Sommario

- Crittografia Classica
  - Cifrari di Cesare
  - Cifrari Vigènere
  - Esempi di crittoanalisi
  - Il cifrario perfetto (one time pad)

Metodi di cifratura

- La tecnologia fondamentale su cui si basano tutte le applicazioni per la sicurezza delle reti e dei computer è la **crittografia**
- I due approcci principali sono:
  - **Cifratura convenzionale**, anche detta "cifratura simmetrica"
  - **Public-key Encryption**, detta anche "cifratura asimmetrica"

Crittografia moderna

- **1977:** ***Data Encryption Standard (DES)*** adottato dal U.S. Federal Information Processing per cifrare informazioni "unclassified"
- **1976:** ***Diffie ed Hellman,*** introdussero il rivoluzionario concetto della crittografia a chiave pubblica. La sicurezza si basa sull'impossibilità di risolvere problemi basati sul "logaritmo discreto"
- **1978:** ***Rivest, Shamir, Adleman*** **(RSA)**, forse lo schema più conosciuto: la sicurezza è basata sull'intrattabilità della fattorizzazione di interi molto grandi

Breakable Encryption

- E' possibile avendo tempo e risorse
- Forza bruta di solito impraticabile
- Le stime sono basate sulla tecnologia attuale
- Semplicemente perché uno schema è basato su un problema difficile, non è detto che un crittoanalista provi a risolverlo attaccando il problema

Algoritmo di Encryption

- Trasformazione:

  $C=E(P)$

  $P=D(C)$

  $P=D(E(P))$

- La sicurezza dell'algoritmo risiede nella segretezza delle chiavi (l'algoritmo può essere noto):

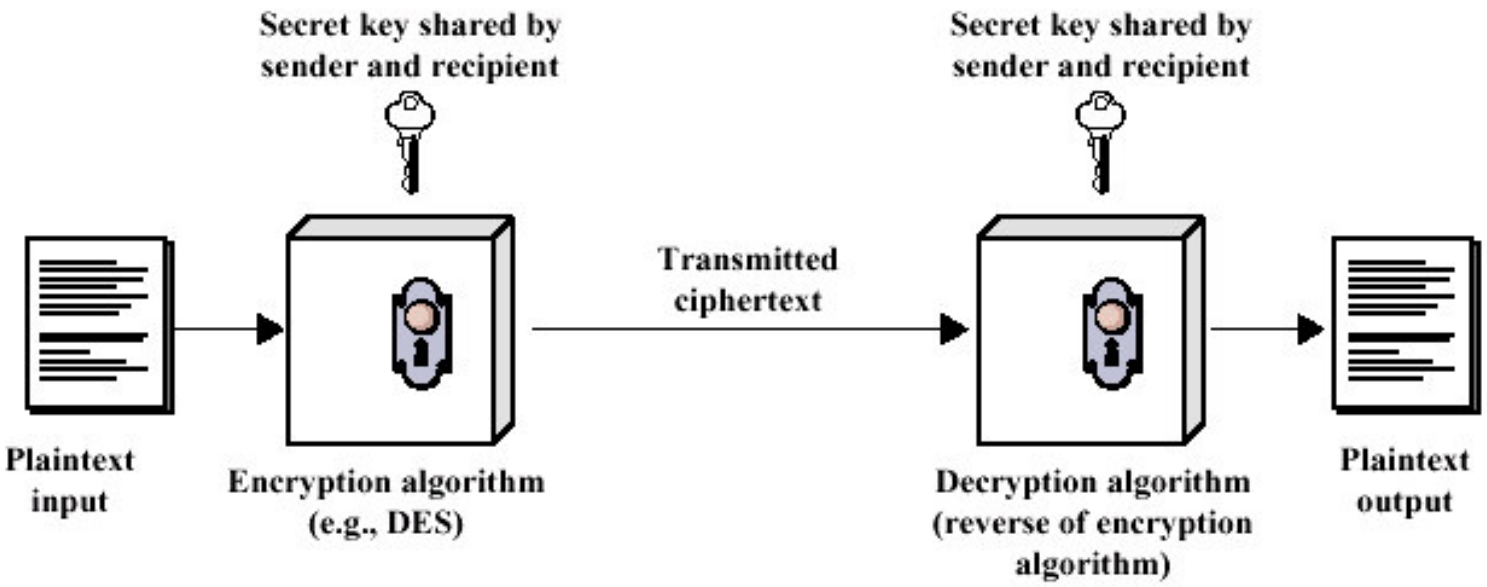
  Symmetric: $C=E(k,P)$ $P=D(k,C)$

  Asymmetric: $C=E(k_1,P)$ $P=D(k_2,C)$

Conventional Encryption

- L'unica forma di encryption precedente al 1976
- Algoritmo composto da 5 elementi
  - **Plaintext**: Il messaggio (i dati) originale
  - **Encryption algorithm**: Applica diverse sostituzioni e trasformazioni al plaintext
  - **Secret key**: Input per l'algoritmo. Le sostituzioni e le trasformazioni dipendono da questa chiave
  - **Ciphertext:** Messaggio incomprensibile prodotto in output. Dipende dal plaintext e dalla secret key
  - **Decryption algorithm**: Algoritmo di encryption eseguito all'inverso. Usa il ciphertext la secret key per produrre il plaintext originale

Modello del Conventional Encryption



Requisiti & debolezze

- Requisiti
  - Un **robusto** algoritmo di encryption
  - Un **processo sicuro** per il mittente ed il destinatario di **ottenere le secret keys**
- Tipi di attacco
  - **Cryptanalysis** (crittanalisi)
  - **Forza bruta**

Tipi di cifrari

- **Cifrari a sostituzione** – rimpiazzano bit, caratteri o blocchi di caratteri con dei sostituti
- **Cifrari a trasposizione** – riordinano i bit o i caratteri nei dati

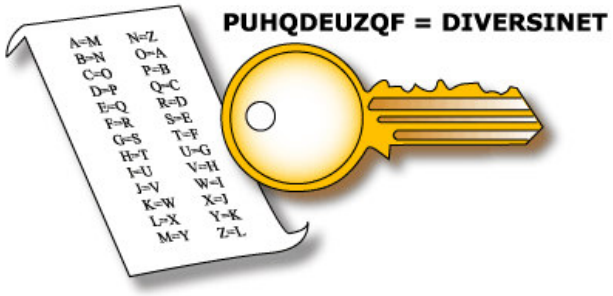
Strumenti del Cryptanalyst

- Dati sulla frequenza delle lettere
- Liste di prefissi e suffissi
- Liste di coppie e triple di lettere
- Liste di campioni comuni

Cifrario di Cesare

plain: **abcdefghijklmnopqrstuvwxyz**

key: **defghijklmnopqrstuvwxyzabc**

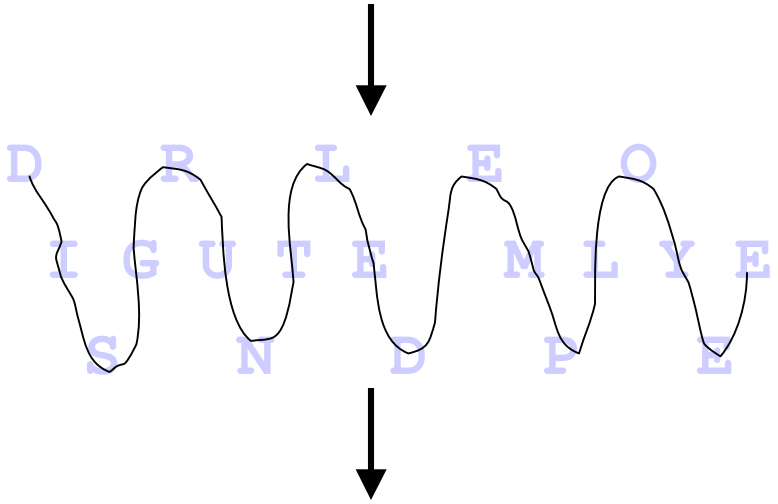


plain: **MEET ME AFTER THE TOGA PARTY**

cipher: **PHHW PH DIWHU WKH WRJD SDUWB**

Cifrario "Rail-Fence" (staccionata)

DISGRUNTLED EMPLOYEE



DRLEOIGUTE MLYESNDPE

Cryptosystem

- Quintupla $(\mathcal{E}, \mathcal{D}, \mathcal{M}, \mathcal{K}, C)$
  - $\mathcal{M}$ insieme di plaintexts
  - $\mathcal{K}$ insieme delle key
  - C insieme dei ciphertexts
  - $\mathcal{E}$ insieme delle funzioni di encryption
    $e: \mathcal{M} \times \mathcal{K} \rightarrow C$
  - $\mathcal{D}$ insieme delle funzioni di decryption
    $d: C \times \mathcal{K} \rightarrow \mathcal{M}$

Esempio

- Esempio: Cifrario di Cesare
  - $\mathcal{M}$ = { sequenze di lettere }
  - $\mathcal{K}$ = { i | i è un intero e $0 \leq i \leq 25$ }
  - $\mathcal{E}$ = { E_k | $k \in \mathcal{K}$ e per ogni lettera m, $E_k(m) = (m + k) \bmod 26$ }
  - $\mathcal{D}$ = { D_k | $k \in \mathcal{K}$ e per ogni lettera c, $D_k(c) = (c - k) \bmod 26$ }
  - $C = \mathcal{M}$

Attacchi

- Un soggetto che ha lo scopo di rompere il cryptosystem è l'avversario (*adversary)*
  - Si assume che l'avversario sappia qual è l'algoritmo usato, ma non conosce la chiave
- Tipi di attacco:
  - *conoscenza del ciphertext*: l'avversario conosce soltanto il ciphertext; lo scopo è risalire al plaintext, eventualmente anche alla key
  - *conoscenza del plaintext*: l' avversario conosce il ciphertext ed il relativo plaintext; lo scopo è risalire alla chiave
  - ….

Le basi degli attacchi

- Attacchi matematici
  - basati sull'analisi dei fondamenti matematici
- Attacchi statistici
  - fanno assunzioni riguardo la distribuzione delle lettere, coppie di lettere (digrams), triple di lettere (trigrams), etc.
    - Detti *modelli del linguaggio*
  - esaminano i ciphertext ed associano proprietà basandosi su assunzioni

Cifrari mono-alfabetici

- Possono essere prodotti con tabelle ad accesso diretto ("direct table lookup") (semplice nei campi algebrici)
- Il tempo di encrypt/decrypt varia direttamente con la lunghezza
- Può essere tradito dalla frequenza delle lettere

Esempio

- Ciphertext:

  HQFUBSWLRQLVDPHDQVRIDWWDLWL
  WJVHFXUHFRPSXWDWLRQRYHULQVH
  FXUHFKDQQHOVEBXVLQJHQFUBSWL
  RQZHGLVXLVHWHKPHVVDJH

- Plaintext:

  ENCRYPTIONISAMEANSOFATTAING
  SECURECOMPUTATIONOVERINSECU
  RECHANNELSBYUSINGENCRYPTION
  WEDISGUISETHEMESSAGE

Frequenza delle lettere

- Frequenza delle vocali dell'Inglese

vocale	A	E	I	O	U
percent	7.49	14.0	6.67	7.37	3.0

- Frequenze nel Ciphertext (104 lett.)

vocale	A	E	I	O	U
percent	0	0.96	0.96	0.96	4.81

Conson.	H	L	V	Q	W
percent	13.5	11.5	9.62	9.62	8.65

Sicurezza dei Cifrari Mono-alfabetici

- Sono sicuri?
  - 26! cifrari possibli (alfabeto Inglese)
  - Con i computer moderni sarebbero richiesti 10 anni con attacchi di forza bruta, ma…..
  - nei messaggi lunghi

    le frequenze delle lettere "tradiscono" il testo

Osservazioni significative

1. Una encryption basata su un problema difficile non è sicura per la sola difficoltà del problema
2. Un algoritmo di encryption deve essere regolare – questa è la sua debolezza
3. Una misura di sicurezza deve essere forte abbastanza da resistere all'avversario per l'intera "vita" dei dati

Cifrari Poli-alfabetici

- "Appiattiscono" le distribuzioni delle frequenze
- Nascondono le coppie di lettere
- Nascondono i prefissi e i suffissi
- Esempio: (usando il modulo moltiplicativo)
  Per le posizioni dispari: f(i)=(3*i) mod 26
  Per le posizioni pari: f(i)=((5*i)+13) mod 26

Cifrari Mono-alfabetici con chiave

- Permutazione della chiave (la chiave non ha lettere che si ripetono)

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

- Modulo Moltiplicativo (la chiave è il moltiplicatore)
  - f(i) = (3*i) mod 26
  - f('K') = 3*10 mod 26 = 4 = 'E'

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

Cifrari con trasposizione

- Invece che sostituire i caratteri, li permutano
- Gli Spartani usavano bastoni con diametri fissi e strisce di pergamena
  - Scrivevano da un capo all'altro delle strisce
  - Leggevano il ciphertext lungo le strisce
  - (Funziona meglio con 2 matite)
- In termini moderni, si usano le matrici

Cifrari a trasposizione delle colonne

- La chiave è il numero di colonne della matrice

T	H	I	S	I	S	A	M	E	S
S	A	G	E	T	O	S	H	O	W
H	O	W	A	C	O	L	U	M	N
A	R	T	R	A	N	S	P	O	S
I	T	I	O	N	W	O	R	K	S

- Ciphertext: TSHAI HAORT IGWTI SEARO ITCAN SOONW ASLSO MHUPR EOMOK SWNSS

Analisi delle trasposizioni di colonne

- Semplice, ma efficace
- Il lavoro per ogni carattere è costante; quello totale è proporzionale alla lunghezza del messaggio
- Richiede l'intero messaggio

  nel buffer di encryption
- La frequenza delle lettere

  è simile ai cifrari mono-alfabetici
- E' possibile usare tabelle per le frequenze di digrammi e trigrammi

Come rompere i cifrari per trasposizione di colonne

- Problema: quali sono le colonne adiacenti
- Separare in strisce e cercare i digrammi e i trigrammi

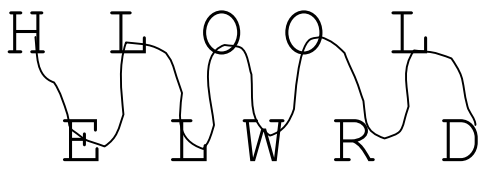
T		I		
S	I	T	A	
H	G	C	S	
A	W	A	L	E
I	T	N	S	O
H	I	S	O	M
A	S	O	M	O
O	E	O	H	K
R	A	N	U	S
T	R	W	P	W
	O		R	N
				S
				S

Cifrari che combinano più tecniche

- Usano insieme la sostituzione e la trasposizione
- Le sostituzioni confondono le informazioni
- La permutazioni sparpagliano le informazioni
- Se ben usate, l'una può migliorare l'altra
- Tutti i cifrari moderni combinano le due tecniche

Cifrari per trasposizione

- Un altro esempio (cifrario Rail-Fence)
  - Plaintext: `HELLO WORLD`
  - Riordinato come



  - Ciphertext: `HLOOL ELWRD`

Attaccare il cifrario

- Anagrammando
  - Se le frequenze degli 1-gram sono simili alle frequenze dell'alfabeto Inglese, ma le frequenze degli n-gram non lo sono, è probabile l'uso della trasposizione
  - Si riordinano le lettere fino a formare n-gram con frequenze maggiori

Esempio

- Ciphertext: `HLOOLELWRD`
- Frequenze dei 2-grams che iniziano per H
  - HE 0.0305
  - HO 0.0043
  - HL, HW, HR, HD < 0.0010
- Frequenze dei 2-grams che terminano per H
  - WH 0.0026
  - EH, LH, OH, RH, DH ≤ 0.0002
- Allora, E segue H

Esempio

- Ordinando in modo che H ed E siano adiacenti

```
HE
LL
O W
OR
LD
```

- Leggendo da sinistra a destra e poi verso il basso, si ottiene il plaintext originale

Cifrari per sostituzione

- Un altro esempio (cifrario di Cesare)
  - Plaintext: `HELLO WORLD`
  - Ogni lettera è scambiata con la terza lettera che la segue (X diventa A, Y=B, Z=C,…)
    - La Key è 3, di solito indicata con 'D'
  - Ciphertext: `KHOOR ZRUOG`

Attaccare il cifrario

- Ricerca esaustiva
  - Se lo spazio delle chiavi è abbastanza piccolo, si provano tutte le chiavi fino a trovare quella giusta
  - Il cifrario di Cesare ha 26 possibili key
- Analisi statistiche
  - Si confrontano con i modelli degli 1-gram (dell'alfabeto Inglese)

Attacco statistico

- Si calcola la frequenza di ogni lettera del ciphertext:

G	0.1	H	0.1	K	0.1	O	0.3
R	0.2	U	0.1	Z	0.1		

- Si applica il modello per gli 1-gram in Inglese (frequenze riportate nella prossima slide)

Frequenze dei caratteri

a	0.080	h	0.060	n	0.070	t	0.090
b	0.015	i	0.065	o	0.080	u	0.030
c	0.030	j	0.005	p	0.020	v	0.010
d	0.040	k	0.005	q	0.002	w	0.015
e	0.130	l	0.035	r	0.065	x	0.005
f	0.020	m	0.030	s	0.060	y	0.020
g	0.015					z	0.002

Analisi statistica

- $f(c)$ frequenza del carattere c nel ciphertext
- $\varphi(i)$ correlazione delle lettere nel ciphertext con le relative lettere in Inglese, assumendo che la chiave sia i
  - $\varphi(i) = \Sigma_{0 \leq c \leq 25} f(c)p(c - i)$ e quindi,

    $\varphi(i) = 0.1p(6 - i) + 0.1p(7 - i) + 0.1p(10 - i) + 0.3p(14 - i) + 0.2p(17 - i) + 0.1p(20 - i) + 0.1p(25 - i)$
    - $p(x)$ è la frequenza del carattere x in Inglese

Correlazioni: $\varphi(i)$ per $0 \leq i \leq 25$

i	$\varphi(i)$	i	$\varphi(i)$	i	$\varphi(i)$	i	$\varphi(i)$
0	0.0482	7	0.0442	13	0.0520	19	0.0315
1	0.0364	8	0.0202	14	0.0535	20	0.0302
2	0.0410	9	0.0267	15	0.0226	21	0.0517
3	0.0575	10	0.0635	16	0.0322	22	0.0380
4	0.0252	11	0.0262	17	0.0392	23	0.0370
5	0.0190	12	0.0325	18	0.0299	24	0.0316
6	0.0660					25	0.0430

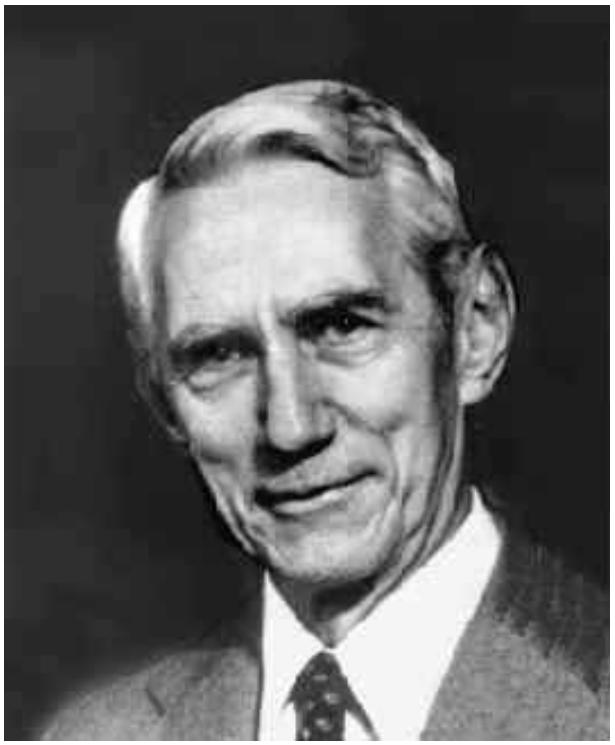
Il risultato

- Le chiavi più probabili, in base ai valori φ:
  - $i = 6, \varphi(i) = 0.0660$
    - plaintext `EBIIL TLOLA`
  - $i = 10, \varphi(i) = 0.0635$
    - plaintext `AXEEH PHKEW`
  - $i = 3, \varphi(i) = 0.0575$
    - plaintext `HELLO WORLD`
  - $i = 14, \varphi(i) = 0.0535$
    - plaintext `WTAAD LDGAS`
- Soltanto per $i = 3$ la frase ha senso in Inglese
  - Dunque, la chiave è 3 (o 'D')

Problema di Cesare

- La chiave è troppo piccola
  - Può essere trovata con una ricerca esaustiva
  - Frequenze statistiche non nascoste bene
    - Somigliano troppo alle lettere regolari dell'alfabeto Inglese
- Ingrandire la chiave
  - Più lettere nella stessa chiave
  - L'idea è di livellare le frequenze statistiche per rendere più difficile la crittoanalisi

Claude Shannon (1916 - 2001)



- *A Mathematical Theory of Communication* (1948), dà forma alla cosiddetta Teoria dell'Informazione
- Descriveva delle metodologie per misurare le informazioni, utilizzando la quantità di disordine in un dato sistema, assieme al concetto di entropia
- La *Magna Charta* dell'era dell'informatica
- Si ritirò a 50 anni

Teoria di Shannon

- La teoria di Shannon descrive
  - La diffusione
  - La confusione
  - La sicurezza ottenibile (Unconditional Security)
- Diffusione
  - Una trasformazione fissata può dare una buona encryption alle prime iterazioni, ma alla lunga può fallire

Claude Shannon

- Concetto di entropia – è equivalente alla restrizione del contenuto informativo in un messaggio
- Seconda legge della termodinamica – l'**entropia** è il grado di disordine di un sistema
- Tante frasi possono essere molto condensate senza che perdano il loro significato
- Shannon ha dimostrato che in una conversazione disturbata, è comunque possibile inviare un segnale senza che esso venga distorto

Claude Shannon

- Se il messaggio è codificato in modo che si **corregga da solo**, i segnali saranno ricevuti con la stessa precisione di un canale senza interferenze
- Un linguaggio ha un **codice di correzione degli errori integrato**

http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html

Teoria dell'Informazione

- La teoria dell'informazione misura la **quantità di informazione** contenuta in un messaggio, a partire dal numero di bit necessari per codificare tutti i possibili messaggi in una codifica ottimale
- Il campo SEX in un database: soltanto un bit di informazione (Male:0; Female:1)
- Codificato in ASCII – più spazio occupato, *ma nessuna informazione in più*

Teoria dell'Informazione

- La **quantità di informazione** in un messaggio è misurata formalmente con l'**entropia** del messaggio
- L'**Entropia** è una funzione della distribuzione di probabilità sull'insieme di tutti i possibili messaggi

Teoria dell'Informazione

- L'**Entropia** di un dato messaggio è definito con la media pesata su tutti i possibili messaggi *X*:

$$H(X) = \sum_{X} p(X) \log_2 \left(\frac{1}{p(X)} \right)$$

Es. di applicazione della T.I.

p(male) = p(female) =1/2, allora

$$H(X) = \frac{1}{2}(\log_2 2) + \frac{1}{2}(\log_2 2)$$

$$= \frac{1}{2} + \frac{1}{2} = 1$$

C'è soltanto un bit di informazione nel campo SEX del database

Teoria dell'Informazione

- I file di testo possono essere ridotti di circa il 40% senza perdita di informazione
- Poiché *1/p(x)* diminuisce come *p(x)* cresce, una codifica ottimale utilizza codici piccoli per i messaggi con frequenza maggiore, codici lunghi per messaggi poco frequenti
- *Codice Morse*

  E •, T –, J • – – –, Z – – • •

Teoria dell'Informazione

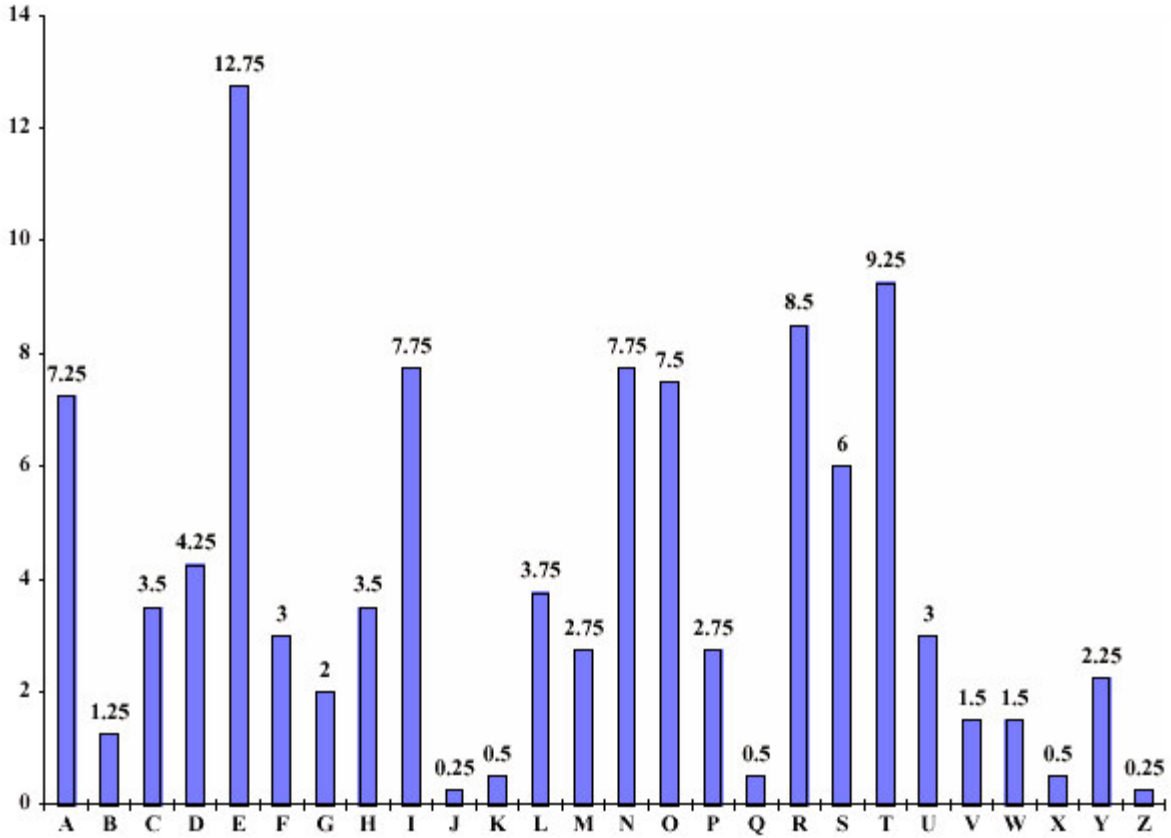
- L'entropia di un messaggio misura la sua incertezza. Il numero di bit che possono essere estratti quando il messaggio è nascosto nel ciphertext
- L'**Inglese** è una lingua molto ridondante
- occurring frequently => crng frqnly

Ridondanza dell'Inglese

- Si possono cancellare le vocali e le consonanti doppie

mst ids cn b xprsd n fwr ltrs,
bt th xprnc s mst nplsnt

Frequenza di ogni lettera nell'alfabeto Inglese



Sostituzione poli-alfabetica

- Trasformare alcune lettere in diverse altre, in base alla posizione nel plaintext
- In questo modo si appiattisce la frequenza di distribuzione
  - La cryptanalysis è più difficile
- Ha lo stesso effetto di avere diversi cifrari di Cesare che dipendono dalla posizione

Cifrari di Vigènere

- Simile al cifrario di Cesare, ma usa una intera frase
- Esempio
  - Messagggio `THE BOY HAS THE BALL`
  - Key `VIG`
  - Cifratura usando il cifrario di Cesare per ogni lettera:

```
key    VIGVIGVIGVIGVIGV
plain  THEBOYHASTHEBALL
cipher OPKW WECIYOPKWIRG
```

Termini utili

- *Periodo:* lunghezza della chiave
  - Nel precedente esempio, il periodo è 3
- *tableau*: tabella usata per cifrare e decifrare
  - Il cifrario di Vigènere ha le lettere chiavi in alto, mentre le lettere del plaintext sulla sinistra
- *poli-alfabetico*: la chiave è composta da diverse lettere
  - Il cifrario di Cesare è mono-alfabetico

Parti rilevanti della tabella

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

- Sono riportate solo alcune righe e colonne rilevanti
- Esempi:
  - key V, lettera T: nella colonna di V, alla riga di T si trova la lettera O
  - Key I, lettera H: nella colonna di I, alla riga di H si trova la lettera P

Cryptanalysis di Vigenère

- Per i cifrari poli-alfabetici bisogna conoscere:
  - Il numero di alfabeti usati
  - La key di ognuno
- La cryptanalysis è più difficile perché non basta controllare di quanto è "slittata" ogni frequenza

Kasiski

- Intuizione:
  - Alcune sequenze di lettere appaiono più volte nel testo
  - Un ciphertext che usa Vigenère avrà le sequenze di lettere (encrypted) mostrate in più punti
- Così:
  - Notare le distanze fra le sequenze di almeno 3 lettere ripetute nel ciphertext
  - Un buon candidato per il numero di alfabeti (dimensione della chiave) è il MCD delle distanze

Plaintext:	TOBEORNOTTOBE
Key:	NOWNOWNOWNOWN
Ciphertext:	GCXRCNACPGCXR

IC di Friedman

- L'**Indice di Coincidenza** $(I_c(x))$ è definito come
  - La probabilità che 2 elementi a caso delle stringhe *x* da *n* lettere siano uguali
- Così se $f_0, f_1, f_2 \ldots f_{25}$ sono il numero di presenze delle lettere A, B, C, … Z in una stringa, si ha che:

$$I_c(x) = \frac{\sum_0^{25} \binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_0^{25} f_i(f_i - 1)}{n(n-1)}$$

IC di Friedman

- La probabilità che la lettera *i* sia ripetuta in due punti della stringa è p_i^2
- Una stringa casuale avrà tutte le p_i uguali e fissate a 1/26
- In realtà, nella lingua Inglese ogni lettera ha una differente probabilità di essere ripetuta

IC di Friedman

- Allora, con queste considerazioni e usando i valori precedenti si ha che
  - Per le stringhe casuali: $I_c(x) = 0.038$
  - Per le stringhe in Inglese: $I_c(x) = 0.065$
- Ogni trasformazione operata da ogni alfabeto non dovrebbe cambiare l'I_c

IC di Friedman

- Questo conduce al metodo di Friedman
  - Si ipotizza che la chiave sia lunga *m*
  - Si divide il ciphertext in *m* parti
  - Si calcola il I_c per ogni parte di testo
  - Se l'ipotesi iniziale è corretta, allora per ogni parte di testo l'I_c sarà circa 0.065; altrimenti, l'I_c sarà circa 0.038

One-Time Pad

- Un cifrario di Vigenère con una chiave casuale lunga almeno quanto il messaggio
  - Si dimostra che è inviolabile
  - Perché? Sia dato come ciphertext `DXQR`. E' ugualmente probabile che corrisponda al plaintext `DOIT` (key `AJIY`) e al plaintext `DONT` (key `AJDY`) e a qualsiasi altra sequenza di 4 lettere
  - NB: la chiave deve essere casuale, altrimenti si può attaccare il cifrario cercando di ricostruire la chiave
    - Le approssimazioni (come ad esempio le sequenze pseudo-randomiche) non sono casuali!!

Cifrario di Vernam

(istanza del one time pad)

- Usa lunghe sequenze di numeri senza ripetizioni, combinate con il plaintext
- Il ciphertext non rivela la chiave
- Metodo

1. Rappresentazione binaria di P `101101`
2. Xor binario di numeri a caso `101111`
3. Output binario del ciphertext `000010`

Sommario

- Cifrari a sostituzione e trasposizione: basi per i moderni algoritmi simmetrici
- Crittoanalisi dei cifrari a trasposizione e sostituzione: ancora utile per metodologia e problematiche che propongono
- One time pad: il solo cifrario sicuro, ma…