**07- CRITTOGRAFIA e HASHING**

**Crittografia classica**

Con l’eccezione dello schema “one-time pad”, nessun algoritmo di cifratura è incondizionatamente sicuro. Poiché la crittografia è la scrittura segreta basata su di un codice condiviso fra gli interlocutori, il livello di segretezza di un testo cifrato dipende da due fattori:

* dal cifrario (algoritmo) utilizzato: tenuto segreto nella crittografia classica
* dalla complessità della chiave di cifratura che determina il modo in cui un messaggio viene cifrato

Caratteristiche dei cifrari “classici”:

Utilizzano algoritmi simmetrici: crittografia sui testi, stessa chiave per cifrare e decifrare.

Possono essere basati su:

- Tecniche di sostituzione: per ogni lettera del testo ne uso un altra, avrò bisogno di un cifrario che mi dice a cosa corrisponde ogni lettera cambiata, chiave è il parametro di questa funzione

- Tecniche di trasposizione: ho un testo, non cambio le lettere ma le mescolo, detto in altre parole è un anagramma

**Algoritmi simmetrici** → Nei sistemi “simmetrici”, viene utilizzata una sola chiave, detta segreta. La chiave è utilizzata come parametro unidirezionale e invertibile. Poiché si utilizza una funzione invertibile, il destinatario dovrà soltanto elaborare nuovamente il crittogramma, richiamando l’inverso della funzione di cifratura avente come parametro la medesima chiave utilizzata dal mittente. Visto che mittente e destinatario utilizzano la medesima chiave per crittografare e decifrare il messaggio, la crittografia simmetrica si basa sulla capacità del mittente e del destinatario di mantenere segreta la chiave di cifratura.

Tipi di algoritmi simmetrici "classici":

* codifica (simboli diversi)
* cifrari per sostituzione (simboli uguali)
* cifrari per trasposizione
* cifrari prodotto: cifrari che integrano sostituzione e trasposizione

Per tantissimi anni abbiamo usato un crittografia "stupida", cioè la codifica cambiando i simboli; poi cifrari per sostituzione, trasposizione e infine cifrari prodotto.

**Tecniche di sostituzione** → esempio: Svetonio, nella “Vita duodecim Caesarum”, riferisce che: Ottaviano Augusto era solito sostituire ogni lettera con la successiva (“B pro A, C pro B ac deinceps, eadem ratione, sequentes litteras ponit”); Giulio Cesare lo ruotava di tre posti (“D pro A et perinde reliquas commutet”).

Le rotazioni di Cesare e di Augusto sono le remote progenitrici degli algoritmi classici a chiave segreta: i cifrari a sostituzione.

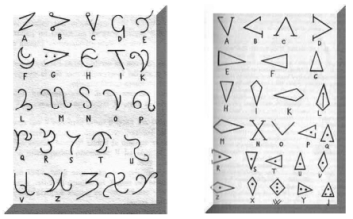
**Tecniche di trasposizione** → esempio: Galileo Galilei, per salvaguardare le sue scoperte astronomiche, le annunciava per mezzo di anagrammi. Vincenzo Viviani, discepolo dello scienziato scrive che il Maestro nel 1610 descrisse le fasi di Venere con il seguente anagramma: “Haec immatura a me iam frustra leguntur o, y” il quale per volontà dell’Imperatore Rodolfo II fu così svelato: “Cynthiae figuras aemulatur mater amorum”. Questi anagrammi sono i progenitori di una seconda categoria di algoritmi classici: i cifrari a trasposizione.

**Codifica e Cifratura** → Si utilizza un alfabeto sostitutivo per il messaggio originale. Ci si assicura che tutti i destinatari autorizzati a leggere il messaggio abbiano una copia dell’alfabeto sostitutivo o un modo per poterlo ricostruire (es: tramite “chiave”). Questa forma (nelle sue molteplici varianti) è stato il metodo alla base della cifratura per migliaia di anni:

– Alfabeto sostitutivo con simboli diversi (codifica)

– Alfabeto sostitutivo con stessi simboli (cifratura)

**Codifica (simboli diversi)**



Codifica con simboli diversi era primo passo di iniziazione per i gruppi che facevano uso di cifrari. Sostituivi alle lettere dei simboli diversi.

Debolezza sta nelle vocali: in italiano il 30/40% delle lettere in una frase sono vocali. Non è un metodo robusto

**Sostituzione (stessi simboli)**

Cifratura → l’alfabeto sostitutivo per il messaggio originale è costituito da stessi simboli mediante:

* Sostituzione
* Trasposizione

Si possono utilizzare anche combinazioni:

* Sostituzione+Trasposizione (cifrari prodotto)
* Codifica parziale+Sostituzione

**Cifrari mono-alfabetici**

Algoritmi di shifting (es: Cesare) → l’alfabeto utilizzato usa lettere spostate in avanti (shiftate) di n posizioni

CHIAVE → è il numero di posizioni di shift: n

y = [x + n] mod 26 (26: numero lettere alfabeto)

Esempio:

Chiave n=4: A→E, B→F, C→G, D→H, …, Z→D

MODENA → QSHIRE

In un alfabeto di 26 lettere, vi sono solo 25 possibilità di rotazione, e quindi solo 25 possibili chiavi.

La crittoanalisi a forza bruta molto semplice. Basta provare tutte le 25 combinazioni (in media bastano 13).

Non si cambiano i simboli ma si sostituiscono, per tanti anni abbiamo sostituito con altri simboli di altri alfabeti ma si può fare con gli stessi simboli usando una chiave (crittografia di cesare).

Un semplice shift per intenderci, in quello di cesare era uno shift circolare. Cambio la chiave, cambio cifrario

Bassa robustezza (ho 25 chiavi se le provo tutte alla fine la trovo).

Cifrari affini → Generalizzazione degli algoritmi di shifting (noto anche come cifrario di Cesare generalizzato)

y = [a x + n] mod 26

con MCD(a, 26) = 1

Esempio: a=9, n=2

affine → CVVWPM

Abbiamo reso sempre più complesso questo cifrario: al posto di un solo parametro ne mettiamo due (a ed n nella formula)

Cifrario basato su alfabeto sostitutivo (base) → Si modifica l’alfabeto con un alfabeto sostitutivo con sostituzione casuale delle lettere: la chiave è l’intero alfabeto sostitutivo. L'idea è avere una chiave di 26 caratteri: un alfabeto sostitutivo proprio. Molto più robusto ma si porta dietro anche una chiave più complessa e pesante, non te lo ricordi a memoria. La chiave posso cambiarla quando voglio (ogni giorno, ogni settimana, ogni mese)

Esempio:

Alfabeto: abcdefghijklmnopqrstuvwxyz

Alf. Sost.: DKVQFIBJWPESCXHTMYAUOLRGZN

Testo: ifwewishtoreplaceletters Testo cifrato: WIRFRWAJUHYFTSDVFSFUUFYA

Cifrario basato su alfabeto sostitutivo con chiave → La modifica è riassunta da una “parola chiave” o “frase” che non deve avere lettere ripetute (se ci sono, si eliminano quelle dopo la prima).

Esempio:

Chiave: XRAYFILMS – lunghezza=9

Alfabeto e alfabeto sostitutivo con chiave

abcdefghijklmnopqrstuvwxyz

XRAYFILMSBCDEGHJKNOPQTUVWZ

Esempio:

Alfabeto: abcdefghijklmnopqrstuvwxyz

Alf. sost. base: DKVQFIBJWPESCXHTMYAUOLRGZN

Alf. sost (chiave):XRAYFILMSBCDEGHJKNOPQTUVWZ

“FACCIA” → IDVVWD (base)

IXAASX (chiave)

**NOTA**: Hanno stessa robustezza.

Perché l’alfabeto sostitutivo con chiave è preferibile all’alfabeto sostitutivo base?

Alfabeto: abcdefghijklmnopqrstuvwxyz

Alf. sost base: DKVQFIBJWPESCXHTMYAUOLRGZN

Alf. sost chiave: XRAYFILMSBCDEGHJKNOPQTUVWZ

Maggiore semplicità di trasmissione della chiave tra i due interlocutori. La semplicità di gestione è e sarà uno degli elementi fondamentali per la scelta di un cifrario.

L’alfabeto sostitutivo è decisamente più robusto dello shifting di Cesare in quanto in un alfabeto di 26 lettere, vi sono 26! possibili combinazioni e quindi 26! possibili chiavi invece di 25. Un attacco di forza bruta richiederebbe migliaia di anni per provare le 26! combinazioni anche se ciascun tentativo richiedesse un microsecondo. Ottima cosa: abbiamo trovato un cifrario sicuro, anche a prova di attacco di forza bruta!

Finito? Purtroppo, no!

Purtroppo, è possibile utilizzare la conoscenza della lingua del messaggio per semplificare il problema della crittoanalisi → crittonalisi statistica o per frequenza.

Crittoanalisi statistica → eventualmente integrata con crittoanalisi per dizionario che considera le specificità del testo cifrato rispetto alla tematica trattata: ogni settore (medico, informatico, militare, legale, politico, industriale, ecc.) ha sue parole chiave, forme, utilizzo o meno di parole straniere, ecc.

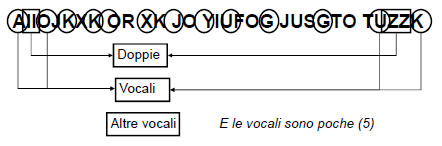
Gli Arabi (nel IX secolo) per primi “superarono” i metodi di cifratura che prevedevano la sostituzione mono-alfabetica mediante meccanismi di analisi di frequenze delle lettere. Il problema è che la sostituzione mono-alfabetica non modifica le frequenze relative delle lettere. Quindi per la crittoanalisi:

* Si calcolano le frequenze delle lettere nel testo cifrato
* Si confrontano con i valori noti di quella lingua
* Si provano a sostituire le lettere sulla base delle frequenze corrispondenti: molte funzionano, alcune no, ma una lingua e un testo hanno altre “debolezze”: bigrammi, trigrammi, parole chiave, affinità o repulsione tra lettere, ecc.

Usando la crittoanalisi statistica integrala, La conoscenza della lingua del messaggio, oltre alla frequenza delle lettere, consente di determinare: diverse frequenze delle combinazioni di lettere: bigrammi e trigrammi più comuni; bigrammi o trigrammi impossibili; lettere doppie; affinità delle lettere: in italiano, ‟q” è sempre seguita da ‟u”; le parole tendono a terminare per vocale; prima e dopo una doppia c’è una vocale; non affinità delle lettere: in italiano, ‟n” non precede mai ‟p”; le vocali non sono mai doppie; spazi significativi nelle frasi, anche se questo problema fu presto superato eliminando gli spazi o introducendoli ogni gruppo fisso di lettere per disorientare il crittoanalista.

Esempio: scoprire il significato della frase cifrata mediante una crittoanalisi basata sulla conoscenza della lingua italiana e dell’argomento (congiura), anche se non si conosce il cifrario e la chiave

AIIOJKXK OR XK JO YIUFOG JUSGTO TUZZK

 → uccidere il re di scozia domani notte

Soluzione: banale algoritmo di shifting con N=6

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

AIIOJKXK OR XK JO YIUFOG JUSGTO TUZZK → UCCIDERE IL RE DI SCOZIA DOMANI NOTTE

Sarebbe stata più semplice una crittoanalisi a forza bruta.

Purtroppo, un cifrario di shifting non si usa da millenni

La debolezza dei cifrari monoalfabetici sta nel fatto che, anche se cambiano una lettera con un’altra, non possono cambiare la “identità” di ciascuna lettera: frequenza della lettera, frequenza di bigrammi e trigrammi, affinità con altre lettere, non affinità con altre lettere, tendenza alla posizione in una parola.

Poi, oltre la sintassi c’è la semantica (soggetto-verbo-complemento) a favorire il crittoanalista

Esempio (alfabeto inglese):

e = 12,7%

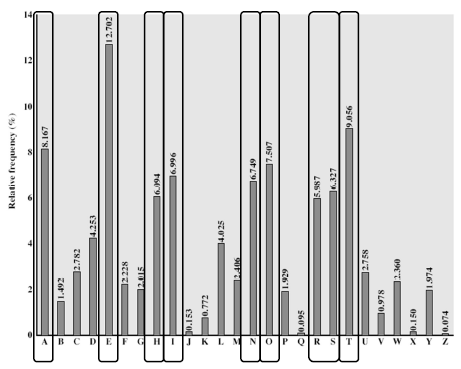
t = 9,1%

a = 8,2%

o = 7,5%

Altre lettere (es., j, q, x, z) sono poco frequenti; pertanto, scoprire le lettere più e meno frequenti è relativamente semplice, e da lì le combinazioni possibili si riducono:

* Il bigramma più frequente è “th”
* Il trigramma più frequente è “the”

Frequenza lettere inglesi

Le possibilità potrebbero essere:

- Concentrarsi sulle 4 vocali (e, a, o, i) e 5 consonanti più frequenti (t, n, s, h, r) consente di scoprire circa il 70% del contenuto di un messaggio.

- Analogamente, concentrarsi sulle 4 lettere meno frequenti (j, q, x, z) consente di eliminare molte combinazioni.

- CAVEAT: per testi brevi, l’analisi di frequenza può non funzionare perfettamente e qualche iterazione è necessaria; per testi molto brevi, non funziona.

Esempio: decifrare il seguente testo cifrato (intercettato durante la guerra del Vietnam) mediante “crittoanalisi delle frequenze”:

UZQSOVUOHXMOPVGPOZPEVSGZWSZOPFPESXUDBMETSXAIZVUEPHZHMDZSHZOWSFPAPPDTSVPQUWYMXUZUHSXEPYEPOPDZSZUFPOMBZWPFUPZHMDJUDTMOHMQ

Contare le frequenze relative del testo cifrato:

- P 13.33

- Z 11.67

- S,U 8.33

- O 7.50

- M 6.67

- H 5.83

- D, E 5.00

- V, X 4.17

- F, W 3.33

- Q, T 2.50

- A, B, G, Y 1.67

- I, J 0.83

- C, K, L, N, R 0.00 (non compaiono)

Dal confronto delle occorrenze calcolate con la tabella, si deduce che:

- P 13.33 e 12.70 pertanto P → e

- Z 11.67 t 9.05 pertanto Z → t

- Le lettere {S, U, O, M, H} hanno frequenza relativa alta e potrebbero appartenere all’insieme {a, h, i, n, o, r, s}

- Le lettere {A, B, G, Y, I, J} hanno frequenza relativa bassa e potrebbero appartenere all’insieme {b, j, k, q, v, x, z}

A questo punto, si può procedere in vari modi: verificando i bigrammi, trigrammi, per tentativi

Dai primi, si deduce che Wh in quanto ZW è il bigramma più frequente nel testo cifrato (compare 3 volte)

Verificando la sequenza ZWSZ, non si sa se forma una parola, ma se lo fosse e avendo “th\_t” si può ipotizzare che S→ a (tentativo supportato dalla frequenza).

Abbiamo ottenuto: P → e , S → a, W → h, Z → t, ovvero:

UtQaOVUOHXMOeVGeOteEVaG that OeFeEaXUDBMETaXAIt

VUEeHtHMDtaHtOhaFeAeeDTaVeQUthYMXUtUHaX

EPYEeOeDtatUFeOMBtheFUetHMDJUDTMOHMQ

Possibile obiettivo: individuare le altre vocali “i”, “o” che dall’analisi di frequenza dovrebbero essere in {S, U, O, M, H}, infatti: U→i, M→o

it QaOViOHXoOeVGeOteEVaG that OeFeEaXiDBoETaXAIt

ViEeHtHoDtaHtOhaFeAeeDTaVeQithYoXUtiHaX

EPYEeOeDtatiFeOoBtheFietHoDJUDToOHoQ

Con ulteriori tentativi sulla struttura della frase (soggetto-verbo) e statistiche, si arriva a Q→w, O→s:

it was VisHXoseVGesteEVaG that seFeEaXiDBoETaXAIt

ViEeHtHoDtaHtshaFeAeeDTaVewithYoXUtiHaX

EPYEeseDtatiFesoBtheFietHoDJUDTosHow

Sapendo che era un messaggio USA durante la guerra del Vietnam, si può ipotizzare altro? [Consiglio: guardare in fondo]

it was VisHXoseVGesteEVaG that seFeEaXiDBoETaXAIt

ViEeHtHoDtaHtshaFeAeeDTaVewithYoXUtiHaX

EPYEeseDtatiFesoB the FietHoDJUDTosHow

Con ulteriori tentativi, si arriva al testo crittoanalizzato:

it was disclosed yesterday that several informal but direct contacts have been made with political representatives of the vietcong in moscow

Esempio (alfabeto italiano):

Le vocali E, A, O, I sono le lettere più frequenti (arrivano a circa il 45% del testo).

Seguono le consonanti: L, N, R, S, T, C

Tra i bigrammi più frequenti vi è QU seguito da vocale, ma anche MP, ST, e le doppie CC, RR, TT, SS, LL.

Trigrammi frequenti sono anche: CHE, CHI, GHE, GHI, STR.

Due lettere uguali sono necessariamente consonanti doppie, che sono sicuramente precedute e seguite da vocali.

Decifrare il seguente testo italiano cifrato mediante “crittoanalisi delle frequenze”:

GNDASBLLNIOEPCBCBVBUBTNTBOBSLESZLELSNAZMDZBSNPCBSZBSDBIZDDNQGNPZIBGELLNAZGZLEZDSBHHEGLEUZIBBSQAPZBDVEFBHZEZDPQEDEUBTBGLZBDQZNUZPPNMSBGSNONSIZDDNNQGBGELLNTZCEGBSSBLEDNABTEDNUNZLNIOZOBPPBLZNDNDNMMNGUNUNMDZBGLZHCZSNOEPPENPPNSNHEPZBSUZLBUBHCZNUNSNQGABTESNBTEPLSBIBNPLB

Decifrare il seguente testo italiano cifrato mediante “crittoanalisi delle frequenze”:

GND ASBLLNIOE PCBCBVBU BTNTB OBSLESZLE LSN AZMDZ B SN PCBSZBS. DB IZDDNQGNPZIB GELLN, AZGZLE ZD SBHHEGLE UZ IBBSQA, PZ BDVE, FBHZE ZD PQEDE UBTBGLZ B DQZ N UZPPN: “MSBG SN, ONS IZDDN N QGB GELLN TZ CE GBSSBLE DN ABTEDN UNZ LNIOZ OBPPBLZ N DN DNMMNGUN UNMDZ BGLZHCZ SN. OEPPE NPPNSN HEPZ BSUZLB UB HCZNUNSN QG ABTESN B TEPLSB IBNPLB?”

Soluzione: Nel frattempo Sharazad aveva partorito tre figli a re Shariar. La milleunesima notte, finito il racconto di Maaruf, si alzò, baciò il suolo davanti a lui, e disse: «Gran re, per mille e una notte ti ho narrato le favole dei tempi passati e le leggende degli antichi re. Posso essere così ardita da chiedere un favore a vostra maestà?»

**Cifrari misti e poli-alfabetici**

Soluzioni alla crittoanalisi per frequenza:

* Nomenclatore
* Cifrari omofonici
* Cifrari polialfabetici
  + Cifrario di Leon Battista Alberti
  + Cifrario di Vigenère
* Cifrari poligrafici
  + Cifrario di Playfair (bigramma)
  + La grand ciffre dei Rossignols (trigramma+nomenclatore)

Nomenclatore: Esempio nella congiura della Regina Maria Stuarda di Scozia → le sue lettere erano cifrate con un algoritmo che combinava cifratura e codifica: non solo le lettere venivano permutate, ma venivano usati dei simboli unici per riferirsi a certe parole d’uso comune (finì male…). I nomenclatori furono usati per secoli con un numero di simboli sempre maggiore → Problemi di gestione: rischio di furto, evidenza del ruolo della persona trovata in possesso di un nomenclatore, modifica, …

Cifratore per sostituzione omofonica: Si utilizza per la stessa lettera un numero di valori proporzionale alle sua frequenza: ad esempio, per “a” che ha una frequenza 11% si utilizzano 11 valori.

- Si annulla la debolezza principale relativa alla frequenza

- Non si annulla la debolezza dell’identità delle lettere

‟La grand ciffre” (1626): Ideato dai Rossignols (padre e figlio) ai tempi di Luigi XIV (“Re Sole”) e poi utilizzato per 200 anni. Usava un libro codice con un simbolo (numero di 3 cifre) per ogni lettera frequente e sillaba della lingua francese (587 numeri) e una cifratura dei trigrammi più frequenti della lingua francese.

Cifrario Playfair (1854): Cifra bigrammi invece di lettere singole come nel caso del cifrario mono-alfabetico. Più sicuro rispetto alla crittoanalisi per frequenza in quanto ci sono 26\*26 bigrammi invece di 26 lettere. Ritenuto altamente sicuro e utilizzato dall’esercito inglese fino alla 1ª Guerra Mondiale e, in qualche caso (non top-secret), anche nella 2ª Guerra. In seguito, con l’avvento dei computer si è dimostrato meno sicuro di quanto ipotizzato.

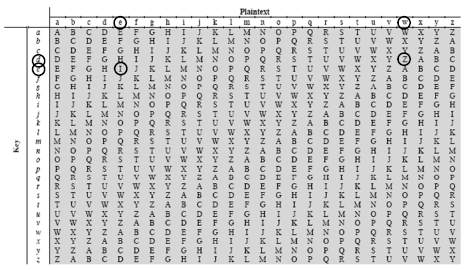
Algoritmi polialfabetici: Utilizzano più alfabeti di sostituzione

Esempio: Utilizzo di 4 alfabeti: A1, A2, A3, A4

La prima lettera viene cifrata con la prima lettera di A1, la seconda con la prima lettera di A2 e così via. La quinta, di nuovo con A1. Riesce ad eliminare l’evidenza delle frequenze caratteristiche delle lettere più popolari.

→ Leon Battista Alberti (1467): uso di due alfabeti

Cifrario di Vigenère (1586): In pratica, è un cifrario di Cesare multiplo.

Si utilizza una chiave di d lettere: K = k1 k2 ... kd dove l’i-ma lettera specifica l’i-mo alfabeto da utilizzare. Si utilizza un alfabeto per ogni lettera da cifrare ripetendo la chiave se (come tipicamente accade) il testo è più lungo di d lettere. Per semplificare le operazioni di cifratura/decifratura si può utilizzare la cosiddetta Tavola di Vigenère. →

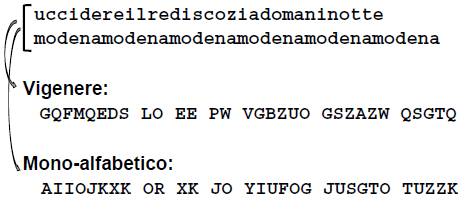
Esempio, Chiave: deceptive

Chiave: deceptivedeceptivedeceptive

Testo: wearediscoveredsaveyourself

Critto: ZICVTWQNGRZGVTWAVZHCQYGLMGJ

Confronto:



La robustezza di Vigenere rispetto alla crittoanalisi per frequenza è evidente anche da una semplice ispezione dei due messaggi.

Cifrari polialfabetici: Il cifrario di Vigenère è stato considerato sicuro per più di tre secoli (anche se non molto adottato per la lentezza di gestione). Verso la metà del 1800, Babbage e Kasiski individuarono il punto debole: “puntare prima a identificare la lunghezza della chiave”.

Principio: i simboli a distanza pari alla lunghezza della chiave sono cifrati con lo stesso alfabeto. In pratica, una lunghezza di chiave 9 fa sì che le lettere 1, 10, 19, 28 siano cifrate con lo stesso alfabeto. E’, infatti, poco probabile che queste ripetizioni siano dovute alla semplice casualità.

Passi di crittoanalisi:

1. Identificare la lunghezza della chiave L
2. Suddividere il testo cifrato in L sottotesti
3. Applicare un algoritmo di crittoanalisi per frequenza a ogni sottotesto

1. Calcolare la lunghezza della chiave:

*Primo metodo*

Si scrive il testo cifrato in orizzontale

Si scrive lo stesso testo shiftato di 2, 3, …, m posizioni, dove m è la massima lunghezza ipotizzata della chiave

V V H Q R F A A L H J M N B B V G …

V V H Q R F A A L H J M N B B V G …

V V H Q R F A A L H J M N B B V G …

Si segna un \* ogni volta che una lettera sulla prima striscia è uguale a quella della seconda striscia e così via per spostamenti successivi.

Si ottengono i seguenti dati:

- spostamenti: 2 3 4 5 6 7

- corrispondenze: 14 14 16 14 24 12 → ipotesi: lungh. chiave=6

*Metodo di Kasiski*

Ricercare sequenze di caratteri identici nel testo cifrato → significa che con alta probabilità corrispondono allo stesso testo in chiaro.

Calcolare la distanza t tra gli inizi di queste sequenze: allora t è un multiplo della lunghezza L della chiave

Identificare tutte le sequenze identiche nel testo cifrato e calcolare le varie distanze: t1, t2, …, tk

Lunghezza chiave: L = MCD(t1, t2, …, tk) ovvero L = n MCD()

Esempio (1):

RUNRUNRUNRUNRUNRUNRUNRUNRUNRUNRUNRUNRUN

tobeornottobethatisthequestionwhetheris



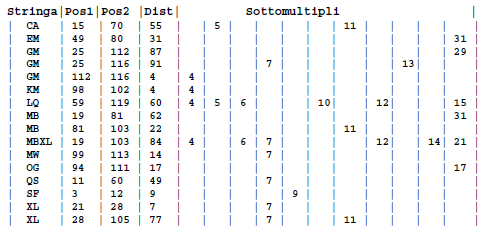
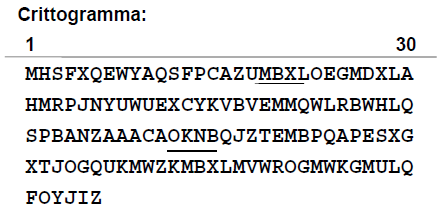
Ogni volta che la stringa RUNR cifra il testo to be si ottiene lo stesso testo cifrato KIOV.

Ogni volta che la stringa UN cifra il testo th si ottiene lo stesso testo cifrato NU

L = MCD(9, 6)= 3 è un multiplo della chiave

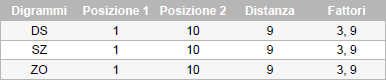
(in questo caso semplice, 3 corrisponde alla lunghezza della chiave)

Esempio (2):



Testo cifrato: DSZOSPXSRDSZO

Analisi del testo cifrato alla ricerca dei digrammi:



Lunghezza chiave: 3

Sotto-testi:

1. DOXDO

2. SSSS

3. ZPRZ

Crittoanalisi di frequenza per ciascun sotto-testo

1. DOXDO

2. SSSS

3. ZPRZ

(Impossibile da svolgere in quanto i testi sono troppo brevi)

Chiave: KEY

Testo in chiaro: tobeornottobe

XOR (eXclusive OR) binario

L’operatore OR restituisce il risultato VERO quando almeno uno dei due operandi è vero (in altre parole restituisce FALSO solo quando entrambi sono falsi). L’operatore XOR restituisce il risultato VERO se, e solo se, uno solo dei suoi operandi è VERO (in altre parole restituisce FALSO sia quando entrambi gli operandi sono falsi sia quando sono entrambi veri).

0 ⊕ 0 → 0

0 ⊕ 1 → 1

1 ⊕ 0 → 1

1 ⊕ 1 → 0

E’, di fatto, equivalente a una somma modulo 2.

Il simbolo di operatore XOR in crittografia viene applicato anche in contesti non binari come SOMMA MODULARE. Esempio:

S ⊕ I → A equivale a: (18 + 8) mod 26 = 0

E ⊕ H → L equivale a: (4 + 7) mod 26 = 11

Esempio:

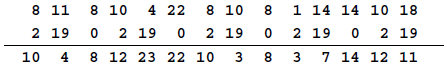
Testo in chiaro: ilikewikibooks

Chiave: cta

Convertire testo (lettere): 8 11 8 10 4 22 8 10 8 1 14 14 10 18

Convertire la chiave: 2 19 0

XOR (mod 26)



Il risultato è convertito nel testo cifrato: keimxwkdiehoml

Cifrario auto-chiave

Idealmente, per avere un crittogramma sicuro, bisogna avere una chiave lunga quanto il testo da cifrare. Sempre Vigenère propose il cifrario auto-chiave (a chiave auto-generatrice) → la chiave scelta è concatenata con il testo da cifrare per avere una chiave di esecuzione sufficientemente lunga e non ripetuta.

Esempio: Chiave deceptive

Chiave: deceptivewearediscoveredsav

Testo: wearediscoveredsaveyourself

Crittogr.: ZICVTWQNGKZEIIGASXSTSLVVWLA

Anche questo cifrario è soggetto a caratteristiche di frequenza che consentono la crittoanalisi statistica

**OTP**

OTP: unico strumento di cifratura incondizionatamente sicuro.

Uso del “blocco usa e getta” (one-time pad) → chiave lunga quanto il messaggio (quindi, mancanza di ricorrenze). Idea originaria:

– Ogni pagina del blocco contiene una lista di numeri casuali (es., chiavi di 30 caratteri)

– Per cifrare un messaggio, il mittente deve staccare tanti fogli quanti ne sono necessari a cifrare tutto il testo (es., 11 fogli per un messaggio di 330 caratteri)

Mittente e destinatario devono avere una copia identica del blocco e usarlo in modo perfettamente sincronizzato.

Teoricamente, ogni sistema basato su ripetizione è critto-analizzabile → One Time Pad non usa ripetizione.

Poiché ogni lettera del messaggio originale è accoppiata a un numero casuale differente, il testo cifrato risultante è incondizionatamente sicuro rispetto a attacchi linguistici, di ricorrenza o di dizionario.

OTP è incondizionatamente sicuro ammesso che:

* Il PAD non è mai riutilizzato
* I numeri pseudo-casuali sono statisticamente impredicibili

Libro come fonte di one-time pad: un libro è un fonte di “simboli casuali” sufficientemente lunga per implementare nella pratica la tecnica del blocco “usa e getta”. Mittente e destinatario devono mettersi d’accordo su: libro, pagine, righe, frequenza lettera.

Alla base di vari romanzi ad esempio ne “Il codice Rebecca” (Ken Follett) si usava il libro di Daphne du Maurier (“Rebecca”) come fonte delle chiavi OTP per le spie della II Guerra Mondiale.

→ OTP e superpotenze:

*Unione Sovietica*:

Ha utilizzato OTP per proteggere tutte le ambasciate, polizia segreta, esercito, governo dal 1930 in poi. Tutti hanno le loro chiavi e le comunicazioni intercettate sono sempre gruppi di 5 cifre.

*Stati Uniti*:

Sigsaly → Sistema di comunicazione telefonico (soprannome, The Green Hornet) utilizzato dal Presidente e alti generali per comunicare con gli Alleati.

Sigtot → Il sistema produceva due nastri (identici) perforati in modo del tutto casuale. I nastri contenevano la chiave costituita da 5 cifre parallele che venivano mescolate (XOR) con il Testo per produrre il crittogramma. I nastri venivano distrutti dopo l’utilizzo.

Problemi dell’OTP: la sicurezza dell’OTP si basa sulla effettiva casualità della chiave: se la sequenza è veramente casuale, il crittogramma non è soggetto ad attacchi, altrimenti …

Quindi, ci sono vari problemi di realizzazione:

* Il problema di generare enormi quantità di chiavi effettivamente casuali (possono essere necessarie sequenze di milioni di caratteri al giorno)
* E’ estremamente complesso riuscire a distribuire a tutti gli interlocutori blocchi sufficientemente lunghi di chiavi per trasmettere tutti i messaggi necessari
* Vi è, infine, il problema della protezione del “pad” da copie e furti

Mai riutilizzare il pad: si ipotizzi di utilizzare due volte lo stesso pad K:

C1=M1 ⊕ K

C2=M2 ⊕ K

Se il crittoanalista intercetta i messaggi crittografati con lo stesso pad, può sfruttare la seguente proprietà:

C1 ⊕ C2 = (M1 ⊕ K) ⊕ (M2 ⊕ K) = M1 ⊕ M2

Mediante tecniche di crittoanalisi statistica, se i messaggi sono sufficientemente lunghi, è possibile ricavare M1 e M2

Progetto Venoma della CIA: intercettazione di enormi quantità di messaggi cifrati dell’URSS, con l’auspicio

che si riutilizzassero i pad. Cosa che era stata fatta …

I cifrari tendono a ricreare approssimazioni dell’OTP, inizialmente in modo meccanico, poi in modo elettronico. Stare attenti ai “snake oil algorithm” (cirari commerciali) che dichiarano di essere non crittoanalizzabili, ma la cui robustezza non è dimostrata.

Con i computer si possono generare “blocchi” di chiavi molto lunghi mediante generazione di numeri pseudo-casuali (es: funzione random()).

Cosa significa pseudo?

* La generazione è algoritmica
* Serve un numero iniziale: seme

E’ possibile implementare un cifrario OTP in modo relativamente semplice, ma … generazione algoritmica → non è un OTP con caratteristiche di inviolabilità!

“Pseudo-OTP guarantees pseudo-security”

**Cifrari per trasposizione**

L’obiettivo della sostituzione è la confusione (nascondere le lettere originali) per rendere difficile il compito di determinare come un messaggio viene trasformato in un testo cifrato.

Con la trasposizione, la crittografia mira alla diffusione, spargendo le informazioni del messaggio nel testo cifrato, in pratica, le trasposizioni cercano di spezzare gli schemi stabiliti, riorganizzando i simboli di un messaggio cifrato, allo scopo di evitare la crittoanalisi delle adiacenze

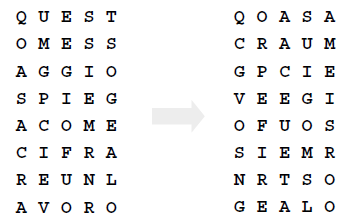
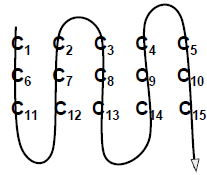
Esempio storico: Scitale

Cilindro, tipicamente di legno, usato inizialmente dagli spartani per scambiarsi messaggi segreti. Sullo citale veniva avvolta una striscia (di cuoio o stoffa) su cui si scriveva il messaggio in orizzontale. Il crittogramma si otteneva svolgendo la striscia. Il crittogramma poteva essere interpretato solo se il destinatario avesse posseduto uno scitale uguale a quello di partenza.

Trasposizione colonnare:

Si scrivono i caratteri di un testo in chiaro su righe di cinque caratteri, una sotto l’altra.

Il testo cifrato viene ottenuto leggendo le colonne dall’alto verso il basso.



**Cifrari prodotto**

E' una combinazione di tecniche. La sostituzione e la trasposizione sono le pietre angolari della crittografia. Tuttavia, la crittografia degli ultimi decenni si basa su una combinazione di più tecniche (es: Crittografia sovietica nella II Guerra Mondiale: Alfabeto sostitutivo + One-time Pad). Una combinazione di due o più cifrature è detta *cifratura prodotto.*

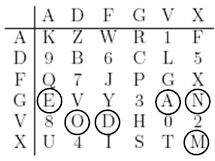
* Con due sostituzioni si ottiene un cifrario per sostituzione più complesso
* Con due trasposizioni si ottiene un cifrario per trasposizioni più complesso
* Ma con una sostituzione seguita da una trasposizione si ottiene un cifrario completamente nuovo e molto più robusto

→ Questa combinazione rappresenta il ponte tra la crittografia classica e quella moderna.

Esercito tedesco: 1a Guerra Mondiale Algoritmo “ADFGVX”

Esempio di algoritmo “prodotto” che combina sostituzione e trasposizione

1. Fase di sostituzione: a ciascun elemento della Tabella (sempre uguale, da cui il nome ADFGVX) si fa corrispondere una coppia di lettere:

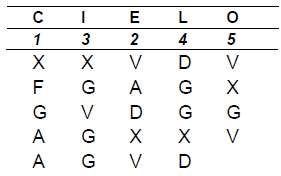
Es: testo in chiaro: "modena centro"



XX VD VF GA GX GV DG GA GX XV AG VD

2. Fase di trasposizione (1): basata su chiave

Esempio di chiave (parola/e con lettere diverse): CIELO → da cui si ricava la lunghezza (in colonne) del blocco e l’ordinamento (lessicografico) delle colonne



2. Fase di trasposizione (2): si scrivono le lettere in colonna, seguendo la numerazione determinata dallo ordinamento lessicografico delle lettere della chiave

Testo cifrato: XFGAA VADXV XGVGG DGGXD VXGV ( = modena centro)

L’algoritmo “ADFGVX” era molto più valido dei precedenti , ma ha margini di miglioramento in quanto:

* effettua un solo passo di sostituzione e di trasposizione
* la tabella di sostituzione è fissata
* la tabella di sostituzione non dipende dalla chiave

Telegramma di Zimmermann (1a Guerra Mondiale): La 1a guerra mondiale è la prima dopo l'invenzione del telefono e della radio. Questi mezzi di comunicazione sono croce e delizia dei militari perché consentono comunicazioni a velocità prima impensabile, ma sono esposti all’intercettazione (mentre catturare un corriere era impresa occasionale). Il più noto dei messaggi intercettati fu il “telegramma Zimmermann” con il quale i

Tedeschi offrivano un’alleanza ai Messicani in chiave anti-USA. Questo messaggio fu uno dei fattori che spinsero gli USA a entrare in guerra nel 1917. La 1a guerra mondiale chiarì a tutti gli Stati la fondamentale importanza della crittologia nelle trasmissioni.

Messaggi non scritti: code talkers

Nel momento in cui si diffondono i nuovi mezzi di comunicazione a distanza, un’alternativa alla crittografia nello scambio di messaggi tattici è stata l’utilizzo dei code talkers. Gli Stati Uniti ne hanno fatto grande uso, utilizzando soprattutto:

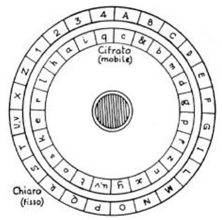
– Chactow (nella I Guerra Mondiale)

– Navajo (nella II Guerra Mondiale e nella guerra di Corea)

Esempio: Navajo (US Army, II Guerra Mondiale) Guadalcanal, Tarawa, Peleliu, Iwo Jima: utilizzati soprattutto nelle guerre dei Marines nel Pacifico dal 1942 al 1945.

**L'epoca delle macchine**

Disco cifrante di Alberti (XV secolo)

Dispositivo costituito da due dischi di rame concentrici di diametro diverso, infilati su un perno, sui cui lati è riportato un alfabeto.

Il disco esterno riproduce l’alfabeto nell’ordine consueto ed è fisso.

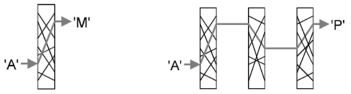
Il disco interno riproduce un alfabeto in ordine casuale (alfabeto cifrante).

Per effettuare la cifratura, il disco mobile viene fissato su una data lettera (chiave di messaggio): si ottiene così una corrispondenza ”lettera in chiaro (disco esterno) – lettera cifrata (disco interno).

**Macchine cifranti** (1908-1945)

Una volta compresi i vantaggi della cifratura prodotto, vi era il problema di automatizzare multipli step di sostituzione e trasposizione che sarebbero stati molto lenti e complessi da realizzare a mano correttamente. → Realizzazione di *rotor machines* e poi concatenati

→ CHIAVE: posizone iniziale dei rotori



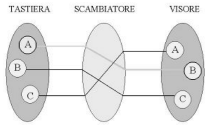
Due rotori: periodo di 26x26 = 676

Tre rotori: periodo di 26x26x26 = 17576

Poco prima dei computer c'erano: ENIGMA (Germania) macchina cifrante più famosa; ma anche altri stati ne avevano una, Converter M-209-b (US), Type X (UK), Purple (Giappone).

**Enigma**, la cifrante di Scherbius:

quando un operatore utilizzava Enigma, digitava le lettere che costituivano il messaggio sulla tastiera della macchina e i meccanismi interni della stessa trasformavano quel testo in un altro cifrato (un visore con varie lampadine indicava la lettera da inserire nel testo cifrato).



Lo scambiatore è l’elemento più importante della macchina Enigma. Costituito da un disco di gomma attraversato da una rete di fili provenienti dalla tastiera. Questi fili entrano nello scambiatore e, dopo un percorso formato da vari incroci, emergono dalla parte opposta. Lo schema interno del rotore determina un alfabeto cifrante utilizzabile per una semplice cifratura a sostituzione monoalfabetica.

Il passo successivo prevedeva di far ruotare il disco dello scambiatore di un ventiseiesimo di giro dopo la cifratura di ogni lettera. In questo modo l’alfabeto cifrante cambiava dopo ogni lettera: si passò così dalla cifratura monoalfabetica a quella polialfabetica. Tuttavia, il meccanismo presentava il problema della ripetizione (con conseguente rischio di cifratura debole).

Pertanto vennero introdotti un secondo e un terzo scambiatore:

- Il secondo compie una rotazione parziale soltanto dopo che il primo rotore ha compiuto un intero giro

- Il terzo fa lo stesso rispetto al secondo rotore

Per migliorarlo ulteriormente inserirono dei rotori rimovibili e lo Stecker: pannello a prese multiple posto tra tastiera e primo rotore; permetteva di inserire cavi muniti di spinotti, che avevano l’effetto di scambiare due lettere prima della loro immissione nel rotore.

Combinando tutti gli elementi, si può calcolare il numero di chiavi che Enigma poteva impiegare:

- posizionamento di 3 rotori scelti su 5: 60

- orientamenti dei rotori: 26x26x26 = 17.576

- possibili abbinamenti di 12 lettere su 26 dovuti allo stecker: 100.391.791.500

TOTALE: 105.869.167.644.240.000 (sarebbe stata inviolabile se usata appropriatamente)

Crittoanalisi di Enigma: poca attenzione iniziale da parte degli Inglesi, forti della loro superiorità militare, ai messaggi cifrati. Le prime intercettazioni di messaggi cifrati da Enigma furono effettuate in Polonia, che si sentiva minacciata dalla Germania (anni ’30) e anche dalla Russia. Fu potenziato il lavoro di intercettazione e raccolta dei crittogrammi Enigma da parte dell’Ufficio Cifre polacco, che reclutò molti matematici, oltre che linguisti, tra cui Marian Rejewski.

Egli basò la sua strategia sul fatto che la ripetizione (come noto) è nemica della sicurezza, perché crea degli schemi: la più ovvia ripetizione di Enigma era quella della chiave di messaggio, cifrata e trasmessa due volte di seguito all’inizio di ogni testo trasmesso. Rejewski sfruttò il legame presente con l’assetto della chiave giornaliera usata per regolare il nuovo assetto tra la I e la IV lettera, tra la II e la V, tra III e la VI.

Altri errori umani: riutilizzo di chiavi (già violate), utilizzo di chiavi per lunghi periodi, messaggi di saluto e chiusura standard, … [SONO LEZIONI ANCHE PER LA CRITTOGRAFIA MODERNA]

Intercettando moltissimi messaggi, Rejewski riuscì a risalire a delle concatenazioni tra le lettere. Capì, inoltre, che il numero dei collegamenti dipendeva solo dagli scambiatori e non dal pannello a prese multiple stecker (usato raramente): in questo modo vennero escluse molte combinazioni di chiavi da provare per decrittare il messaggio → era sufficiente controllare le 17576 posizioni dei rotori.

Per velocizzare il procedimento, Rejewski progettò congegni meccanici detti bombe. Poco prima di essere invasa, la Polonia passò tutta la documentazione e le relative tecnologie agli inglesi.

Quelli di Bletchley park, sede della principale unità di crittoanalisi della Gran Bretagna durante la 2a Guerra mondiale, (ri)costruirono la “Bomba” (precedentemente polacca) per decifrare i crittogrammi di Enigma. Mediante concorso assunsero, tra gli altri, Alan Turing (lo scienziato della “macchina di Turing”, del “test di Turing”, ecc.): uno dei grandi padri nobili dell’informatica.

Il lavoro di crittoanalisi di Alan Turing: struttura rigida di alcuni messaggi intercettati, esempio: bollettini metereologici, Mein Fuehrer.

Metodo intuitivo per l’elaborazione di un’ipotesi detta crib.

Confronto crib-crittogramma: corrispondenza lettera in chiaro-lettera cifrata.

Associazione di un numero a ogni coppia di lettere rappresentante uno dei possibili stati della macchina che permetteva di ottenere quella cifratura.

Esempio:



Individuazione di percorsi chiusi di lettere, detti loop, sfruttando le corrispondenze trovate. Ad esempio:

E → 5 → T → 3 → I → 21 → A → 23 → E

Possibile rappresentazione di questi loop con dei circuiti elettrici e successiva meccanizzazione della ricerca di ordine e posizione corrette dei rotori.

Costruzione della “Bomba”: macchina Enigma aperta

Necessità di collegare più macchine Enigma insieme: modifica della macchina originale, in modo che l’output di una costituisca l’input di un’altra. Ogni colonna di ogni batteria costituiva una macchina Enigma aperta.

“Colossus”: il primo computer

Il Colossus è stato il primo computer in grado di forzare i codici sviluppati dalla macchina Enigma. Realizzato grazie alle precedenti esperienze della Bomba (macchina elettromeccanica). Colossus è stato il rimo ad usare le valvole termoioniche (ben 1500) al posto dei relè, aumentando di molto la potenza di calcolo. Il prototipo, Colossus Mark I, venne assemblato a Bletchley Park nel febbraio del 1944. Il migliorato Colossus Mark II fu installato nel giugno del 1944 (subito prima dello sbarco in Normandia). Altri dieci Colossus furono costruiti prima della fine della guerra e poi distrutti.

Colossus fu sviluppato per violare “con forza bruta” i messaggi codificati con la cifratura Lorenz SZ 40/42 (detta Tunny), più robusta di Enigma, usata solo nelle comunicazioni tra Hitler e i capi di Stato maggiore. Colossus confrontava continuamente due flussi di dati mediante operazioni booleane: il primo flusso era rappresentato dal crittogramma letto ad alta velocità da una striscia di carta; l’altro flusso era generato internamente e corrispondeva a una simulazione elettronica della macchina di Lorenz con una certa configurazione. Dopo si cambiava configurazione fino a quando il confronto non superava una determinata soglia. In tal caso, l’output veniva stampato mediante una macchina da scrivere elettrica.

Alcuni successi storici:

Battaglia di Capo Matapan (1941): la decrittazione da parte degli Inglesi di alcuni messaggi cifrati della marina tedesca circa l’esatta posizione della flotta italiana, ne causò la disfatta.

Sbarco in Normandia (1944): gli Americani, in grado di leggere gran parte dei messaggi degli alti comandi tedeschi, ebbero conferma che Hitler aveva creduto alla falsa notizia di uno sbarco alleato nei pressi di Calais. Organizzarono così lo sbarco in Normandia sicuri di incontrare una resistenza più limitata.

Turing nonostante tutto viene incarcerato nel 1952 per omosessualità, condannato a subire pesanti dosaggi ormonali che ebbero effetti devastanti sul suo fisico. Si uccise nel 1954 mordendo una mela avvelenata con cianuro di potassio. Il 24 dicembre 2013, dopo una petizione sottoscritta da molti scienziati, la regina Elisabetta II elargì la grazia postuma per Alan Turing. Scelta del logo della Apple: tra realtà e leggenda.

