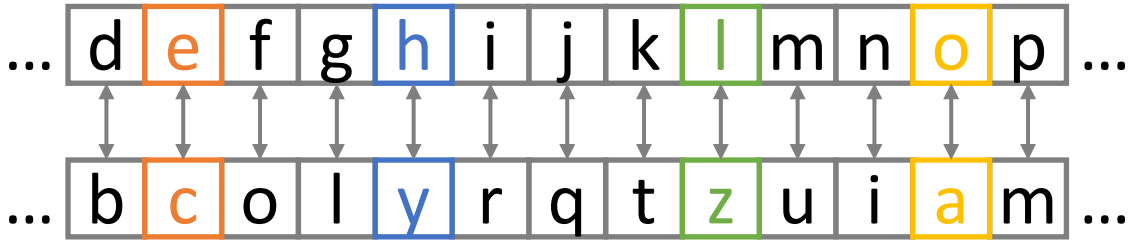
Substitution Ciphers

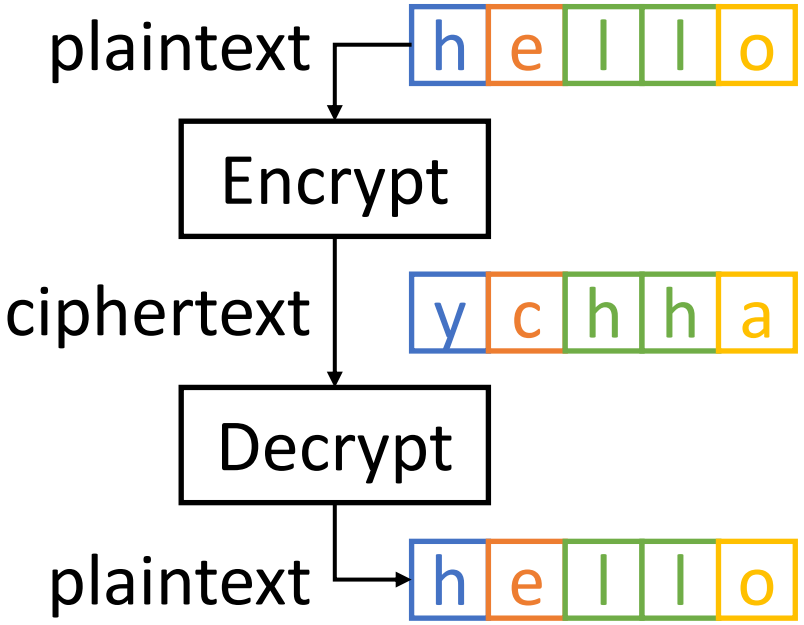
Elements of Applied Data Security M

Livia Manovi – livia.manovi@unibo.it

Substitution Ciphers

Each plaintext character (or group of characters) is replaced with a different ciphertext symbol. The receiver deciphers the text by performing the inverse substitution.





Substitution Ciphers

- Historical ciphers rely on the substitution of letters in the plaintext with other letters based on a predetermined key or rule.
- The replacement remains consistent throughout the message.
- Limited key space implies vulnerability to brute force attacks.
- Patterns in the frequency distribution of letters or characters can be exploited to break the cipher.
- Despite their lack of security by modern standards, historic ciphers hold significant importance.

Assignment

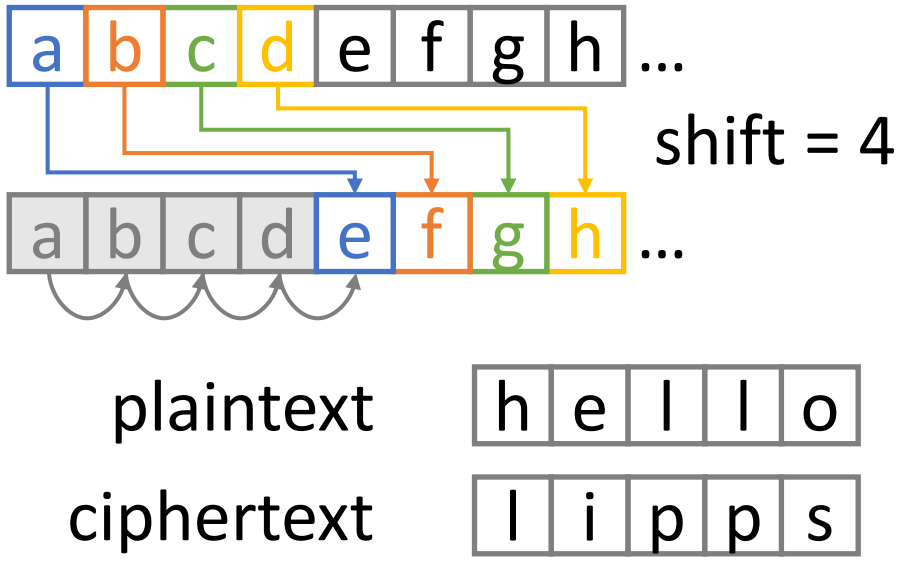
1. Task 1: Breaking a Caesar Cipher
2. Task 2: Breaking a Simple Substitution Cipher

Task 1: Caesar Cipher

Caesar Cipher

The method is named after Julius Caesar, who used it in his private correspondence. Each letter in the plaintext is replaced by a letter shifted by some fixed number of positions down the alphabet.

- Same characters for plaintext and ciphertext.
- Very simple encryption rule: only 26 possibilities!



Breaking a Caesar Cipher

- **Brute force**:
  - The English alphabet is 26 letters long, meaning that only 26 shifts are possible. Hence, you can try all possibilities and check whether the resulting plaintext makes sense.

Task 1

- Inputs:
  - Ciphertext as a text file: `ciphertext_caesar.txt`.
    - Ciphertext is a Wikipedia page encrypted with a Caesar Cipher
    - Only lower-case letters are considered
    - spaces and special characters are unchanged

`ciphertext_caesar.txt`

```
aucom dofcom wuymul (12 dofs 100 vw - 15 gulwb 44 vw) qum u liguh
ayhyluf uhx mnunymguh. u gygvyl iz nby zclmn nlcogpcluny, wuymul
fyx nby liguh ulgcym ch nby auffcw qulm vyzily xyzyuncha bcm
jifcncwuf lcpuf jigjys ch u wcpcf qul, uhx movmykoyhnfs vywugy
xcwnunil zlig 49 vw ohncf bcm ummummchuncih ch 44 vw. by jfusyx u
wlcncwuf lify ch nby ypyhnm nbun fyx ni nby xygcmy iz nby liguh
lyjovfcw uhx nby lcmy iz nby liguh ygjcly.
ch 60 vw, wuymul, wlummom, uhx jigjys zilgyx nby zclmn
nlcogpcluny, uh chzilguf jifcncwuf uffcuhwy nbun xigchunyx liguh
jifcncwm zil mypyluf syulm. nbycl unnygjnm ni ugumm jifcncwuf
jiqyl qyly ijjimyx vs guhs ch nby myhuny, ugiha nbyg wuni nby
siohayl qcnb nby jlcpuny mojjiln iz wcwyli. wuymul limy ni vywigy
```

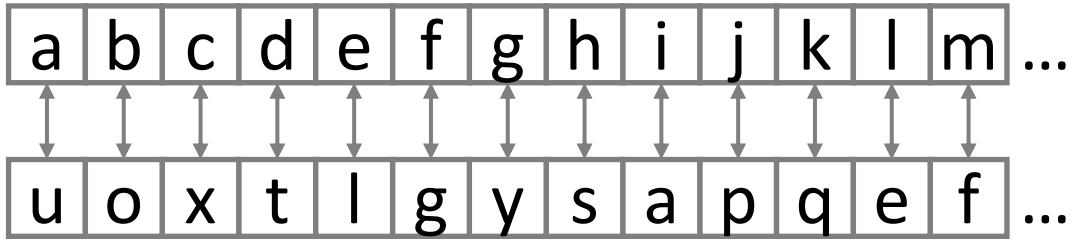
- Outputs:
  - **Key** that is the shift to apply to the alphabet to decrypt the ciphertext.
  - **Plaintext** decrypted from the ciphertext.

Task 2: Simple Substitution

Simple Substitution Cipher

Each plaintext character is replaced with a different ciphertext character.

- As for the Caesar Cipher, plaintext and ciphertext share the same set of characters (the alphabet).
- Mapping from plaintext to ciphertext can be any of the $26! \sim 10^{26} \sim 2^{88}$ possibilities



Breaking a Simple Substitution Cipher

- **Brute force**
  - assuming 1ns for each try, it would take $> 10^9$ years to break it!
  - Nowadays machines cannot explore $26!$ candidates.
- Substitution preserves the underlying statistics, enabling the deduction of the plaintext through **frequency analysis** of the ciphertext letters.
  - For reasonably large pieces of text (with enough characters to be statistically relevant), a possible procedure can be to replace:
    - the most common ciphertext character with the most common character in the plaintext
    - the second most common ciphertext character with the second most common character in the plaintext
    - and so on

Task 2

- Inputs:
  - Ciphertext as a text file: `ciphertext_simple.txt`.
    - As before, ciphertext is the encryption a Wikipedia page with all lower-case letters and special characters unchanged
  - An English text `The-Adventure-of-the-Dancing-Men.txt` to estimate of the English letter distribution.

`ciphertext_simple.txt`

```
gihoaz zijlla nvhbblb (hkwri 30, 1916 dzpwohwq 24, 2001) jhn hb
hczwrghb chsvzchsrgrhb, zizgswrghi zberbzzw, glckoszw ngrzbsrns
hba gwqkslewhkvzw tbljb hn svz "dhsvzw ld rbdlwchsrlb svzlwq".
vz jhn svz drwns sl azngwrpz svz pllizhb ehszn (zizgswlbrg
grwgorsn) svhs hwz znnzbsrhi sl hii arershi zizgswlbrg grwgorsn,
hba vz poris svz drwns chgvrbz izhwbrbe azxrgz, svon dlobarbe
svz drzia ld hwsrdrgrhi rbsziirezbgz. vz rn gwzarsza hilbenraz
ezlwez plliz dlw ihqrbe svz dlobahsrlbn ld svz rbdlwchsrlb
hez.hn h 21-qzhw-lia chnszw'n azewzz nsoazbs hs svz
chnnhgvonzssn rbnsrsosz ld szgvblileq (crs), vz jwlsz vrn svznrn
azclbnswhsrbe svhs zizgswrghi hkkirghsrlbn ld pllizhb hiezpwh
gloia glbnswogs hbq ilerghi boczwrghi wzihsrlbnvrk, svzwzpq
znshpirnvrbe svz svzlwq pzvrba arershi glckosrbe hba arershi
```

- Outputs:
  - **Substitution rule** to apply to the alphabet to decrypt the ciphertext.
  - **Plaintext** decrypted from the ciphertext.

Deadline

Tuesday, March 26 at 12PM (noon)