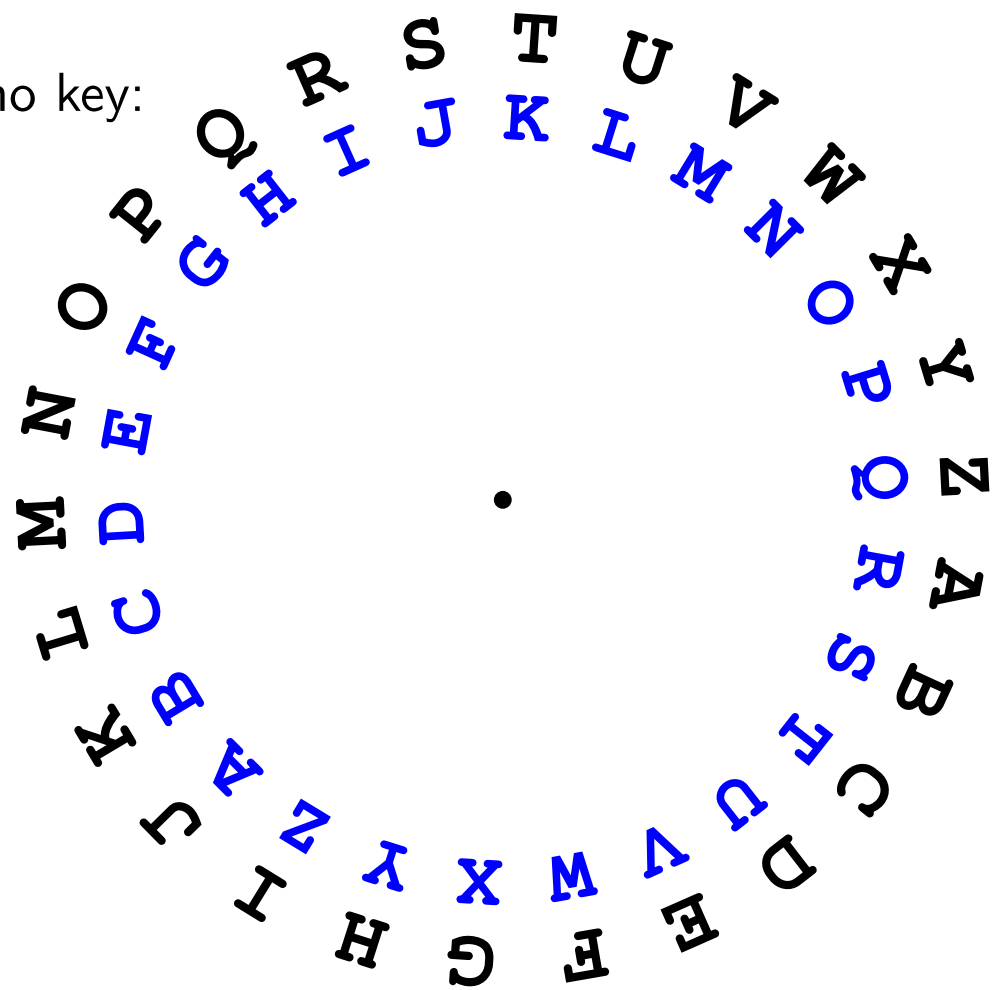
  Key is the shifting distance, e.g. 3 means A $\mapsto$ D;



Where are the keys?

- For the substitution cipher the key is the knowledge of the full alphabet in symbols, e.g. A $\mapsto$ ✡
  (This does not help you other than learning that there was no A in the sample.)
- For the Caesar cipher there is no key: knowing the system is knowing everything.–
  But we can turn the Caesar cipher into a keyed cipher:

  Key is the shifting distance, e.g. 3 means A $\mapsto$ D;
  17 means A $\mapsto$ R.

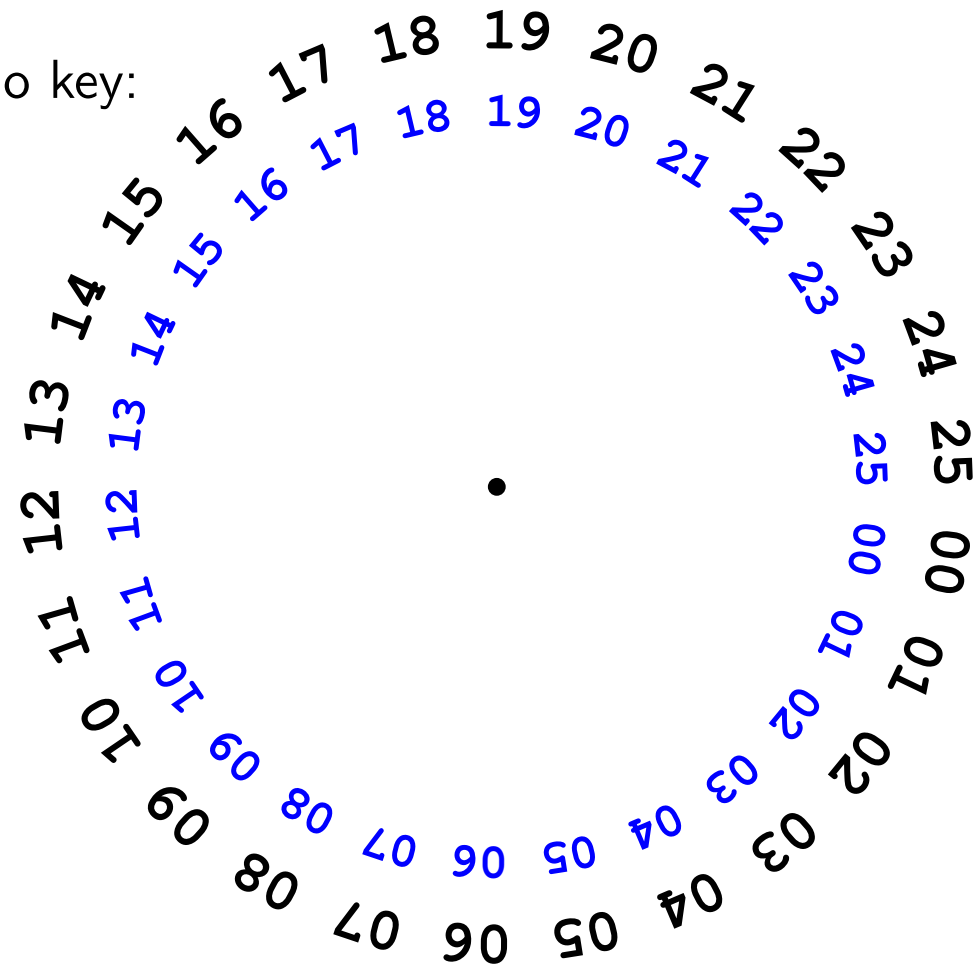


Where are the keys?

- For the substitution cipher the key is the knowledge of the full alphabet in symbols, e.g. A $\mapsto$ ✡
  (This does not help you other than learning that there was no A in the sample.)
- For the Caesar cipher there is no key: knowing the system is knowing everything.–
  But we can turn the Caesar cipher into a keyed cipher:

  Key is the shifting distance, e.g. 3 means A $\mapsto$ D;
  17 means A $\mapsto$ R.

  Easier to compute with integers modulo 26.

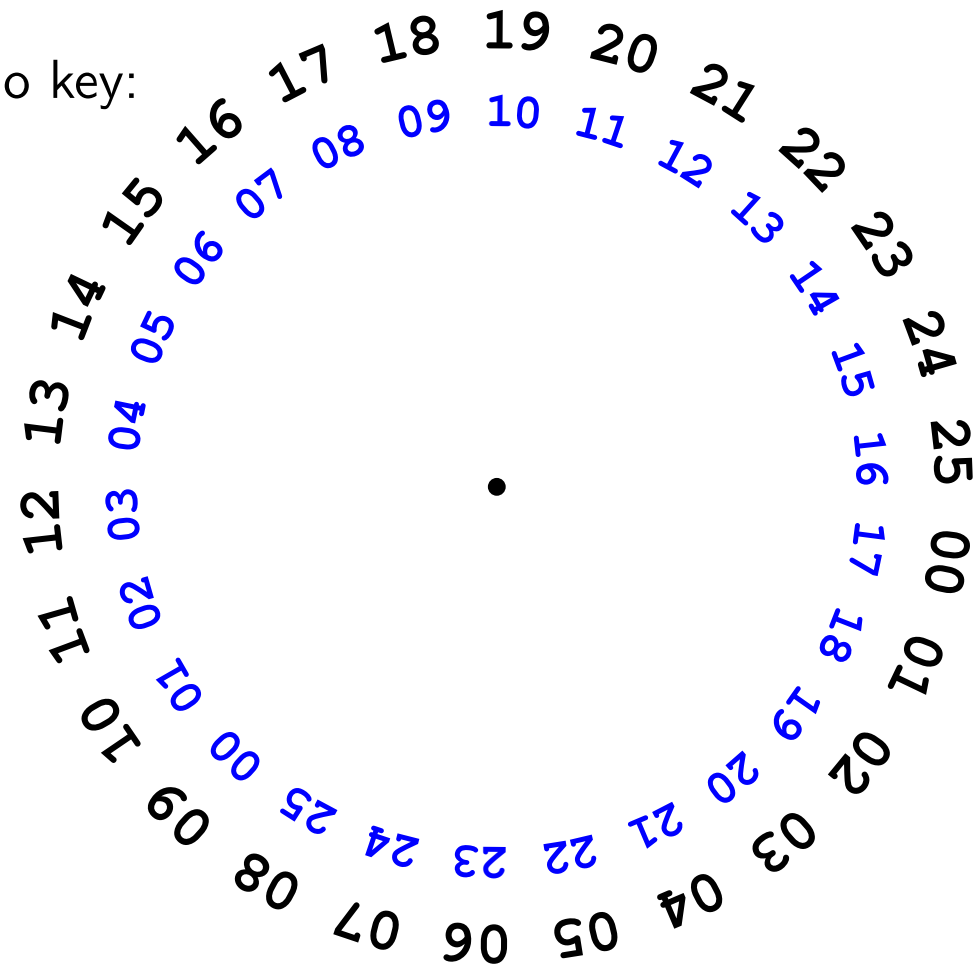


Where are the keys?

- For the substitution cipher the key is the knowledge of the full alphabet in symbols, e.g. A $\mapsto$ ✡
  (This does not help you other than learning that there was no A in the sample.)
- For the Caesar cipher there is no key: knowing the system is knowing everything.–
  But we can turn the Caesar cipher into a keyed cipher:

  Key is the shifting distance, e.g. 3 means $\mathtt{A} \mapsto \mathtt{D}$;
  17 means $\mathtt{A} \mapsto \mathtt{R}$.

  Easier to compute with integers modulo 26.



How many keys are there?

- There are 26 different shifting distances,
  so only 26 keys for Caesar.
  Easy to try each (at least with a computer).
  Probably only one will give sensible text.

How many keys are there?

- There are 26 different shifting distances,
  so only 26 keys for Caesar.
  Easy to try each (at least with a computer).
  Probably only one will give sensible text.
- With 26 symbols the substitution cipher has
  $26! = 26 \cdot 25 \cdot 24 \cdots 3 \cdot 2 \approx 10^{26.6}$ different keys.
  Still possible to try on a supercomputer with some time.

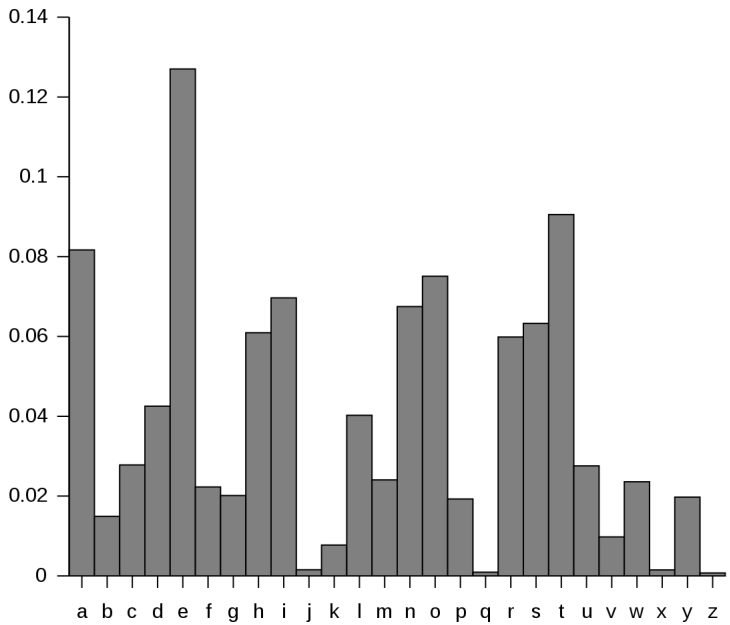
How many keys are there?

- There are 26 different shifting distances,
  so only 26 keys for Caesar.
  Easy to try each (at least with a computer).
  Probably only one will give sensible text.
- With 26 symbols the substitution cipher has
  $26! = 26 \cdot 25 \cdot 24 \cdots 3 \cdot 2 \approx 10^{26.6}$ different keys.
  Still possible to try on a supercomputer with some time.
- The set from which the keys are drawn is called the keyspace.
  A minimum requirement for security is that the keyspace is
  too large to search.

Cryptanalysis

For sufficiently long texts, frequency analysis is much more powerful than key search.

The frequency distribution of letters in English[1] shows a very strong peak at E, and strong peaks at T, A, O, I, and N. Can look for his pattern of peaks in Caesar to get most-likely shifting distance.



Also helps to find candidates for most common symbols in substitution cipher.

[1] Source: Wikipedia

One-time pad

- Let $m \in \{0,1\}^\ell$, i.e., a message is a string of ℓ bits.
  Let $k \in \{0,1\}^\ell$, chosen uniformly at random.
  Then $c = m + k$, where addition is done modulo 2 in each position.
  (In more mathematical notation: $m, k \in \mathbb{F}_2^\ell, c = m + k$.)
- The one-time pad is information-theoretically secure –
  there is no information about the plaintext in the ciphertext.
  $c_i = 0$ can come from $m_i = k_i = 0$ or from $m_i = k_i = 1$.
  $c_i = 1$ can come from $m_i = 0, k_i = 1$ or from $m_i = 1, k_i = 0$.

One-time pad

- Let $m \in \{0,1\}^\ell$, i.e., a message is a string of ℓ bits.
  Let $k \in \{0,1\}^\ell$, chosen uniformly at random.
  Then $c = m + k$, where addition is done modulo 2 in each position.
  (In more mathematical notation: $m, k \in \mathbb{F}_2^\ell, c = m + k$.)
- The one-time pad is information-theoretically secure –
  there is no information about the plaintext in the ciphertext.
  $c_i = 0$ can come from $m_i = k_i = 0$ or from $m_i = k_i = 1$.
  $c_i = 1$ can come from $m_i = 0, k_i = 1$ or from $m_i = 1, k_i = 0$.
- This requires the key to be as long as the message –
  the "two-time" pad is insecure.
  This makes the scheme unusable for most situations.

One-time pad

- Let $m \in \{0,1\}^\ell$, i.e., a message is a string of ℓ bits.
  Let $k \in \{0,1\}^\ell$, chosen uniformly at random.
  Then $c = m + k$, where addition is done modulo 2 in each position.
  (In more mathematical notation: $m, k \in \mathbb{F}_2^\ell, c = m + k$.)
- The one-time pad is information-theoretically secure –
  there is no information about the plaintext in the ciphertext.
  $c_i = 0$ can come from $m_i = k_i = 0$ or from $m_i = k_i = 1$.
  $c_i = 1$ can come from $m_i = 0, k_i = 1$ or from $m_i = 1, k_i = 0$.
- This requires the key to be as long as the message –
  the "two-time" pad is insecure.
  This makes the scheme unusable for most situations.
- Possible practical usage: agree on a book to use as key;
  letter in the book determines the shifting distance of that letter.
  Using key $k =$`THISISTHESTORYOFLITTLERED..`

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
          |
TUVWXYZABCDEFGHIJKLMNOPQRS
```

One-time pad

- Let $m \in \{0,1\}^\ell$, i.e., a message is a string of ℓ bits.
  Let $k \in \{0,1\}^\ell$, chosen uniformly at random.
  Then $c = m + k$, where addition is done modulo 2 in each position.
  (In more mathematical notation: $m, k \in \mathbb{F}_2^\ell, c = m + k$.)
- The one-time pad is information-theoretically secure –
  there is no information about the plaintext in the ciphertext.
  $c_i = 0$ can come from $m_i = k_i = 0$ or from $m_i = k_i = 1$.
  $c_i = 1$ can come from $m_i = 0, k_i = 1$ or from $m_i = 1, k_i = 0$.
- This requires the key to be as long as the message –
  the "two-time" pad is insecure.
  This makes the scheme unusable for most situations.
- Possible practical usage: agree on a book to use as key;
  letter in the book determines the shifting distance of that letter.
  Using key $k =$`THISISTHESTORYOFLITTLERED..`

```
  HELLOBOB      ABCDEFGHIJKLMNOPQRSTUVWXYZ
+ THISISTH            |
----------      TUVWXYZABCDEFGHIJKLMNOPQRS
  ALTDWTHI
```

  Problem: Key is no longer uniformly distributed.

One-time pad

- Let $m \in \{0,1\}^\ell$, i.e., a message is a string of ℓ bits.
  Let $k \in \{0,1\}^\ell$, chosen uniformly at random.
  Then $c = m + k$, where addition is done modulo 2 in each position.
  (In more mathematical notation: $m, k \in \mathbb{F}_2^\ell, c = m + k$.)
- The one-time pad is information-theoretically secure –
  there is no information about the plaintext in the ciphertext.
  $c_i = 0$ can come from $m_i = k_i = 0$ or from $m_i = k_i = 1$.
  $c_i = 1$ can come from $m_i = 0, k_i = 1$ or from $m_i = 1, k_i = 0$.
- This requires the key to be as long as the message –
  the "two-time" pad is insecure.
  This makes the scheme unusable for most situations.
- Possible practical usage: agree on a book to use as key;
  letter in the book determines the shifting distance of that letter.
  Using key $k =$`THISISTHESTORYOFLITTLERED..`

```
  HELLOBOB      ABCDEFGHIJKLMNOPQRSTUVWXYZ
+ THISISTH          |
----------      HIJKLMNOPQRSTUVWXYZABCDEFG
  ALTDWTHI
```

  Problem: Key is no longer uniformly distributed.

One-time pad

- Let $m \in \{0,1\}^\ell$, i.e., a message is a string of ℓ bits.
  Let $k \in \{0,1\}^\ell$, chosen uniformly at random.
  Then $c = m + k$, where addition is done modulo 2 in each position.
  (In more mathematical notation: $m, k \in \mathbb{F}_2^\ell, c = m + k$.)
- The one-time pad is information-theoretically secure –
  there is no information about the plaintext in the ciphertext.
  $c_i = 0$ can come from $m_i = k_i = 0$ or from $m_i = k_i = 1$.
  $c_i = 1$ can come from $m_i = 0, k_i = 1$ or from $m_i = 1, k_i = 0$.
- This requires the key to be as long as the message –
  the "two-time" pad is insecure.
  This makes the scheme unusable for most situations.
- Possible practical usage: agree on a book to use as key;
  letter in the book determines the shifting distance of that letter.
  Using key $k =$`THISISTHESTORYOFLITTLERED..`

```
  HELLOBOB      ABCDEFGHIJKLMNOPQRSTUVWXYZ
+ THISISTH                  |
----------      IJKLMNOPQRSTUVWXYZABCDEFGH
  ALTDWTHI
```

  Problem: Key is no longer uniformly distributed.

Viginère

- Use codeword as key, e.g., k =CRYPTO.
  Encryption works the same as for one-time pad with key
  CRYPTOCRYPTOCRYPTOCRYPTOCRYPTOCRYPTOCRYPTOCRYPTOCRYPTOCRYPTO...

```
  THISISABETTERWAYTOENCRYPTTHANCAESAR
+ CRYPTOCRYPTOCRYPTOCRYPTOCRYPTOCRYPT
----------------------------------
  VYGHBGCSCIMSTNYNMCGEAGRDVKFPGQCVQPK
  |     |     |     |     |     |
```

  Letter 1, 7, 13, 19, ... use Caesar with A $\mapsto$ C

```
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
|       |                       |   |
C 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
```

Viginère

- Use codeword as key, e.g., $k =$CRYPTO.
  Encryption works the same as for one-time pad with key
  CRYPTOCRYPTOCRYPTOCRYPTOCRYPTOCRYPTOCRYPTOCRYPTOCRYPTOCRYPTO...

```
  THISISABETTERWAYTOENCRYPTTHANCAESAR
+ CRYPTOCRYPTOCRYPTOCRYPTOCRYPTOCRYPT
--------------------------------------
  VYGHBGCSCIMSTNYNMCGEAGRDVKFPGQCVQPK
  |     |     |     |     |     |
```

  Letter 1, 7, 13, 19, ... use Caesar with A $\mapsto$ C

```
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
|       |                       |   |
C 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
```

- Keyspace much larger than for Caesar, but for long m and known $|k|$ frequency analysis on letters $1, |k|+1, 2|k|+1, 3|k|+1, \ldots$ finds first shifting distance; on letters $2, |k|+2, 2|k|+2, 3|k|+2, \ldots$ finds second shifting distance; etc.
- For unknown $|k|$ either bruteforce length by looking for clear peaks in letter frequencies when skipping 1,2, 3, ... letters or find repeating combinations of letters; those appear likely at a multiple of $|k|$.