Sistemi di Elaborazione: Sicurezza su Reti **Anno Acc. 1998-99**
PROF. ALFREDO DE SANTIS

CRITTOGRAFIA CLASSICA

a cura di Elisabetta Martone *2 Marzo 1999*

Indice

1 Cenni storici

Da che mondo è mondo, si è sempre sentito il problema di mantenere riservati, ovvero protetti da sguardi indiscreti, i messaggi scritti. Ordini militari, messaggi diplomatici, confessioni segrete, informazioni delicate: tutte queste cose non sono certo invenzioni recenti, cosí come non lo sono i traditori o i corrieri poco fidati. Per questo sin dall'etá piú antica, gli uomini si accorsero che per mantenere la riservatezza di un messaggio, non bastava affidarsi ad un corriere di fiducia o contare sulla lealtá dei propri alleati, ma occorreva qualcosa di piú concreto, soprattutto qualcosa che rendesse possibile la comprensione del messaggio solo al legittimo destinatario.

Uno dei metodi piú bizzarri per trasmettere le informazioni segrete era utilizzato nell'antica Persia e ci viene raccontato da Erodoto. Esso consisteva nel rapare i capelli di uno schiavo e nel tatuargli il messaggio sulla testa. Lo schiavo veniva poi mandato a destinazione dopo che gli erano ricresciuti i capelli e il messaggio veniva recuperato rapandoglieli nuovamente. Piú che di crittografia in questo caso si puó parlare di steganografia, ovvero dell'arte di nascondere i messaggi.

Uno dei piú antichi metodi crittografici sviluppati in occidente è un ingegnoso sistema meccanico dovuto agli spartani e descrittoci da Plutarco. Detto "*scytala*" , esso consisteva in un bastone costruito in due esemplari identici su cui si avvolgeva un nastro di pergamena. Il messaggio veniva scritto nel senso della lunghezza del bastone; il nastro veniva poi sciolto e spedito al destinatario che possedeva una seconda copia del bastone. Riavvolgendo il nastro, in modo da far combaciare i lembi, compariva il messaggio in chiaro.

Il metodo di cifratura piú vecchio, citato da Polibio, consisteva nel disporre i 26 caratteri dell'alfabeto inglese in una matrice 5×5.

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

Nella matrice possono essere disposti 25 caratteri, pertanto o ne mettiamo due nella stessa casella, o ne eliminiamo uno. A ciascuna lettera del messaggio possiamo far corrispondere un indice di riga e uno di colonna, cosí al messaggio

ATTACCARE

corrisponde la sequenza

1,1 4,4 4,4 1,1 1,3 1,3 1,1 4,2 1,5.

Polibio suggeriva di mandare tanti messaggeri quanti erano i caratteri del messaggio, contenenti nella mano sinistra tante torce quanto era l'indice di riga e nella mano destra tante torce quanto era l'indice di colonna. Non sappiamo se questo schema sia stato veramente usato, ma è importante perché è il primo schema in cui si passa da un alfabeto a un altro con cardinalitá minore. Il primo sistema crittografico veramente usato è quello di cui si serviva Cesare per mandare le lettere a Cicerone. Cesare sostituiva ciascun carattere con quello che lo seguiva di tre posti nell'alfabeto, cosí la codifica per il messaggio

ATTACCO

era

DWWDFFR.

La funzione di cifratura utilizzata in questo caso era la seguente: facendo corrispondere a ciascuna lettera da A a Z un intero tra 0 e 25, se x era il messaggio in chiaro, il messaggio cifrato era:

$$y = x + 3 \bmod 26.$$

Questo schema è il primo esempio di **cifrario a sostituzione**, ossia in cui ogni carattere viene sostituito da un altro. Anche Augusto usava un cifrario simile a quello usato da Cesare, ma sostituiva ciascun carattere con quello che lo seguiva di un solo posto e codificava l'ultimo carattere con AA. Un cifrario di questo tipo è peró facilmente attaccabile, come suggerito in un famoso romanzo di Edgar Allan Poe, "Lo scarabeo d'oro", che è il primo che analizza i metodi di attacco ai cifrari (crittoanalisi).

Sono del 1500 le prime formalizzazioni teoriche della crittografia dovute principalmente alla mente eclettica di Leon Battista Alberti, crittologo della Corte Papale di Roma. Egli, in particolare, scrisse un trattato nel quale esaminava i pro e i contro dei vari metodi crittografici in uso all'epoca, illustrava il metodo corretto per la decifrazione dei cifrari monoalfabetici ed infine esponeva un nuovo metodo di cifratura di sua invenzione che riteneva particolarmente sicuro. In effetti il metodo dell'Alberti altro non era che un precursore dei cosiddetti cifrari a *sostituzione polialfabetica* ritenuti assolutamente inviolabili fino alla seconda metá del XIX secolo.

Nel '700 cominciarono a diffondersi sistemi crittografici basati su ausili meccanici, come i regoli cifranti di Vigenère, che facilitando l'uso dei metodi polialfabetici, contribuirono a renderne piú solida la fama. Nel 1800 tuttavia, l'invenzione del telegrafo sconvolse il modo di concepire la trasmissione dei messaggi e dunque anche la teoria della loro protezione crittografica.

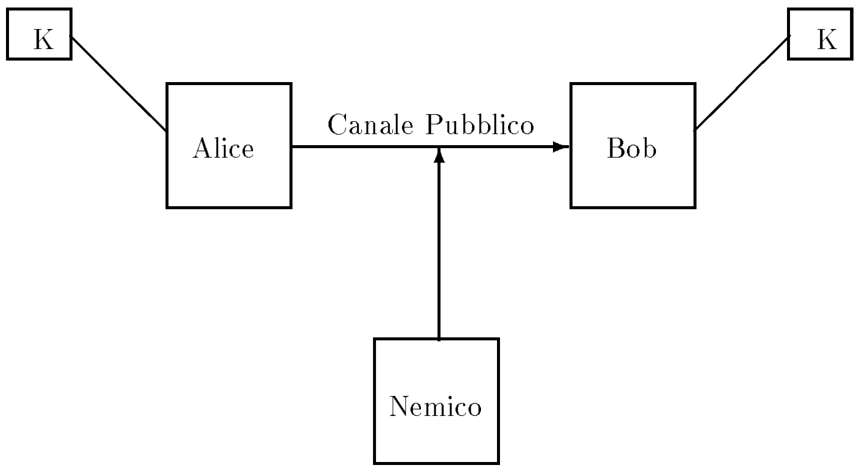
Ai sicuri, ma lenti sistemi polialfabetici, si cominciarono a preferire i piú vulnerabili ma veloci metodi a *permutazione*, alle volte scoprendo risultati giá esposti secoli prima da Cardano. Nel frattempo la teoria andava avanti: e mentre Kerckhoffs chiariva le differenze fra la crittografia militare e quella diplomatica, ed identificava correttamente il principio secondo il quale la sicurezza di un cifrario doveva stare nella sola chiave e non nel procedimento di cifratura, Kasiski metteva, infine, a punto, il metodo generale per decrittare quei cifrari polialfabetici dell'Alberti ritenuti inviolabili per tre secoli.

All'inizio del nostro secolo il progresso scientifico permise di costruire macchine cifranti completamente automatizzate; Vernam, un impiegato dell'AT&T, inventó un ingegnoso sistema di protezione crittografica. Con l'introduzione di sistemi crittografici a chiave pubblica, come il DH e l'RSA, entriamo in una nuova fase che va sotto il nome di crittografia moderna. Notizie piú approfondite sulla storia della crittografia possono essere attinte in [1].

2 Sistemi Crittografici

Anche oggi l'obiettivo fondamentale della crittografia è quello di consentire che due o piú persone possano trasmettersi messaggi attraverso un canale pubblico (non sicuro) in modo tale che, anche se un nemico li intercetta, non è in grado di comprenderli. La crittografia utilizza

strumenti propri della teoria della complessitá, teoria degli algoritmi, teoria dell'informazione, teoria dei numeri, studiando tecniche che permettono di mascherare il linguaggio in modo da renderlo inintellegibile a chi non possieda un'opportuna "chiave" di lettura.
Lo schema piú semplice di un *Crittosistema* è il seguente:



Definizione 1 *Un Crittosistema è una quintupla $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$ con le seguenti proprietá:*

1. *$\mathcal{P}$ é un insieme finito dei possibili testi in chiaro*
2. *$\mathcal{C}$ è un insieme finito dei possibili testi cifrati*
3. *$\mathcal{K}$, lo spazio delle chiavi, è un insieme di possibili chiavi*
4. *Per ogni $k \in \mathcal{K}$ c'è una funzione di cifratura $e_k \in \mathcal{E}$ e una corrispondente funzione di decifratura $d_k \in \mathcal{D}$.*
   Le funzioni:
   $$e_k : \mathcal{P} \rightarrow \mathcal{C} \qquad d_k : \mathcal{C} \rightarrow \mathcal{P}$$
   sono tali che $d_k(e_k(x)) = x$ per ogni testo in chiaro $x \in \mathcal{P}$.

Alice e Bob, innanzitutto, scelgono una chiave a caso $k \in \mathcal{K}$, lo fanno incontrandosi in un posto segreto dove non sono osservati dal nemico o, alternativamente, quando, essendo in posti diversi, possono accedere ad un canale privato. Supponiamo che il messaggio in chiaro sia una stringa

$$x = x_1 x_2 \cdots x_n$$

per qualche intero $n \geq 1$, dove ogni simbolo del testo in chiaro $x_i \in \mathcal{P}$, $1 \leq i \leq n$. Ogni x_i è cifrato usando la regola di cifratura e_k specificata dalla predeterminata chiave k. Quindi, Alice calcola $y_i = e_k(x_i)$, $1 \leq i \leq n$, e la risultante stringa cifrata

$$y = y_1 y_2 \cdots y_n$$

è spedita sul canale pubblico. Quando Bob riceve $y_1 y_2 \cdots y_n$ lo decifra usando la funzione di decifratura d_k, ottenendo l'originale testo in chiaro $x_1 x_2 \cdots x_n$.

Chiaramente, la funzione di cifratura e_k è una **funzione iniettiva** perché altrimenti la decifratura potrebbe avvenire in modo ambiguo. Per esempio, se $y = e_k(x_1) = e_k(x_2)$ dove $x_1 \neq x_2$, allora Bob non ha modo di sapere se y, decifrato, fornisce x_1 oppure x_2. Notiamo che se $\mathcal{P} = \mathcal{C}$ segue che ogni funzione di cifratura é una *permutazione*, nel senso che il testo cifrato è ottenuto come permutazione dei caratteri del testo in chiaro.

Se un crittosistema deve essere di uso comune, deve soddisfare alcune proprietá, tra cui:

1. Ogni funzione di cifratura e_k ed ogni funzione di decifratura d_k devono essere calcolabili efficientemente.
2. Un nemico, guardando la stringa cifrata y, non deve essere capace di determinare la chiave k che è stata usata, oppure la stringa del testo in chiaro x. Una assunzione da fare è che il nemico conosce sempre l'algoritmo di cifratura ma non la chiave. Questa assunzione viene fatta per due motivi:
   - per trovarsi nel caso peggiore
   - perché i sistemi che saranno presi in considerazione possono essere utilizzati da utenti diversi, pertanto è improbabile che l'algoritmo possa restare segreto.

3 Principali Cifrari

Analizziamo adesso i cifrari che sono stati piú utilizzati nel corso dei secoli e che oggi conservano esclusivamente valenza storica.

3.1 Cifrari a sostituzione

In questi sistemi di cifratura, ogni simbolo del testo in chiaro viene trasformato in un simbolo dell'alfabeto del testo cifrato. I segni di punteggiatura e gli spazi non vengono cifrati, in modo tale che il testo diventa un'unica stringa di caratteri alfabetici. Questi cifrari si dividono in due categorie: *monoalfabetici* e *polialfabetici*. Nei monoalfabetici, ogni carattere, una volta sostituito con un altro carattere, conserverá tale sostituzione durante tutta la fase di cifratura; per esempio, se la parola BACCA sará cifrata in EFGGF, allora la lettera A sará sempre F, la C sará G e cosí via. Nei sistemi polialfabetici, invece, lo stesso carattere puó essere cifrato in modo differente. Un esempio di cifrario a sostituzione monoalfabetica è quello di Cesare. Nei cifrari che vedremo l'alfabeto sará di 26 caratteri posti in corrispondenza biunivoca con gli interi da 0 a 25 come mostrato nella Tabella 1:

A	B	C	D	E	F	G	H	I	J	K	L	M
0	1	2	3	4	5	6	7	8	9	10	11	12

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
13	14	15	16	17	18	19	20	21	22	23	24	25

- Tabella 1 -

Vediamo alcune varianti degli schemi a sostituzione.

3.2 Cifrario Affine

In questo cifrario la chiave è una coppia di interi in $\mathbf{Z}_{26}$

$$k = (a, b) \text{ con } a, b \in \mathbf{Z}_{26}$$

e la funzione di cifratura è del tipo

$$e_k(x) = ax + b \bmod 26$$

dove $a, b \in \mathbf{Z}_{26}$. Queste funzioni sono dette *funzioni affini*, da cui il nome del cifrario. Vediamo sotto quali condizioni una funzione di questo tipo va bene per la cifratura. Come abbiamo giá detto, una funzione di cifratura deve essere iniettiva, se così non fosse Bob non potrebbe decifrare il messaggio in modo univoco. Pertanto la chiave (a, b) deve essere scelta in modo che l'equazione $y = ax + b \bmod 26$ abbia un'unica soluzione così che la funzione sia iniettiva.

Dimostreremo che condizione necessaria e sufficiente affinchè $y = ax + b \bmod 26$ abbia un'unica soluzione è che $\gcd(a, 26) = 1$. Vediamo un esempio.

Supponiamo di scegliere come chiave la coppia $a = 2$ e $b = 0$ e che il messaggio sia $y = 0$; allora la funzione di cifratura è

$$e_k(x) = 2x$$

e quindi per decrifrare il messaggio l'equazione da risolvere è

$$2x = 0 \bmod 26.$$

Le soluzioni sono $x = 0$ e $x = 13$, quindi due messaggi in chiaro distinti corrispondono allo stesso messaggio cifrato, questo perchè il $\gcd(a, 26) = 2$, di conseguenza la scelta $a = 2$ non è ammissibile perchè vogliamo che la congruenza

$$ax = 0 \bmod 26$$

abbia un'unica soluzione.

Vogliamo risolvere l'equazione $y = ax + b \bmod 26$ ma questa è equivalente all'equazione $ax = y - b \bmod 26$, siccome $y \in \mathbf{Z}_{26}$ anche $y - b \in \mathbf{Z}_{26}$ per cui è equvalente risolvere la congruenza $ax = y \bmod 26$.

Teorema 1 *$ax = y \bmod 26$ ha un'unica soluzione per ogni y se e solo se $\gcd(a, 26) = 1$.*

Dimostrazione Supponiamo che $ax = y \bmod 26$ abbia un'unica soluzione per ogni y e supponiamo per assurdo che $\gcd(a, 26) = d > 1$, risolvendo l'equazione per $y = 0$ giungiamo ad un'assurdo, infatti $ax = 0 \bmod 26$ ha due soluzioni distinte in $\mathbf{Z}_{26}$: $x = 0$ e $x = \frac{26}{d}$. Quindi abbiamo provato che esiste un y per cui $ax = y \bmod 26$ non ammette soluzione unica.

Sia $\gcd(a, 26) = 1$, siano x_1 e x_2 soluzioni distinte di $ax = y \bmod 26$ quindi

$$ax_1 = ax_2 \bmod 26$$

da cui

$$a(x_1 - x_2) = 0 \bmod 26$$

quindi

$$26 | a(x_1 - x_2)$$

dato che $\gcd(a, 26) = 1$ si ha $26|(x_1 - x_2)$ quindi

$$x_1 = x_2 \bmod 26$$

così x_1 e x_2 sono la stessa soluzione. $\square$

Calcoliamo la funzione di decifratura d applicata ad un messaggio cifrato y. Siccome

$$y = ax + b \bmod 26$$

si ha che

$$y - b = ax \bmod 26$$

da cui

$$a^{-1}(y - b) \bmod 26 = x$$

quindi

$$d(y) = a^{-1}(y - b) \bmod 26.$$

Il numero di possibili chiavi per il cifrario affine è dato dal numero di coppie (a, b). Poiché b puó assumere 26 possibili valori ci resta da determinare il numero di possibili valori per a, cioé il numero di valori tra 0 e 25 primi con 26, che è dato dalla cardinalitá di $\mathbf{Z}^*_{26}$, che puó essere determinata tramite la funzione di Eulero $\phi(26) = (13 - 1)(2 - 1) = 12$.
Pertanto il numero di possibili chiavi è 26×12 (troppo basso per scoraggiare una ricerca esaustiva).

3.3 Cifrario di Vigenère

Il cifrario di Vigenère è un ottimo esempio di cifrario a sostituzione di tipo polialfabetico ed è dovuto agli studi di Blaise de Vigenère, che visse nel diciassettesimo secolo. Sia m un intero positivo fissato. Definiamo $\mathcal{P} = \mathcal{C} = \mathcal{K} = (\mathbf{Z}_{26})^m$. Per una chiave $k = (k_1, k_2, \ldots, k_m) \in (\mathbf{Z}_{26})^m$, definiamo la funzione di cifratura come

$$e_k(x_1, x_2, \ldots, x_m) = (x_1 + k_1, x_2 + k_2, \ldots, x_m + k_m)$$

e quella di decifratura come

$$d_k(y_1, y_2, \ldots, y_m) = (y_1 - k_1, y_2 - k_2, \ldots, y_m - k_m),$$

dove tutte le operazioni sono fatte in $\mathbf{Z}_{26}$.

Utilizzando la corrispondenza $A \leftrightarrow 0$, $B \leftrightarrow 1$, $\ldots$, $Z \leftrightarrow 25$ della Tabella 1, possiamo associare ogni chiave k con una stringa alfabetica di lunghezza m.
L'idea del cifrario è di codificare m caratteri alfabetici alla volta: il messaggio viene partizionato in blocchi di m caratteri e ogni blocco è cifrato con la stessa chiave. Ai blocchi di $n < m$ caratteri vengono aggiunti i mancanti $m - n$ in modo arbitrario, e tale che non alteri il significato del messaggio. Lo spazio delle chiavi è 26^m. Questo cifrario è stato ritenuto sicuro per circa due secoli, in quanto la chiave non è costituita solo dai valori k_i, ma anche dal numero delle chiavi m, quindi bisogna determinare prima il numero di sottochiavi e poi il valore di ciascuna.

Per semplicitá vediamo la cifratura con un esempio.

Esempio 1 *Supponiamo di voler cifrare il messaggio* ATTACCOALLALBA *con la chiave* SALERNO*. Si procede in questo modo: si divide il messaggio in gruppi di sette lettere e si cifrano i due gruppi che si ottengono con la chiave. La caratteristica di questo metodo è che due lettere uguali del messaggio in chiaro possono essere cifrate in modo differente, infatti:*

messaggio in chiaro	A	T	T	A	C	C	O	A	L	L	A	L	B	A
chiave	S	A	L	E	R	N	O	S	A	L	E	R	N	O
messaggio cifrato	S	T	E	E	T	P	C	S	L	W	E	C	O	O

Come si puó vedere, le due T *sono cifrate con caratteri differenti. Per facilitare i conti veniva utilizzata una tabella di 26 righe e di 26 colonne in cui si ricercavano le coppie (carattere del testo in chiaro-carattere della chiave), rispettivamente sulla prima riga e sulla prima colonna; il carattere del testo cifrato era dato dall'elemento individuato dall'incrocio della riga e della colonna ottenute. Per esempio la prima* A *di* ATTACCO *e la* S *di* SALERNO*, si trovano rispettivamente sulla prima riga e sulla diciannovesima colonna: il carattere del testo cifrato è* S.

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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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

- Tabella di Vigenère -

Questo cifrario è tanto piú resistente agli attacchi crittoanalitici quanto piú lunga è la chiave.

3.4 Cifrario di Hill

Il cifrario di Hill è un altro crittosistema polialfabetico e fu inventato nel 1929 da Lester Hill. Sia m un intero positivo, definiamo $\mathcal{P} = \mathcal{C} = (\mathbf{Z}_{26})^m$. L'idea è di costruire m combinazioni lineari contenenti ognuna m caratteri alfabetici del testo in chiaro e produrre, con ciascuna combinazione lineare, un singolo elemento del testo cifrato. La chiave utilizzata in questo cifrario è una matrice K invertibile di dimensione $m \times m$. Per $x = (x_1, \ldots, x_m) \in \mathcal{P}$ e $K \in \mathcal{K}$, calcoliamo $y = e_k(x) = (y_1, \ldots, y_m)$ come segue:

$$(y_1, y_2, \ldots, y_m) = (x_1, x_2, \ldots, x_m) \begin{pmatrix} k_{1,1} & k_{1,2} & \cdots & k_{1,m} \\ k_{2,1} & k_{2,2} & \cdots & k_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ k_{m,1} & k_{m,2} & \cdots & k_{m,m} \end{pmatrix}.$$

Usando la notazione matriciale:

$$y = xK.$$

Diciamo che il testo cifrato è ottenuto dal testo in chiaro per mezzo di una *trasformazione lineare*, di blocchi di m caratteri. In ultimo vediamo come avviene la decifratura, cioé come x puó essere calcolato a partire da y. Chiaramente, si dovrá usare la matrice inversa K^{-1}, per cui il testo in chiaro x si otterrá come

$$x = yK^{-1}.$$

Esempio 2 *Supponiamo che la chiave sia*

$$K = \begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix}.$$

Poiché la chiave è una matrice 2×2 *dobbiamo cifrare il messaggio in blocchi di due caratteri. Supponiamo che il messaggio da cifrare sia CASA. Al blocco CA corrisponde la coppia (2,0) e il testo cifrato è dato da:*

$$(2,0) \begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix} = (22, 16).$$

Al blocco SA corrisponde la coppia (18,0) e il testo cifrato è dato da

$$(18,0) \begin{pmatrix} 11 & 8 \\ 3 & 7 \end{pmatrix} = (198, 144) = (16, 14).$$

Quindi la cifratura di casa è WQQO che corrisponde a (22,16,16,14).
Per quanto riguarda la decifratura occorre calcolare l'inversa della matrice K *che è:*

$$K^{-1} = \begin{pmatrix} 7 & 18 \\ 23 & 11 \end{pmatrix}$$

e poi calcolare:

$$(22,16) \begin{pmatrix} 7 & 18 \\ 23 & 11 \end{pmatrix} = (2,0)$$

e

$$(16,14) \begin{pmatrix} 7 & 18 \\ 23 & 11 \end{pmatrix} = (18,0).$$

3.5 Cifrario a permutazione

Tutti i crittosistemi che abbiamo discusso finora coinvolgono la sostituzione: caratteri del testo in chiaro sono sostituiti, nel testo cifrato, con caratteri diversi. L' idea di un cifrario a permutazione è di non modificare i caratteri del testo in chiaro, ma di cambiare le loro posizioni. Formalmente, sia m un intero positivo fissato. Sia $\mathcal{P} = \mathcal{C} = (\mathbf{Z}_{26})^m$ e sia $\mathcal{K}$ l'insieme di tutte le permutazioni di $\{1, \ldots, m\}$, lo spazio delle chiavi avrà quindi cardinalità $m!$. Per una chiave π (permutazione), definiamo la funzione di cifratura come

$$e_\pi(x_1, \ldots, x_m) = (x_{\pi(1)}, \ldots, x_{\pi(m)})$$

e la funzione di decifratura come

$$d_\pi(y_1, \ldots, y_m) = (y_{\pi^{-1}(1)}, \ldots, y_{\pi^{-1}(m)}),$$

dove π^{-1} è la permutazione inversa di π.

Esempio 3 *Sia* ATTACCOALLALBAX *il messaggio da cifrare e la permutazione (2,3,1) la nostra chiave. Questo significa che:*

$$\pi(1) = 2 \qquad \pi(2) = 3 \qquad \pi(3) = 1$$

La chiave scelta ci dice di dividere il testo in chiaro in gruppi di tre caratteri e di permutarli nell'ordine stabilito, cioé la prima lettera occuperá il secondo posto della tripla, la seconda lettera il terzo posto, la terza il primo. Per esempio, per la prima tripla ATT*, avremo* TAT*. Cosí procedendo il messaggio cifrato sará:*

TATCACLOAALALXBA.

Notiamo che il numero di chiavi disponibili con questo metodo è $m!$, quante sono le possibili permutazioni di m caratteri. I cifrari a permutazione sono un caso particolare del cifrario di Hill. Data una permutazione π dell'insieme $\{1, \ldots, m\}$, possiamo infatti definire la *matrice delle permutazioni* $K_\pi = (k_{i,j})$ di dimensioni $m \times m$ in accordo alla formula

$$k_{i,j} = \begin{cases} 1 & \text{se } j = \pi(i) \\ 0 & \text{altrimenti} \end{cases}$$

(La *matrice di permutazione* è una matrice in cui ogni riga e ogni colonna contiene esattamente un "1" e tutti gli altri valori sono "0"). Riconsiderando l'Esempio 3, la matrice che si ottiene è:

$$K_\pi = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Analogamente al cifrario di Hill, per decifrare il messaggio, dobbiamo calcolare l' inversa della matrice di permutazione che chiamiamo K_π^{-1} che è ancora una matrice di permutazione.

3.6 Cifrario di Playfair

Questo sistema fa uso di una matrice 5×5, che rappresenta la chiave, in cui ci sono 25 lettere distinte dell'alfabeto disposte in ordine arbitrario; l'unica lettera non presente è la Q, in quanto è la lettera che è quasi sempre seguita dalla U e quindi facilmente riconoscibile. Ogni lettera individua ed è individuata dal suo numero di riga e dal suo numero di colonna. Il testo in chiaro deve essere di lunghezza pari, dovrà essere diviso in coppie di caratteri e ciascuna coppia deve avere caratteri distinti. Nel caso in cui il testo sia di lunghezza dispari si aggiunge un carattere fittizio che non alteri il senso del messaggio, per ottenere la paritá. Analogo stratagemma si utilizza se una delle coppie ha i due caratteri uguali: si inserisce come secondo carattere uno che non altera la comprensione del testo. Con queste caratteristiche ogni coppia di caratteri rientra sicuramente in uno fra questi tre casi:

1. le due lettere della coppia si trovano sulla stessa riga;
2. le due lettere si trovano sulla stessa colonna;
3. le due lettere non si trovano né sulla stessa riga né sulla stessa colonna.

La cifratura è eseguita su una coppia di lettere per volta. Ogni coppia del testo in chiaro rientra in uno dei tre casi precedenti, ciascuno dei quali va trattato in modo differente:

1. le lettere della coppia del testo in chiaro vanno cifrate con le lettere che le seguono sulla stessa riga (con la convenzione "circolare" che l'ultima lettera della riga sia seguita dalla prima). Per esempio se la matrice è:

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

   per la coppia WI si ha HS, WH diventa HI, etc. .

2. le lettere della del testo in chiaro vanno cifrate con le lettere immediatamente sottostanti (sempre con la convenzione "circolare"); ad esempio IG diventa BO e VB diventa IG.

3. In questo caso, che è il piú frequente, le due lettere individuano sulla matrice un "rettangolo" di cui due vertici rappresentano la coppia di caratteri del messaggio in chiaro e gli altri due i corrispondenti caratteri del messaggio cifrato. La convezione che si utilizza nella cifratura è che la prima lettera del testo risultante è quella che si trova sulla stessa riga della prima lettera della coppia in chiaro. Per esempio, NC diventa PA, mentre CN diventa AP.

La chiave di questo cifrario è la matrice di lettere, e costituisce una chiave un po' troppo difficile da ricordare. Possiamo ovviare ad una tale difficoltà usando uno stratagemma, gli utenti che vogliono comunicare si accordano su una frase da ricordare a memoria, la prima occorrenza di ogni lettera nella frase è inserita nella matrice, le lettere dell'alfabeto che non compaiono nella frase sono inserite successivamente in ordine lessicografico. Per esempio se la frase fosse: "cogli l'attimo fuggente", la matrice sarebbe:

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

Alcune varianti del cifrario utilizzano matrici 3×9 oppure 4×6.

3.7 Stream cipher

Nei crittosistemi studiati fino ad ora successivi elementi del testo in chiaro sono cifrati mediante la stessa chiave k, cioé la stringa cifrata y è ottenuta nel modo seguente :

$$y = y_1 y_2 \ldots = e_k(x_1) e_k(x_2) \ldots .$$

Crittosistemi di questo tipo sono spesso definiti "cifrari a blocchi". Un approccio alternativo consiste nell'utilizzare gli "stream cipher". L'idea di base consiste nel generare una sequenza, detta "keystream"

$$z = z_1 z_2 \ldots$$

e nell'utilizzarla per cifrare la stringa

$$x = x_1 x_2 \dots$$

del testo in chiaro nel modo seguente

$$y = y_1 y_2 \dots = e_{z_1}(x_1) e_{z_2}(x_2) \dots$$

dove

$$z_i = f_i(k, x_1, \dots, x_{i-1}).$$

L'elemento z_i della sequenza è utilizzato per cifrare x_i, ossia $y_i = e_{z_i}(x_i)$. Pertanto per cifrare la stringa $x_1 x_2 \dots$ bisogna calcolare

$$z_1, y_1, z_2, y_2 \dots.$$

Analogamente per decifrare la stringa $y_1 y_2 \dots$ bisogna calcolare

$$z_1, x_1, z_2, x_2 \dots.$$

Per esempio se $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbf{Z}_{26}$ le funzioni di cifratura e decifratura possono essere le seguenti:

$$e_{z_i}(x_i) = y_i = x_i + z_i \bmod 26$$

$$d_{z_i}(y_i) = x_i = y_i - z_i \bmod 26.$$

I cifrari a blocchi possono essere considerati un caso particolare degli stream cipher, in cui $z_i = k$ per ogni $i \geq 1$. Gli stream cipher sono spesso definiti in termini binari, cioé $\mathcal{P} = \mathcal{C} = \mathcal{K} = \mathbf{Z}_2$. In tal caso le funzioni di cifratura e decifratura possono essere le seguenti:

$$e_{z_i}(x_i) = x_i + z_i \bmod 2$$

$$d_{z_i}(y_i) = y_i - z_i \bmod 2.$$

Se associamo a "0" il valore booleano "falso", ed a "1" il valore booleano "vero" allora l'addizione modulo 2 corrisponde allo XOR, pertanto la cifratura e la decifratura possono essere implementate molto efficientemente in hardware. Un metodo per generare la sequenza keystream a partire da m valori $k_1, \dots, k_m$ consiste nel fissare i valori per $z_1 \dots z_m$ come $z_i = k_i$ per ogni $i = 1, \dots, m$, e nel calcolare i valori successivi in base ad una relazione di ricorrenza lineare di grado m :

$$z_{i+m} = \sum_{j=0}^{m-1} c_j z_{i+j} \bmod 2,$$

dove $c_0 \dots c_{m-1} \in \mathbf{Z_2}$ sono costanti predeterminate. Questa ricorrenza è detta lineare perché z_{i+m} è funzione lineare dei termini precedenti, ed ha grado m perché ciascun termine dipende dagli m precedenti. In tal caso la chiave consiste dei $2m$ valori

$$(k_1, \dots, k_m, c_0, \dots, c_{m-1}),$$

dove $k_i = z_i$ per $i = 1 \dots m$.

Esempio 4 *Supponiamo che sia* $m = 4$ *e che la keystream sia generata nel modo seguente :*

$$z_{i+4} = z_i + z_{i+1} \bmod 2$$

dove $i \geq 1$. *Se ad esempio si considerano i seguenti valori iniziali:* $1, 0, 0, 0$ *la keystream corrispondente è*

$$1, 0, 0, 0, 1, 0, 0, 1, 1, 0, ...$$

Un altro aspetto interessante di questo metodo è che la keystream puó essere generata in hardware in modo molto efficiente utilizzando un *feedback shift register*
Consideriamo un registro con m stadi. Il vettore $(k_1, \ldots, k_m)$ viene usato per l' inizializzazione. Ad ogni istante vengono effettuate contemporaneamente le seguenti operazioni:

1. k_1 si presenta come nuovo bit della keystream ;

2. $k_2, \ldots, k_m$ vengono shiftati a sinistra di un posto;

3. viene calcolato il nuovo valore di k_m, che è :

$$\sum_{j=0}^{m-1} c_j k_{j+1}.$$

(questo é il "feedback lineare").

Osserviamo che il feedback lineare é portato avanti attraverso certi stadi del registro (come specificato dalle costanti c_j che assumono valore " 1"), e viene computata una somma modulo 2 (che corrisponde ad uno XOR). Ció è illustrato nella Figura 1, in cui viene descritto il feedback shift register che genera la keystream relativa all'esempio precedente.

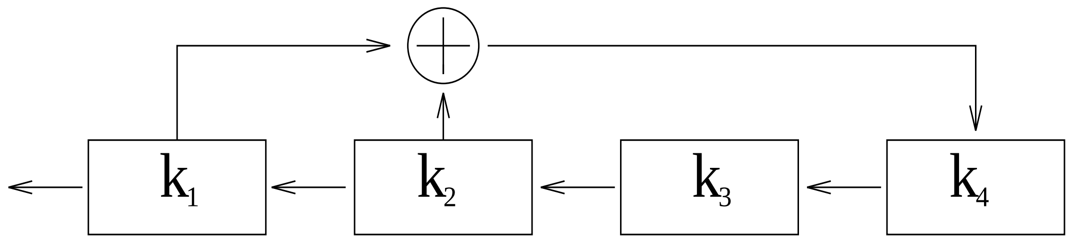


Figura 1: Schema di un feedback shift register

3.8 Cifrario autoclave

Questo cifrario puó essere considerato un caso particolare dello Stream cipher in cui la sequenza degli z_i è calcolata nel modo seguente: $z_i = k_i$ per $i \leq m$, $z_i = x_{i-m}$, per $i > m$. Le funzioni di cifratura e decifratura sono le seguenti:

$$y_i = x_i + z_i \bmod 26$$

$$x_i = y_i - z_i \bmod 26.$$

Vediamo come realizzare ció, attraverso un esempio.

Esempio 5 *Supponiamo di voler cifrare il messaggio* ATTACCOALLALBA *con la chiave* SALERNO*: dopo aver cifrato la prima parte con la chiave* SALERNO*, anziché riutilizzarla come accade in Vigenère, si procede con la cifratura utilizzando il messaggio in chiaro come continuazione della chiave.*

Messaggio in chiaro	A	T	T	A	C	C	O	A	L	L	A	L	B	A
Chiave	S	A	L	E	R	N	O	A	T	T	A	C	C	O
Messaggio cifrato	S	T	E	E	T	P	C	A	E	E	A	N	D	O

3.9 Cifrario a dizionario

Sono dei cifrari a sostituzione in cui, peró, ad essere sostituite sono delle parole intere anziché un solo simbolo alla volta. Si compila una specie di dizionario nel quale le parole in chiaro vengono tradotte in una lingua di fantasia, inventata apposta per servire agli scopi della crittografia. Alle parole in chiaro del nostro dizionario corrisponderanno parole di codice segrete; queste ultime, di solito, sono sequenze di un numero fisso di cifre, oppure sequenze di un numero fisso di lettere. I dizionari possono essere a *lista semplice* o a *lista doppia*. Nei primi si ha a disposizione nel dizionario soltanto la trasformazione da un messaggio in chiaro ad uno cifrato, o viceversa (per esempio, inglese-italiano) mentre nei secondi, nello stesso dizionario, si hanno entrambe le trasformazioni (per esempio, inglese-italiano-inglese). Nei dizionari a lista semplice le parole rispettano l'ordinamento lessicografico, una volta effettuata la cifratura (decifratura) della prima parola in chiaro, se la seconda parola in chiaro segue (precede) lessicograficamente la prima, allora la corrispondente parola cifrata seguirá (precederá) la prima parola cifrata, e cosí per tutte le altre. Ció è stato fatto per facilitare la decifratura, rendendo più veloce la ricerca.
Il vantaggio di questo cifrario è che il messaggio sembra una frase comune e non un messaggio cifrario; lo svantaggio risiede nella "grandezza" della chiave: un libricino.

Esempio 6 *Supponiamo che il messaggio in chiaro sia* ATTACCO ALLA ALBA*. Per cifrare questo testo si può anche pensare di ottenerne un altro che comunque abbia senso compiuto, come ad esempio* VEDIAMOCI A PRANZO*.*

3.10 One-time pad

Nel 1917 il signor Gilbert Vernam, impiegato della compagnia AT&T, inventó un ingegnosissimo sistema di protezione crittografica, per comunicazioni su telegrafo, dei testi codificati in binario. Egli costruí per prima cosa un dispositivo in grado di leggere contemporaneamente due nastri in input e generare a partire da essi un nastro di output tale che ciascun foro fosse generato mediante un XOR dei due corrispondenti fori sui nastri input. Dopodiché prese un nastro su cui era perforata una sequenza di caratteri casuale ed un nastro su cui era perforato un testo reale e li passó nella sua macchina. Il risultato fu un nastro completamente inintellegibile, ovvero cifrato.

Lo schema di crittografia di Vernam è uno schema one-time pad, un tale schema richiede che:

- la chiave sia usata una sola volta (da qui il nome),
- deve essere lunga almeno quanto il testo in chiaro,

- fra i bit che compongono la chiave non deve esserci alcuna relazione,
- la chiave deve essere generata casualmente.

In pratica se il testo in chiaro è $X = 0110$ e la chiave è $K = 1100$, applicando il metodo di Vernam otteniamo il seguente testo cifrato:

$$Y = X \oplus Y = 1010$$

la decrifratura si ottiene nel seguente modo:

$$X = Y \oplus K = 0110.$$

Notiamo che abbiamo applicato la stessa chiave ed effettuato la stessa operazione sia per la cifratura che per la decifratura, ciò caratterizza un sistema crittografico reversibile, questo è uno dei molti aspetti notevoli del cifrario di Vernam. Per ciò che concerne la sicurezza, a tutt'oggi, questo è l'unico metodo ad essere *perfetto*, ossia costituisce un cifrario assolutamente indecifrabile in senso stretto.

Un cifrario si dice *perfetto* se gode della seguente proprietà :
per ogni $\hat{X}$ e $\hat{Y}$ risulta:

$$Pr(X = \hat{X}) = Pr(X = \hat{X} \mid Y = \hat{Y}).$$

La proprietà di cui sopra si chiama sicurezza perfetta.

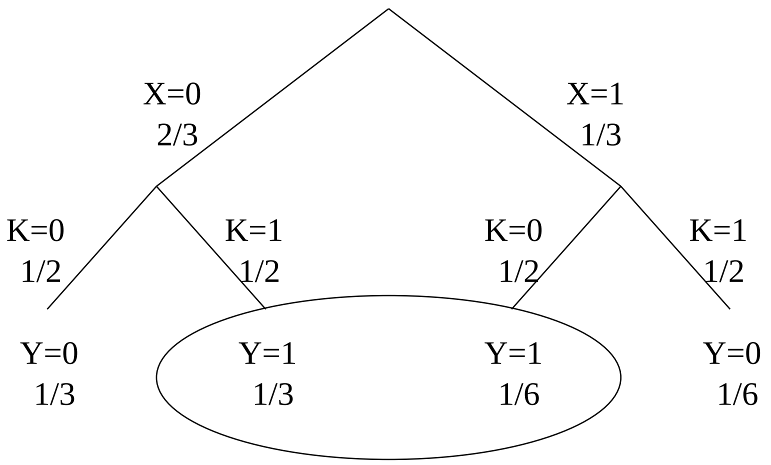
Per un cifrario che gode della sicurezza perfetta, l'indecisione nello stabilire qual'è il testo in chiaro X senza conoscere il testo cifrato Y è la stessa che si ha su X conoscendo il testo cifrato Y.

Facciamo un esempio col primo bit della sequenza in chiaro, supponiamo che l'intruso sappia che tale bit assuma valore 0 con probabilità $\frac{2}{3}$, e 1 con probabilità $\frac{1}{3}$ ovvero:
$Pr(X = 0) = \frac{2}{3}$ e $Pr(X = 1) = \frac{1}{3}$

Sotto l'ipotesi che la chiave sia casuale si ha:

$$Pr(K = 0) = Pr(K = 1) = \frac{1}{2}$$

Ora, sapendo tutto ció, proviamo a fare delle ipotesi su come possa essere cifrato questo bit, e con che probabilitá.



Ora calcoliamo:

$$Pr(X = 0 \mid Y = 1) = \frac{Pr(X = 0 \wedge Y = 1)}{Pr(Y = 1)} = \frac{1/3}{1/2} = \frac{2}{3}$$

L'intruso sapeva che $Pr(X = 0) = 2/3$, quindi per lui, conoscere il testo cifrato non fornisce ulteriori informazioni su X.
La maggior forza del one-time pad è nel fatto che la chiave è usata una sola volta, infatti se cifriamo usando la stessa chiave K due testi in chiaro X_1 e X_2 otteniamo: $Y_1 = X_1 \oplus K$ e $Y_2 = X_2 \oplus K$.

Quindi se l'intruso computa:

$$Y_1 \oplus Y_2 = (X_1 \oplus K) \oplus (X_2 \oplus K) = X_1 \oplus X_2$$

ciò può sembrare di non molta importanza, ma vengono prodotte delle regolarità statistiche che possono condurre a individuare i due messaggi originali.

Diamo ora una dimostrazione formale della sicurezza perfetta del one-time pad.

Teorema 2 $Pr(X = x \mid Y = y) = Pr(X = x)$

Dimostrazione Assumiamo che X e Y siano di n bit. Dalla regola della probabilità condizionata abbiamo

$$Pr(X = x \mid Y = y) = \frac{Pr(X = x \wedge Y = y)}{Pr(Y = y)}$$

Calcoliamo pertanto la probabilità congiunta

$$\begin{aligned} Pr(X = x \wedge Y = y) &= Pr(X = x \wedge K = x \oplus y) \\ &= Pr(X = x) \cdot Pr(K = x \oplus y) \\ &= Pr(X = x) \cdot 2^{-n} \end{aligned}$$

e la probabilità

$$\begin{aligned} Pr(Y = y) &= \textstyle\sum_x Pr(X = x \wedge Y = y) \\ &= \textstyle\sum_x Pr(X = x) \cdot 2^{-n} \\ &= 2^{-n} \cdot \textstyle\sum_x Pr(X = x) \\ &= 2^{-n}. \end{aligned}$$

Quindi

$$Pr(X = x \mid Y = y) = \tfrac{Pr(X=x) \cdot 2^{-n}}{2^{-n}} = Pr(X = x). \qquad \square$$

Le proprietà che caratterizzano il one-time pad sono estremamente restrittive, volendole rispettare si ottiene un sistema scomodo da usare in pratica, considerando che le ingombranti chiavi andrebbero generate in anticipo rispetto al loro uso previsto, e conservate in luogo sicuro. Sono questi i motivi per cui questo sistema non viene usato che per casi eccezionali, come la famosa hot-line tra Washington e Mosca.

Un'altro problema è che il one-time pad è modificabile; un intruso può cambiare Y così che il messaggio M decifrato sia differente dal messaggio spedito. Non ci sono modi per il destinatario di controllare che il mittente abbia spedito proprio il messaggio ricevuto.

Ci sono delle varianti che possono evitare di utilizzare delle chiavi così grandi, ma che fanno perdere la perfezione al sistema perchè introducono delle dipendenze statistiche. Un esempio è quello di prendere una chiave in un grosso testo, come la Bibbia, specificando un punto di inizio qualunque, tutti i caratteri da quel punto in poi, formeranno la chiave. La dipendenza statistica è insita proprio nel fatto che le parole devono avere senso compiuto. Inoltre la difficoltà per i crittoanalisti, oltre alla conoscenza della chiave (punto di inizio nel testo), sta anche nel capire qual è il testo utilizzato.

Il problema con le chiavi corte, che dunque devono essere riutilizzate ciclicamente nel corso del messaggio, è che producono, in uscita, delle regolarità statistiche che possono essere usate dai crittoanalisti per forzare il cifrario.

3.11 Sicurezza perfetta

Abbiamo visto nella sezione precedente, come Vernam riuscí a realizzare un cifrario perfetto. Ragionando con i moderni termini della teoria dell'informazione, si comprende come la "somma" di una chiave casuale ad un testo in chiaro annulli l'ordine intrinseco nel testo stesso, producendo un cifrario a sua volta casuale, nel quale l'*entropia* H, che rappresenta la quantitá di informazione contenuta in un messaggio X, è massima.

Definizione 2 *L'entropia $H(X)$ di una variabile aleatoria X con distribuzione di probabilitá $p_1, p_2, \ldots, p_m$, è uguale a*

$$-\sum_{i=1}^{m} p_i \log p_i.$$

e definisce la quantitá di informazione contenuta in X.

In termini di entropia possiamo definire un sistema di cifratura come segue: sia X la variabile aleatoria associata al messaggio in chiaro, sia Y la variabile aleatoria associata al messaggio cifrato e sia K la variabile aleatoria associata alla chiave, tutte con distribuzione di probabilità arbitraria. Con questa definizione la (3.10) si puó scrivere:

$$H(X) = H(X|Y).$$

Teorema 3 *In un sistema crittografico con sicurezza perfetta, la lunghezza della chiave deve essere tanto grande quanto quella del messaggio in chiaro.*

Dimostrazione Supponiamo che X sia il messaggio in chiaro che Alice vuole mandare a Bob, che Y sia il messaggio cifrato che Bob riceve e che la chiave utilizzata sia K. Ricordiamo che X, Y e K sono variabili casuali. Affinché le tre variabili realizzino un sistema crittografico perfetto si deve avere :

1. CIFRATURA: il testo cifrato è univocamente determinato dal testo in chiaro e dalla chiave. Formalmente
$$H(Y|XK) = 0$$
2. DECIFRATURA: il testo in chiaro è univocamente determinato dal testo cifrato e dalla chiave. Formalmente
$$H(X|YK) = 0$$

3. SICUREZZA PERFETTA: la conoscenza del testo cifrato non dá alcuna informazione sul testo in chiaro. Formalmente

$$H(X|Y) = H(X)$$

La *mutua informazione* $I(KX|Y)$ puó essere scritta come
$I(KX|Y) = H(K|Y) - H(K|XY)$ oppure come
$I(KX|Y) = H(X|Y) - H(X|YK)$
da cui otteniamo:

$$H(K|Y) = H(K|XY) + H(X|Y) - H(X|YK). \tag{1}$$

dato che $H(X|YK) = 0$, otteniamo:

$$H(K|Y) = H(X|Y) + H(K|XY). \tag{2}$$

In più, dato che l'incertezza sulla chiave è piú grande dell'incertezza che si ha sulla chiave conoscendo anche il messaggio cifrato Y, cioé $H(K) \geq H(K|Y)$, possiamo scrivere:

$$\begin{aligned} H(K) &\geq H(K|Y) \\ &= H(X|Y) + H(K|XY) && \text{per la (2)} \\ &= H(X) + H(K|XY) && \text{per la sicurezza perfetta} \\ &\geq H(X) && \text{siccome } H(K|XY) \geq 0 \end{aligned}$$

Ció significa che l'incertezza che si ha sulla chiave è maggiore o uguale della incertezza che si ha sul messaggio X. Se supponiamo che X sia distribuito uniformemente, allora:

$$H(X) = \log |\mathcal{X}|$$

da cui la lunghezza di K è maggiore o uguale della lunghezza di X, cioé

$$\log |\mathcal{K}| \geq H(K) \geq H(X) = \log |\mathcal{X}|.$$

3.12 Macchine cifranti

Alla fine del XVIII secolo, Thomas Jefferson inventó una macchina cifrante, detta *cilindro di Jefferson* che è rimasta in uso fino alla seconda guerra mondiale. Essa consiste in un cilindro montato su un asse e diviso in 36 dischi uguali che possono ruotare liberamente attorno all'asse. Sul bordo esterno di ciascun disco sono scritte le 26 lettere dell'alfabeto, equispaziate l'una dall'altra. L'ordine con cui sono scritte le lettere non è quello naturale e varia da disco a disco.

Il messaggio in chiaro va cifrato a blocchi di 36 lettere ciascuno; si compone tale messaggio ruotando opportunamente i dischi e il messaggio cifrato si ottiene scegliendo una qualunque delle 25 "righe" rimanenti. Decifrare il messaggio è facile: basta "scrivere" sul cilindro il cifrato e scegliere, sulle restanti 25 righe, la frase che ha senso compiuto. Poiché il cilindro è formato da 36 dischi, la probabilitá di avere piú di una frase di senso compiuto è bassissima.

Una variante é quella di fissare a priori la riga su cui sará possibile trovare il messaggio in chiaro.

Nello stesso periodo in cui Vernam ideava il suo sistema crittografico, si andava affermando una classe di macchine cifranti basate su principi meccanici o elettromeccanici, le cosiddette *macchine a rotori*, che rimasero per anni le dominatrici incontrastate della scena crittografica. Il funzionamento di una macchina a rotori è concettualmente molto semplice: esso si basa su una catena di dischi mobili (i rotori, appunto) affacciati l'uno sull'altro, ciascuno dei quali costruito di materiale isolante e dotato su entrambe le facce di 26 contatti elettrici. Ogni contatto di una faccia corrisponde ad uno ed un solo contatto sull'altra faccia, ma la corrispondenza non è diretta, bensí "incrociata" in qualche modo. Ad ogni contatto corrisponde un carattere dell'alfabeto. La cifratura avviene applicando una tensione sul primo rotore al contatto corrispondente al carattere da cifrare.

La decifratura avviene similmente, basta applicare una tensione al contatto dell'ultimo rotore, corrispondente al carattere del testo da decifrare. Supponiamo che i dischi siano tre in successione. Applicando una tensione al contatto 1 del primo disco, questi la trasferisce ad un differente contatto del secondo disco per via del suo collegamento elettrico interno, il secondo fa lo stesso con il terzo cosí che alla fine il contatto alimentato in uscita corrisponde ad una lettera differente da quella applicata all'ingresso. Ora i tre dischi possono ruotare a passi discreti attorno al proprio asse e sono collegati tra loro da un meccanismo, l'*odometro*, grazie al quale, non appena viene creata la codifica di una lettera del testo in chiaro, l'ultimo disco ruota di $\frac{1}{26}$ di giro;ció ha l'effetto di modificare la struttura cifrante della macchina e quindi, in pratica, le permette di applicare al prossimo carattere una trasformazione differente da quella applicata al carattere precedente. Quando l'ultimo disco ha compiuto un'intera rotazione, quello alla sua sinistra avanza di un passo ed il ciclo riprende, ma con una struttura che non si ripete mai come prima. In pratica dunque, la macchina modifica il suo cifrario ad ogni carattere e non riapplica la stessa trasformazione se non dopo un numero straordinariamente alto di caratteri cifrati.

Tra le varie macchine del genere, costruite tra le due guerre mondiali, la tedesca *ENIGMA* è quella che ha avuto un ruolo di primaria importanza nella cifratura e decifratura di messaggi militari nel corso della seconda guerra mondiale. In questa macchina le due funzioni sono realizzate da quattro rotori, di cui l'ultimo realizza l'*involuzione* del segnale elettrico, necessaria per poter decifrare i messaggi in arrivo.

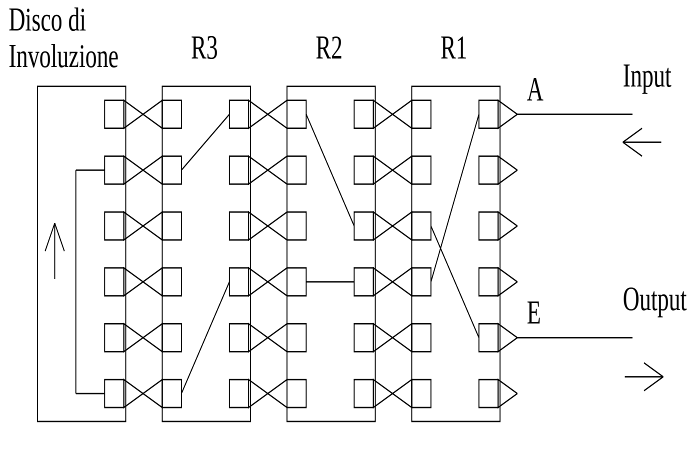


Figura 2: Schema parziale della macchina Enigma.

Per effettuare l'operazione di cifratura si procede in questo modo: il testo in chiaro viene inserito nella macchina un carattere alla volta, a partire dal rotore R1 (vedi figura 1). Tramite i collegamenti interni il carattere viaggia fino all'ultimo rotore, cambiando il suo "valore" ad ogni passaggio di rotore. Giunto nell'ultimo disco, il segnale viene invertito e rispedito verso il primo rotore, sempre attraverso collegamenti che ne modificano il "valore". All'uscita del rotore R1 si ha il carattere del testo cifrato. Per rendere piú sicura la macchina, per ogni carattere cifrato, il rotore R1 ruota di $\frac{1}{26}$ di giro; quando R1 ha effettuato una rotazione completa, allora è R2 che ruota, e cosí via.

La tecnica adottata dall'ENIGMA puó essere definita una sostituzione polialfabetica di periodo molto elevato, tanto da far pensare alla sua perfezione.

Questa macchina ha rivestito un ruolo fondamentale nella storia della computer science perché fu grazie ad essa che Turing, con l'aiuto dei piú importanti scienziati dell'epoca, riuscí a "rompere" l'ENIGMA, progettando e costruendo quello che è considerato il primo computer della storia: il COLOSSUS.

4 Crittoanalisi

La crittoanalisi è la disciplina che studia come forzare i cifrari costruiti dai crittografi, cioé come determinare la chiave che è stata utilizzata per cifrare i messaggi. Prima di affrontare la crittoanalisi di alcuni dei sistemi visti in precedenza, dobbiamo differenziare i diversi livelli di attacchi ad un crittosistema.

Il primo metodo, quello piú semplice, detto di **forza bruta**, rompe i crittosistemi facendo una ricerca esaustiva nello spazio delle chiavi; esso, ovviamente, puó essere applicato efficientemente solo quando lo spazio delle chiavi è piccolo.

Definizione 3 *Siano* $a \geq 1$ *e* $m \geq 2$ *due interi. Se* $\gcd(a,m) = 1$, *allora* a *e* m *sono relativamente primi. Il numero di interi in* $\mathbf{Z}_m$ *che sono relativamente primi ad* m *è denotato da* $\phi(m)$ *(questa funzione é detta funzione di Eulero).*

Definizione 4 *Detti* **p** *e* **q** *due interi, primi, tali che* $m = pq$, $\phi(m) = (p-1)(q-1)$.

Esempio 7 *Nei cifrari affini (vedi Sezione 3.2) il messaggio cifrato* y *si ottiene come:*

$$y = ax + b \bmod 26$$

per trovare le chiavi a *e* b *si possono avere* 26 *possibili scelte per* b *e* 12 *per* a *(perché* $|\ \mathbf{Z}^*_{26}\ | = \phi(26) = (13-1)\cdot(2-1) = 12$*); quindi occorre fare al piú* $26 \cdot 12$ *tentativi per trovare la chiave.*

Si suppone che l'intruso possa solo ascoltare i messaggi e non modificarli, i metodi di attacco ai crittosistemi sono i seguenti:

1. **Ciphertext-only**: consiste nell'ascoltare i messaggi mandati da Alice a Bob, dopodiché il nemico cerca di rompere il sistema.
2. **Known plaintext**: il nemico conosce alcune coppie (x_i, y_i), rispettivamente messaggio in chiaro e messaggio cifrato.
3. **Chosen plaintext**: è simile all'attacco precedente, solo che il nemico, oltre a conoscere le coppie (x_i, y_i), puó addirittura scegliere i messaggi in chiaro di cui vuole ottenere i testi cifrati.

4. **Chosen ciphertext**: è simile al chosen plaintext, solo che il nemico ha a disposizione la macchina che decifra i messaggi.

Molte tecniche di crittoanalisi utilizzano proprietá statistiche dei linguaggi naturali. Esistono tavole che riportano le frequenze di ciascun carattere all'interno di un testo. Nella lingua italiana, per esempio, la lettera 'E' è la piú frequente, seguita dalla 'I'. Inoltre esistono tavole con le frequenze dei 'digrammi' (coppie di lettere) e dei 'trigrammi' (terne di lettere). Un esempio di tavola delle frequenze [2], calcolate su di un testo di 1000 caratteri, è la seguente :

A	B	C	D	E	F	G	H	I	J	K	L	M
103	9	43	38	126	8	20	11	116	0	0	66	26

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
66	87	32	6	67	61	61	30	15	0	0	0	9

- Tabella 2 -

Vedremo nelle prossime sezioni come utilizzare queste statistiche.

4.1 Crittoanalisi del Cifrario Affine

Questo attacco puó essere portato a qualunque cifrario a sostituzione. Sia x il testo in chiaro ed y il testo cifrato. Per questo sistema (vedi Sezione 3.2) la regola di cifratura è:

$$y = ax + b \bmod 26.$$

Il primo passo che il crittoanalista fa è calcolare le frequenze di ogni lettera nel messaggio y. Supponiamo che le frequenze del messaggio in codice siano:

T	appare	8	volte
D	"	6	"
K	"	5	"
P	"	5	"
H	"	5	"
⋮	⋮	⋮	⋮

Il crittoanalista puó supporre che la 'T', probabilmente, è la codifica della lettera piú frequente, cioé la 'E', e che la 'D' sia la codifica della 'I'. In base a queste ipotesi di tipo statistico, tenendo presente che la 'T' corrisponde alla diciannovesima lettera, la 'E' alla quarta, la 'D' alla terza e la 'I' alla ottava, si ottiene il seguente sistema di equazioni:

$$\begin{cases} 4a + b &= 19 \\ 8a + b &= 3 \end{cases}$$

La soluzione di questo sistema è:

$$\begin{cases} a &= 22 \\ b &= 9 \end{cases}$$

Tale soluzione non è ammissibile perché $\gcd(a, 26) \neq 1$, quindi l'ipotesi fatta non é giusta. Il crittoanalista allora considera una nuova possibilitá: 'T' è la codifica di 'E' e 'K' è la codifica di 'I'. Il sistema diventa :

$$\begin{cases} 4a + b &= 19 \\ 8a + b &= 10 \end{cases}$$

Il sistema fornisce $a = 24$, ancora non ammissibile. Supponendo invece che 'T' sia la codifica di 'E' e 'P' quella di 'I', si ottiene la soluzione

$$\begin{cases} a & = 25 \\ b & = 0 \end{cases}$$

che è ammissibile. Per provare che questa è effettivamente la chiave, si cifra il testo con essa: se si ottiene un testo che ha senso compiuto allora la chiave individuata è quella giusta.

4.2 Crittoanalisi del Cifrario di Vigenère

In questo attacco si suppone che il testo in chiaro $X = (x_1, x_2, \ldots, x_n) \in (\mathbf{Z}_{26})^n$ sia in linguaggio naturale. Nella cifratura del crittosistema di Vigenère, il testo cifrato $Y = (y_1, y_2, \ldots, y_n) \in (\mathbf{Z}_{26})^n$ è ottenuto tramite la chiave $K = (k_1, k_2, \ldots, k_m) \in (\mathbf{Z}_{26})^m$ dividendo il testo in chiaro in blocchi di m caratteri ed effettuando la somma modulo 26 carattere per carattere, tra il primo blocco di m caratteri di X e gli m caratteri della chiave K, poi tra il secondo blocco di m caratteri di X e gli m caratteri di K, e così via. Il messaggio in chiaro deve essere abbastanza più lungo della chiave, in modo che le regolarità statistiche prodotte nel messaggio cifrato dall'applicazione ripetuta della stessa chiave, ci consentiranno di scoprire la chiave K con buona probabilità . Per calcolare la chiave dobbiamo prima calcolare la sua lunghezza e poi i valori numerici assunti da ciascuna componente. Il primo problema che affrontiamo è il calcolo della lunghezza della chiave.

Kasiski lo ha risolto fornendo un metodo che va alla ricerca di alcune regolaritá che si ripetono all'interno del testo cifrato. Il *test di Kasiski* é basato sull'osservazione che due segmenti identici del testo in chiaro vengono cifrati allo stesso modo solo se la porzione di chiave atta a cifrarli è la stessa per entrambi. Di contro, se troviamo due segmenti identici del testo cifrato, ciascuno di lunghezza almeno tre, allora ci sono buone probabilitá che essi corrispondano a pezzi identici del testo in chiaro. Il test di Kasiski cerca le coppie di segmenti identici nel testo cifrato di lunghezza almeno tre e memorizza la distanza tra le posizioni iniziali dei due segmenti. Se si ottengono le distanze d_1, d_2, ..., allora si puó dedurre che m (lunghezza della chiave) divida il gcd dei d_i. Il limite di questo metodo è che fornisce solo la probabile lunghezza della chiave e non il suo valore.

Un secondo attacco al cifrario di Vigenére si basa sul calcolo delle statistiche relative al testo cifrato. Esso consiste in due passi: nel primo viene calcolata la lunghezza della chiave, nel secondo il suo valore.

Definizione 5 *Supponiamo che $X = x_1\, x_2 \ldots x_n$ sia una stringa di n caratteri dell'alfabeto A,B,..., Z. L'indice di coincidenza di X, denotato con $IC(X)$, è definito come la probabilitá che due elementi scelti a caso in X siano uguali.*

Ricordiamo, a questo punto, che la probabilitá si puó calcolare come

$$\frac{numero\ eventi\ favorevoli}{numero\ eventi\ possibili}$$

e osserviamo che, data una stringa $X = x_1\, x_2 \ldots x_n$, il numero di modi di scegliere due caratteri è $\binom{n}{2}$. Se indichiamo con $f_0, f_1, \ldots, f_{25}$ il numero delle occorrenze dei caratteri da A a Z in X allora il numero di modi di scegliere due caratteri in X entrambi uguali ad A è $\binom{f_0}{2}$, entrambi uguali a B è $\binom{f_1}{2}$, da cui otteniamo che il numero di casi favorevoli è :

$$\binom{f_0}{2} + \binom{f_1}{2} + \ldots + \binom{f_{25}}{2}.$$

Allora risulta:

$$IC(X) = \frac{\sum_{i=0}^{25} f_i(f_i - 1)}{n(n-1)}. \tag{3}$$

Supponiamo che il messaggio in chiaro X sia in lingua italiana, dal momento che si conosce la distribuzione tipica dei caratteri in un testo, cioé si sa che la 'E' è la piú probabile , la 'I' è la seconda piú probabile,...etc secondo la Tabella 2, si puó calcolare il valore medio di $IC(X)$. Dette $p_0, p_1, \ldots, p_{25}$ le probabilitá delle occorrenze delle lettere 'A', 'B',. . ., 'Z', riportate nella Tabella 2, abbiamo che la probabilitá che scelte a caso due posizioni in un testo qualunque, le lettere occupanti tali posizioni siano due A è p_0^2, due B p_1^2,...etc. Allora si ha che :

$$IC(X) \approx \sum_{i=0}^{25} p_i^2 = 0.075.$$

Se supponiamo che X venga cifrato usando una chiave $K = k$ allora l'indice di coincidenza del messaggio in codice è lo stesso di X perché i vari caratteri in X sono sostituiti sempre allo stesso modo, shiftando ciascuno dei caratteri del valore k, e quindi le coincidenze nel messaggio in codice sono le stesse del messaggio in chiaro. Questo ci aiuta perché, se prendendo il messaggio cifrato e calcolandone l'indice di coincidenza, otteniamo un valore vicino a quello del testo in chiaro che è 0.075 allora possiamo ipotizzare che il testo sia stato cifrato usando sempre la stessa chiave.

Osserviamo peró che in un testo i cui caratteri sono scelti a caso si ha:

$$\begin{cases} p_0 & = \frac{1}{26} \\ p_1 & = \frac{1}{26} \\ & \vdots \\ p_{25} & = \frac{1}{26} \end{cases} \tag{4}$$

Quindi il valor medio dell'indice di coincidenza è :

$$\sum_{i=0}^{25} \left(\frac{1}{26}\right)^2 = 0.038$$

che è sufficientemente distante da 0.075; questo ci permette di dire che se l'indice di coincidenza del messaggio in codice è diverso da 0.075 e vicino a 0.038 allora la chiave usata per cifrare il messaggio in chiaro non è la stessa. Pertanto si puó, in questo modo, determinare correttamente la lunghezza della chiave. Detta m la lunghezza della chiave, l'analisi procede come segue:

- ipotizziamo $m = 1$. Calcoliamo l'indice di coincidenza dell'intero messaggio cifrato $C = c_1 c_2 \ldots$; se tale valore è lontano da quello atteso, allora

- ipotizziamo $m = 2$. In tal caso il testo cifrato viene scritto come

$$\begin{array}{cc} k_1 & k_2 \\ \downarrow & \downarrow \\ c_1 & c_2 \\ c_3 & c_4 \\ c_5 & c_6 \\ \vdots & \vdots \end{array}$$

dove la chiave usata per la cifratura è k_1k_2, con k_1 che viene usata per cifrare tutti i caratteri della prima colonna, mentre k_2 tutti quelli della seconda. In questo caso calcoliamo due indici di coincidenza, uno per i caratteri contenuti nella prima, l'altro per i caratteri contenuti nella seconda colonna; se entrambi i valori sono abbastanza vicini a quello atteso, allora questa ipotesi è accettabile, cioé la chiave avrá, con buona probabilitá, lunghezza 2.

Se nemmeno questa scelta di m è giusta, allora proviamo per m=3 e cosí via.

Una volta individuata la lunghezza della chiave potremmo attaccare il crittosistema usufruendo della sola lunghezza della chiave provando, con la forza bruta, tutte le possibili chiavi. Se, per esempio, la lunghezza della chiave fosse 7, potremmo provare tutte le 26^7 chiavi possibili. Risulta evidente che una tale soluzione é di fatto impraticabile per cui utilizzeremo una tecnica statistica che generalizza l'indice di coincidenza a due stringhe.

Definizione 6 *Siano $X = x_1x_2\dots x_n$ e $Y = y_1y_2\dots y_{n'}$ due stringhe di lunghezza n ed n', rispettivamente. L'indice mutuo di coincidenza di X e Y, denotato con $\mathbf{MI}(X,Y)$ è definito come la probabilitá che, preso un carattere da X e uno da Y, essi siano uguali. Dette $f_0, f_1, \dots, f_{25}$ e $f'_0, f'_1, \dots, f'_{25}$ il numero di occorrenze di A, B, C, ..., Z in X ed Y, rispettivamente, allora*

$$MI(X,Y) = \frac{\sum_{i=0}^{25} f_i f'_i}{nn'}.$$

Possiamo calcolare il valor medio di MI nel caso in cui conosciamo le distribuzioni di X e Y. Supponiamo che X_1 e Y_1 siano due messaggi ottenuti cifrando rispettivamente X ed Y con le chiavi k_1 e k_2 in $\mathbf{Z}_{26}$, allora possiamo calcolare $MI(X_1, Y_1)$ calcolando le probabilitá di avere due caratteri uguali. Avendo, per esempio, $k_1 = 1$ e $k_2 = 4$, si puó dire che la probabilitá di avere in X_1 una A ed in X_2 una A è pari alla probabilitá di avere Z in X e W in Y, mentre la probabilitá di avere GG è uguale alla probabilitá di avere FC, ...etc. Indicando con $p_0, p_1, \dots, p_{25}$ le probabilitá riportate nella Tabella 2 e ricordando che l'aritmetica usata per gli indici è modulo 26, abbiamo che la probabilitá di avere A in X_1 e A in Y_1 è :

$$p_{-k_1}p_{-k_2}$$

quella di avere B in X_1 e B in Y_1 è :

$$p_{1-k_1}p_{1-k_2}$$

e cosí via. Ma allora il valor medio è:

$$MI(X,Y) \approx \sum_{i=0}^{25} p_{i-k_1}p_{i-k_2} \tag{5}$$

Osserviamo che dalla (5), il valor medio dell'indice mutuo di coincidenza dipende solo dalla differenza tra le due chiavi. Infatti, ponendo $h = i - k_1$ otteniamo:

$$\sum_{i=0}^{25} p_{i-k_1}p_{i-k_2} = \sum_{h=0}^{25} p_h p_{h+(k_1-k_2)}.$$

Questo risultato ci dice che i possibili indici mutui di coincidenza sono 26. Dato che, preso un certo valore l, si ha:

$$\sum_{h=0}^{25} p_h p_{h-l} = \sum_{h=0}^{25} p_h p_{h+l}$$

si deduce che il valore atteso per l'indice mutuo di coincidenza dipende solo da $l = k_1 - k_2$ e puó assumere solo 14 valori. Tali valori possono essere tabulati in due colonne di cui la prima rappresenta la differenza tra le due chiavi e la seconda rappresenta il valore atteso dell'indice mutuo di coincidenza. La Tabella per la lingua italiana è la seguente:

shift relativo	valore atteso di MI
0	0.075
1	0.033
2	0.034
3	0.034
4	0.047
5	0.027
6	0.032
7	0.026
8	0.027
9	0.023
10	0.024
11	0.027
12	0.015
13	0.021

- Tabella 3 : Valori attesi di MI -

Come possiamo osservare solo il valore relativo a $l = 0$ è 0.075, mentre gli altri oscillano nell'intervallo (0.015, 0.047).

A questo punto possiamo vedere un esempio di come vengono usati questi valor medi di MI con la lunghezza della chiave pari a 3. Il messaggio cifrato verrá scritto come

$$\begin{array}{ccc} k_1 & k_2 & k_3 \\ \downarrow & \downarrow & \downarrow \\ c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 \\ \vdots & \vdots & \vdots \end{array}$$

dato che la chiave k_1 cifra $c_1 c_4 c_7 \ldots$, k_2 cifra $c_2 c_5 \ldots$, k_3 cifra $c_3 c_6 \ldots$.

Se lo shift tra k_1 e k_2 è 0, cioè k_1 - k_2 = 0, allora $MI(a, b)$, con a prima colonna e b seconda, è vicino al valore atteso 0.075; se lo shift è diverso da zero, ad esempio è 1, cioè k_1 - k_2 = 1, allora $MI(a, b')$, con $b' = c_2{+}1\ c_5{+}1 \ldots$ la colonna ottenuta da b shiftando tutti i suoi caratteri di una posizione, è vicino al valore atteso 0.075. Analogamente si procederá con gli altri shift 2, 3, $\cdots$ 25. Per questi 26 valori ce ne sará uno solo per cui MI sará vicino a 0.075 da cui possiamo ipotizzare di aver trovato il valore dello shift. In questo modo è possibile calcolare lo shift relativo tra il primo e il secondo carattere della chiave, tra il primo e il terzo, e cosí via, ottenendo un sistema di $m - 1$ equazioni in cui tutte le chiavi sono espresse in funzione della prima, k_1, che è incognita.

A questo punto basterá provare le 26 possibili scelte per k_1 per avere il valore corretto della chiave.

Di seguito presentiamo lo pseudocodice di un algoritmo che realizza i due passi della crittoanalisi appena descritta.

CALCOLO DELL'INDICE DI COINCIDENZA

```
FUNZ_COINCIDENZA (Y,lung_max)
(1) for lung_chiave = 1 to lung_max
(2)         INIZIALIZZA(FREQ)
(3)         for riga= 0 to lung_chiave - 1
(4)                 for colonna = riga to lung_mess step riga
(5)                         car_to_int = cifrato[colonna] - 65
(6)                         FREQ[riga,car_to_int] = FREQ[riga,car_to_int] + 1
(7)         for riga = 0 to lung_chiave - 1
(8)                 for colonna = 0 to 25
(9)                         lung[riga] = lung[riga] + FREQ[riga,colonna]
(10)        INIZIALIZZA(INDICE)
(11)        for riga = 0 to lung_chiave - 1
(12)                for colonna = 0 to 25
(13)                if FREQ[riga,colonna] ≥ 2
(14)                        INDICE[lung_chiave,riga] = INDICE[lung_chiave,riga] + ( FREQ[riga,colonna]*(FREQ[riga,colonna] - 1))
(15)                INDICE[lung_chiave,riga] = INDICE[lung_chiave,riga] / (lung[riga] * (lung[riga] - 1))
(16)        cont = 0
(17)        for colonna = 0 to lung_chiave - 1
(18)                if  | INDICE[lung_chiave,colonna] - ATTESO | < ERRORE
(19)                        cont = cont + 1
(20)        if cont = lung_chiave
(21)                min = indice piu' vicino al valore atteso
(22)                SCOSTAMENTO[lung_chiave] = min
(23) CALCOLO DEL MINIMO INDICE DI COINCIDENZA TRA GLI SCOSTAMENTI
(24) STAMPA DELLA PROBABILE LUNGHEZZA DELLA CHIAVE
(25) INIZIALIZZA(vettore[,])
```

La funzione FUNZ_COINCIDENZA calcola l'indice di coincidenza per tutte le possibili sottostringhe in cui puó essere suddiviso il messaggio cifrato Y in base alla lunghezza della chiave, che puó assumere al piú lunghezza lung_max. Ottenuti tutti i possibili valori si calcola il piú piccolo scostamento che si ha tra il valore atteso e quelli risultanti. Questi scostamenti sono memorizzati in un vettore, in modo tale che ad ogni valore della lunghezza della chiave corrisponda uno scostamento. Lo scostamento piú piccolo corrisponde al valore piú probabile per la lunghezza della chiave.

Nel dettaglio, le linee da (3) a (6) calcolano le frequenze dei caratteri per ogni sottostringa ottenuta dal messaggio cifrato, conservando questi valori in un vettore bidimensionale FREQ. Ad ogni passo del ciclo di linea (1) questo vettore viene inizializzato a 0 (linea (2)), in modo tale che le frequenze siano sempre relative al singolo valore della lunghezza della chiave.

Le linee da (7) a (9) calcolano la lunghezza delle sottostringhe, che comunque non possono differire per piú di un elemento.

La linea (10) inizializza il vettore che contiene i valori degli indici di coincidenza calcolati per i singoli valori della lunghezza della chiave. Essi ovviamente cambiano ad ogni passo del ciclo in linea (1), da cui la necessitá di inizializzare.

Nelle linee da (11) a (15) è calcolato effettivamente il valore degli indici di coincidenza, realizzando la formula 3.

A questo punto si verifica che tutti i valori degli indici di coincidenza ottenuti, si avvicinino con buona approssimazione al valore ATTESO; se ció accade, cioé la variabile cont assume lo stesso valore di lung_chiave allora si passa al calcolo dello scostamento piú piccolo, memorizzando tale valore in un vettore chiamato SCOSTAMENTO.

Alla fine, quando tutti i valori possibili di lung_chiave sono stati analizzati, si calcola il minimo scostamento e di conseguenza la probabile lunghezza della chiave.

CALCOLO DELL'INDICE MUTUO DI COINCIDENZA

```
FUNZ_MUTUO_INDICE(lung_chiave)
(1) for riga= 0 to lung_chiave - 1
(2)         for colonna = riga to lung_mess step riga
(3)                 car_to_int = cifrato[colonna] - 65
(4)                 FREQ[riga,car_to_int] = FREQ[riga,car_to_int] + 1
(5)         for riga = 0 to lung_chiave - 1
(6)         for colonna = 0 to 25
(7)                 lung[riga] = lung[riga] + FREQ[riga,colonna]
(8) min = 1
(9) forriga = 1 to lung_chiave - 1
(10)    for colonna = riga + 1 to lung_chiave
(11)                forshift = 0 to 25
(12)                   tmp = 0
(13)                   for somma = 0 to 25
(14)                         if FREQ[riga,somma] ≠ 0 AND FREQ[colonna,(somma + shift) mod 26] ≠ 0
(15)                               tmp = tmp + (FREQ[riga,somma] * FREQ[colonna,(somma + shift) mod 26])
(16)                         MIC[riga,colonna,shift] = tmp / (lung[riga]* lung[colonna])
(17)                calcolo del MIC che si avvicina maggiormente al valore atteso
(18)                COSTRUZIONE DEL SISTEMA PER IL CALCOLO EFFETTIVO DELLA CHIAVE
```

La FUNZ_MUTUO_INDICE prende in input dalla FUNZ_COINCIDENZA la probabile lunghezza della chiave e restituisce il valore delle possibili chiavi.

Nel dettaglio, dalla linea (1) alla (7) si calcolano le frequenze dei caratteri che compaiono nelle lung_chiave sottostringhe in cui il messaggio cifrato viene suddiviso e le relative lunghezze.

Dalla linea (9) alla (16) viene implementata la formula 5. Nella linea (17) il programma calcola il minimo scarto tra il valore ATTESO e quello dell'indice mutuo di coincidenza.

Nella linea (18), infine, viene risolro il sistema di equazioni lineari che calcola il valore della chiave in funzione di k_1.

Esempio 8 *Analizziamo ora la crittoanalisi di un testo cifrato con il sistema di Vigenère.*

Il testo cifrato è:

```
RLEYFBDOQSMCATCEZCBAPTHRJPCGRONVZMCHZOEBPKRNRVVCNHFEEACOZNGS
SIOGHFUIZCOKIGIUKONGFEIRUPCFVOTVCBBERDRZMFSCCSXEESFUEYFJVNGF
BIEQWRLEYZJMIRBRLAFWBLNGFBKTBOSVSGFJEGRFTZENDSVNQSSTOEGPVFVU
VIAQWGZUZSUIAHBQIOZCOKOEWPRDRGUIARIORMCWBTOFHJVRNRBCLNZUIACO
SKERWMGOAHFTHRWWZCBBHZUAUFCEQIFIIISQRRPVFIEARBJAZQPIPVITVNFW
CZLROMCOPQIZODIFJTNHSRSSCSDAMWPEERGFXNVWMGUAHPZNPIJZLYOHFCRG
TREYOEUAEWDFMVBDZACSSIICWHCINFQFIACNVDVZBXOQCWVLRFJMENZMFNGO
ORNQCTZDVBVFVBZBJCVOOCAPEVRDVGUVNQSSJIRFBCLRBURRFWJENHCWZGBZ
GZEVBOLOIWTVNVZBTOFHJVRNTPIMNHBUAYRFGOFWUFDVHSVGECTJIGCSIEAH
JJCRBEVACDPXGVOURAQIFDOAHJTOAHJXUVZVEOQSUKOVZTRNZOSKIACMRLGF
PTOAJPTEYCNSAERBZLESTVGBBFUAVAPCTVQPTUMNPCIVBGZLNQIVIAJFIOYC
GRNACTFMVUMZAESBLNNGFXAGOMTHRBPEEPVJRLCFJDOISEVRYCQLRPVFJINR
JWRBBUVCBAFGEESTVMCWPUIFIMVMHFBUIZWMRNBQIVGHOSUAACBJEGHFETEW
PEEACOCOQWTTEEBBKOFHPRUAHBCCBBUIAFGFXNBWOHURZMRLHBHREIOTKATW
PXAVOERGYWBCTEWNFNGWEZNBAFGIHCTTUECFUISCSDACWVTOZIOVPRFVEBHC
OGEMNPCAPCTKAFOMVCBBVEPRBEZOYSOKORQPETVBVFPBWTZRBAQVIADPXGVS
JEVNZMFNPSMCIVBFITRSJEIFDJRNNHFJEPCOUOYCTJAGISRDRRVVMBBUZEVZ
```

La crittoanalisi ci suggerisce, per prima cosa, di calcolare l'indice di coincidenza del testo cifrato che ci forniŕa la probabile lunghezza della chiave.

Detta m la probabile lunghezza della chiave, l'indice di coincidenza per $m = 1$ è 0.0450. Con $m = 2$ si hanno due indici di coincidenza che sono 0.0463 e 0.0438. Per $m = 3$ gli indici sono 0.0431, 0.0459 e 0.0456. Per $m = 4$ gli indici di coincidenza sono 0.0448, 0.0421, 0.0495, 0.0437. Per $m = 5$ abbiamo: 0.0710, 0.0721, 0.0805, 0.0684, 0.0759 e come si vede, gli indici sono tutti molto "vicini" al valore atteso (0.075). Questo fatto ci dice che la lunghezza probabile della chiave è 5.

A questo punto è necessario calcolare il mutuo indice di coincidenza che ci fornirá effettivamente, tramite la risoluzione di un sistema di equazioni, la probabile chiave. Nella tabella 3 sono riportati tutti i valori dell'indice mutuo di coincidenza. Tale tabella ha nelle prime due colonne l'indice dei caratteri della chiave per i quali si effettua il calcolo di MI, mentre nella terza colonna sono riportati tutti i 26 valori calcolati, uno per ogni possibile shift. Come si puó notare, i valori racchiusi tra rettangolini sono quelli che maggiormente si avvicinano al valore atteso 0.075. Avendo ipotizzato che la chiave ha lunghezza 5, avremo $\binom{5}{2} = 10$ equazioni e cioé:

$$k_1 - k_2 = 16$$

$$k_1 - k_3 = 25$$

$$k_1 - k_4 = 12$$

$$k_1 - k_5 = 13$$

$$k_2 - k_3 = 9$$

$$k_2 - k_4 = 22$$

$$k_2 - k_5 = 23$$

$$k_3 - k_4 = 13$$

$$k_3 - k_5 = 14$$

$$k_4 - k_5 = 1$$

Le precedenti equazioni non sono linearmente indipendenti. Pertanto sarà sufficiente risolvere le prime quattro, dato che le altre sono combinazioni lineari di queste. Ad esempio, sottraendo la prima equazione dalla seconda si ottiene la quinta equazione:

$$k_1 - k_3 - k_1 + k_2 = 25 - 16$$

$$k_2 - k_3 = 9$$

Calcolando tutte le equazioni in funzione di k_1, otteniamo la chiave

$$(k_1,\ k_1 - 16,\ k_1 - 25,\ k_1 - 12,\ k_1 - 13)$$

Sostituendo a k_1 tutti i possibili valori in $\mathbf{Z}_{26}$ otteniamo 26 differenti chiavi tra cui la chiave BRANO *che ci fornisce il messaggio in chiaro che è:*

```
QUELRAMODELLAGODICOMOCHEVOLGEAMEZZOGIORNOTRADUECATENENONINTE
RROTTEDIMONTITUTTOASENIEGOLFIASECONDADELLOSPORGEREEDELRIENTR
AREDIQUELLIVIENQUASIAUNTRATTOARESTRINGERSIEAPRENDERCORSOEFIG
URADIFIUMETRAUNPROMONTORIOADESTRAEUNAMPIACOSTIERADALLALTRAPA
RTEEILPONTECHEIVICONGIUNGELEDUERIVEPARCHERENDASAMCORPIUSENSI
BILEALLOCCHIOQUESTATRASFORMAZIONEESEGNIILPUNTOINCUIILLAGOCES
SAELADDARICOMINCIAPERRIPIGLIARPOINOMEDILAGODOVELERIVEALLONTA
NANDOSIDINUOVOLASCIANLACQUADISTENDERSIERALLENTARSIINNUOVIGOL
FIEINNUOVISENILACOSTIERAFORMATADALDEPOSITODITREGROSSITORRENT
ISCENDEAPPOGGIATAADUEMONTICONTIGUILUNODETTOILSANMARTINOLALTR
OCONVOCELOMBARDAILRESEGONEDAIMOLTICOCUZZOLIINFILACHEINVEROLO
FANNOSOMIGLIAREAUNASEGATALCHENONECHIALPRIMOVEDERLOPURCHESIAD
IFRONTECOMEPERESEMPIODISULEMURADIMILANOCHEGUARDANOASETTENTRI
ONENONLODISCERNATOSTOAUNTALCONTRASSEGNOINQUELLALUNGAEVASTAGI
OGAIADAGLIALTRIMONTIDINOMEPIUOSCUROEDIFORMAPIUCOMUNEPERUNBUO
NPEZZOLACOSTASALECONUNPENDIOLENTOECONTINUOPOISIROMPEINPOGGIE
INVALLONCELLIINERTEEINISPIANATESECONDOLOSSATURADEDUEMONTIEIL
```

i	j	Valore di MI
1	2	.0325 .0415 .0422 .0436 .0385 .0444 .0388 .0390 .0347
		.0350 .0404 .0315 .0419 .0398 .0370 .0380 $\boxed{.0703}$ .0314
		.0346 .0356 .0436 .0269 .0327 .0298 .0381 .0371
1	3	.0326 .0341 .0345 .0365 .0245 .0367 .0284 .0393 .0394
		.0373 .0358 .0432 .0439 .0399 .0382 .0363 .0334 .0315
		.0355 .0449 .0384 .0518 .0403 .0313 .0370 $\boxed{.0738}$
1	4	.0380 .0407 .0370 .0381 .0295 .0330 .0415 .0361 .0423
		.0411 .0330 .0411 $\boxed{.0705}$ .0364 .0324 .0361 .0460 .0301
		.0321 .0316 .0397 .0355 .0354 .0423 .0390 .0403
1	5	.0401 .0393 .0379 .0353 .0345 .0273 .0357 .0461 .0371
		.0439 .0420 .0288 .0412 $\boxed{.0737}$ .0352 .0350 .0401 .0401
		.0328 .0387 .0311 .0403 .0368 .0348 .0370 .0340
2	3	.0361 .0328 .0311 .0389 .0334 .0533 .0355 .0390 .0286
		$\boxed{.0741}$ .0328 .0437 .0325 .0415 .0272 .0406 .0284 .0378
		.0428 .0382 .0446 .0380 .0463 .0358 .0395 .0260
2	4	.0465 .0302 .0369 .0320 .0391 .0410 .0361 .0488 .0354
		.0447 .0351 .0440 .0297 .0429 .0318 .0309 .0336 .0327
		.0442 .0347 .0362 .0328 $\boxed{.0721}$ .0344 .0412 .0318
2	5	.0355 .0419 .0339 .0436 .0320 .0408 .0423 .0371 .0470
		.0334 .0434 .0374 .0414 .0295 .0400 .0296 .0317 .0375
		.0328 .0434 .0355 .0322 .0314 $\boxed{.0711}$ .0330 .0415
3	4	.0443 .0393 .0421 .0358 .0426 .0318 .0269 .0392 .0318
		.0378 .0321 .0363 .0372 $\boxed{.0724}$ .0348 .0354 .0342 .0533
		.0364 .0391 .0324 .0373 .0358 .0315 .0419 .0360
3	5	.0353 .0453 .0415 .0367 .0310 .0374 .0296 .0307 .0446
		.0303 .0350 .0321 .0321 .0376 $\boxed{.0779}$ .0343 .0343 .0357
		.0470 .0397 .0478 .0344 .0388 .0369 .0329 .0399
4	5	.0382 $\boxed{.0736}$ .0368 .0349 .0367 .0414 .0390 .0434 .0293
		.0336 .0420 .0351 .0427 .0329 .0388 .0361 .0427 .0327
		.0366 .0317 .0326 .0402 .0367 .0450 .0345 .0319

- Tabella 4 -

Riferimenti

[1] D. Kahn, *The codebreakers: the story of secret writing.* Macmillan, New York 1967.

[2] A. Lange e E.A. Soudart, *Traité de Cryptographie*, Parigi 1925 (la traduzione inglese è pubblicata dall'Aegean Park Press).