2 Crittoanalisi

2.1 Principio di Kerckhoffs

Uno dei principi base della Crittografia è la seguente

Definizione 2.1. **(Principio di Kerckhoffs)**
La sicurezza di un crittosistema dipende solo dalla segretezza della chiave.

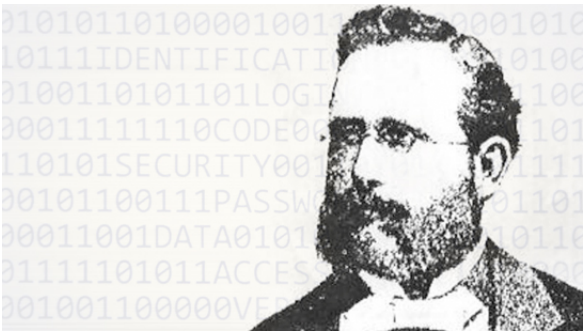

Figura 2.1: Auguste Kerckhoffs (1835 – 1903)

Esistono diversi **modelli di attacco** ai crittosistemi, però tutti hanno a comune un solo obiettivo: determinare la chiave. La differenza principale tra tali modelli è rappresentato dalle informazioni che l'avversario ha a disposizione quando esso effettua un attacco.

I modelli di attacco sono classificati come segue:

- **Attacco basato sulla conoscenza del testo cifrato**: l'avversario conosce una stringa y di testo cifrato.
- **Attacco basato sulla conoscenza del testo in chiaro**: l'avversario conosce una stringa x di testo in chiaro e la corrispondente stringa y di testo cifrato.
- **Attacco basato sulla conoscenza del testo in chiaro scelto**: l'avversario ha accesso temporaneo alla macchina di cifratura. Quindi, può scegliere un'opportuna stringa di testo in chiaro x e calcolarne la corrispondente stringa di testo cifrato y.

Esempio 2.2. Durante la II guerra mondiale gli inglesi posizionarono delle mine in certe località, sapendo che i tedeschi, trovate quelle mine, avrebbero criptato i luoghi e li avrebbero rimandati al quartier generale. Questi messaggi crittografati sono stati utilizzati dai crittoanalisti del Bletchley Park (il sito dell'unità principale di crittoanalisi del Regno Unito in quegli anni) per decifrare lo schema di crittografia tedesco.

Esempio 2.3. Nel Maggio del 1942, i crittoanalisti della Marina statunitense intercettarono un messaggio crittografato dai giapponesi che furono in grado di decodificare parzialmente.
Il risultato indicava che i giapponesi stavano pianificando un attacco su AF, dove AF era un frammento di testo cifrato che gli Stati Uniti non erano in grado di decodificare.
Per altri motivi, gli Stati Uniti credevano che l'obiettivo fossero le Isole di Midway. Sfortunatamente, i loro tentativi di convincere i pianificatori di Washington furono futili; la convinzione generale era che le Midway non potessero essere il bersaglio.
I crittoanalisti della Marina hanno escogitato il seguente piano: hanno incaricato le forze statunitensi di inviare un falso messaggio alle Midway e i giapponesi hanno immediatamente riferito ai loro superiori che *"AF è a corto di acqua"*.
I crittoanalisti della Marina avevano ora la loro prova che AF corrispondeva a Midway, e gli Stati Uniti spedirono tre portaerei in quella posizione. Il risultato fu che le Isole di Midway furono salvate ed i giapponesi subirono perdite significative. Questa battaglia fu un punto di svolta nella guerra tra Stati Uniti e Giappone nel Pacifico.
I crittoanalisti della Marina hanno effettuato un attacco con testo in chiaro, poichè erano in grado di influenzare i giapponesi (anche se in modo indiretto) per crittografare la parola "Midway". Se lo schema di crittografia giapponese fosse stato protetto contro gli attacchi con testo in chiaro scelto, questa strategia da parte dei crittoanalisti statunitensi non avrebbe funzionato (e la storia sarebbe potuta andare in modo molto diverso!).

- **Attacco basato sulla conoscenza del testo cifrato scelto**: l'avversario ha accesso temporaneo alla macchina di decifratura. Quindi, può scegliere un'opportuna stringa di testo cifrato y e calcolarne la corrispondente stringa di testo in chiaro x.

Esempio 2.4. Come nel caso degli attacchi basati sulla conoscenza del testo in chiaro scelto, non ci aspettiamo che le parti oneste decodifichino il testo cifrato arbitrario scelto da un avversario.
Tuttavia, potrebbero esserci degli scenari in cui un avversario potrebbe essere in grado di influenzare ciò che viene decodificato e apprendere alcune informazioni parziali sul risultato.

Esempio 2.5. Ad esempio:

1. Nell'**Esempio 2.3** è concepibile che anche i crittoanalisti statunitensi abbiano provato a inviare messaggi crittografati ai giapponesi e quindi a monitorare il loro comportamento. Tale comportamento (ad esempio il movimento di forze e poi simili) avrebbe potuto fornire informazioni importanti sul testo in chiaro sottostante.

2. Immagina un utente che invia messaggi crittografati alla propria banca. Un avversario può essere in grado di inviare i testi cifrati per conto di tale utente; la banca decodifica i testi cifrati e l'avversario potrebbe imparare qualcosa sul risultato. Ad esempio, se un testo cifrato corrisponde a un testo in chiaro non formattato (ad esempio un messaggio incomprensibile o semplicemente uno non formattato correttamente), l'avversario può essere in grado di dedurlo dalla reazione della banca (cioè lo schema della comunicazione successiva).

2.2 Crittoanalisi del cifrario mediante sostituzione

Il cifrario mediante sostituzione non è sicuro, poiché può essere decifrato con il **metodo della ricerca esaustiva della chiave** (o anche **forza bruta**), cioè andando a provare tutte le chiavi possibili e vedendo quale fornisce una frase di senso compiuto; infatti abbiamo solo 26 chiavi possibili (di cui una è quella banale), ed è facile determinare quella giusta provando tutte le chiavi.

Il seguente metodo può essere utilizzato per tutti i cifrari che hanno numeri non eccessivamente grandi di chiavi possibili.

Sia **XTQIFYN** il testo cifrato, convertito in numeri 23 19 16 8 5 24 13; si scelga casualmente una chiave, per esempio $K = 4$ e si applichi la funzione di decifratura corrispondente:

$$\begin{aligned} d_K(23) &= (23-4) \mod 26 = 19 \\ d_K(19) &= (19-4) \mod 26 = 15 \\ d_K(16) &= (16-4) \mod 26 = 12 \\ d_K(8) &= (8-4) \mod 26 = 4 \\ d_K(5) &= (5-4) \mod 26 = 1 \\ d_K(24) &= (24-4) \mod 26 = 20 \\ d_K(13) &= (13-4) \mod 26 = 9 \end{aligned}$$

che sarà quindi **tpmebuj** che non ha senso compiuto.

Applicando nuovamente lo stesso procedimento con $K = 5$, si avrà

$$\begin{aligned} d_K(23) &= (23-5) \mod 26 = 18 \\ d_K(19) &= (19-5) \mod 26 = 14 \\ d_K(16) &= (16-5) \mod 26 = 11 \\ d_K(8) &= (8-5) \mod 26 = 3 \\ d_K(5) &= (5-5) \mod 26 = 0 \\ d_K(24) &= (24-5) \mod 26 = 19 \\ d_K(13) &= (13-5) \mod 26 = 8 \end{aligned}$$

che convertito è **soldati** che ha senso compiuto e rappresenta quindi il messaggio in chiaro di partenza.

2.3 Crittoanalisi del cifrario affine e cifrario mediante permutazione

Per la crittoanalisi del cifrario affine si utilizza la **Tabella delle occorrenze** delle 26 lettere dell'alfabeto. È noto che in qualsiasi alfabeto esistono lettere che si ripetono più spesso di altre. Si viene a creare quindi una tabella delle probabilità con cui le lettere della lingua italiana appaiono nelle parole di senso compiuto.

a	**0,117**		**n**	**0,069**
b	**0,009**		**o**	**0,098**
c	**0,045**		**p**	**0,031**
d	**0,037**		**q**	**0,005**
e	**0,118**		**r**	**0,064**
f	**0,009**		**s**	**0,049**
g	**0,016**		**t**	**0,056**
h	**0,015**		**u**	**0,030**
i	**0,113**		**v**	**0,021**
j	**0,000**		**w**	**0,000**
k	**0,000**		**x**	**0,000**
l	**0,065**		**y**	**0,000**
m	**0,025**		**z**	**0,000**

Tabella 2: Tabella delle occorrenze.

Si cercano le lettere che appaiono più frequentemente nel testo cifrato e, osservando la tabella delle frequenze, si associano a quelle più frequenti nella lingua corrispondente, nel caso specifico italiana.

Si crea quindi un sistema $y = e_k(x) = ax + b$ dato dalle coppie (x_i, y_i), dove y_i sono le lettere del testo cifrato più frequenti e le x_i sono le presunte lettere del testo in chiaro. Lo scopo è quello di trovare $K = (a, b)$, verificando che $mcd(a, 26) = 1$. La coppia utilizzata è corretta se il testo ha senso compiuto.
Per esempio, il testo da decifrare è

XKLJSUJNXDYBUJTYOBLRGXUJDXKJWLTYXSX
ALYQXLAAJYRGXWXUXNNXKBRGJLKLVXYQXRG
XRBNXKLUJNXDYLJSJYQXKYBUXSOBL

Si cerchino le due coppie (x_1, y_1) e (x_2, y_2) tali che $ax_1 + b = y_1$ e $ax_2 + b = y_2$ per determinare $K = (a, b)$. I caratteri sono distribuiti come nella tabella seguente

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**
3	**6**	**0**	**3**	**0**	**0**	**4**	**0**	**0**	**10**	**5**	**10**	**0**	**5**	**2**
P	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**				
0	**3**	**5**	**4**	**2**	**6**	**1**	**2**	**18**	**9**	**0**				

Tabella 3: Tabella delle occorrenze dell'esempio.

Rifacendoci alla tabella delle frequenze si supponga che **X** sia **e** e che **J** sia **a**. Si ha quindi $4a + b = 23$ e $a0 + b = 9$; la congruenza $4a \equiv 14 \pmod{26}$ ha due soluzioni $a = 10$ e $a = -3 = 23$. Poiché la prima è sicuramente non corretta (il massimo comun divisore è diverso da 1) proviamo con la seconda sfruttando la sostituzione $y = 23x + 9$. Sapendo che $23^{-1} \equiv 17 \pmod{26}$, si ha che $x = 17(y - 9)$ e decifrando si ottiene

eriaxfaqecvufaovhuigbeface raniovexeivpeiddavgbenefeqqerugba
iriwevpegbeguqerifaqecviaxavpervufexhui

che però sembra non corretta.

Si consideri ora la sostituzione **L** con **a**, allora $4a + b = 23$ e $a0 + b = 11$; la congruenza $4a \equiv 12 \pmod{26}$ ha due soluzioni: $a = 3$ e $a = 16$; la seconda non è sicuramente corretta mentre, provando la sostituzione $y = 3x + 11$, si ha che $x = 9(y - 11)$.

Decifrando si ha come messaggio in chiaro:

eraildisegnodiunboachedigerivaunelefanteaffinchevedess
erochiaramentechecoseradisegnailinternodelboa

cioè *"era il disegno di un boa che digeriva un elefante. Affinché vedesse chiaramente che cos'era disegnai l'interno del boa"* da *Il piccolo principe* di Antoine de Saint-Exupèry.

Il metodo di crittoanalisi per il cifrario mediante permutazione è analogo a quello del cifrario affine; la permutazione non modifica la frequenza delle singole lettere e l'analisi delle frequenze è necessariamente simile a quella di un testo in chiaro. Suddividendo, quindi, il testo cifrato in porzioni di testo si cercano parole di senso compiuto; una volta trovata la regolarità, questa può essere estesa al resto del messaggio. Pertanto un attacco di forza bruta nel caso peggiore dovrebbe provare tutte le $n!$ possibili permutazioni con n lunghezza del messaggio.

2.4 Crittoanalisi del cifrario di Vigènere

La crittoanalisi del cifrario di Vigenére si basa principalmente su due tecniche: il **Test di Kasiski** e l'**Indice di coincidenza**. Entrambe vengono utilizzate per calcolare la lunghezza della stringa che rappresenta la chiave.

Il metodo Kasiski prende il nome dal maggiore prussiano Friedrich Kasiski, che nel 1863 pubblicò un metodo di decifratura della tavola di Vigenére.

Figura 2.2: Friedrich Wilhelm Kasiski (1805 – 1881)

Il Test di Kasiski venne descritto da Friedrich Kasiski nel 1863, tuttavia venne utilizzato anche prima da Charles Babbage, intorno al 1854.

È basato sull'osservazione che **due segmenti identici di testo in chiaro vengono decriptati nello stesso modo se, e solo se, la distanza tra questi è un multiplo della lunghezza della chiave**.
Viceversa, se osserviamo che **due segmenti di testo cifrato, ognuno di lunghezza almeno 3, sono identici, è probabile che corrispondano a segmenti identici di testo in chiaro**.

Il Test di Kasiski cerca porzioni identiche di testo cifrato di lunghezza almeno 3 e memorizza la distanza tra le posizioni di partenza. Si ottengono quindi diverse distanze e vale che la lunghezza della chiave m divide il massimo comun divisore delle distanze misurate.

L'Indice di coincidenza fu introdotto nel 1920 da Wolfe Friedman.

Figura 2.3: Wolfe Frederick Friedman (1891 – 1969)

Definizione 2.6. **(Indice di coincidenza)**
Sia $x = x_1x_2\cdots x_n$ una stringa di n lettere. L'**Indice di coincidenza** di x, denotato con $I_c(x)$, è la probabilità che $x_h = x_k$ per $h \neq k$.

Siano $f_0, f_1 \cdots f_{25}$ le frequenze di A, B, ..., Z rispettivamente nel testo cifrato. Il numero di coppie di lettere uguali è $\sum_{i=0}^{25} \binom{f_i}{2}$.
Quindi l'Indice di coincidenza è

$$I_c(x) = \frac{\sum_{i=0}^{25} \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)}.$$

Vale che

$$\frac{f_i - 1}{n-1} \approx \frac{f_i}{n} \approx p_i$$

dove i p_i sono dati nella **Tabella 2**, quindi

$$I_c(x) \approx \sum_{i=0}^{25} {p_i}^2 = 0,075.$$

Pertanto l'Indice di coincidenza nel testo cifrato a partire da un testo in chiaro scritto in italiano deve essere approssimativamente $0,075$.

Ora, si supponga di avere un testo cifrato $y = y_1 y_2 ... y_n$, con $y_i \in \mathbb{Z}_{26}$ con $i \in \{1, ..., n\}$ costruito usando il cifrario di Vigenère.
Siano $y'_1\ y'_2 \cdots y'_m$ m sottostringhe di y rappresentate in colonne in una tabella di dimensioni $m \times \frac{n}{m}$:

$$\begin{array}{ccccccc} y'_1 & = & y_1 & y_{m+1} & y_{2m+1} & \cdots \\ y'_2 & = & y_2 & y_{m+2} & y_{2m+2} & \cdots \\ \vdots & & \vdots & \vdots & \vdots & \vdots \\ y'_m & = & y_m & y_{2m} & y_{3m} & \cdots \end{array}$$

Se $y'_1\ y'_2 \cdots y'_m$ sono costruiti in questo modo e m è la lunghezza della chiave, allora ogni $I_c(y'_i)$ è approssimativamente uguale a $0,075$. D'altra parte, se m non è la lunghezza della chiave, allora la sottostringa y'_i sembrerà molto più casuale e il valore di $I_c(y'_i)$ si scosterà visibilmente da $0,075$.
Si osservi che una stringa completamente casuale avrà un indice di coincidenza $I_c(x)$ approssimativamente uguale a $0,038$, per l'equiprobabilità di ogni lettera, e i due valori sono sufficientemente lontani da darci la possibilità di determinare la lunghezza della chiave grazie a questo metodo (o confermare il Test di Kasiski).

Si supponga ora di aver determinato il valore corretto di m e si voglia determinare ora la chiave. Descriviamo un metodo semplice ma efficace per fare ciò.

Sia $1 \leq i \leq m$ e siano $f_0, f_1 \cdots f_{25}$ le frequenze di A, B,...,Z nella stringa y'_i. Sia inoltre $n' = \frac{n}{m}$ la lunghezza della stringa y'_i; la distribuzione di probabilità delle 26 lettere in y'_i è

$$\frac{f_0}{n'}, \cdots, \frac{f_{25}}{n'}$$

Si noti che la sottostringa y'_i ottenuta dalla cifratura di un sottoinsieme del testo in chiaro mediante sostituzione con chiave k_i. Quindi, dobbiamo sperare che

$$\frac{f_{k_i}}{n'}, \cdots, \frac{f_{25+k_i}}{n'}$$

sia "vicina" alla distribuzione ideale con $p_0, \cdots, p_{25}$.

Supposto $0 \leq g \leq 25$ sia

$$M_g = \sum_{i=0}^{25} \frac{p_i f_{i+g}}{n'}.$$

Se $g = k_i$ allora $M_g \approx 0,075$ simile all'indice di coincidenza. Se $g \neq k_i$, allora M_g sarà di solito significativamente più piccolo di 0,075. Questa tecnica quindi dovrebbe permetterci di determinare il corretto valore di k_i per ogni i, con $1 \leq i \leq m$.

Di seguito è fornito un testo cifrato secondo Vigènere che viene decifrato utilizzando il Test di Kasiski e l'Indice di coincidenza.

DLSVHR**LTFBLP**JVTNPTKQSAXZTWXOCTWZZKW**UPTKW**
IFGIIFEGJMLEKLVSNWLIRKUEMVTRLDALRVIUNUDXS
RTRBGOGJKJZYTQVTLFTYZXVMVLODXWEAFAADUWNAOO
MMFEUJCTTYJINLRRAGWOKIJTGVAWWBROYTGDWEAXR
KWXOJWDLYZBMLZRAMWRVKGDZVWUNOUMFEGCQVTHFZ
FPUVQDNAZVGXKSIKEGMIAYWLMAEKDXALYGIJRKIMA
WZVZJZXVI**UPTKW**DPMYMSWRZVLZXEWDLHZBSKOFV**WO**
KCTSEOXZ**WOKCTS**XGCMKTGGWKEGTWEPGHCAWGJCVTA
EIYCGEZMAKKIYWORBSLVZKUZYLTELXVIUTTHCWNKEB
GAGJAAOGCTWFRKQEPIRXSYTVLWWBZTDLMXQ
GOOXQWSGMMEBAVTD**LTFBLP**IFVLCUZTKZRZBGPXR

Per determinare la lunghezza della chiave si utilizza il Test di Kasiski. Cercando le ripetizioni nel testo si ha che

- il gruppo di lettere **LTFBLP** è ripetuto nel testo a distanza di 415 lettere;
- il gruppo di lettere **UPTKW** è ripetuto nel testo a distanza di 220 lettere;
- il gruppo di lettere **WOKCTS** è ripetuto a distanza di 10 lettere.

Quindi è ipotizzabile che la lunghezza delle chiave sia

$$mcd(415, 220, 10) = 5.$$

Sapendo la lunghezza della chiave, le lettere nelle posizioni $n + 5k$ dove $n \in \{0, 1, 2, 3, 4\}$ sono cifrate con lo stesso cifrario mediante sostituzione. Analizzando le frequenze dei gruppi di lettere e confrontandoli con l'analisi delle frequenze nelle lettere italiane, è possibile capire quale è la sostituzione delle lettere. Per esempio, nel caso delle lettere di posto $1 + 5k$ che sono

DRLNSWWUI FLSRVAUSG JVYVWAAFT NGJWYEWDM MGUFVFDGK AAAJAJUDS LDSWSWSKK EAVYMYSUE UWGAWESWD GWEDLLKG

dall'analisi delle frequenze risulta che il cambio è $S \to a$.

Procedendo analogamente per le posizioni $2 + 5k$ $3 + 5k$, $4 + 5k$, $5k$ si trova che il testo in chiaro è

lamez zanot tedel venti april emill eotto cento quara ntase tteun acqua zzone diluv ialea ccomp agnat odasc rosci difol goree daimp etuos isoff idiv entos ubiss avala solit ariae selva ggiam ompra cemis olasi tuata sulle coste occid ental idibo rneoe ilcui nomeb astav ainqu eitem piasp arger eilte rrore acent olegh ealli ntorn olabi tazio nedel latig redel lamal esiap ostac omeaq uilas udiun agran rupet aglia taapi ccosu lmare acinq uecen topas sidal leult imeca panne delvi llagg iodig jehaw emque llano tteco ntroi lsoli toera

cioè

La mezzanotte del 20 aprile 1847, un acquazzone diluviale, accompagnato da scrosci di folgore e da impetuosi soffi di vento subissava la solitaria e selvaggia Mompracem, isola situata sulle coste occidentali di Borneo, e il cui nome bastava in quei tempi a spargere il terrore a cento leghe all'intorno. L'abitazione della Tigre della Malesia, posta come aquila su di una gran rupe tagliata a picco sul mare, a cinquecento passi dalle ultime capanne del villaggio di Gjehawem, quella notte, contro il solito, era

da *La tigre della Malesia* di Emilio Salgari e la chiave è **SLGRI**.

2.5 Crittoanalisi del cifrario di Hill

Il cifrario di Hill può essere difficile da violare con la sola conoscenza del testo cifrato, ma è più semplice se si conosce una porzione del testo in chiaro corrispondente.

Si supponga che si conosca il valore di m che è l'ordine possibile delle matrici che rappresentano la chiave del cifrario. Si supponga inoltre che si abbiano almeno m parti di testo in chiaro e del corrispondente testo cifrato, cioè

$$x_j = x_{j,1}, x_{j,2} \cdots x_{j,m}$$
$$y_j = y_{j,1}, y_{j,2} \cdots y_{j,m}$$

dove $y_j = e_K(x_j)$ e $1 \leq j \leq m$.

Siano $X = (x_{i,j})$ e $Y = (y_{i,j})$ delle matrici quadrate di ordine m e sia K la chiave che si intende determinare. Poiché vale che $Y = XK$, se X è invertibile, allora $K = X^{-1}Y$.

Per cui un esempio può essere $m = 2$ e che il testo in chiaro **friday** sia cifrato con **PQCFKU**. Quindi si ha che

$$e_k \begin{pmatrix} 5 \\ 17 \end{pmatrix} = \begin{pmatrix} 15 \\ 16 \end{pmatrix}, \; e_k \begin{pmatrix} 8 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

allora

$$K = \begin{pmatrix} 5 & 8 \\ 17 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 15 & 2 \\ 16 & 5 \end{pmatrix} = \begin{pmatrix} 9 & 2 \\ 1 & 15 \end{pmatrix} \begin{pmatrix} 15 & 2 \\ 16 & 5 \end{pmatrix} = \begin{pmatrix} 11 & 2 \\ 21 & 25 \end{pmatrix}.$$