Cryptology = science of secrecy.

How

encipher a plaintext into a ciphertext to protect its secrecy. The recipient deciphers the ciphertext to recover the plaintext. A cryptanalyst shouldn't complete a successful cryptanalysis. Attacks [6]:

- known ciphertext : access only to the ciphertext
- known plaintexts/ciphertexts: known pairs (plaintext,ciphertext); search for the key
- chosen plaintext: known cipher, chosen cleartexts; search for the key

Short history

- J. Stern [8]: 3 ages:
 - craft age : hieroglyph, bible, ..., renaissance, → WW2
 - technical age : complex cipher machines
 - paradoxical age : PKC

Evolves through maths' history, computing and cryptanalysis:

- manual
- electro-mechanical
- by computer

Polybius's square

Polybius, Ancient Greece: communication with torches

	1	2	3	4	5
1	а	b	С	d	е
2	f	g	h	ij	k
3	ı	m	n	0	р
4	q	r	s	t	u
5	٧	W	Χ	У	Z

TEXT changed in 44,15,53,44. Characteristics

- encoding letters by numbers
- shorten the alphabet's size

encode a character x over alphabet A in y finite word over B. Polybius square : $\{a, \ldots, z\} \rightarrow \{1, \ldots, 5\}^2$.

History – ancient Greece

500 BC: scytale of Sparta's generals



Secret key: diameter of the stick

History - Caesar



Change each char by a char 3 positions farther
A becomes d, B becomes e...
The plaintext TOUTE LA GAULE becomes wrxwh od jdxoh.

Why enciphering?

- Yesterday :
 - for strategic purposes (the enemy shouldn't be able to read messages)
 - by the church
 - diplomacy
- Today, with our numerical environment
 - confidentiality
 - integrity
 - authentication

Goals of cryptology

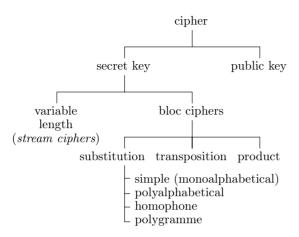
Increasing number of goals:

- secrecy: an enemy shouldn't gain access to information
- authentication: provides evidence that the message comes from its claimed sender
- signature: same as auth but for a third party
- minimality: encipher only what is needed.

The tools

- Information Theory : perfect cipher
- Complexity: most of the ciphers just ensure computational security
- Computer science: all make use of algorithms
- Mathematics: number theory, probability, statistics, algebra, algebraic geometry,...

Ciphers Classification



Symmetrical ciphers

Made of [1]:

plaintext alphabet : $A_{\mathcal{M}}$

lack ciphertext alphabet : $\mathcal{A}_{\mathcal{C}}$

• keys alphabet : A_K

• encipher; application $E: \mathcal{A}_{\mathcal{K}}^{\star} \times \mathcal{A}_{\mathcal{M}}^{\star} \to \mathcal{A}_{\mathcal{C}}^{\star}$;

 $_{\bullet}$ decipher; application $\mathit{D}:\mathcal{A}_{\mathcal{K}}^{\star}\times\mathcal{A}_{\mathcal{C}}^{\star}\rightarrow\mathcal{A}_{\mathcal{M}}^{\star}$

E and *D* are such that $\forall K \in \mathcal{A}_{\mathcal{K}}^{\star}, \forall M \in \mathcal{A}_{\mathcal{M}}^{\star}$:

$$D(K, E(K, M)) = M$$

Monoalphabetical ciphers

Monoalphabetical cipher: bijection between letters from A_M and A_C . If both alphabets are identical: permutation.

Example : Caesar. {a,...,z} \equiv {A,...,Z} \equiv {0,...,25} = \mathbb{Z}_{26}

Caesar cipher is additive.

Encipher: $\forall x \in \mathbb{Z}_{26}, x \mapsto x+3 \mod 26$ Decipher: $\forall y \in \mathbb{Z}_{26}, y \mapsto y-3 \mod 26$

Multiplicative cipher

We consider : $x \mapsto t \cdot x \mod 26$ for $t \in \mathbb{N}$. Acceptable values of t are s.t. $\gcd(t,26) = 1 \Leftrightarrow t \nmid 26$. $\varphi(26)$ acceptables values $\{1,3,5,7,9,11,15,17,19,21,23,25\}$ Other values don't ensure the uniqueness of the deciphering (e.g. 2)

а	b	С	d	е	f	g	h	i	j	k	- 1	m	
0	1	2	3	4	5	6	7	8	9	10	11	12	
n	0	р	q	r	S	t	u	٧	W	Х	У	Z	
13	14	15	16	17	18	19	20	21	22	23	24	25	
0	2	4	6	8	10	12	14	16	18	20	22	24	

To decipher, we require the existence of t^{-1} modulo 26. We use the extended Euclidean algorithm which provides Bezout coefficients i.e. $x, y \in \mathbb{N}$ st. $d = \gcd(a, b) = ax + by$. From Bezout coefficients, one can deduce t^{-1} modulo 26:

$$\gcd(t,26) = 1 \Leftrightarrow \exists x,y \in \mathbb{N} : tx + 26y = 1 \Leftrightarrow x \equiv t^{-1} \mod 26$$

Iterative computation

```
Extended Euclidean(q,r) with q < r Q \leftarrow (1,0); R \leftarrow (0,1); while r \neq 0 do t \leftarrow q \mod r; T \leftarrow Q - \lfloor q/r \rfloor R; (q,r) \leftarrow (r,t); (Q,R) \leftarrow (R,T); end return (q,Q); q: gcd value and Q provides the coeffs. end
```

Extended Euclidean (11, 26)

q	r	t	Q	$\lfloor q/r \rfloor$	R	Τ
11	26	11	(1,0)	0	(0, 1)	(1,0)
26	11	4	(0,1)	2	(1,0)	(-2, 1)
11	4	3	(1,0)	2	(-2,1)	(5, -2)
4	3	1	(-2,1)	1	(5, -2)	(-7, 3)
3	1	0	(5, -2)	3	(-7,3)	(26, -11)
1	0		(-7,3)		(26, -11)	

pgcd(11, 26) = 1 and Bezout's coefficients are (-7, 3). The mult, inverse of 11 mod 26 = -7 = 19.

Affines Ciphers

When combining 26 additive ciphers and 12 multiplicative ones, we get **affine** ciphers:

given s and $t \in \mathbb{N}$, encipher with : $x \mapsto (x + s) \cdot t \mod 26$. The key is the pair (s, t) and the deciphering is done by applying successively the previous methods.

There are 26.12=312 possible affine ciphers. Far from the 26!=403291461126605635584000000 possible ones.

Ciphers defined by keyword

To get all possible monoalphabetical ciphers by :

- a keyword like, for instance CRYPTANALYSIS;
- a key letter like e.

Remove multiple occurrences of the same letter in the keyword -here CRYPTANLSI- then

abcdefghijklmnopqrstuvwxyz VWXZCRYPTANLSIBEDFGHJKMOQU

Cryptanalysis

Shannon: a small proportion of letters provides more information than the remaining 2/3 of the text.

By applying a frequency analysis on the letters then of bigrams, ... in the ciphertext.

Solving $ax \equiv b \mod n$

We have used the method for solving the integer equation $ax \equiv b \mod n$. There are two cases :

- $gcd(a, n) = 1 : ax \equiv b \mod n \Leftrightarrow x \equiv a^{-1}b \mod n$ with a^{-1} given by the extended Euclidean algorithm.
- $gcd(a, n) = d \neq 1$ splits into two new cases :
 - $ightharpoonup d \nmid b$, the equation has no solution;
 - ▶ $d|b|ax \equiv b \mod n \Leftrightarrow da'x \equiv db' \mod dn'$. We divide lhs and rhs by d and we solve $a'x \equiv b' \mod n'$. We get a set of solutions : $\{x = a'^{-1}b' + kn' : 0 \le k < d\}$.

Conclusion

Monoalphabetical ciphers aren't robust against a frequency analysis.

We need ciphers for which the statistical distribution of the letters tend to be a uniform one.

1.st attempt: use a crypto transformation which associates a set of distinct letters in the ciphertext to the plaintext letters.

We get what is called polyalphabetical ciphers

Vigenère's cipher (1586)

In a **polyalphabetical cipher**, plaintext characters are transformed by means of a key $K = k_0, \ldots, k_{j-1}$ which defines j distinct functions f_0, \ldots, f_{j-1} s.t.

$$\forall i, \ 0 < j \leq n \quad f_{k_l} : \mathcal{A}_M \mapsto \mathcal{A}_C, \forall l, \ 0 \leq l < j$$
$$c_i = f_{k_i \mod j}(m_i)$$

Idea: use *j* distinct monoalphabetical ciphers.

Vigenère's square

abcdefghijklmnopgrstuvwxyz abcdefghijklmnopgrstuvwxyz MNOPQRSTUVWXYZABCDEFGHIJKL ZABCDEFGHIJKLMNOPQRSTUVWXY

ABCDEFGHIJKLMNOPQRSTUVWXYZ NOPQRSTUVWXYZABCDEFGHIJKLM BCDEFGHIJKLMNOPQRSTUVWXYZA OPQRSTUVWXYZABCDEFGHIJKLMN CDEFGHIJKLMNOPQRSTUVWXYZAB PQRSTUVWXYZABCDEFGHIJKLMNO DEFGHIJKLMNOPQRSTUVWXYZABC QRSTUVWXYZABCDEFGHIJKLMNOP EFGHIJKLMNOPQRSTUVWXYZABCD RSTUVWXYZABCDEFGHIJKLMNOPQ FGHIJKLMNOPQRSTUVWXYZABCDE STUVWXYZABCDEFGHIJKLMNOPQR GHIJKLMNOPQRSTUVWXYZABCDEF TUVWXYZABCDEFGHIJKLMNOPQRS HIJKLMNOPQRSTUVWXYZABCDEFG UVWXYZABCDEFGHIJKLMNOPQRST IJKLMNOPQRSTUVWXYZABCDEFGH VWXYZABCDEFGHIJKLMNOPQRSTU JKLMNOPQRSTUVWXYZABCDEFGHI WXYZABCDEFGHIJKLMNOPQRSTUV KLMNOPQRSTUVWXYZABCDEFGHIJ XYZABCDEFGHIJKLMNOPQRSTUVW LMNOPQRSTUVWXYZABCDEFGHIJK YZABCDEFGHIJKLMNOPQRSTUVWX

polyalphabetique KSYSSGTUUTZXVKMZ VENUSVENUSVENUSV

Cryptanalysis...

... becomes more difficult; we tend to a uniform distribution.

But, if we re-arrange the ciphertext in a matrix with as many columns as the key length, all the letters in the same column come from the same monoalphabetical cipher.

Cryptanalysis works as follows:

- (1) find the key length
- (2) apply the previous methods

2 tests to find the key length: Kasiski and Friedman.

Homophone Ciphers

Goal: smooth the frequency distribution of the letters. The ciphertext alphabet contains several equivalents for the same plaintext letter.

We thus define a multiple representation substitution. Thus, letter e from the plaintext, instead of being always enciphered by a 4 could be replaced for instance by 37, 38, 39,

These different cryptographic units corresponding to the same plaintext character are called homophones.

letter	frequency	letter	frequency
а	0,26,27,28,29,30	n	13,68,69,70,71,72
b	1	0	14,73,74,75,76
С	2,31,32,33,34	р	15,77,78
d	3,35,36	q	16
е	4,37,,54	r	17,79,80,81,82
f	5,55	S	18,83,84,85,86,87
g	6,56	t	19,88,89,90,91,92,93
h	7,57	u	20,94,95,96,97
i	8,58,59,60,61,62	V	21
j	9	W	22
k	10	Х	23
I	11,63,64,65,66	у	24,98
m	12,67	Z	25

Transposition

Implements a permutation of the plaintext letters $A_C = A_M$.

$$\forall i, \quad 0 \leq i < 0 \quad f: \mathcal{A}_M \to \mathcal{A}_M \ \eta: \mathbb{Z}_n \to \mathbb{Z}_n \ c_i = f(m_i) = m_{n(i)}$$

Simple array transposition

Given a passphrase, we define a numerical key:

T₁₈ R₁₄ A₁ N₈ S₁₅ P₂ O₁ S₁ T₁ I₂ O₁ N₃ S₁ I₃ M₇ P₂ L₆ E₂
We encipher, «*le chiffrement est l'opération qui consiste à transformer un texte clair, ou libellé, en un autre texte inintelligible appelé texte chiffré ou chiffré» [5].*

18	14	1	8	15	12	10	16	3	19	4	11	9	17	5	7	13	6	2
1	е	С	h	i	f	f	r	е	m	е	n	t	е	s	t	- 1	0	р
é	r	а	t	i	0	n	q	u	i	С	0	n	s	i	s	t	е	à
t	r	а	n	S	f	0	r	m	е	r	u	n	t	е	х	t	е	С
1	a	i	r	0	u	- 1	i	b	е	- 1	- 1	é	е	n	u	n	а	u
t	r	е	t	е	Х	t	е	i	n	i	n	t	е	- 1	- 1	i	g	i
b	1	е	а	р	р	е	- 1	é	t	е	х	t	е	С	h	i	f	f
r	é	0	u	C	r	V	g	t	0	а	r	а	m	m	е			

Vernam cipher (1917)

Is the one-time pad a «perfect» cipher?

A and B share a true random sequence of n bits : the secret key K.

A enciphers M of n bits in $C = M \oplus K$. B deciphers C by $M = K \oplus C$.

Example

M = 0011, K = 0101 $C = 0011 \oplus 0101 = 0110$ $M = K \oplus C$.

Non-reusability: for every new message, we need a new key.

Why a new key?

... To avoid revealing information on the \oplus of plaintexts.

Eve can sniff $C = \{M\}_K$ and $C' = \{M'\}_K$ and computes :

$$C \oplus C' = (M \oplus K) \oplus (M' \oplus K) = M \oplus M'$$

Given enough ciphertexts, she's able to recover a plaintext by a frequency analysis and with the help of a dictionnary [4].

If we respect the above requirements, Vernam cipher guarantees the condition of perfect secrecy.

Condition (perfect secrecy)

$$Pr(M = m \mid C = c) = Pr(M = m)$$

Intercepting C doesn't reveal any information to the cryptanalyst

Why is it secure?

Vernam ciphers provides **perfect secrecy**. We have three classes of information :

- plaintexts *M* with proba. distribution $Pr(M)/\sum_{M} Pr(M) = 1$
- ciphertexts C with proba. distribution $Pr(C)/\sum_{C} Pr(C) = 1$
- keys with proba. distribution Pr(K) s.t. $\sum_{K} p(K) = 1$

 $Pr(M \mid C)$ = proba that M has been sent knowing that C was received (C is the corresponding ciphertext of M). The perfect secrecy condition is defined as

$$Pr(M \mid C) = Pr(M)$$

The interception of the ciphertext does not provide any information to the crypto-analyst.

Conclusion

Perfect secrecy but difficult to achieve

- generate truly random sequences
- store them and share them with the recipients

example of use: «red phone».

Product and iterated ciphers

Improvement: combine substitutions and transpositions

A cipher is **iterated** if the ciphertext is obtained from repeated applications of a round function to the plaintext At each round, we combine a round key with the plaintext.

Definition

In an iterated cipher with r rounds, the ciphertext is computed by repeated applications of a **round function** g to the plaintext:

$$C_i = g(C_{i-1}, K_i)$$
 $i = 1, ..., r$

 C_0 the plaintext, K_i round key and C_r the ciphertext. Deciphering is achieved by inverting the previous equation. For a fixed K_i , g must be invertible.

Special case, Feistel ciphers.

Feistel ciphers

A **Feistel cipher** with block size 2n and r rounds is defined by :

$$g: \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^m \to \{0,1\}^n \times \{0,1\}^n$$

 $X, Y, Z \mapsto (Y, F(Y, Z) \oplus X)$

g function of $2n \times m$ bits into 2n bits and \oplus denoting the n bit XOR **Operation mode**

Given a plaintext $P = (P^L, P^R)$ and r round keys K_1, \ldots, K_r , the ciphertext (C^L, C^R) is obtained after r rounds. Let $C^L_0 = P^L$ and $C^R_0 = P^R$ and we compute for $i = 1, \ldots, r$

$$(C_{i}^{L}, C_{i}^{R}) = (C_{i-1}^{R}, F(C_{i-1}^{R}, K_{i}) \oplus C_{i-1}^{L})$$

with $C_i = (C_i^L, C_i^R)$ and $C_r^R = C^L$ and $C_r^L = C^R$ The round keys K_1, \ldots, K_r , are obtained by a key scheduling algorithm on a master key K.

DES

- NBS launches a competition in 1973
- DES (Data Encryption Standard) proposed by IBM in 1975
- adopted in 1977
- security evaluation every 4 years
- replaced by AES or Rijndael [2]
- enciphering example of DES in STINSON's book [9]

DES usage

DES was (is?) widely used (banks, computer security systems with DES as the main component).

Feistel cipher with special properties.

Operation

DES receives as an input :

- a message M of 64 bits;
- a key K of 56 bits.

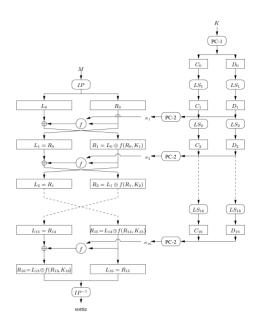
and outputs a ciphertext of 64 bits.

DES algorithms first applies to M an initial permutation IP which provides M', a perutation of M.

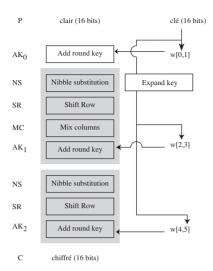
M' is then cut into two 32 bits words:

- L₀ the left part of M'
- R_0 its right part.

DES then applies 16 iterates of function f combining substitutions and transpositions.



AES or Rijndael[3]



State matrix & nibbles

1 nibble = 4 bits word (I/O of SAES components)

$$\frac{b_0b_1b_2b_3 | b_8b_9b_{10}b_{11}}{b_4b_5b_6b_7 | b_{12}b_{13}b_{15}b_{15}} = \frac{S_{0,0} | S_{0,1}}{S_{1,0}}$$

key representation :

$$\underbrace{k_0k_1\ldots k_7}_{w[0]} \quad \underbrace{k_8\ldots k_{1!}}_{w[1]}$$

Operations in $GF(16) \simeq \mathbb{F}_2[x]/x^4 + x + 1$

- $m(x) = x^4 + x + 1$ is an irreducible of \mathbb{F}_2
- elements : nibble $b_0b_1b_2b_3 \stackrel{\theta}{\leftrightarrow} b_0x^3 + b_1x^2 + b_2x + b_3$
- addition : by adding the coefficients : $(x^3 + x + 1) + (x^2 + 1)$
- multiplication: product of polynomials mod m(x)
- byte encoding : in a quadratic extension $\mathbb{F}_{16}[z]/z^2 + 1$
- beware ! $z^2 + 1$ is not invertible in GF(16)
- Reminder: to find the multiplicative inverse of an element:
 Extended Euclidean on polynomials

$$(x+1,m) = (x^3 + x^2 + x)(x+1) + 1 \cdot m$$

Inverses in \mathbb{F}_{16}

S-box used in Nibble substitution

•
$$\forall$$
 nibble : $\begin{array}{ccc} b_0b_1 & b_2b_3 \\ \hline row & column \end{array}$: 00 01 $\stackrel{\$}{\rightarrow}$ 01 00

$$\frac{0001 \mid 0001}{1100 \mid 1110} \xrightarrow{S} \frac{0100 \mid 0100}{1100 \mid 1111} = \frac{4 \mid 4}{c \mid f}$$

S-box algebraically

- 1. init the S-box with the nibbles arranged in a 1D array row by row
- 2. convert each nibble in a polynomial
- 3. invert each nibbble in \mathbb{F}_{16}
- 4. associate to the inverse its ploynomial in $\mathbb{F}_{16}[y]/y^4 1 = N(y)$
- 5. compute $a(y)N(y) + b(y) \mod y^4 + 1$ with $a = y^3 + y + 1$ et $b = y^3 + 1$

Normally
$$S(0011) = 1011 \equiv S(3) = b$$

Other transformations

Shift row : transposition of the nibble bits : $b_0b_1\ b_2b_3 \mapsto b_2b_3\ b_0b_1. \frac{4\mid 4}{c\mid f} \mapsto \frac{4\mid 4}{f\mid c}$

• Mix columns : modifies the polynomial representation of the state's rows $\frac{N_i \mid .}{N_j \mid .}$; we associate $c(z) = N_i z + N_j \in \mathbb{F}_{16}[z]/z^2 + 1$; compute $c(z).(x^2z + 1)$ mod $z^2 + 1$.

Example

For
$$4f \leftrightarrow 0100\ 11111 \mapsto c(z) = x^2z + x^3 + x^2 + x + 1$$
: $(x^3 + x^2 + 1)z + (x^3 + x^2) = N_k z + N_\ell \leftrightarrow 1101\ 1100$ because $z^2 = 1, x^4 = x + 1$ and $x^5 = x^2 + x$.

Mix columns (matrix)

We work directly on the state:

$$\begin{pmatrix} 1 & x^2 \\ x^2 & 1 \end{pmatrix} \cdot \begin{pmatrix} S_{0,0} & S_{0,1} \\ S_{1,0} & S_{1,1} \end{pmatrix} =_{\mathbb{F}_{16}} \begin{pmatrix} 1 & 4 \\ 4 & 1 \end{pmatrix} \cdot \begin{pmatrix} S_{0,0} & S_{0,1} \\ S_{1,0} & S_{1,1} \end{pmatrix}$$

Example

$$\begin{pmatrix} 1 & x^2 \\ x^2 & 1 \end{pmatrix} \cdot \begin{pmatrix} x^2 & x^2 \\ x^3 + x^2 + x + 1 & x^3 + x^2 \end{pmatrix} = \begin{pmatrix} x^3 + x^2 + 1 & 1 \\ x^3 + x^2 & x^3 + x^2 + x + 1 \end{pmatrix} = \begin{pmatrix} d & 1 \\ c & f \end{pmatrix}$$

Key scheduling

• initialisation :
$$w[0] = k_0 \dots k_7$$
 $w[1] = k_8 \dots k_{15}$
• $2 \le i \le 5$
$$\begin{cases} w[i] = w[i-2] \oplus \mathsf{RCON}(i/2) \oplus \mathsf{SubNib}(\mathsf{RotNib}(w[i-1])) & i \text{ even} \\ w[i] = w[i-2] \oplus w[i-1] & i \text{ odd} \end{cases}$$

With

- RCON[i]=RC[i]0000
- $RC[i] = x^{i+2} \in \mathbb{F}_{16} (RC[1] = x^3 \leftrightarrow 1000)$
- RotNib(N_0N_1)= N_1N_0
- SubNib $(N_0N_1)=S(N_0)S(N_1)$ where S denotes the S-box

Example

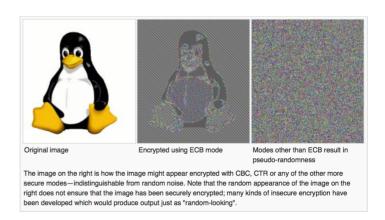
with $w[0]w[1] = 0101\ 1001\ 0111\ 1010$, we have $w[2] = 1101\ 1100$, $w[3] = 1010\ 0101$, $w[4] = 0110\ 1100$ and $w[5] = 1100\ 1010$

Why SAES?

- introduced in [7] for academic purposes
- simpler than AES and can be used by hand
- allows to illustrate cryptanalysis
- has all the features of AES

Block ciphers modes of operation

Modes of operation pictured



http://en.wikipedia.org/wiki/Block_cipher_mode_of_operation

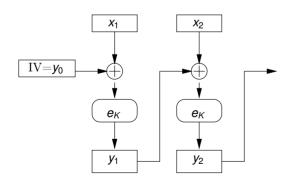
ECB: electronic codebook mode

The one previously used; given a plaintext, each block x_i is enciphered with the key K, and provides the ciphertext $y_1 y_2 ...$

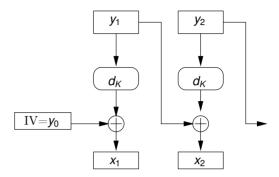


CBC: cipher block chaining mode

Each ciphertext y_i is XORed with next plaintext x_{i+1}



CBC – Deciphering

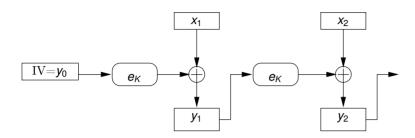


OFB (output feedback mode) and CFB (cipher feedback mode)

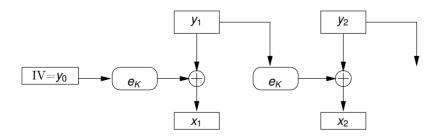
Encipher each plaintext block by successive XORing with keys coming from the application of a secret key cipher:

- **OFB**: sequence of keys comes from the repeated enciphering started on an initial value IV. We let z_0 =IV and we compute the sequence $z_1z_2...$ by $z_i = e_K(z_{i-1})$. The plaintext is then enciphered by $y_i = x_i \oplus z_i$
- **CFB**: We start with y_0 =IV and the next key is obtained by enciphering the previous ciphertext $z_i = e_K(y_{i-1})$. Otherwise, everything works like in OFB mode.

CFB enciphering



CFB deciphering



MAC-MDC

For Message Authentication Code (Modification Detection Code), or message fingerprint (MAC=MDC+IV \neq 0).

Possible with CBC and CFB.

We start with IV=0. We build the ciphertext $y_1 ldots y_n$ with the key K in CBC mode. MAC is the last block y_n . Alice sends the message $x_1 ldots x_n$ and the MAC y_n .

Upon reception of $x_1 ldots x_n$, Bob builds $y_1 ldots y_n$ by using the secret key K and verifies that y_n is the same than the received MAC.

