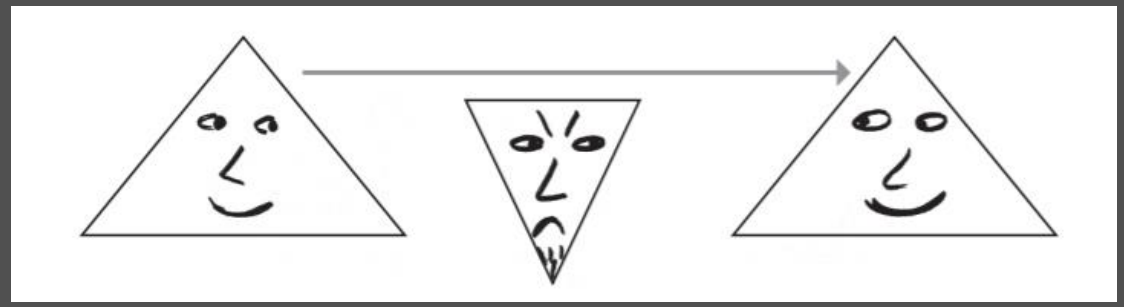
ELEMENTI DI CRITTOGRAFIA

Liceo classico e linguistico Tito Lucrezio Caro-Roma

Prof.sse Erminia Izzo, Francesca Ruzzi con la collaborazione di Claudia D'Armiento

Che cos'è la crittografia?

- E' la scienza che studia come rendere segreta e sicura la comunicazione tra due persone o entità nascondendo il significato del messaggi.
- **Crittografia** significa letteralmente «***scrittura segreta***».
- Con questo termine si intende oggi un insieme di tecniche che consentono di trasmettere messaggi mantenendoli segreti a tutti, tranne ad alcune persone che possiedano la chiave per comprenderli.



Proprietà della crittografia

- **Segretezza**

il messaggio non deve essere leggibile a terzi.

- **Autenticazione**

il destinatario deve poter essere sicuro del mittente.

- **Integrità**

il destinatario deve poter essere sicuro che il messaggio non sia stato modificato.

- **Attendibilità**

il mittente non deve poter negare di aver inviato il messaggio.

Un po' di definizioni….

- La **cifratura** è l'operazione con la quale si nascondono le informazioni; essa viene effettuata tramite un procedimento chiamato cifrario.
- **Il testo in chiaro** è il messaggio da cifrare.
- Il **testo cifrato** è il messaggio trasformato in modo da non essere più leggibile tramite una semplice lettura.
- La **decifrazione** è la riconversione di un testo cifrato nella sua forma originaria, cioè nel testo in chiaro.
- Il **cifrario** è il procedimento (algoritmo) che consente di crittare e decrittare i testi.

- Il processo di cifratura deve essere biunivoco, in modo permettere un processo inverso che ritrovi il messaggio originale.
- Chi riceve il messaggio deve essere in grado di interpretarlo e cioè di decifrarlo.
- Il mittente ed il destinatario si devono essere messi d'accordo prima su come "cifrare" e "decifrare" e scegliere un metodo efficace in modo che per gli altri sia sostanzialmente impossibile cifrare e decifrare un messaggio.
- La crittografia fornisce metodi effettivi per effettuare cifratura e decifratura dei messaggi.
- Il processo di trasformazione dal messaggio in chiaro al messaggio cifrato e viceversa è spesso noto, ma si basa su una informazione specifica (detta "**chiave**"), senza la quale non si è in grado di operare.
- I metodi di cifratura si sono estremamente evoluti nell'arco della storia.

Utilizzo tradizionale della crittografia

- Gli usi tradizionali riguardavano quasi esclusivamente gli ambiti militari e di spionaggio/controspionaggio

- Sono riportati numerosissimi esempi di uso di sistemi crittografici nel corso di guerre, battaglie, rivoluzioni, cospirazioni, complotti, ...

Utilizzi moderni della crittografia

- L'uso più importante della crittografia in ambito "civile" è quella della sicurezza delle comunicazioni in rete

- Più in particolare le applicazioni di commercio elettronico sono quelle in cui maggiormente è sentita la necessità della sicurezza e della segretezza (scambio di dati sensibili, quali il numero di carta di credito, numero di conti bancari, ecc.)
- Un altro utilizzo importante è quello della firma digitale e dell'autenticazione dei documenti, che ha applicazioni nella pubblica amministrazione (e-government) e in generale negli aspetti burocratici (contratti, domande, moduli, vari documenti ufficiali, ecc.)

Che cos'è la crittoanalisi?

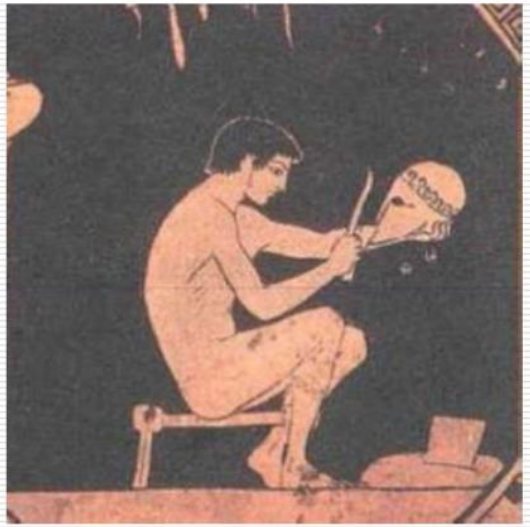
- La **crittoanalisi** è l'arte della "rottura" dei codici e dei cifrari.
- La crittoanalisi studia come decifrare un messaggio senza esserne "autorizzati".
- La **decrittazione** è la riconversione di un testo cifrato nella sua forma originaria, cioè nel testo in chiaro , senza essere in possesso della chiave.
- La crittoanalisi ha il ruolo fondamentale di far capire quanto un sistema di cifratura/decifratura sia sicuro.

scritture segrete

- CRITTOGRAFIA (altera il messaggio)
  - TRASPOSIZIONE (permuta le lettere del messaggio secondo una qualche regola reversibile) Es. scitala e griglie quadrate
  - SOSTITUZIONE
    - CODICE (sostituisce le parole)
    - CIFRATURA (sostituisce le lettere)
      - Monoalfabetica (ogni elemento del testo chiaro viene *sostituito* sempre dallo stesso elemento secondo una tabella di sostituzione unica) Es Cifrario di Cesare
      - Polialfabetica (un elemento del testo chiaro viene sostituito da un elemento che non è sempre lo stesso, ma da determinarsi per mezzo di una chiave segreta da concordarsi) Es. Cifrario di Vigenere
- STEGANOGRAFIA (occulta il messaggio)

Esempi di steganografia

La scritta veniva applicata sul capo rasato di uno schiavo, quindi si attendeva che i capelli ricrescessero e si inviava il messaggero. All'arrivo presso il destinatario, questi rasava nuovamente lo schiavo e leggeva il messaggio.



Erodoto (libro V delle Storie)

INCHIOSTRI SIMPATICI:
da Plinio il Vecchio (I dC) a Umberto Eco (Il Nome della Rosa), si narra di una metodologia di scrittura a base di limone o latice di titimabo, che appaiono invisibili, ma ricompaiono una volta che il testo venga esposto a una fonte di calore.

ITALIA:
nel XVI secolo un modo interessante di recapitare i messaggi era di scriverli con aceto sul guscio di un uovo sodo; il guscio, poroso, permetteva all'aceto di passare e, una volta recapitato il messaggio, era sufficiente sgusciare l'uovo per leggerlo!

La crittografia nella storia
(dall'antichità al 1975)

➢Metodi antichi

- La scitala spartana
- La scacchiera di Polibio
- Il codice atbash
- Il codice di Cesare

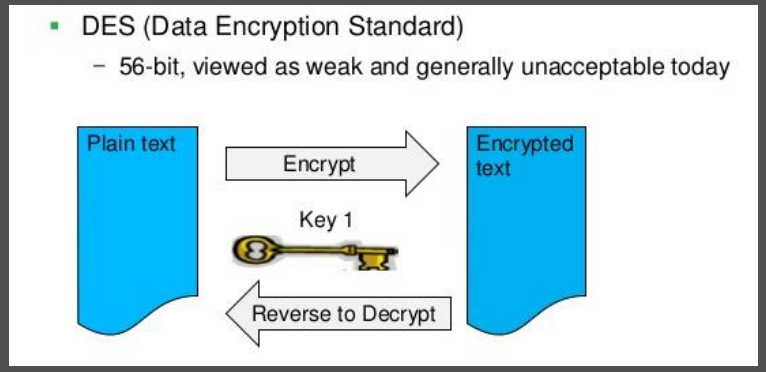
Il disco cifrante di Leon Battista Alberti

➢Rinascimento

- Blaise Vigenère

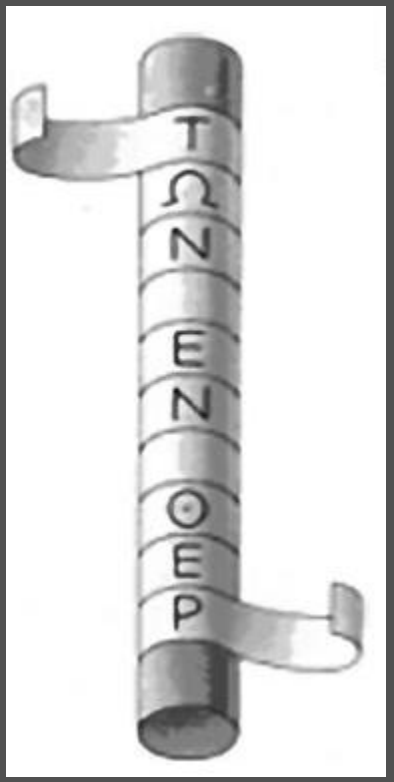
➢XX secolo

- La macchina Enigma (usata dai tedeschi durante la seconda guerra mondiale)
- Il DES (Data Encryption Standard)



La scitala spartana
(400 anni a.c.)

- La scitala (skytale) era un dispositivo con cui gli spartani cifravano e decifravano messaggi segreti (Plutarco, Vite parallele)
- Il messaggio veniva inciso su un nastro di cuoio o una striscia di papiro arrotolato attorno a un bastone (la scitala), in modo tale che ogni giro della striscia riportasse un solo carattere: srotolata la striscia, le lettere risultavano mescolate e il messaggio non più leggibile.
- Solo l'utilizzo di un bastone identico a quello del mittente (la chiave) consentiva al destinatario di ricostruire il testo in chiaro.



La scacchiera di Polibio
(150 anni a.c.)

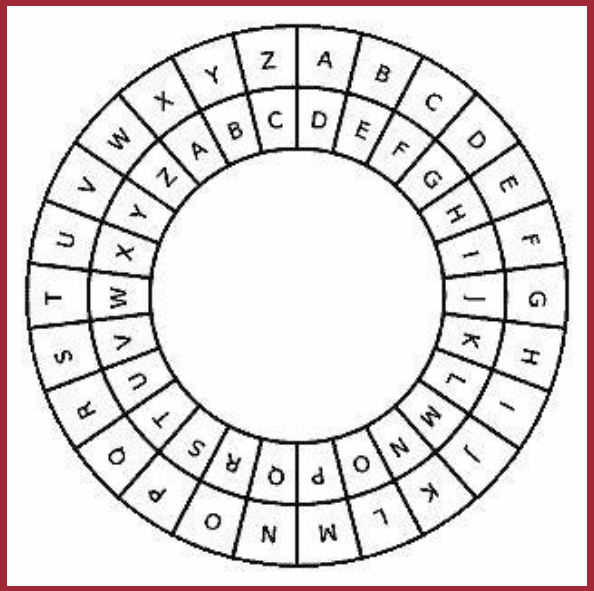
Nel libro X delle Storie, (circa 200-118 a. C.) Polibio attribuisce ai suoi contemporanei Cleoxeno e Democleito l'introduzione di un sistema di telecomunicazione (telegrafo ottico) trasmesso con due gruppi di 5 torce .
La scacchiera di Polibio è un cifrario per sostituzione che associa ad ogni lettera dell'alfabeto in chiaro una coppia ordinata di numeri.

	1	2	3	4	5
1	a	b	c	d	e
2	f	g	h	i, j	k
3	l	m	n	o	p
4	q	r	s	t	u
5	v	w	x	y	z

Esempio B è 12, P è 35, V è 51

- La scacchiera originale è costituita da una griglia composta da 25 caselle ordinate in 5 righe ed altrettante colonne
- Le lettere dell'alfabeto vengono inserite da sinistra a destra e dall'alto in basso
- Le righe e le colonne sono numerate: tali numeri sono gli indici o le «coordinate» delle lettere costituenti il messaggio in chiaro

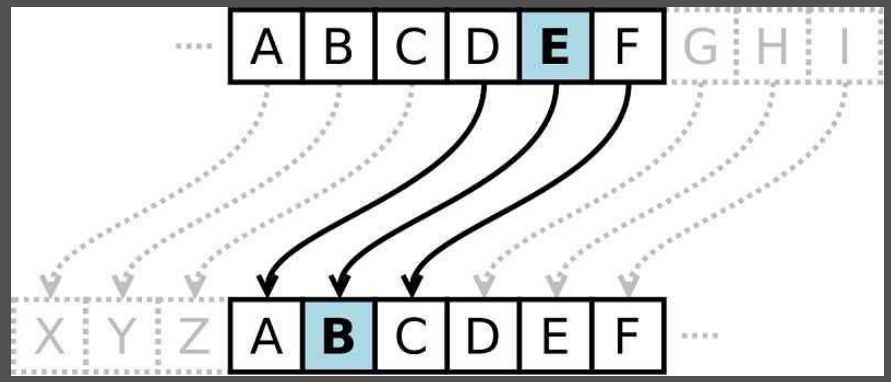
Il cifrario di Giulio Cesare



Svetonio, nella Vita dei dodici Cesari, racconta che Giulio Cesare utilizzava un sistema di cifrazione molto semplice: ogni lettera chiara viene sostituita con quella che si trova tre posti dopo nell'alfabeto.

esempio: la lettera A è sostituita dalla D, la B dalla E e così via..

La decifrazione è altrettanto semplice, basta sostituire ogni lettera quella che si trova tre posti prima.



Più in generale si dice codice di Cesare un codice nel quale la lettera del messaggio chiaro viene spostata di un numero fisso k di posti, con k (**chiave**) compreso tra 1 e 26 se ci si riferisce all'alfabeto internazionale.

In generale il crittosistema di Cesare è un cifrario in cui la stessa lettera è codificata sempre con la stessa lettera (Sostituzione monoalfabetica)

Sostituzione monoalfabetica

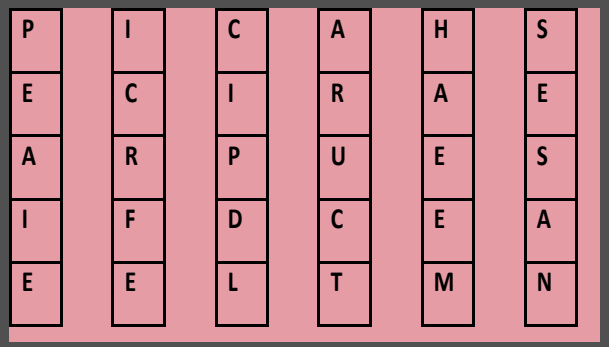
- Il caso più generale è quello in cui l'alfabeto cifrato è una permutazione di quello in chiaro
- La chiave della cifra è la lista delle cifre in ordine alfabetico; se si cifra con lettere dello stesso alfabeto sarà un alfabeto disordinato impossibile da ricordare a memoria
- Un modo semplice per ricordare la permutazione è quello di usare una frase (chiave) le cui lettere fanno da inizio della lista (escludendo quelle ripetute), lasciando poi le altre lettere a seguire

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

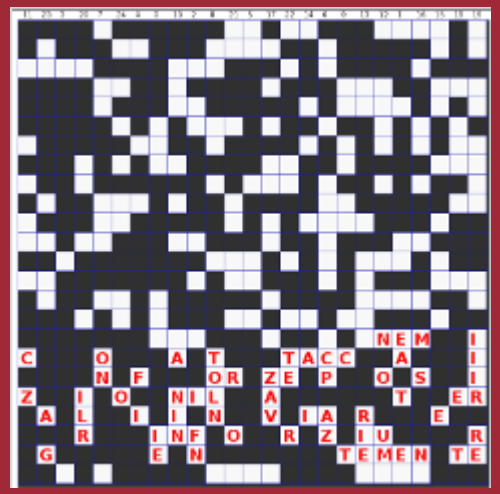
in questo esempio la frase è **NEL MEZZO DEL CAMMIN DI NOSTRA VITA**

Le cifrature per trasposizione

- I sistemi di trasposizione letterale consistono nel *rimescolare* i caratteri del testo chiaro secondo una qualche regola reversibile.
- Le trasposizioni più semplici consistono nell'invertire il testo intero o le singole parole, soluzioni ovviamente debolissime dal punto di vista della segretezza.
- Sistemi più impegnativi richiedono di disporre il testo su una matrice rettangolare, righe e colonne; la trasposizione più semplice consiste allora nel disporre il testo su n righe e quindi scambiare righe con colonne.
- Un po' più sofisticata l'idea di rimescolare le colonne in base a una parola chiave segreta (trasposizione con chiave) o usando un qualche dispositivo come le griglie.



Le cifrature per trasposizione



- Abbiamo già conosciuto il più antico esempio di cifra per trasposizione, la *scitala* . Sistemi a trasposizione sono stati usati ancora nel XIX e XX secolo nella I e nella II guerra mondiale.
- L'uso dei sistemi a trasposizione declinò bruscamente con la nascita delle prime macchine cifranti, quasi tutte basate su sistemi di sostituzione letterale. Alla fine del XX secolo con l'avvento dei computer la trasposizione ha conosciuto una seconda giovinezza, a livello informatico è infatti facile realizzare sistemi di trasposizione anche molto complessi; oggi la trasposizione è usata soprattutto in combinazione (sovracifratura) con la sostituzione come nei cifrari DES e AES, dove però la trasposizione è a livello di blocchi di bit non più di lettere alfabetiche.

griglia Rasterschlüssel 44, usata dai Tedeschi alla fine della II guerra mondiale.

Le cifrature per trasposizione

Nei sistemi di trasposizione letterale il testo cifrato è un anagramma del testo originale; matematicamente si parla di una permutazione, e il numero di permutazioni possibili è dato dal fattoriale: $P_n = n! = n \times (n-1) \times (n-2) \ldots 1$

In realtà questa è una stima per eccesso, valida solo se le n lettere sono tutte distinte, mentre è più probabile, quanto più lungo è il testo, che vi siano lettere ripetute; in tal caso le permutazioni possibili sono molte di meno; in generale vale la formula, per un testo di n lettere, con m lettere ripetute, la prima n1 volte, la seconda n2 volte, ecc.: $P_n = n!/(n1! \times n2! \ldots)$

Le cifrature per trasposizione: la sicurezza

Dalla formula precedente si ricava che i sistemi a trasposizione semplice per testi brevi o griglie piccole sono molto deboli e relativamente facili da decrittare essendo piccolo il numero di permutazioni; ma a differenza di quanto avviene per i sistemi a sostituzione, quelli a trasposizione sono tanto più sicuri quanto più lungo è il testo; i sistemi a doppia trasposizione poi sono quasi inattaccabili; hanno però il difetto di essere piuttosto complicati e delicati da usare a mano. Ovviamente oggi con il computer sono praticabili trasposizioni multiple anche molto complesse.

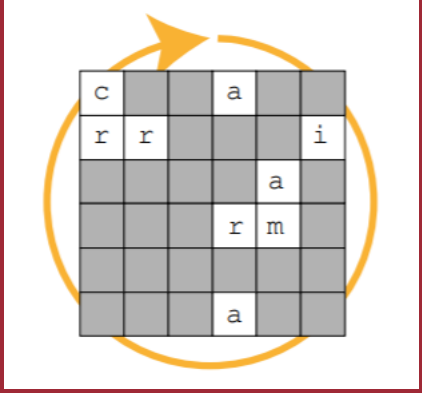
Una soluzione molto usata è quella di alternare sostituzioni e trasposizioni in modo da aumentare la sicurezza; cifrari misti di questo tipo sono anche i già citati DES e AES.

Le griglie quadrate

Le griglie sono uno dei sistemi per trasposizione più antichi e semplici; consistono di un cartoncino o lamina metallica o di legno, con una serie di fori. Il messaggio chiaro viene scritto nei fori secondo una qualche regola e il cifrato si ottiene leggendo i caratteri secondo una diversa regola, oppure muovendo la griglia secondo un ordine stabilito. Per decifrare si segue il procedimento inverso.

Storicamente le griglie quadrate furono popolari nel XIX secolo, una griglia 6x6 viene usata da Jules Verne nel suo *Mathias Sandorf*; nel 1881 furono usate dai Tedeschi anche nella Grande Guerra.

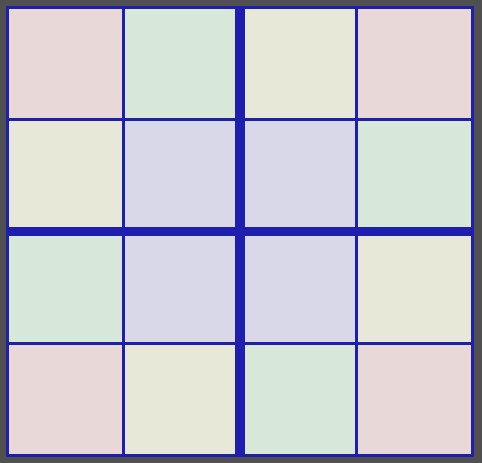
Le griglie quadrate di rotazione



- I fori devono essere la metà (griglia a due rotazioni) oppure un quarto (griglia a quattro rotazioni) del numero totale di caselle disponibili e devono essere disposti in modo da coprire tutte le caselle del quadrato una e una sola volta in quattro successive rotazioni.
- Per cifrare si dispone la griglia nella posizione iniziale sopra un foglio di carta, e si scrive il messaggio nei fori disponibili fino ad esaurimento, quindi si ruota la griglia di 180° (due rotazioni) o 90° (quattro rotazioni) e si continua fino ad esaurimento del messaggio; se il messaggio è più breve del numero totale di caselle si possono riempire le caselle avanzate con nulle; se è più lungo si può ricominciare dalla posizione iniziale, oppure rovesciare la griglia ...
- Più sicure e più usate sono le griglie a quattro rotazioni
- La griglia deve essere vista come formata da quattro sottogriglie di ordine dimezzato, destinate a sovrapporsi l'un l'altra secondo 4 rotazioni di un angolo retto; queste sottogriglie devono essere formate da un numero divisibile per 4, se si vogliono distribuire uniformemente i fori; altrimenti qualsiasi quadrato di ordine pari è accettabile.

Le griglie quadrate di rotazione 4 × 4

vedi il disegno a lato dove le posizioni sovrapponibili sono nello stesso colore; nella prima sottogriglia possiamo mettere il foro in una delle 4 caselle disponibili; nella seconda sottogriglia possiamo mettere il foro in una delle 3 caselle; dobbiamo infatti escludere la casella in posizione sovrapponibile al primo foro; analogamente nella terza sottogriglia abbiamo solo 2 caselle libere per il foro e nella quarta ne resta solo una.

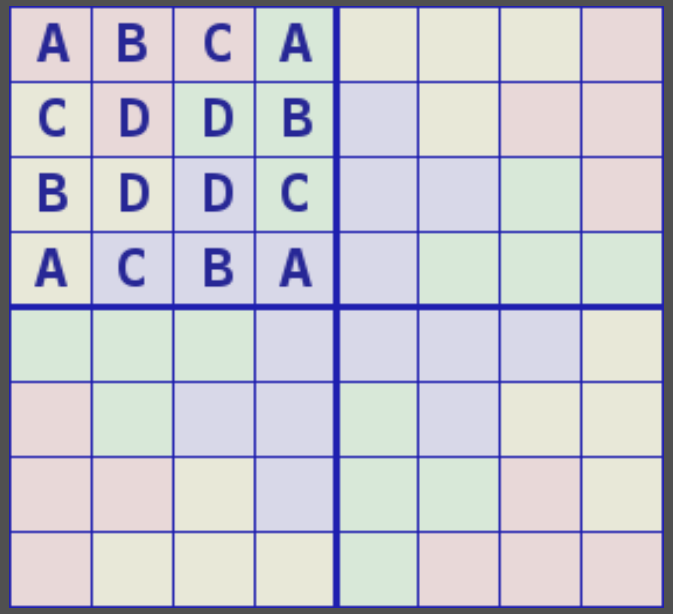


In conclusione le griglie possibili sono 4×3×2×1=4!=24.

24 è un numero crittograficamente piccolo, anzi minuscolo, e la sicurezza di una griglia 4×4 è pressoché nulla.
Ma una griglia può essere ruotata in quattro modi diversi, cambia solo la posizione iniziale e quindi le griglie effettivamente diverse sono solo sei; di queste due sono simmetriche, le altre quattro sono una la simmetrica (il ribaltamento) di un'altra;
in totale basterebbero quindi quattro griglie materiali per avere a disposizione tutte le griglie 4×4 possibili.

Le griglie quadrate di rotazione 8×8

vedi il disegno a lato dove le posizioni sovrapponibili sono nello stesso colore e con la stessa lettera; anche qui dividiamo la griglia in 4 sottogriglie 4×4. La prima sottogriglia può avere il primo foro in una delle 4 posizioni sovrapponibili di tipo A; il secondo foro in una delle 4 di tipo B; il terzo foro in una delle 4 di tipo C e il quarto foro in una della 4 di tipo D. In totale 4^4.



La seconda sottogriglia può analogmente avere i fori in 3 posizioni per ogni tipo quindi in totale 3^4, la terza sottogriglia di 2^4 e la quarta di una sola casella libera per ogni tipo. Riordinando i fattori il numero di disposizioni dei fori è $(4!)^4=24^4=331776$ numero molto maggiore del precedente, ma ancora alla portata di attacchi esaustivi condotti con un computer.

Ammettendo le ripetizioni si ha semplicemente una potenza $4^{16}=2^{32}=42949672964$

La sostituzione monoalfabetica: la sicurezza

Esistono 21!=51090942171709440000 permutazioni possibili, cioè circa $51 \cdot 10^{18}$, ossia più di cinquanta miliardi di miliardi

Una ricerca esaustiva per trovare la permutazione giusta è praticamente impossibile , eppure questo codice è tutt'altro che sicuro...

Usare simboli diversi al posto delle lettere non aumenta la difficoltà di decifratura

Si confronti ad esempio il codice MRSNTA NHHNLLCZFZSR con il codice 11-16-17-12-18-01-00-12-08-08-12-10-10-0321-06-21-17-16
La debolezza del codice sta nelle ripetizioni di simboli

Elementi di crittonalisi: l'analisi delle frequenze

Abbiamo visto che la crittografia per sostituzione monoalfabetica ha delle fragilità tali da mettere in dubbio la sua applicabilità: dall'analisi di doppie, coppie, triple, di lettere amiche e nemiche e in generale di rapporti fra lettere vicine, ci si può difendere con una trasposizione in sovracifratura, cioè fatta in aggiunta alla sostituzione. Resta però la vulnerabilità rispetto all'analisi delle frequenze, dalla quale nessuna trasposizione mette al riparo.
una semplice crittanalisi statistica basata sulle caratteristiche statistiche delle lingue è in genere sufficiente per forzare il testo.

Nella crittoanalisi, **l'analisi delle frequenze** è lo studio della frequenza di utilizzo delle lettere o gruppi di lettere in un testo cifrato. Questo metodo è utilizzato per violare i cifrari classici.

l'analisi delle frequenze

In particolare un primo metodo che si adotta in attività di crittanalisi si basa sul fatto che **in ogni lingua** la frequenza di uso di ogni lettera è piuttosto determinata; questo è vero in modo rigoroso solo per testi lunghi, ma spesso testi anche corti hanno frequenze non molto diverse da quelle previste.
Vediamo come riferimento le frequenze percentuali delle lettere più comuni di due lingue: **italiano ed inglese**

Italiano		Inglese	
E	11,79	E	12,31
A	11,74	T	9,59
I	11,28	A	8,05
O	9,83	O	7,94
N	6,88	N	7,19

Possiamo notare quanto le prime lettere di queste lingue siano presenti in quantità molto maggiore delle altre

Ad esempio sapendo che il testo è in italiano, è facile che l'ultima lettera di ciascuna parola sia una vocale.
• Si cercano i simboli più frequenti nel testo cifrato • Si provano a sostituire con le lettere più frequenti in italiano
• Si cerca di vedere se si riesce a "intravedere" delle parti di parole
• Qualche tentativo può portare a parole improbabili, in tal caso si deve rivedere alcune scelte

Esempio: TRT QRDIAR NTMNFZ N GLPRIN

- Primo tentativo*:* inizio dalle lettere terminali di una parola e associando loro le vocali, in ordine di frequenza: **R = e, N = a , T = i, Z = o,**
- si ottiene **iei QeDIAe aiMNFZ a GLPeIa**
- Rivediamo alcune scelte (la prima parola non ha senso) **T = n**, (è la seconda consonante per frequenza: la prima è L che non sembra adatta), **R = o, Z = e,**
- Otteniamo **non QoDIAo anMaFe a GLPoIa**
- Ora introduciamo le consonanti più frequenti ancora mancanti (**l**, **r**,**t**,**s**,**c**) e reintroduciamo la **i**. Proviamo con **I = l**, **A = i**, **D = t**, **F = r, G = s** , **L=c:**

  non Qotlio anMare a scPola

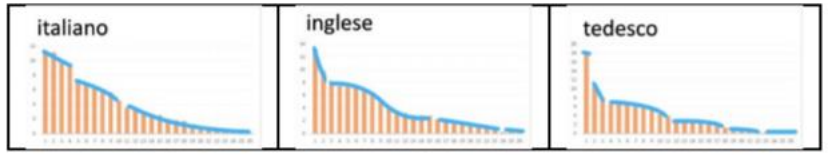
  Posso modificare ancora **D = g** e continuare...

TRT QRDIAR NTMNFZ N GLPRIN

·Lettera	·Occor-renze	·Lettera	·Occor-renze	·Lettera	·Occor-renze
·A	·1	·H	·0	·Q	·1
·B	·0	·I	·2	·R	·4
·C	·0	·L	·1	·S	·0
·D	·1	·M	·1	·T	·3
·E	·0	·N	·4	·U	·0
·F	·1	·O	·0	·V	·0
·G	·1	·P	·1	·Z	·1

Le caratteristiche statistiche delle lingue

Le impronte linguistiche



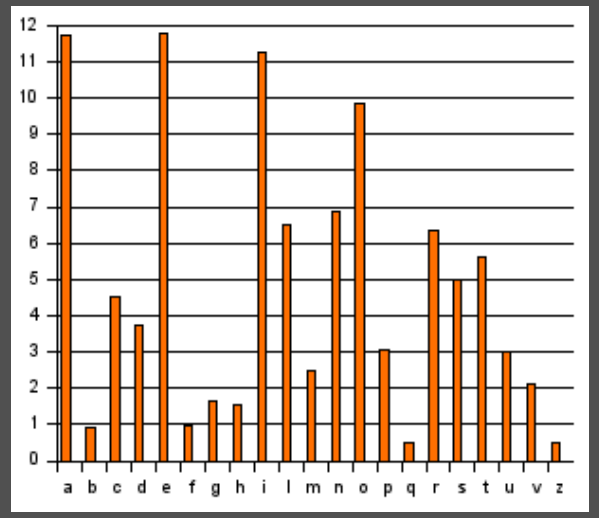
- Ogni lingua si differenzia dalle altre per caratteristiche statistiche quali la frequenza delle singole lettere, dei bigrammi e dei digrammi; nell'insieme queste caratteristiche costituiscono una vera e propria ***impronta digitale*** della lingua.
- Per esempio, l'italiano è una lingua molto vocalizzata, l'inglese, al contrario, presenta un'alta frequenza di consonanti. Il francese è anch'esso molto vocalizzato anche se é ricco di lettere come la ***J***, che in italiano sono poco frequenti. La lettera ***E*** è la più frequente in quasi tutte le lingue europee, in particolare in francese e in tedesco ha una prevalenza nettissima.
- Questa impronta digitale è di grande importanza per la crittanalisi statistica.

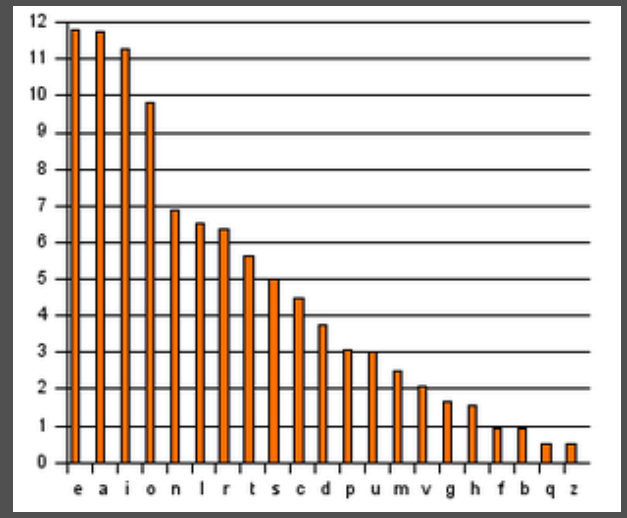
Le frequenze della Lingua Italiana

- È detto da tutti che l'italiano é una lingua musicale; infatti analizzando un testo e calcolando la frequenza delle varie lettere , si può notare che le vocali E , A , O sono le lettere più frequenti, senza che esista una preponderanza netta di una di queste sulle altre, e la loro percentuale, che in media raggiunge il 46,4%, é tra le più alte tra le varie lingue esaminate
- Seguono le consonanti L , N , R , S , T .
- Tra i bigrammi più frequenti troviamo Q-U sempre seguiti da vocale. La lettera H é spesso preceduta dalla lettera C o G per formare i trigrammi CHE , CHI , GHE , GHI . Nella lingua italiana sono quasi assenti le lettere J , K , Y , X ,W salvo nei nomi di persone o località straniere.

Frequenza delle lettere in italiano

·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





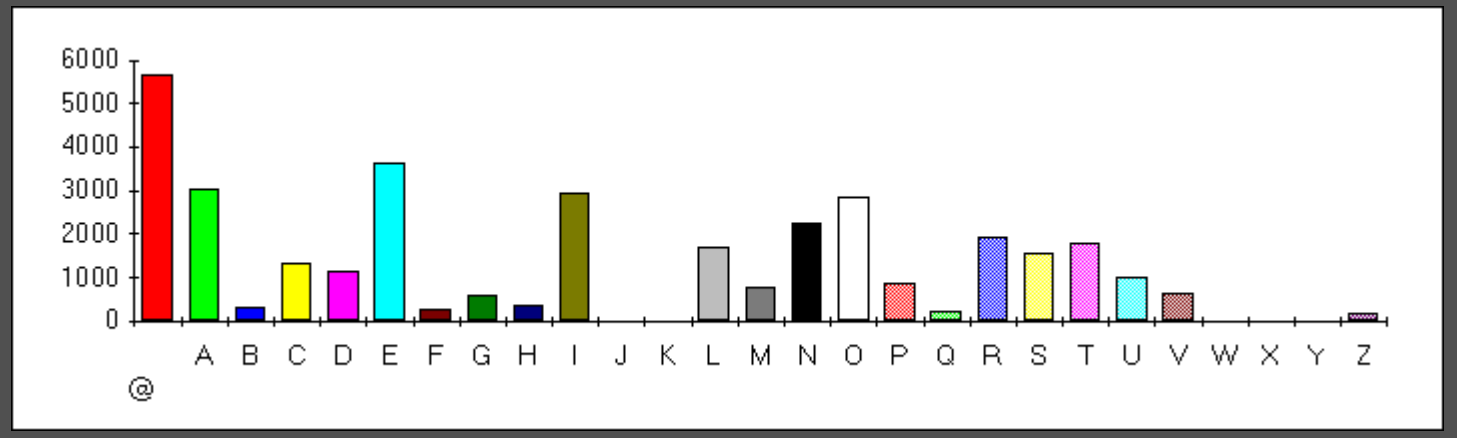
Le frequenze della Lingua Italiana

I seguenti grafici mostrano la frequenza delle varie lettere del primo capitolo del celebre romanzo di Alessandro Manzoni, "I Promessi Sposi":

Bigrammi e doppie ne *I PROMESSI SPOSI*

Bigramma	Percentuale
ER	1,91%
ON	1,68%
EN	1,54%
AN	1,52%
RE	1,44%
TO	1,38%
ES	1,36%
DI	1,31%
AR	1,37%

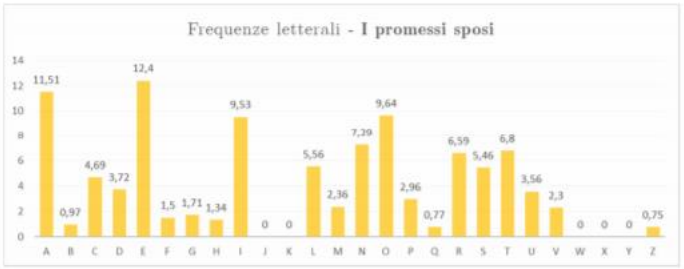
Doppia	Percentuale
LL	0.86%
TT	0.73%
SS	0.64%
CC	0.38%
AA	0.3%
EE	0.3%
BB	0.19%
RR	0.18%
NN	0.18%



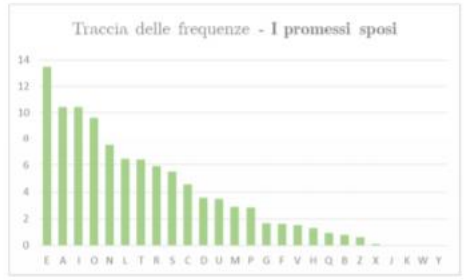
Frequenze letterali ne *I PROMESSI SPOSI*

A	11,51	J	0	S	5,46
B	0,97	K	0	T	6,8
C	4,69	L	5,56	U	3,56
D	3,72	M	2,36	V	2,3
E	12,4	N	7,29	W	0
F	1,5	O	9,64	X	0
G	1,71	P	2,96	Y	0
H	1,34	Q	0,77	Z	0,75
I	9,53	R	6,59		

Le statistiche riportate sopra offrono un ottimo spaccato, dal momento che il romanzo di Manzoni è composto da più di un milione di lettere. L'analisi diventa più veloce se si rappresentano i dati in forma di istogramma:



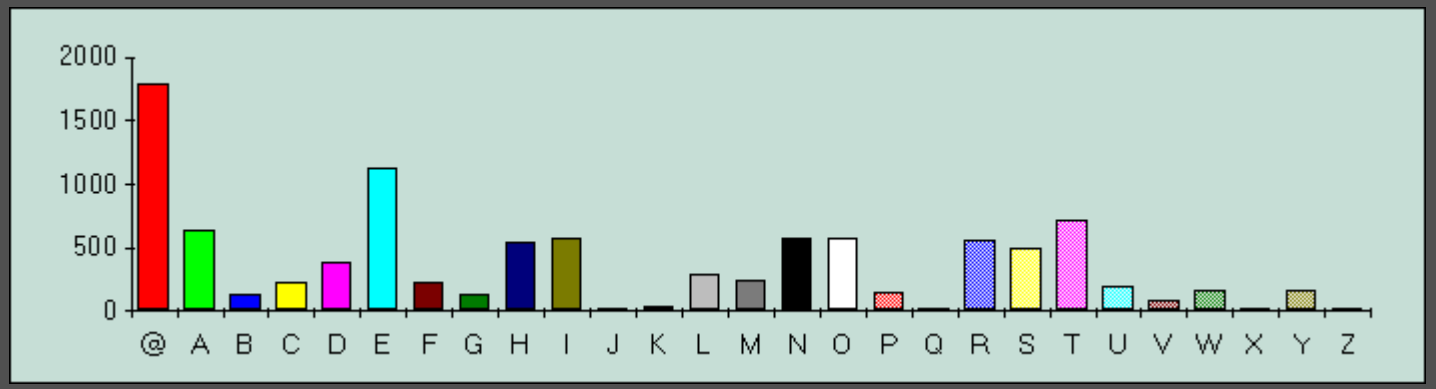
Un altro grafico utile per la decrittazione è la cosiddetta "Traccia della frequenze", che riporta gli stessi dati di prima ma in ordine decrescente.



Le frequenze della Lingua Inglese

Anche per questa lingua, la lettera E è la più frequente, ma essa ha una frequenza di poco superiore a quella della lettera T, che immediatamente la segue. Seguono poi le vocali (in ordine) O,A,I, e le consonanti N,R,S,H, che non hanno frequenze molto diverse l'una dall'altra. Le lettere più frequenti formano la successione mnemonica: ETOANIRSH.

I seguenti bigrammi sono tra i più frequenti:
TH, HE, AN, ER, ON, RE, IN, ED, ND, AT, OF, OR, HA, EN, NT, EA, TO, TI, ST, IT, ecc.,
con frequenze percentuali che variano tra 35 e 10 per mille lettere di un testo normale.
Le lettere J, V, Z sono sempre seguite da vocali.



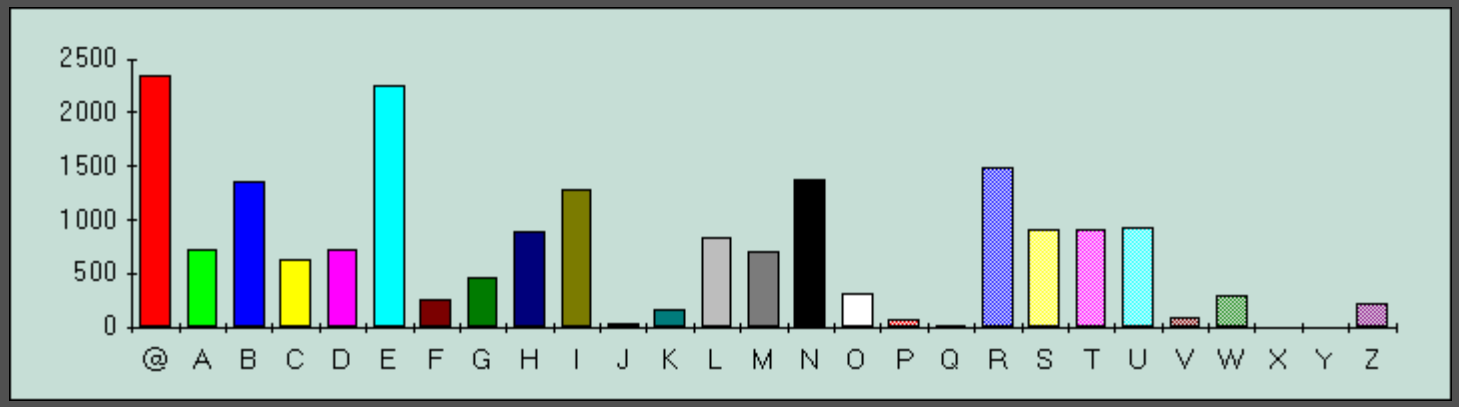
Questo grafico illustra la frequenza delle lettere presenti nel primo capitolo del "Frankenstein" di Mary Shelley

L' inglese ha alcune caratteristiche specifiche: il bigramma TH (3.4%) ed i raddoppiamenti della E e della O (circa 0.2%); altre doppie sono S, L, N, D e più rare M, F, G, P: a queste si possono aggiungere quelle che provengono dalle coppie finale-iniziale di parole consecutive, come EE, MM, TT, RR, WW.

Le frequenze della Lingua Tedesca

Nella lingua tedesca, come nella lingua francese, spicca la lettera E, che ha una forte preponderanza sulle altre vocali e sulla lettera che immediatamente la segue, che è la N. Le lettere più frequenti si possono ricordare con la loro successione mnemonica: ENRITSAUD.

I bigrammi più frequenti sono in ordine decrescente di frequenza: EN, ER, CH, ND, DE, TE, IE, EI, UN, GE, DI, ES, IN, HE, BE, IT.
Una caratteristica del tedesco è di avere numerose parole che terminano per consonante (81%).



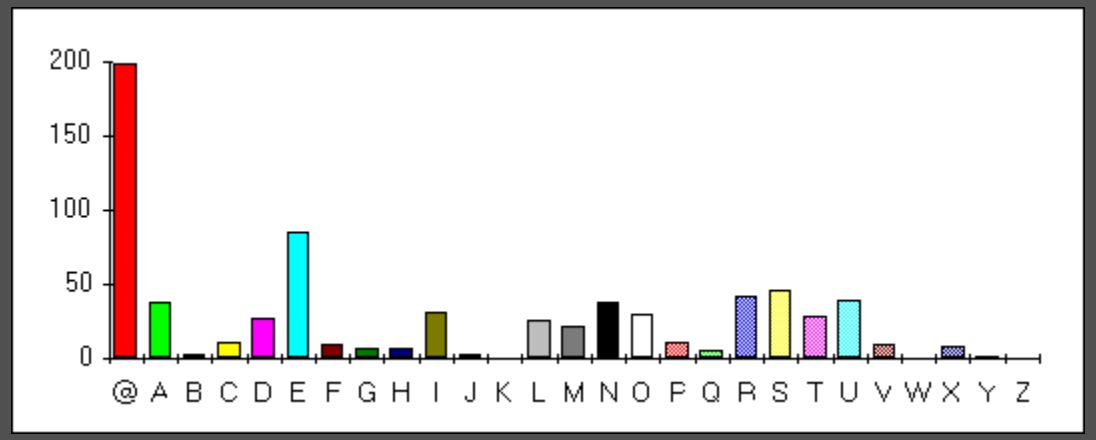
Ecco il grafico che visualizza le frequenze delle lettere presenti nel primo atto ('Nacht') del "Faust" di Goethe

Le vocali, in media, hanno una frequenza percentuale complessiva del 38%. Il gruppo D, N, R, S, T del 35.5%; il gruppo F, K, V, P, J del 5%.
Le consonanti che si raddoppiano più spesso sono L, S, N.

Le frequenze della Lingua Francese

Anche la lingua francese, come quella italiana e tutte quelle che derivano dal latino, é molto vocalizzata, a differenza dell'inglese e del tedesco.
Si può notare che la vocale E é la lettera in assoluto più frequente, con una percentuale molto superiore in confronto alle altre. Segue di norma la lettera N e le altre quattro vocali A , I , O , U .

La percentuale delle vocali è di circa il 44,5%.
Anche le lettere R , L , S , T sono piuttosto frequenti, e la loro percentuale è del 34%.



Ecco il grafico delle frequenze delle varie lettere della poesia «Parfum Exotique» del poeta francese Baudelaire

Una caratteristica della lingua francese é che la lettera E é spesso raddoppiata. Per quanto riguarda i bigrammi, sono frequenti QU , EZ e UX.

Sostituzione polialfabetica

- Un crittosistema per sostituzione polialfabetica non usa sempre lo stesso simbolo per la stessa lettera
- Per ogni posizione del messaggio in chiaro viene usato un alfabeto diverso
- Così si supera (parzialmente…) il problema della debolezza visto per i crittosistemi monoalfabetici
- Infatti ha "meno" senso contare le frequenze di simboli, non essendoci più corrispondenza tra lettere in chiaro e simboli cifrati

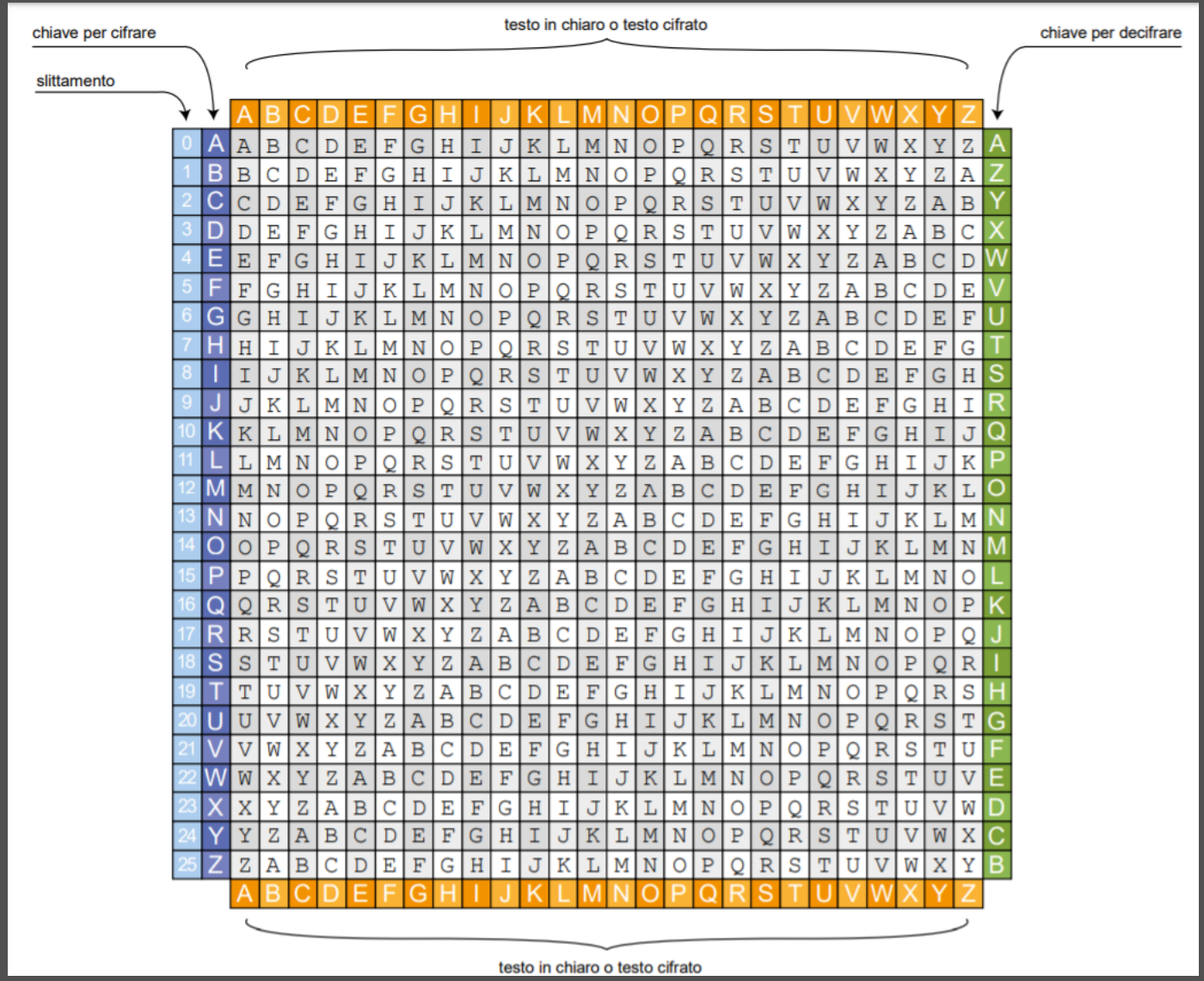
Il cifrario di Vigenère

Blaise de Vigenère pubblicò nel 1586 un trattato di cifre nel quale ne proponeva tra le altre una che ebbe grande fortuna e che è ricordata con il suo nome. Si tratta della più semplice cifra di sostituzione polialfabetica.

Il metodo si può considerare una generalizzazione del cifrario di Cesare; invece di spostare sempre dello stesso numero di posti la lettera da cifrare, questa viene spostata di un numero di posti variabile, determinato in base ad una parola chiave, da concordarsi tra mittente e destinatario, e da scriversi sotto il messaggio, carattere per carattere; la parola è detta verme, per il motivo che, essendo in genere molto più corta del messaggio, deve essere ripetuta molte volte sotto questo, come nel seguente esempio:

- Testo chiaro - ARRIVANOIRINFORZI
- Verme - VERMEVERMEVERMEVE
- Testo cifrato - VVIUZVRFUVDRWAVUM

Il cifrario di Vigenère è un crittosistema simmetrico, che usa la stessa chiave sia per la cifratura sia per la decifratura.



Il testo cifrato si ottiene spostando la lettera chiara di un numero fisso di caratteri, pari al numero ordinale della lettera corrispondente del verme. Di fatto si esegue una somma aritmetica tra l'ordinale del chiaro (A = 0, B = 1, C = 2 ...) e quello del verme; se si supera l'ultima lettera, la Z, si ricomincia dalla A. Matematicamente il Vigenère si riduce a un'addizione modulo 26 (vedi aritmetiche finite).

Per semplificare questa operazione il Vigenère propose l'uso della seguente tavola quadrata, Volendo ad esempio cifrare la prima R di ARRIVANO si individuerà la colonna della R, quindi si scenderà lungo la colonna fino alla riga corrispondente della corrispondente lettera del verme (qui E); la lettera trovata all'incrocio è la lettera cifrata (qui V); la seconda R invece sarà cifrata con la lettera trovata sulla riga della R di VERME, e cioé con la I.

Il vantaggio rispetto alle cifre mono-alfabetiche è evidente: la stessa lettera del testo chiaro non è sempre cifrata con la stessa lettera; e questo rende più difficile l'analisi statistica del testo cifrato e la decrittazione.

Chi riceve il messaggio per decifrarlo deve semplicemente usare il metodo inverso (sottrarre invece che sommare); riferendosi all'esempio di sopra si avrà:

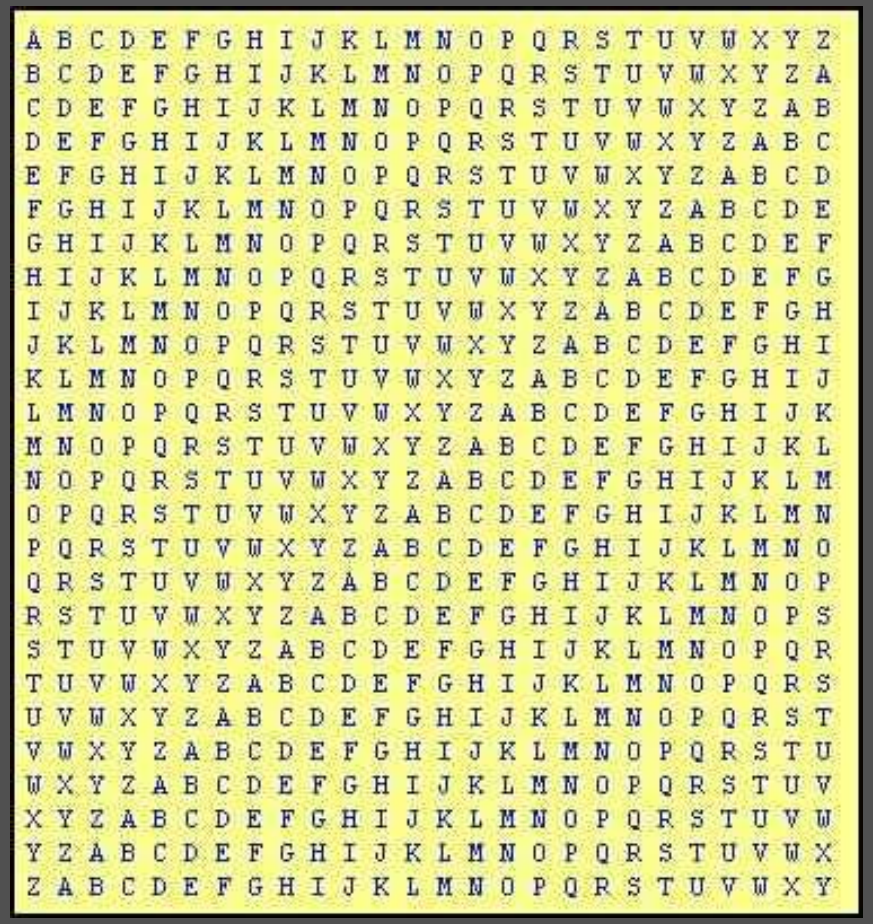
Testo cifrato - VVIUZVRFUVDRWAVUM Verme - VERMEVERMEVERMEVE Testo chiaro - ARRIVANOIRINFORZI

si potrà decifrare la seconda V ricercandola nella riga della corrispondente lettera del verme, la E; la colonna dove si trova la V ha al primo posto in alto la lettera chiara, la R.

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
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

Per utilizzare la tavola e cifrare un primo messaggio è necessario dunque scegliere una chiave, ad esempio HTML. A questo punto la tavola appena vista si riduce di qualche riga, cinque per essere esatti, costituite dalla prima e dalle quattro righe che iniziano con le lettere della chiave. La tavola che segue sarà quindi sufficiente per eseguire la cifratura:

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
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
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
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
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



Il cifrario di Vigenère: la sicurezza

Come si è detto questo cifrario ha goduto per tre secoli la fama di essere un cifrario inattaccabile; nel 1863 il colonnello prussiano Friedrich Kasiski pubblicò un primo metodo di decrittazione. In realtà Anche se questa cifratura è molto più sicura di quella di Cesare, è anch'essa facilmente crackabile.

La debolezza del Vigenère sta nell'essere, di fatto, un insieme di n cifrari di Cesare, dove n è la lunghezza della chiave; se il crittoanalista riesce a determinare la lunghezza della chiave (nel nostro caso, n) la decrittazione diventa molto semplice.

Il cifrario di Vigenère: la sicurezza

Per far ciò si possono utilizzare metodi statistici per trovare n, e successivamente applicare l'analisi delle frequenze a ciascun alfabeto cifrante. La parte più complicata sta dunque nello scoprire la lunghezza della chiave cifrante anche se la cosa non è impossibile. Nel testo cifrato infatti, se la chiave utilizzata è breve, vi saranno probabilmente delle serie di lettere ripetute. Queste serie di lettere, se abbastanza lunghe (5-6 caratteri), saranno generate probabilmente dalla stessa parola in chiaro. Basterà allora calcolare la distanza tra una parola e l'altra e la lunghezza della ripetizione per risalire alla lunghezza n della chiave. Così facendo si può capire quali lettere usano il primo alfabeto cifrante, quali il secondo, e così via procedendo poi con l'analisi delle frequenze per ciascun alfabeto.

Il cifrario di Vigenère: la sicurezza

Per evitare questo problema, una soluzione consiste nell'usare una chiave di dimensioni simili a quella del testo per rendere impossibile uno studio statistico del testo criptato. Questo tipo di sistema di cifratura è detto **Sistema a chiave usa e butta**. Il problema di questo metodo è la lunghezza della chiave di cifratura (più il testo da criptare è lungo, più la chiave deve essere voluminosa), che impedisce la sua memorizzazione e include una probabilità d'errore nella chiave maggiore (un solo errore rende il testo indecifrabile, ecc.).

Questo cifrario, più che una variante, può essere considerato una versione più avanzata del Cifrario di Cesare: quello infatti, per cifrare una stringa sostituisce ogni lettera con quella dell'alfabeto che si trova di n posti davanti (dove n è la nostra chiave).
Il cifrario di Vigenère invece, utilizza praticamente n cifrari di cesare: infatti ogni lettera della stringa viene spostata di un valore differente.
Per stabilire di quanto ogni lettera deve essere "spostata", si utilizza sempre una chiave, che in gergo viene definita "Verme".
Vediamo quindi un esempio:
Stringa: CIFRARIODIVIGENERE
Verme: INFORMATICALABINF
Codice: KVKFRDIHLKVTGFVRWS
Il verme è appunto una parola, e la prima lettera della stringa, deve essere spostata un numero di volte pari alla posizione della prima lettera del verme nell'alfabeto.
Un esempio (ed anche un metodo) facile per capire subito come funziona questo cifrario, è visualizzare la Tavola di Vigenère (o quadrato di Vigenère):

ripetizioni. La distanza fra queste due ripetizioni corrisponde alla lunghezza della chiave o ad un suo multiplo.
Questo ci permette, sopratutto in testi lunghi, di trovare diverse sequenza ripetute che quindi ci permettono di capire (quasi sicuramente) che la lunghezza della chiave è il massimo comun divisore tra le distanze tra sequenze ripetute o al massimo un suo minimo.
Una volta capita la lunghezza della chiave, per ritornare al messaggio ci bastera decifrare n messaggi cifrati con un Cifrario di Cesare.
Quindi, i passaggi in generale sono:
1.Trovare ed elencare i gruppi di lettere ripetuti con la posizione della lettera iniziale rispetto al crittogramma
2.Calcolo degli intervalli fra le ripetizioni (facendo delle sottrazioni)
3.Scomposizione degli intervalli nei rispettivi sottomultipli (es: 55 i sottomultipli sono 5 e 11)
4.Trovare i fattori che ricorrono più frequentemente tenendo a mente che i gruppi di lettere più grandi sono meno probabilmente causali

l risultato finale della cifratura è:

WKAGHWUNPYDLANDL

La forza di questo cifrato sta nel fatto che il numero di chiavi è enorme, quindi gli attacchi praticati con la forza (ad esempio provare tutte le possibili combinazioni) non sono praticabili. Per quasi trecento anni rimase infatti inviolato ma poi cadde di fronte all'analisi del colonnello prussiano Friedrich Kasiski, che nel 1863 pubblicò un libro che conteneva un metodo di decrittazione della tavola di Vigenère e del suo cifrario polialfabetico con chiave ripetuta. Considerando infatti l'esempio precedente:

PROVADICIFRATURA – testo in chiaro
HTMLHTMLHTMLHTML – chiave
WKAGHWUNPYDLANDL – testo cifrato

Le due **R** del testo in chiaro vengono cifrate la prima con una **T** la seconda con una **M** come deve essere in un cifrario polialfabetico. Ma le ultime due **A** vengono invece cifrate con la stessa lettera, la **L**. Il motivo è evidente: le due **A** si trovano a quattro caratteri di distanza l'una dall'altra e quattro è proprio la lunghezza della chiave. Di fatto il codice di Vigenère si riduce qui a cinque codici di Cesare intercalati.

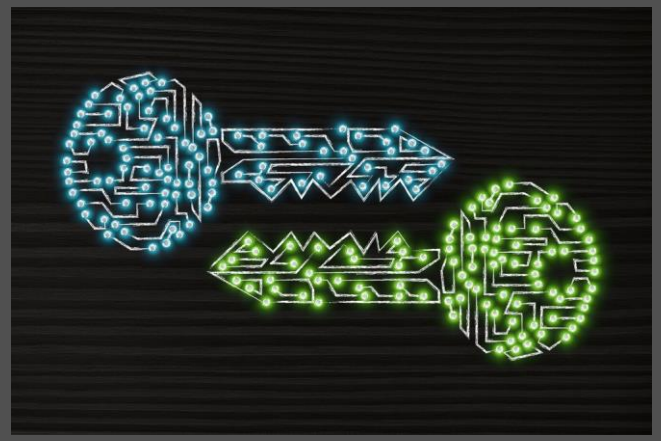
L'attacco di Kasiski si basa quindi sull'osservazione che in un crittogramma alla Vigenère si trovano spesso sequenze identiche di caratteri a una certa distanza l'una dell'altra. Questo avviene evidentemente per il motivo esposto sopra, il fatto cioè di utilizzare ciclicamente la stessa chiave. Se per esempio usando una chiave di quattro lettere come sopra, e si scrive due volte la stessa parola a 20 caratteri di distanza questa sarà cifrata in modo identico essendo 20 un multiplo della lunghezza della chiave che è cinque.

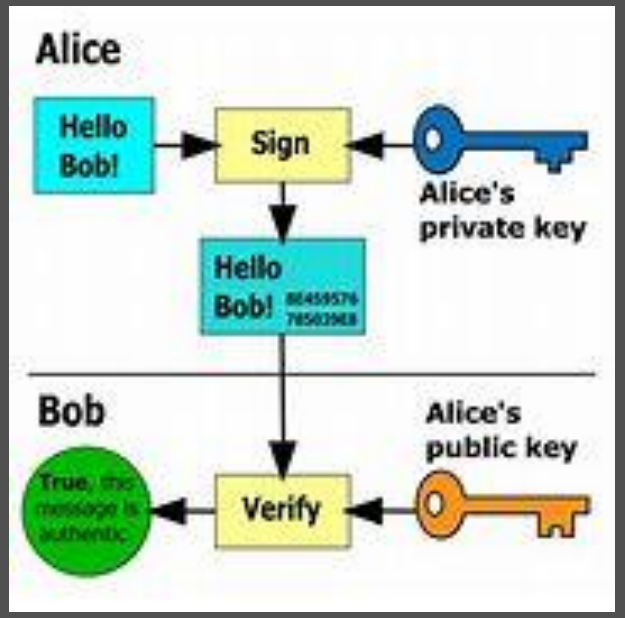
Se allora si individuano tutte le sequenze ripetute (e in un testo lungo o in più testi se ne troveranno molte) allora è pressoché certo che il massimo comune divisore tra le distanze tra sequenze identiche è la lunghezza della chiave, o tutt'al più un suo multiplo. Una volta individuata la lunghezza della chiave (supponiamo sia quattro), il messaggio si riduce a quattro messaggi intercalati, tutti cifrati con un codice di Cesare ed è allora molto facile completarne la decifratura.

La conclusione è che la cifra di Vigenère è affidabile solo quando la chiave è di lunghezza comparabile a quella del testo e viene cambiata molto spesso, cosa che comporta problemi pratici non indifferenti (trasmissione e cambiamento della chiave richiedono un canale di comunicazione assolutamente sicuro).

LEZIONE SUCCESSIVA

http://www.dia.uniroma3.it/~dispense/merola/critto/tesine/vigenere.pdf




Alice
Hello Bob!
Sign
Alice's private key
Hello Bob!
Bob
Alice's public key
Verify

scritture segrete

CRITTOGRAFIA (altera il messaggio)

TRASPOSIZIONE
(permuta le lettere del messaggio secondo una qualche regola reversibile)
Es. scitala e griglie quadrate

SOSTITUZIONE

CODICE
(sostituisce le parole)

CIFRATURA
(sostituisce le lettere)

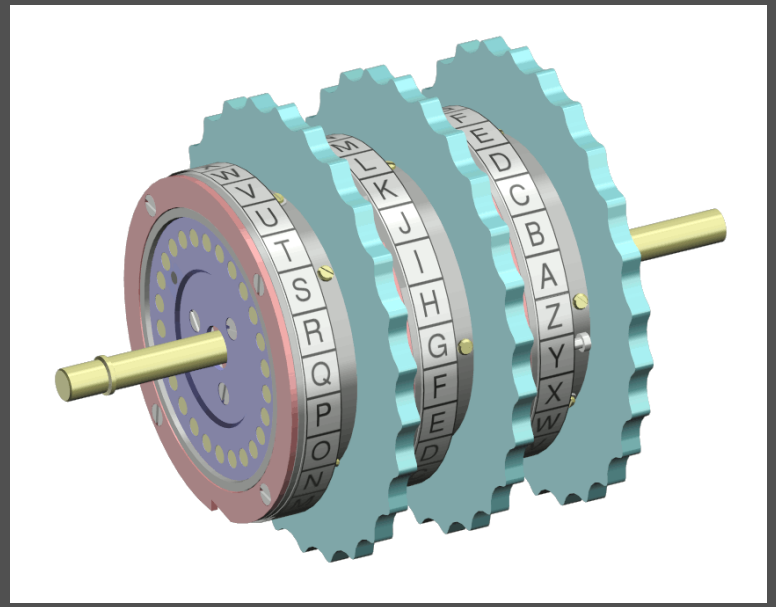
Monoalfabetica
(ogni elemento del testo chiaro viene *sostituito* sempre dallo stesso elemento secondo una tabella di sostituzione unica)
Es Cifrario di Cesare

Polialfabetica
(un elemento del testo chiaro viene sostituito da un elemento che non è sempre lo stesso, ma da determinarsi per mezzo di una chiave segreta da concordarsi)
Es. Cifrario di Vigenere

STEGANOGRAFIA (occulta il messaggio)

- Le macchine decifranti

Schema dei rotori della macchina Enigma



La storia di Enigma

- **La storia di Enigma**

- Alla fine della prima guerra mondiale è apparsa impellente la necessità di criptare i messaggi (anche se le tecniche di cifratura esistevano già da molto tempo). È stato un olandese residente in Germania, il dott. **Arthur Scherbius** che mise a punto la macchina **Enigma** per decriptare dei messaggi a scopo commerciale.

  Il modello **A** della macchina (*Chieffrienmaschinen Aktien Gesellschaft*) fu presentato nel 1923 al *Congresso Postale Internazionale* di Berna. Il prezzo di questa macchina all'epoca (l'equivalente di 30000 euro ad oggi), ma fu un puro fallimento. L'idea era però lanciata e la marina militare tedesca riprese il progetto nel 1925 e confidò il suo sviluppo al servizio di cifratura (*Chiffrierstelle*) del ministero delle guerra tedesco. Il modello *Enigma M3* fu infine adottato dalla *Wehrmacht* (l'esercito tedesco) il 12 gennaio 1937.

  Quello che i tedeschi ignoravano, era che i servizi di controspionaggio francesi e polacchi lavorarono a loro volta, dal 1930, ad un metodo di decifratura messaggi simile. Il comandante **Gustave Bertrand** dei servizi segreti francesi, reclutò a questo scopo **Hans Thilo Schmidt** (nome in codice **Asche**), che lavorava all'epoca per la *Chiffrierstelle*.

  Allo scoppio della seconda guerra mondiale nel 1939, gli alleati sapevano decriptare i messaggi di Enigma. Il 24 luglio 1939, **Marian Rejewski**, responsabile di *Biuro Szyfrow* (il servizio europeo più avanzato nelle ricerche sulla cifratura tedesca) consegna un modello della macchina Enigma al comandante Bertrand e a **Alistar Denniston**, capo del servizio di decifratura dell'*Intelligence Service* (IS) britannico.

  Con l'intensificarsi della guerra la frequenza delle decifratura aumentò. Tra i mesi di ottobre e giugno del 1939, più di 4000 messaggi cifrati furono decifrati dai servizi segreti francesi. Queste operazioni portavano oramai un nome: *Operazione Z* per i francesi e *Code Ultra* per gli inglesi.

  Nell'agosto del 1939 gli inglesi installarono a Bletchley Park (80 Km da Londra) i servizi di Codice e di Cifrario. Poco meno di 12000 scienziati e matematici inglesi, polacchi e francesi lavoravano per crackare il codice di Enigma. Fra questi matematici, troviamo uno degli inventori dell'informatica moderna: **Alan Turing**, che dirigeva i lavori. I messaggi decriptati a Bletcheley Pard arrivavano su dei nastro trasportatore alla *Huts 6*, poi, alla postazione per essere tradotti (due postazioni per squadra):

  Uno per i messaggi in ritardo;

  Uno per il materiale urgente.

  I messaggi tradotti della *Luftwaffe* erano trasmessi ai 3A e quelli dell'esercito ai 3M (A = aviazione; M = militare). Si attribuivano in seguito delle *Z* in funzione dell'importanza dei messaggi (*1Z*: importanza bassa; *5Z*: estremamente urgente). Le informazioni erano riassunte e inviate in 3 esemplari:

  Uno al SIS di Broadway;

  Uno al servizio del ministero appropriato o a Withehall;

  Uno al generale interessato sul terreno.

  Gli inglesi riuscirono così a decifrare questi messaggi codificati. Solo per la *Kriegsmarine* (Marina militare tedesca), che usava delle misure di cifratura diverse, la decifratura risultò più difficile. La cattura sull'*U-110* di un Enigma e soprattutto le sue istruzioni permisero un progresso importante. Questo permetteva di conoscere le posizioni dei sottomarini e di ridurre l'affondamento di navi dai *U-Boot*. Il primo febbraio 1942, il modello *Enigma M4* fu messo in servizio. Per undici mesi, gli alleati non riuscirono a decriptare nessun messaggio.

  Durante tutta la guerra, furono decriptati più di 18 000 messaggi al giorno, e permisero alle forze alleate di conoscere le intenzioni della Germania. L'ultimo messaggio cifrato fu trovato in Norvegia, firmato dall'**Ammiraglio Doenitz**: «il Führer è morto. Il combattimento continua». I tedeschi non hanno mai dubitato che la loro preziosa macchina potesse essere decriptata.

- Strumenti per la crittografia tradizionale • La crittografia "tradizionale" ha lavorato molto sulle tecniche di cifratura per sostituzione monoalfabetica e polialfabetica • Si usavano anche "classificatori" per codificare parole e frasi • Le tecniche erano pensate per l'uomo o per strumenti meccanici più o meno semplici
- Strumenti per la crittografia moderna • La crittografia moderna è stata fortemente influenzata dall'informatica e dalla matematica • L'informatica – pone nuove sfide rendendo possibile decifrare in modo semplice cifrari ritenuti impossibili – chiede nuove applicazioni e nuove tecniche: ad esempio lo scambio di chiavi, la crittografia a chiave pubblica, la firma digitale, … – fornisce hardware e soprattutto software per la crittografia.La matematica è lo strumento fondamentale per la crittografia – fornisce metodi per cifrare, decifrare, firmare e controllare messaggi – garantisce la sicurezza della crittografia – studia (unitamente all'informatica) come svolgere velocemente le operazioni crittografiche

Chiave segreta o chiave pubblica?

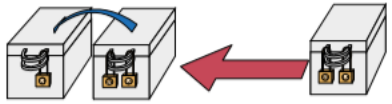
- I cifrari tradizionali sono oggi detti a chiave segreta in quanto richiedono che i due comunicanti debbano preventivamente concordare una chiave segreta.
- A differenza dei cifrari tradizionali, basati su una chiave segreta, questi cifrari usano anche una chiave pubblica, garantita da una qualche autorità.

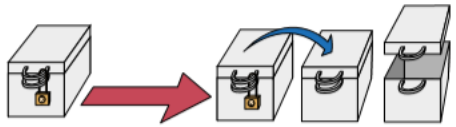
Il superamento del meccanismo della condivisione di chiave venne elaborato nel Novecento: esso viene spiegato quasi sempre attraverso l'analogia del doppio lucchetto, presentata qui sotto.

Una persona [A] deve spedire un messaggio riservato a una persona [B]. Ripone quindi il testo in un bauletto e assicura quest'ultimo con un lucchetto. Nessuno al di fuori di [A] può aprire ora la scatola.

[A] spedisce il bauletto a [B] che però non lo può aprire. Lo sigilla quindi con un secondo lucchetto che nessun al di fuori di [B] può manovrare.

A questo punto [B] rispedisce il tutto ad [A]: questi può togliere il suo lucchetto.





[A] rispedisce un ultima volta la scatola a [B], che può ora aprire il suo lucchetto e quindi la scatola e leggere finalmente il messaggio.

Il meccanismo del doppio lucchetto prevede che il messaggio rimbalzi avanti e indietro tra i due interlocutori, operazione che non crea problemi con le infrastrutture telematiche contemporanee. Come si vede [A] e [B] non condividono alcuna chiave e il messaggio è sigillato in tutti e tre i viaggi.