Cryptography (and Information Security)
6CCS3CIS / 7CCSMCIS

Prof. Luca Viganò

Department of Informatics
King's College London, UK

First term 2020/21

Lecture 3.2: Substitution ciphers — Homophonic substitution ciphers, Playfair cipher, Polyalphabetic substitution ciphers (Vigenère cipher)

Table of contents I

Table of contents I

Table of contents I

Table of contents I

Homophonic substitution ciphers

- Mono-alphabetic ciphers are easy to break because they reflect the frequency data of the original alphabet.
- Countermeasure: provide multiple substitutes (i.e., *homophones*) for a single letter to make frequency analysis more difficult.

Homophonic substitution cipher

- To each $a \in \mathcal{A}$ associate a set $H(a)$ of strings of t symbols, where $H(a), a \in \mathcal{A}$ are pairwise disjoint.
- Replace each a with a randomly chosen string from $H(a)$.
- To decrypt a string c of t symbols, one must determine an $a \in \mathcal{A}$ such that $c \in H(a)$.
- The key for the cipher is the sets $H(a)$.

- **Example:** $\mathcal{A} = \{x, y\}$, $H(x) = \{00, 10\}$, and $H(y) = \{01, 11\}$. The plaintext xy encrypts to one of 0001, 0011, 1001, 1011.
- Cost: data expansion and more work for decryption.

Cryptanalysis still relatively straightforward

- Cryptanalysis relatively straightforward even with homophones:
  - each element of plaintext affects only one element of ciphertext,
  - multiple-letter patterns (e.g., digram frequencies) still survive in the ciphertext.

- Two principal methods used in substitution ciphers to lessen the extent to which the structure of plaintext survives in ciphertext:
  1. encrypt multiple letters of plaintext, e.g., the **Playfair cipher** (or the **Hill cipher**, which we will not discuss here)
  2. use multiple cipher alphabets (polyalphabetic substitution), e.g., **Vigenère cipher**

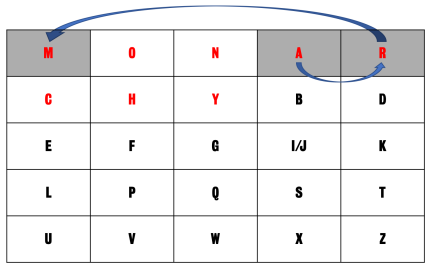
Table of contents I

Playfair cipher

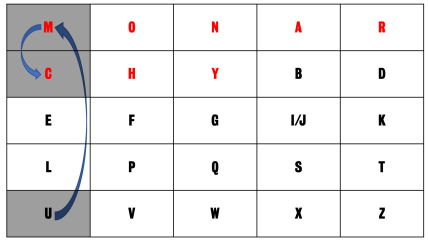
- **Uses a 5×5 matrix of letters constructed using a keyword.**
- Example (solved by Lord Peter Wimsey in Dorothy Sayers's "Have His Carcase"):
  - Pick keyword, here: monarchy.
  - Construct matrix: fill in letters of keyword (minus duplicates) left2right & top2bottom, and remaining letters in alphabetic order, where I and J count as one letter.
  - Plaintext is encrypted two letters at a time:

M	O	N	A	R
C	H	Y	B	D
E	F	G	I/J	K
L	P	Q	S	T
U	V	W	X	Z

1. If a pair is a repeated letter, insert filler like 'X' (e.g., "BALLOON" $\rightsquigarrow$ "BA LX LO ON"). Add an 'X' also at the end, if needed (or any other character, like the 'A' in this example).
2. If both letters fall in the same row, replace each with letter to right, wrapping back to start from end (e.g., "AR" is encrypted as "RM").



3. If both letters fall in the same column, replace each with the letter below it, wrapping to top from bottom (e.g., "MU" is encrypted as "CM").



4. Otherwise each letter is replaced by the letter in the same row and in the column of the other letter of the pair (e.g., "HS" becomes "BP" and "EA" becomes "IM", or "JM", as the encipherer wishes).

Playfair cipher (cont.)

- **Uses a 5×5 matrix of letters constructed using a keyword.**
- Example (solved by Lord Peter Wimsey in Dorothy Sayers's "Have His Carcase"):
  - Pick keyword, here: monarchy.
  - Construct matrix: fill in letters of keyword (minus duplicates) left2right & top2bottom, and remaining letters in alphabetic order, where I and J count as one letter.
  - Plaintext is encrypted two letters at a time:

M	O	N	A	R
C	H	Y	B	D
E	F	G	I/J	K
L	P	Q	S	T
U	V	W	X	Z

1. If a pair is a repeated letter, insert filler like 'X' (e.g., "BALLOON" $\rightsquigarrow$ "BA LX LO ON"). Add an 'X' also at the end, if needed (or any other character, like the 'A' in this example).
2. If both letters fall in the same row, replace each with letter to right, wrapping back to start from end (e.g., "AR" is encrypted as "RM").
3. If both letters fall in the same column, replace each with the letter below it, wrapping to top from bottom (e.g., "MU" is encrypted as "CM").
4. Otherwise each letter is replaced by the letter in the same row and in the column of the other letter of the pair (e.g., "HS" becomes "BP" and "EA" becomes "IM", or "JM", as the encipherer wishes).

Plaintext : TH EQ UI CK BR OW NF OX IU MP SO VE RT HE LA ZY DO G
Plaintext formatted: TH EQ UI CK BR OW NF OX IU MP SO VE RT HE LA ZY DO GA
Ciphertext : PD GL XE DE DA NV OG AV EX OL PA UF DZ CF SM WD HR IN

- To decrypt, use the inverse (opposite) of rules 2 and 3, and the 1st as-is (dropping any extra "X"s that do not make sense in the final message when finished) and the 4th as-is.

Playfair cipher: security

- Security much improved over monoalphabetic: $26 \times 26 = 676$ digrams vs. 26 letters.
- Would need a 676 entry frequency table to analyse, and correspondingly more ciphertext.
- Invented in 1854 by British scientist Sir Charles Wheatstone, but it bears the name of his friend Baron Playfair of St. Andrews, who championed the cipher at the British foreign office).
- Playfair cipher was for a long time considered unbreakable:
  - used as the standard field system by British Army in WWI and still partly used by U.S. Army and other Allied forces during WWII.
- However, breaking it is relatively easy:
  - still leaves much of the structure of the plaintext language intact
  - a few hundred letters of ciphertext are generally sufficient

Table of contents I

Polyalphabetic substitution ciphers

Polyalphabetic substitution cipher

- **Use different monoalphabetic substitutions as one proceeds through the plaintext message.**
  Idea (L.B. Alberti): conceal distribution using family of mappings.
- Two general features common to all such ciphers:
  1. A set of related monoalphabetic substitution rules is used.
  2. A key determines which particular rule is chosen for a given transformation.

- Best known, and perhaps simplest, is the **Vigenère cipher**, by Blaise de Vigenère, from court of Henry III of France (16^{th} century).
  - Set of related monoalphabetic substitution rules consists of the 26 Caesar ciphers with shifts of 0 through 25.
  - Each cipher is denoted by a key letter, which is the ciphertext letter that substitutes for the plaintext letter A.
    Thus, a Caesar cipher with a shift of 3 is denoted by key value D.

Vigenère cipher: idea

A polyalphabetic substitution cipher based on a *tableau* where each row is a Caesar Cipher with incremental shift:

	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
A	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
B	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	A
C	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
D	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
E	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	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	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

Vigenère cipher: details

- Assume
  - a sequence of plaintext letters $P = p_0, p_1, p_2, \ldots, p_{n-1}$,
  - a key consisting of the sequence of letters $K = k_0, k_1, k_2, \ldots, k_{m-1}$.

  Typically $m < n$.
- Sequence of ciphertext letters $C = C_0, C_1, C_2, \ldots, C_{n-1}$:

$$
\begin{aligned}
C &= C_0, C_1, C_2, \ldots, C_{n-1} \\
&= E(K, P) \\
&= E((k_0, k_1, k_2, \ldots, k_{m-1}), (p_0, p_1, p_2, \ldots, p_{n-1})) \\
&= (p_0 + k_0) \mod 26, (p_1 + k_1) \mod 26, \ldots, \\
&\quad (p_{m-1} + k_{m-1}) \mod 26, (p_m + k_0) \mod 26, (p_{m+1} + k_1) \mod 26, \\
&\quad \ldots (p_{2m-1} + k_{m-1}) \mod 26, \ldots \\
C_i &= (p_i + k_{i \bmod m}) \mod 26
\end{aligned}
$$

  - First letter of the key is added to first letter of plaintext, mod 26, second letters are added, and so on through the first m letters of plaintext.
  - For next m letters of the plaintext, the key letters are repeated.
  - Process continues until all of plaintext sequence is encrypted.
- Decryption: $p_i = (C_i - k_{i \bmod m}) \mod 26$

Vigenère cipher: example and strength

- Needs a key as long as message (usually: a repeating keyword).
- Example: if the keyword is “deceptive", the message “we are discovered save yourself" is encrypted as

```
key:          deceptivedeceptivedeceptive
plaintext:    wearediscoveredsaveyourself
ciphertext:   ZICVTWQNGRZGVTWAVZHCQYGLMGJ
```

Expressed numerically:

key	3	4	2	4	15	19	8	21	4	3	4	2	4	15
plaintext	22	4	0	17	4	3	8	18	2	14	21	4	17	4
ciphertext	25	8	2	21	19	22	16	13	6	17	25	6	21	19

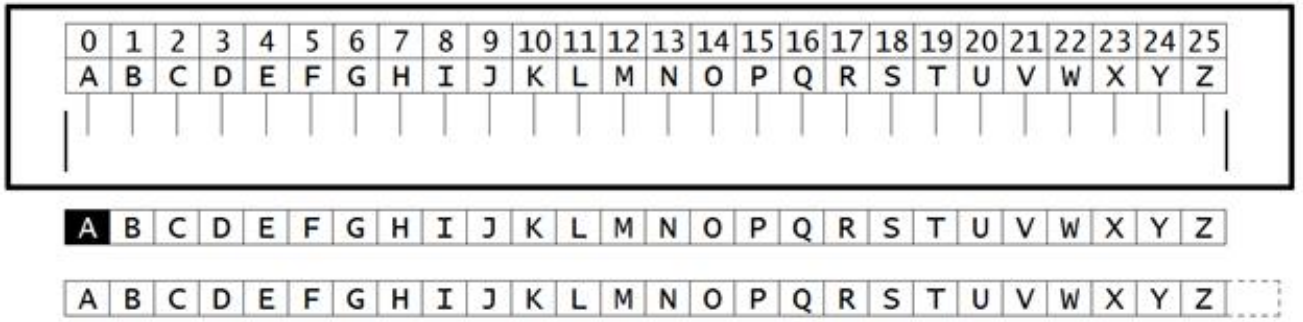
key	19	8	21	4	3	4	2	4	15	19	8	21	4
plaintext	3	18	0	21	4	24	14	20	17	18	4	11	5
ciphertext	22	0	21	25	7	2	16	24	6	11	12	6	9

- Strength: multiple ciphertext letters for each plaintext letter, one for each unique letter of the keyword.
  - Hence, letter frequency information is obscured.
- However, not all knowledge of the plaintext structure is lost.
  - Better than Playfair, but considerable frequency info remains.

Aids for en/de-cryption with polyalphabetic ciphers

- Implementing polyalphabetic ciphers by hand can be very tedious.
- Various aids were devised to assist the process, e.g., **Saint-Cyr Slide** is a simple manual aid:
  - a slide with repeated alphabet,
  - line up plaintext “A" with key letter, e.g., “C",
  - then read off any mapping for key letter.

Can bend round into a cipher disk or expand into a Vigenère Tableau.



(J.-G.-H.-V.-F.-A.-A. Kerckhoffs von Nieuwenhof popularised and named it, after the French National Military Academy where the methods were taught)

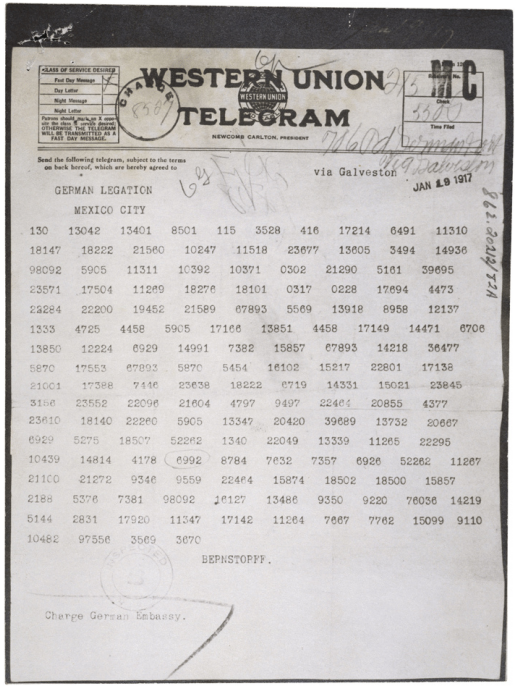
Kasiski Method

- For some centuries the Vigenère cipher was le *chiffre indéchiffrable* (the unbreakable cipher).
- Broken by Charles Babbage (“inventor" of the computer) in 1854 but kept secret (possibly because of the Crimean War).
- Method independently reinvented by Friedrich Kasiski (Prussia, 1863).
  - repetitions in ciphertext give clues to period
  - so find same plaintext an exact period apart which results in the same ciphertext
  - of course, could also be random fluke
  - e.g., repeated “VTW” in previous example
  - suggests size of 3 or 9
  - then attack each monoalphabetic cipher individually using same techniques as before
- See Stalling’s “Cryptography and Network Security” book for a sketch of a method of breaking Vigenère.

Zimmermann telegram: a little bit of history

- However, lack of major advances meant that various polyalphabetic substitution ciphers were used into the 20th century.
- One very famous incident was the breaking of the Zimmermann telegram in WWI which resulted in the USA entering the war.
  - A 1917 diplomatic proposal from the German Empire for Mexico to join in alliance if USA entered WWI against Germany.
  - Intercepted and decoded by British cryptographers of Room 40.

"We intend to begin on the first of February unrestricted submarine warfare. We shall endeavor in spite of this to keep the United States of America neutral. In the event of this not succeeding, we make Mexico a proposal of alliance on the following basis: make war together, make peace together, generous financial support and an understanding on our part that Mexico is to reconquer the lost territory in Texas, New Mexico, and Arizona. The settlement in detail is left to you. You will inform the President of the above most secretly as soon as the outbreak of war with the United States of America is certain and add the suggestion that he should, on his own initiative, invite Japan to immediate adherence and at the same time mediate between Japan and ourselves. Please call the President's attention to the fact that the ruthless employment of our submarines now offers the prospect of compelling England in a few months to make peace." Signed, ZIMMERMANN


WESTERN UNION TELEGRAM

via Galveston

JAN 19 1917

GERMAN LEGATION
MEXICO CITY

130 13042 13401 8501 115 3528 416 17214 6491 11310
18147 18222 21560 10247 11518 23677 13605 3494 14936
98092 5905 11311 10392 10371 0302 21290 5161 39695
23571 17504 11269 18276 18101 0317 0228 17694 4473
23284 22200 19452 21589 67893 5569 13918 8958 12137
1333 4725 4458 5905 17166 13851 4458 17149 14471 6706
13850 12224 6929 14991 7382 15857 67893 14218 36477
5870 17553 67893 5870 5454 16102 15217 22801 17138
21001 17388 7110 23638 18222 6719 14331 15021 23845
3156 23552 22096 21604 4797 9497 22464 20855 4377
23610 18140 22260 5905 13347 20420 39689 13732 20667
6929 5275 18507 52262 1340 22049 13339 11265 22295
10439 14814 4178 6992 8784 7632 7357 6926 52262 11267
21100 21272 9346 9559 22464 15874 18502 18500 15857
2188 5376 7381 98092 16127 13486 9350 9220 76036 14219
5144 2831 17920 11347 17142 11264 7667 7762 15099 9110
10482 97556 3569 3670

BERNSTORFF.

Charge German Embassy.