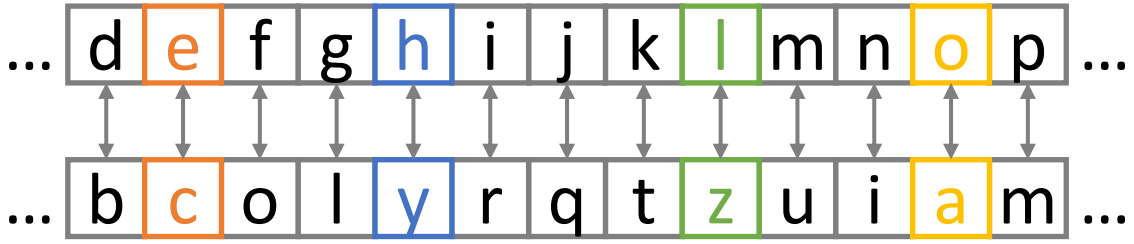
Substitution Ciphers

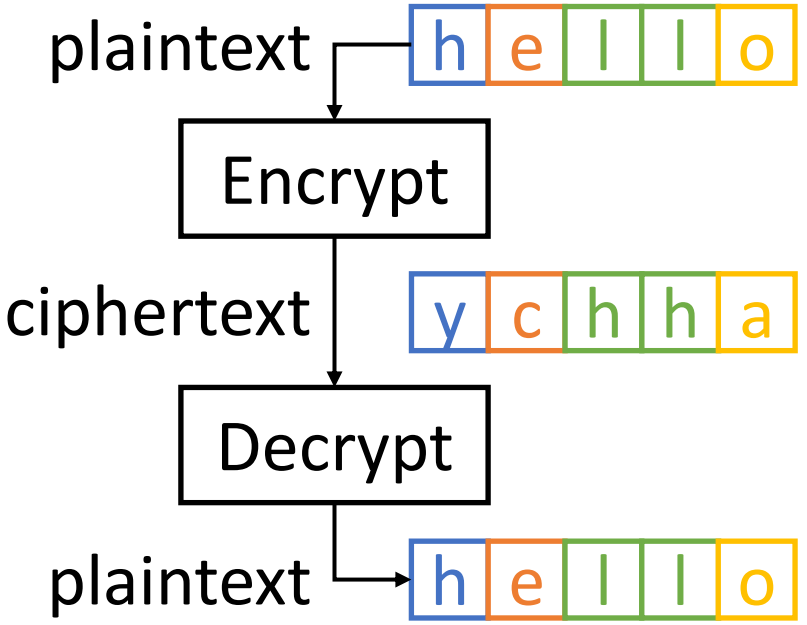
Elements of Applied Data Security M

Lorenzo Capelli – l.capelli@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

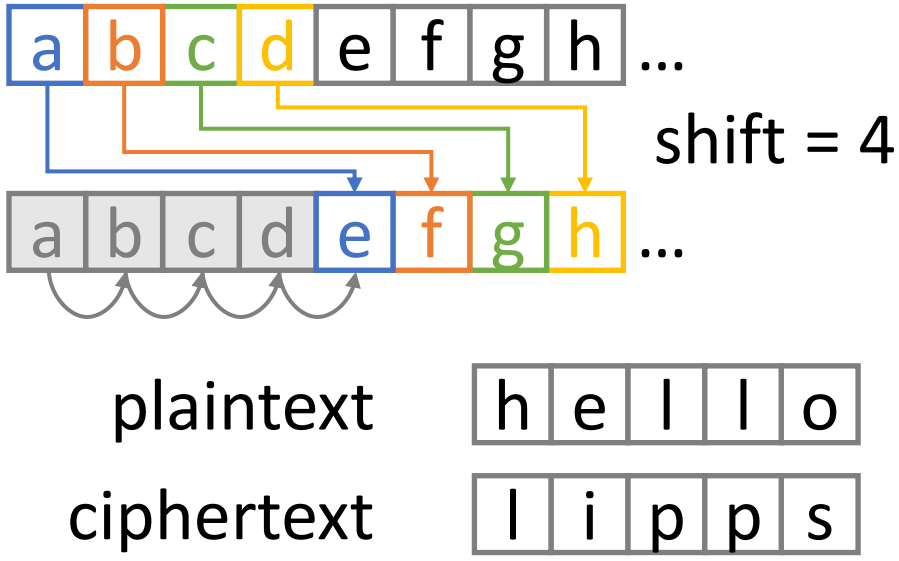
- Task 1: Breaking a Caesar Cipher
- Task 2: Breaking a Simple Substitution Cipher

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`

```
jgew (alsdasf sfv dslaf: jges, hjgfgmfuwv ['jg:es] )
ak lzw ushalsd ualq gx alsdq. al ak sdkg lzw ushalsd
gx lzw dsrag jwyagf, lzw uwfljw gx lzw ewljghgdalsf
ualq gx jgew ushalsd, sfv s khwuasd ugemfw
(emfauahsdalq) fsewv ugemfw va jges ushalsdw. oalz
2,860,009 jwkavwflk af 1,285 ce2 (496.1 ki ea), jgew
ak lzw ugmfljq'k egkl hghmdslwv ugemfw sfv lzw lzajv
egkl hghmdgmk ualq af lzw wmjghwsf mfagf tq
hghmdslagf oalzaf ualq daealk. lzw ewljghgdalsf ualq
gx jgew, oalz s hghmdslagf gx 4,355,725 jwkavwflk, ak
```

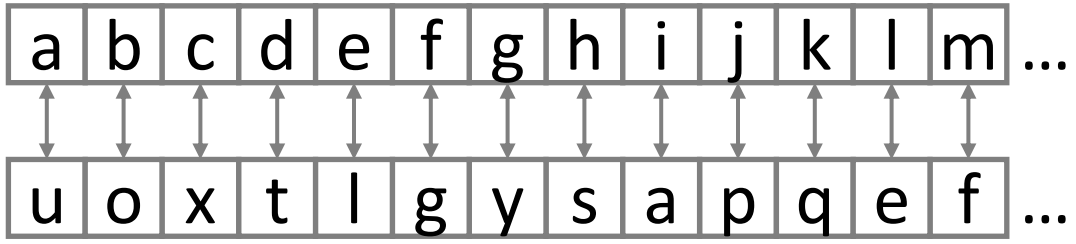
- Outputs:
  - **Key** that is the shift to apply to the alphabet to decrypt the ciphertext.
  - **Plaintext** decrypted from the ciphertext.

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 `wikipedia_cybersecurity.txt` to estimate of the English letter distribution.

`ciphertext_simple.txt`

```
vf pseygdtsbyce, b lxolgvgxgvdf pvycms vl b kmgcdj dr
mfpseygvft vf ucvpc xfvgl dr ywbvfgmng bsm smywbpmj
uvgc gcm pvycmsgmng, vf b jmrvfmj kbffms, uvgc gcm
cmwy dr b qme; gcm "xfvgl" kbe om lvftwm wmggmsl (gcm
kdlg pdkkdf), ybvsl dr wmggmsl, gsvywmgl dr wmggmsl,
kvngxsml dr gcm bodam, bfj ld rdsgc. gcm smpmvams
jmpvycmsl gcm gmng oe ymsrdskvft gcm vfamslm
lxolgvgxgvdf ysdpmll gd mngsbpg gcm dsvtvfbw kmllbtm.
lxolgvgxgvdf pvycmsl pbf om pdkybsmj uvgc
gsbflydlvgvdf pvycmsl. vf b gsbflydlvgvdf pvycms, gcm
xfvgl dr gcm ywbvfgmng bsm smbssbftmj vf b jvrrmsmfg
bfj xlxbwwe hxvgm pdkywmn dsjms, oxg gcm xfvgl
gcmklmwaml bsm wmrg xfpcbftmj. oe pdfgsblg, vf b
lxolgvgxgvdf pvycms, gcm xfvgl dr gcm ywbvfgmng bsm
smgbvfmj vf gcm lbkm lmhxmfpm vf gcm pvycmsgmng, oxg
gcm xfvgl gcmklmwaml bsm bwgmsmj.
```

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

Deadline

Tuesday, March 18th at 12PM (noon)