Cryptography and computer security

Historical cryptography

Cont.

Monoalphabetic substitution - Polygram ciphers

- Polygram ciphers (polygram encryption algorithms)
  - the plaintext character is encrypted into the character group of ciphertext
  - or character group → to character group
- Polybois Square (https://en.wikipedia.org/wiki/Polybius_square)

SQUARE →DCDADEAADBAE

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

Plybius square (2)

- Another example of Polybius square
- K = EAGLES
- P = GHPKFNIG
- C = 13 25 41 32 24 34 31 13

	1	2	3	4	5
1	E	A	G	L	S
2	B	C	D	F	H
3	I	K	M	N	O
4	P	Q	R	T	U
5	V	W	X	Y	Z

G	H	P	K	F	N	I	G
13							

Monoalphabetic substitution - Polygram ciphers

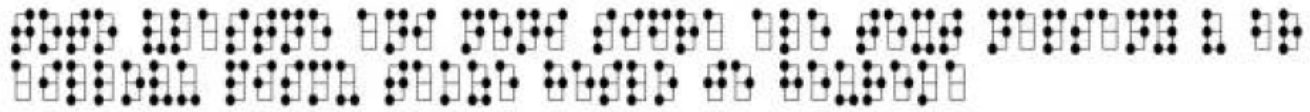
- Bigram to bigram - Playfair, letters from each of the bigrams can appear in three positions in the square: on the same row, on the same column or on a different row and column:
    - If both letters of the bigram are on the same row, they are replaced by the letters to the right of them. If one of the letters is the last one in the row, it is replaced by the first one in the same row.
    - If both letters are in the same column, they are replaced by the letters below them. If one of the letters is the last in a column, it is replaced by the first in the same column.
    - If the two letters are in different rows and columns, each is replaced by the letter at the intersection of the row of that letter and the column of the other letter.
  - See https://en.wikipedia.org/wiki/Playfair_cipher for an example.

Monoalphabetic substitution - different alphabets

- There is no need to replace letters with letters again. We can use arbitrary characters, i.e. the **plain and crypto alphabet can be different**
- Example 1: plaintext (PT) is in English without a space.
- Ciphertext (CT):

```
5 3 ‡ ‡ † 3 0 5 ) ) 6 * ; 4 8 2 6 ) 4 ‡ . ) 4 ‡ ) ; 8 0 6 *
; 4 8 † 8 ¶ 6 0 ) ) 8 5 ; 1 ‡ ( ; : ‡ * 8 † 8 3 ( 8 8 ) 5 *
† ; 4 6 ( ; 8 8 * 9 6 * ? ; 8 ) * ‡ ( ; 4 8 5 ) ; 5 * † 2 :
* ‡ ( ; 4 9 5 6 * 2 ( 5 * - 4 ) 8 ¶ 8 * ; 4 0 6 9 2 8 5 ) ;
) 6 † 8 ) 4 ‡ ‡ ; 1 ( ‡ 9 ; 4 8 0 8 1 ; 8 : 8 ‡ 1 ; 4 8 † 8
5 ; 4 ) 4 8 5 † 5 2 8 8 0 6 * 8 1 ( ‡ 9 ; 4 8 ; ( 8 8 ; 4 (
‡ ? 3 4 ; 4 8 ) 4 ‡ ; 1 6 1 ; : 1 8 8 ; ‡ ? ;
```

- The solution in the short story The Gold-Bug, Edgar Allan Poe, e.g. http://en.wikipedia.org/wiki/The_Gold-Bug
- Example 2:



- This is a coding where Braille blind coding is used to code the letter.

Polyalphabetic substitution

- Multiple substitutions are used for each letter of the plaintext alphabet, i.e. multiple ciphertext alphabets.
- Which ciphertext alphabet is used depends on the key.
- Better hides frequency dependencies, provides better frequency distribution of ciphertext characters.

The Vigenère cipher

- Blaise de Vigenère - The Vigenère cipher
  - le chiffre indéchiffrable
  - published in "*Traicté des Chiffres*" in 1585 a polyalphabetic cipher, which was not broken until 1854 (1863)
  - broken by Ch. Babbage and F. Kasiski respectively.
  - Cryptanalysis, known as the Kasiski test, is based on the assumption that the key is shorter than the plaintext and thus must be used repeatedly.

The Vigenère cipher

```
   ABCDEFGHIJKLMNOPQRSTUVWXYZ
A | ABCDEFGHIJKLMNOPQRSTUVWXYZ |
B | BCDEFGHIJKLMNOPQRSTUVWXYZA |
C | CDEFGHIJKLMNOPQRSTUVWXYZAB |
D | DEFGHIJKLMNOPQRSTUVWXYZABC |
E | EFGHIJKLMNOPQRSTUVWXYZABCD |
F | FGHIJKLMNOPQRSTUVWXYZABCDE |
G | GHIJKLMNOPQRSTUVWXYZABCDEF |
H | HIJKLMNOPQRSTUVWXYZABCDEFG |
I | IJKLMNOPQRSTUVWXYZABCDEFGH |
J | JKLMNOPQRSTUVWXYZABCDEFGHI |
K | KLMNOPQRSTUVWXYZABCDEFGHIJ |
L | LMNOPQRSTUVWXYZABCDEFGHIJK |
M | MNOPQRSTUVWXYZABCDEFGHIJKL |
N | NOPQRSTUVWXYZABCDEFGHIJKLM |
O | OPQRSTUVWXYZABCDEFGHIJKLMN |
...
```

```
M = THISP ROCES SCANA LSOBE EXPRE SSED
K = CIPHE RCIPH ERCIP HERCI PHERC IPHE
C = VPXZT IQKTZ WTCVP SWFDM TETIG AHLH

  C -> CDEFGHIJKLMNOPQRSTUVWXYZAB
  I -> IJKLMNOPQRSTUVWXYZABCDEFGH
  P -> PQRSTUVWXYZABCDEFGHIJKLMNO
  H -> HIJKLMNOPQRSTUVWXYZABCDEFG
  E -> EFGHIJKLMNOPQRSTUVWXYZABCD
  R -> RSTUVWXYZABCDEFGHIJKLMNOPQ

'T' key 'C' maps to 'V'
'H' key 'I' maps to 'P'
'I' key 'P' maps to 'X' etc.
```

Polyalphabetic substitution

- M=C=K=$(Z)_{26}^{n}$, the message m is encrypted in blocks of n characters,
- The key is a string of n characters, k=(k_1 , k_2 ,…,k_n),
- $e_k(m_1, m_2, ..., m_n) = (m_1 + k_1) \bmod 26, ..., (m_n + k_n) \bmod 26,$
- $d_k(c_1, c_2, ..., c_n) = (c_1 - k_1) \bmod 26, ..., (c_n - k_n) \bmod 26.$
- The number of all different keys is 26^n ,

Plaintext: `attackatdawn`

Key: `LEMONLEMONLE`

Ciphertext: `LXFOPVEFRNHR`

$$c_1 = A + L \bmod 26,\ c_2 = T + E \bmod 26...,$$

$$\text{i.e. } c_1 = 0 + 11 \bmod 26,\ c_2 = 19 + 4 \bmod 26...,$$

Autokey Cipher

- Vigenère proposed "the **autokey** cipher" (he wanted to find a way to create a key as long as the plaintext)
- Keyword DECEPTIVE

```
M = WEAREDISCOVEREDSAVEYOURSELF
K = DECEPTIVEWEAREDISCOVEREDSAV
C = ZICVTWQNGKZEIIGASXSTSLVVWLA
```

Hill's cipher

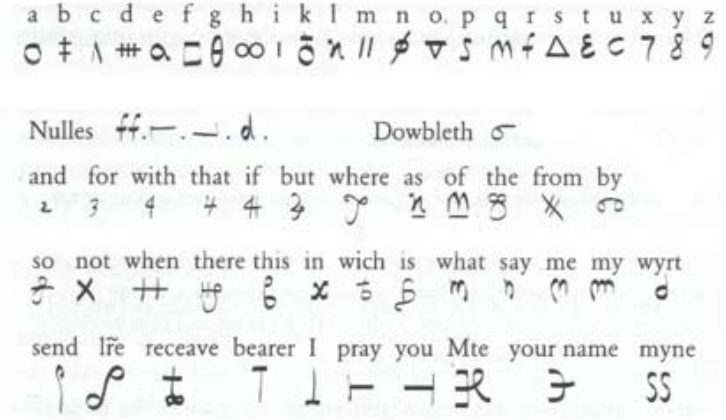
- Polygram cipher (Lester Hill 1929), counting with matrices
- $C = E_k(M) = K*M \bmod 26$
- $M = D_k(C) = K^{-1}*C \bmod 26 = K*K^{-1} *M = M$
- C and M are column vectors of length n (block size n),
- matrix K is the matrix of the key n*n, n represents the number of characters in the group
- $KK^{-1} \bmod 26 = I$, where I is the identity matrix
- Hides frequency dependencies (even of digrams, trigrams)
- https://en.wikipedia.org/wiki/Hill_cipher
- https://www.geeksforgeeks.org/hill-cipher/

Homophonic cipher

- A substitution cipher that assigns to each plaintext character one of a set of possible different ciphertext characters.
- The number of potential substitutes being proportional to the frequency of the letter.
  - For example, the letter 'a' accounts for roughly 8% of all letters in English, so we assign 8 symbols to represent it. Each time an 'a' appears in the plaintext it is replaced by one of the 8 symbols chosen at random, and so by the end of the encryption each symbol constitutes roughly 1% of the ciphertext.
  - The letter 'b' accounts for 2% of all letters and so we assign 2 symbols to represent it. Each time 'b' appears in the plaintext either of the two symbols can be chosen, so each symbol will also constitute roughly 1% of the ciphertext. This process continues throughout the alphabet, until we get to 'z', which is so rare that is has only one substitute.....,
- after encryption, the frequency of each character will be approximately the same.
- https://www.simonsingh.net/The_Black_Chamber/homophonic_cipher.html

Nomenclatures

- Nomenclature
  - is an enhancement of the homophonic cipher - it adds additional code equivalents for the most commonly used words, syllables and names.
  - Nomenclatures gradually grew until they contained thousands of code names and words, providing the basis for the emergence of codebook encryption.
  - The disadvantage of nomenclatures - they were not frequently replaced (one person used his nomenclature for his whole life).
  - The advantage - simplicity and speed of use.



Code books

- Code - a specific cryptographic system that works with linguistic (language) elements. These elements can be selected words, whole sentences or clauses. For example, the code of *an egg* can mean a grenade, ...
  - If the meaning of the codes is publicly known (e.g. radio Q-code, where QPA means "The passcode is ... . ", QTC "How many telegrams do you have to transmit ?"), it is a code in the classical sense.
    - https://en.wikipedia.org/wiki/Q_code
  - Morse code https://en.wikipedia.org/wiki/Morse_code
- If the meaning of the codes is kept secret, it is a special encryption system called a **codebook** (it was even the most used method of encryption during World War I & II).
  - https://en.wikipedia.org/wiki/Code_talker
- Zimmermann telegram (German diplomatic code system 13042)
  - https://en.wikipedia.org/wiki/Zimmermann_Telegram

Code books

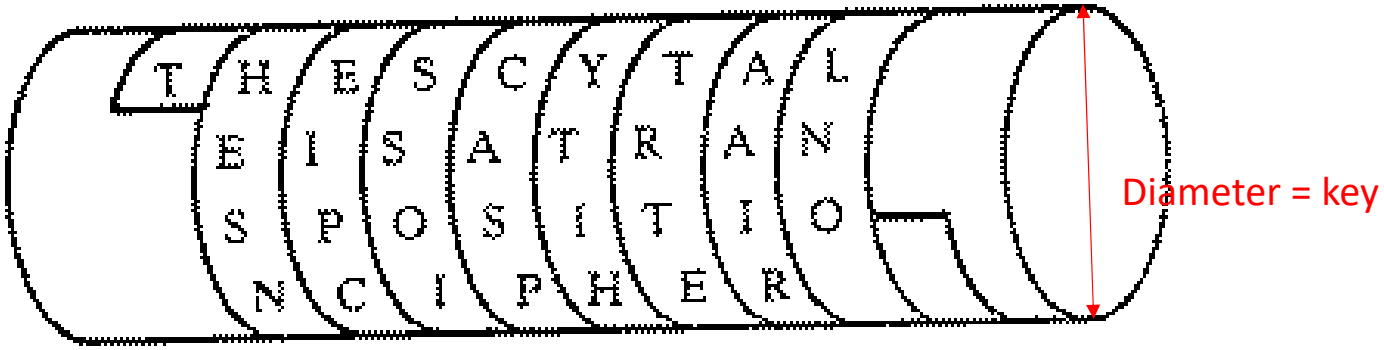
- They are convenient for situations that are anticipated in their preparation, and therefore contain code equivalents only for selected plaintexts (but encryption algorithms are suitable for any situation, since they can convert any plaintext into ciphertext).
- Disadvantages
  - long and demanding preparation of quality code
  - the need to ensure perfect secrecy in the printing of these books
  - (costly) secure and fast distribution of codebooks to end users is required
  - the need to use one type of codebook for a long time (replacing it with a new one is expensive and complicated)
  - the "loss" of a single copy compromises the entire system and it is necessary to switch immediately to a new codebook

Transposition cipher

- **Transposition** or permutation ciphers
- These hide the message by rearranging the letter order
- Without altering the actual letters used
- They can be recognized because ciphertext letter frequencies are the same as plaintext letter frequencies.
- Eg. reverse (mirror) cipher
  - Write the message backwards
  - Plaintext: `ICAMEISAWICONQUERED`
  - Ciphertext: `DEREUQNOCIWASIEMACI`

Transposition cipher

- **Scytale cipher -** an early Greek transposition cipher (500 BC)
- A strip of paper (parchment) was wound round a staff
- Message written along staff in rows, then paper removed
- Leaving a strip of seemingly random letters
- Not very secure as key was width of paper & staff



Rail Fence cipher

- Write plaintext with letters on alternate rows
- Read off ciphertext row by row

  Plaintext: `I A E S W C N U R D`

  `C M I A I O Q E E`

  Ciphertext: `IAESWCNURDCMIAIOQEE`

- To decrypt, write half the letters on one line, half on the second (if there is an odd number of letters, include the “middle” letter on the top line).
- Key – the number of rows

Key Concept for Transposition Ciphers

- In a transposition cipher the key idea is that you:
  - **Write** the plaintext out in columns according to some rule
  - R**ead** the letters off to form the ciphertext according to another rule
- Key used to find order to
  - read off the ciphertext, write in the plaintext, or both
- Encryption key - any permutation
- Decryption key – inverse permutation to encryption key
  - Inverse permutation https://www.geeksforgeeks.org/inverse-permutation/ https://www.youtube.com/watch?v=UY25Nju9yMQ

Columnar transposition

- Columnar transposition is probably the most commonly studied transposition cipher

```
M = THESI MPLES TPOSS IBLET RANSP OSITI ONSXX
K(E): 41532
K(D): 25413
            T H E S I              S T I E H
            M P L E S              E M S L P
            T P O S S              S T S O P
            I B L E T              E I T L B
            R A N S P              S R P N A
            O S I T I              T O I I S
            O N S X X              X O X S N
C= SESESTXTMTIRPOOISSTPIXELOLNISHPPBASN
```

- The plaintext has been padded so that it neatly fits in a rectangle. This is known as a regular columnar transposition.
- An irregular columnar transposition leaves these characters blank, though this makes decryption slightly more difficult.

Row Transposition ciphers

- Group the plaintext and shuffle letters within each group
  - More formally write letters across rows
- Then reorder the columns before reading off the rows
- Always have an equivalent pair of keys (write in (encrypt) vs read off (decrypt))

Row Transposition ciphers

M = THESI MPLES TPOSS IBLET RANSP OSITI ONSXX
K(E): 41532
K(D): 25413

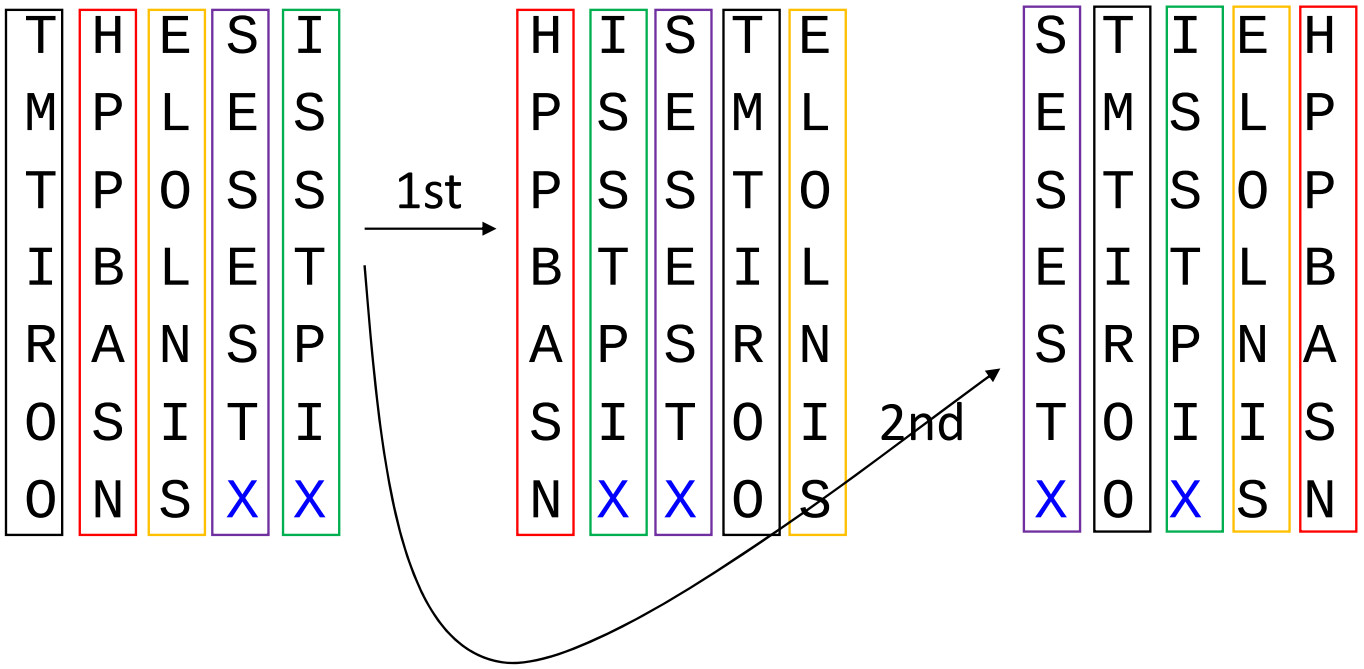
T	H	E	S	I		S	T	I	E	H
M	P	L	E	S		E	M	S	L	P
T	P	O	S	S		S	T	S	O	P
I	B	L	E	T		E	I	T	L	B
R	A	N	S	P		S	R	P	N	A
O	S	I	T	I		T	O	I	I	S
O	N	S	X	X		X	O	X	S	N

C= STIEH EMSLP STSOP EITLB SRPNA TOIIS XOXSN

Different ways how to use (understand) permutations

Plain: THESIMPLESTPOSSIBLETRANSPOSITIONSXX

Key: 4 1 5 3 2



Increasing Cipher Security

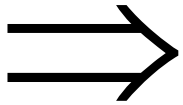
- Ciphers based on just substitutions or transpositions are not secure, because they do not sufficient obscure the underlying language structure
- Hence consider using several ciphers in succession to make harder, but:
  - two substitutions are really only one more complex substitution
  - two transpositions are really only one more complex transposition
  - but a substitution followed by a transposition makes a new „much harder“ cipher

Double Transposition

- Plaintext: attackxatxdawn

	col 1	col 2	col 3
row 1	a	t	t
row 2	a	c	k
row 3	x	a	t
row 4	x	d	a
row 5	w	n	x

Permute rows and columns



	col 1	col 3	col 2
row 3	x	t	a
row 5	w	x	n
row 1	a	t	t
row 4	x	a	d
row 2	a	k	c

- Key is matrix size and permutations: (3,5,1,4,2) and (1,3,2)
- Ciphertext: xtawxnattxadakc

Product ciphers - ADFGVX

- Are substitution-transposition ciphers chained together
- In general are far too hard to do by hand
- However one famous product cipher **ADFGVX** cipher was used in WW1
  - Named since only letters ADFGVX are used
  - https://en.wikipedia.org/wiki/ADFGVX_cipher
  - Used by the German's and broken by the British in WW1
  - Uses a fixed substitution table to map each plaintext letter to a letter pair (row-col index)
  - Then uses a keyed block transposition to split letter pairs up
  - Ciphertext then written in blocks and sent

ADFGVX Product Cipher

- Substitution Table (random)

\\	A	D	F	G	V	X
A	K	Z	W	R	1	F
D	9	B	6	C	L	5
F	Q	7	J	P	G	X
G	E	V	Y	3	A	N
V	8	O	D	H	0	2
X	U	4	I	S	T	M

ADFGVX Product Cipher

- Encryption

Plaintext: PRODUCTCIPHERS

Intermediate Text: FG AG VD VF XA DG XV DG XF FG VG GA AG XG

Keyed Block Columnlar Transposition Matrix

	D	E	U	T	S	C	H
Key	2	3	7	6	5	1	4
Sorted Order	F	G	A	G	V	D	V
	F	X	A	D	G	X	V
	D	G	X	F	F	G	V
	G	G	A	A	G	X	G

Ciphertext: DXGX FFDG GXGG VVVG VGFG CDFA AAXA