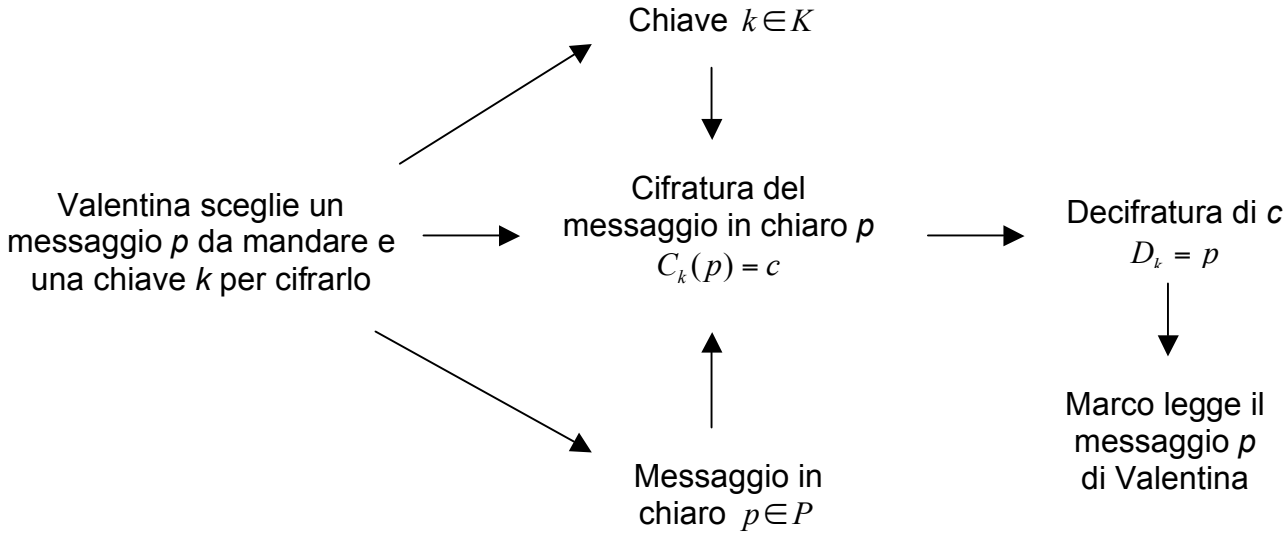
I LEZIONE

Il nostro intento è analizzare in dettaglio i metodi di cifratura che si sono susseguiti nel corso della storia prestando particolare attenzione all'impianto matematico che ne consente la realizzazione. Iniziamo il nostro studio approfondendo il primo esempio di crittografia che la storia ci fornisce. Svetonio, uno storico del II sec d.C., nella sua Vita dei Cesari parla di un sistema utilizzato da Cesare per cifrare i suoi messaggi: egli spostava di tre lettere ogni lettera del messaggio da inviare.

Innanzi tutto ricordiamo che un crittosistema è costituito da:

- l'insieme dei messaggi in chiaro *P* di cui indichiamo gli elementi con la lettera p;
- l'insieme delle chiavi *K* in cui ogni elemento k determina una trasformazione di cifratura C_k e una trasformazione di decifratura D_k che sono una l'inversa dell'altra;
- l'insieme dei messaggi cifrati C i cui elementi sono indicati con la lettera c.

Abbiamo quindi che un crittosistema è determinato da (*P,K,C*) e la comunicazione tra due persone, Valentina e Marco, può essere riassunta dal seguente diagramma:



Nel cifrario di Cesare:

- gli elementi p sono le parole che vogliamo inviare;
- la chiave consiste in fase di cifratura nello spostare di tre posti le varie lettere (C_k) e in fase di decifratura nel rimetterle nella loro corretta posizione (D_k);
- gli elementi c sono il risultato dell'operazione di cifratura.

Se indichiamo con lettere minuscole le 21 lettere dell'alfabeto, ciascuna lettera del nostro messaggio (testo in chiaro) sarà sostituita con la lettera che si trova tre posizioni più avanti, e che per comodità indicheremo con caratteri maiuscoli, ottenendo così un nuovo messaggio (testo cifrato) apparentemente privo di significato

a b c d e f g h i l m n o p q r s t u v z
D E F G H I L M N O P Q R S T U V Z A B C

Ad esempio se il messaggio da inviare è il seguente:

domani non andiamo a scuola

il risultato dopo la cifratura sarà:

GRPDQN QRQ DQGNDPR D VFAROD

Possiamo decidere di generalizzare questo sistema utilizzando una chiave diversa cioè decidendo di spostare le lettere non più di tre posizioni ma di una quantità arbitraria.
Un sistema di questo tipo, in cui l'alfabeto cifrante è ottenuto dall'alfabeto in chiaro spostando di un certo numero di posizioni le lettere, prende il nome di **cifrario di Cesare** o di **cifratura per traslazione**.
Le possibilità per i cifrari di Cesare nel caso della lingua italiana sono solamente 20 perché ovviamente se una lettera si sposta di 21 posizioni, ritorna al punto di partenza. Mentre nel caso dell'alfabeto inglese abbiamo 25 alfabeti cifranti possibili dato che le lettere sono 26.

Abbiamo visto che volendo cifrare un messaggio usando il metodo di Cesare dobbiamo sostanzialmente traslare le lettere di un certa quantità di posizioni (che decidiamo noi e rappresenta la chiave utilizzata per cifrare).
In effetti possiamo interpretare così la procedura che attuiamo: assegnamo ad ogni lettera dell'alfabeto in chiaro un numero corrispondente alla sua posizione (la A occupa la posizione 0 e la Z la posizione 20), dopodichè decidiamo un numero (ad esempio 5) che rappresenta la nostra chiave e lo sommiamo ad ogni posizione così da ottenere che nell'alfabeto cifrante la A corrisponda alla lettera in posizione 5 cioè alla F, la B alla G,…, la R alla Z, la S che occupa la posizione 16 corrisponda alla A, la T alla B,…., la Z alla E.
Il nodo fondamentale di questa procedura sta nel fatto che quando abbiamo deciso la corrispondenza tra la T e la B abbiamo sostanzialmente ragionato cosi:

- la lettera T occupa la posizione 17 (ricorda che partiamo dalla posizione 0),
- 17+5=22 ma le posizioni possibili sono 21 e sono numerate da 0 a 20 quindi il numero 22 non corrisponderebbe a nessuna lettera
- $22 = 1 \cdot 21 + 1$ e abbiamo deciso che la lettera T doveva corrispondere a quella in posizione 1 cioè alla B

Ed è per questo motivo che la lettera S corrisponde alla A (perché S= posizione 16, 16+5=21, $21 = 1 \cdot 21 + 0$ e la lettera A occupa la posizione 0), la U alla C e così via.
Quindi da un punto di vista matematico quando cifriamo con questo metodo operiamo una somma (per traslare) dopodichè se il risultato è maggiore di 21 ci interessiamo solo al resto della divisione.
Questo modo di procedere può sembrare astruso ma in realtà lo utilizziamo quotidianamente ad esempio quando leggiamo l'ora sull'orologio. Infatti quando leggiamo l'orologio sappiamo benissimo che, ad esempio, le 13:00 corrispondono all'1:00 e le 18:00 alle 6:00, e il motivo di questa corrispondenza sta nel fatto che il resto della divisione per 12 di 13 è 1 mentre di 18 è 6.

Cerchiamo di generalizzare il concetto.
Innanzi tutto diamo una definizione cercando di capire che connessione ha con quanto detto finora.

DEF. Sia n un intero positivo fissato. Due numeri $a, b \in Z$ sono ***congrui modulo n*** se e solo se a-b è un multiplo di n, ovvero, espresso in formule, $a \equiv b \ \ (n) \Leftrightarrow (a-b) = n \cdot h$ per qualche $h \in Z$.

Esempi:

1. $25 \equiv 1 \ \ (3)$ perché $25 - 1 = 24 = 3 \cdot \mathbf{8}$
2. $67 \equiv 55 \ \ (6)$ perché $67 - 55 = 12 = 6 \cdot \mathbf{2}$
3. $55 \equiv 1 \ \ (6)$ perché $55 - 1 = 54 = 6 \cdot \mathbf{9}$
4. $-5 \equiv 1 \ \ (6)$ perché $-5 - 1 = 6 \cdot \mathbf{1}$

Osservazioni.

1. Notiamo che gli esempi 2 e 3 ci suggeriscono l'idea di transitività della congruenza. Infatti vale anche che $67 \equiv 1 \ \ (6)$ perché $67 - 1 = 66 = 6 \cdot \mathbf{11}$.

2. Ora possiamo dire che quando dobbiamo criptare un messaggio operiamo modulo 21. Per di più invece di lavorare con tutti i possibili numeri interi lavoriamo solo con quelli compresi tra 0 e 20 (che sono le posizioni possibili). Infatti se, per assurdo, volessimo operare una traslazione di 63 posizioni, per trovare come criptare, ad esempio, la lettera B dovremmo capire a quale numero b è congruo modulo 21 il numero a=64=63+1 (che rappresenta la posizione traslata della lettera B). Ovviamente scoprire che $64 \equiv 43 \ (21)$ anche se è un risultato esatto non ci fornisce nessuna informazione utile perché la lettera numero 43 non esiste! Quindi, in realtà, quando vogliamo capire a "cosa" è congruo un certo numero a modulo n siamo interessati a trovare quell'unico numero b compreso tra 0 e n-1 tale che sia verificata la congruenza.

Mettendo insieme le due osservazioni precedenti possiamo ricavarne che modulo 6 i numeri 67, 55, -5, -23 sono in realtà lo stesso numero e cioè 1 (provare per credere!!!). Perché? Perché il resto della divisione per 6 di questi numeri è sempre congruo a 1 modulo 6. Infatti 67/6 dà resto 1 (perché $67 = 6 \cdot 11 + 1$) e banalmente $1 \equiv 1 \ (6)$ oppure -23/6 dà per resto -5 (perché $-23 = 6 \cdot (-3) - 5$) e $-5 \equiv 1 \ (6)$, e così via.

Quindi, in generale, possiamo concludere che, fissato n, tutti i numeri 0,….,n-1 siano in un certo senso speciali in quanto un qualsiasi altro numero intero (positivo o negativo) sarà congruo ad uno di essi modulo n.
Questo ruolo speciale sta proprio nel fatto che scelto un qualsiasi b tale che $0 \le b \le n-1$ tutti gli altri numeri interi il cui resto della divisione per n è uguale a b possono essere identificati con b stesso nel senso della congruenza.

Restano così definiti n insiemi di numeri ovvero:

DEF. Gli insiemi della forma $\overline{b} = \{a \in Z \mid a \equiv b(n)\}$= interi che divisi per n danno resto b si chiamano ***classi di resto modulo n***. I numeri b sono tali che $0 \le b \le n-1$ e vengono chiamati ***rappresentanti della classe***.

ESEMPIO Possiamo calcolare tutte le classi di resto modulo 4:
$\overline{0} = \{....,-16,-12,-8,-4,0,4,8,12,....\}$= interi che divisi per 4 danno resto 0
$\overline{1} = \{..,-11,-7,-3,1,5,9,....\}$= interi che divisi per 4 danno resto 1
$\overline{2} = \{...,-6,-2,2,6,10,...\}$= interi che divisi per 4 danno resto 2
$\overline{3} = \{..,-9,-5,-1,3,7,11,....\}$ = interi che divisi per 4 danno resto 3.

DEF. L'insieme di queste classi di numeri si indica con Z_n. Cioè $Z_n = \{\overline{0}, \overline{1}, ..., \overline{n-1}\}$.

ESERCIZI
- Specificare tutti gli elementi di Z_{21}.
- A che classe di resto modulo 23 appartiene il numero 74?
- Dire a che classe di resto modulo 5 appartiene il numero -7.
- Indicare la classe di resto modulo 12 di -13.
- Specificare la classe di resto modulo 7 di 63.

Allora, riassumendo in una tabella l'associazione lettera-numero fatta precedentemente,

Lettera	Numero	Lettera	Numero	Lettera	Numero
a	0	h	7	q	14
b	1	i	8	r	15
c	2	l	9	s	16
d	3	m	10	t	17
e	4	n	11	u	18
f	5	o	12	v	19
g	6	p	13	z	20

possiamo notare che l'alfabeto numerico dei messaggi unitari (le singole lettere) è rappresentato da $P = Z_{21}$.
Poiché nel cifrario di Cesare ogni lettera viene sostituita con la lettera che si trova un certo numero di posizioni più avanti abbiamo che l'insieme delle chiavi è $K=\{0,1,...,20\}$ (perché?)
Allora il sistema crittografico per traslazione può essere cosi schematizzato:
data la chiave $k \in K$, la funzione cifrante sarà la seguente:

$$C_k : Z_{21} \to Z_{21}$$
$$p \to p + k \mod 21,$$

mentre la funzione inversa, quella di decifratura, sarà:

$$D_k : Z_{21} \to Z_{21}$$
$$c \to c - k \mod 21.$$

Esercizi di cifratura e decifratura. Utilizzando le congruenze

1. Criptare un messaggio a scelta utilizzando k=7
2. Decriptare il seguente messaggio sapendo che k=16
   DE UHMNH GPFZMD IMDFD Z UMDOOHBMSADS Z DGOZMZNNSGOZ
   (il corso numeri primi e crittografia è interessante)

DECIFRATURA

Se intercettiamo un messaggio che sappiamo essere stato criptato col metodo della traslazione per decifrarlo dovremo scoprire quanto vale *k*, cioè la chiave. In realtà chi ha qualche esperienza di risoluzione di problemi crittografici tipo quelli della settimana enigmistica sa che è possibile decodificare un messaggio pur non conoscendo la chiave *k*.
Nel caso del cifrario di Cesare è possibile pensare di procedere per tentativi. Infatti i cifrari possibili sono 20 nel caso di un testo scritto in lingua italiana (25 se il testo è in inglese). Ma se il testo è molto lungo i tempi di decifratura diventerebbero lenti fornendo magari troppo tardi informazioni utili.
Messi di fronte ad un testo da decifrare si procede con quella che si chiama **analisi del testo**.
Poiché nella lingua italiana la maggior parte delle parole termina con una delle vocali *a,e,i,o* vorrà dire che le lettere finali delle parole del messaggio cifrato dovranno essere una di queste lettere. Purtroppo si fa presto ad eludere questa considerazione spezzando il messaggio in blocchi della stessa lunghezza il che rende complicata la ricostruzione delle singole parole. Però si può ricorrere ad altre considerazioni. Se nel testo ci sono lettere consecutive identiche queste necessariamente devono essere consonanti. Si hanno poi ulteriori informazioni, fornite dalla cosiddetta analisi delle frequenze. In ogni lingua ci sono lettere che compaiono nei testi con maggiore frequenza ed altre più raramente, ad esempio nella lingua italiana le lettere più frequenti sono nell'ordine *e,a,i* mentre le meno usate sono *q,z*. Riportiamo di seguito una tabella riassuntiva.

Lettera	%	Lettera	%	Lettera	%
a	11,74	h	1,54	q	0,51
b	0,92	i	11,28	r	6,38
c	4,50	l	6,51	s	4,98
d	3,73	m	2,52	t	5,63
e	11,79	n	6,88	u	3,02
f	0,95	o	9,83	v	2,10
g	1,65	p	3,05	z	0,49

Altre informazioni si possono reperire dalla frequenza delle doppie, dalla tendenza di certe lettere a non gradire la vicinanza di altre, ecc.

Supponiamo di intercettare il seguente messaggio:

OD LDZZD IUHZZRORVD ID N LDZZNQN FNHFMN

e di voler scoprire cosa significhi.

Riportiamo la frequenza delle lettere nel nostro messaggio:

Lettera	Occorrenze	Lettere	Occorrenze	Lettera	Occorrenze
A	0	H	2	Q	1
B	0	I	2	R	2
C	0	L	2	S	0
D	6	M	1	T	0
E	0	N	5	U	1
F	2	O	2	V	1
G	0	P	0	Z	6

.

La lettera D potrebbe essere una a, una e o una i. Lo stesso dicasi per le lettere N e Z. Tuttavia la Z è quasi sicuramente una consonante doppia perché non si possono avere due vocali consecutive uguali e dunque necessariamente la D, la N, la R e la H sono vocali, perché compaiono in …ZZD…, …ZZN e …HZZR. Molto probabilmente la O è una l visto che abbiamo stabilito che D è una vocale mentre la Z potrebbe essere una r, una n o una t per via della sua alta frequenza.
È evidente che abbiamo notevolmente ridotto il numero dei tentativi da fare per decodificare il messaggio.
Tentiamo con la sostituzione:
N=i, D=a, O=l, H=o, R=u, Z=t,
otteniamo:
la Latta IUottuluVa Ia i LattiQi FioFMi.
È chiaro che la scelta R=u non è proprio felice e dunque prendendo R=o e H=e arriviamo a
la Latta IUettoloVa Ia i LattiQi FieFMi
e con altri tentativi possiamo giungere alla soluzione:
la gatta frettolosa fa i gattini ciechi

È chiaro che il modo di procedere è molto falsato perché stiamo facendo l'analisi di un testo troppo breve, ma l'importante è aver sottolineato come le tante possibilità teoriche possano essere notevolmente ridotte usando informazioni sul linguaggio e procedendo sistematicamente per tentativi.