Crittografia con Python

Corso introduttivo
Marzo 2015

Con materiale adattato dal libro "Hacking Secret Cypher With Python"
di Al Sweigart (http://inventwithpython.com/hacking/index.html)

Prof. Alessandro Bugatti

Attacchi statistici

- Avendo abbastanza testo cifrato (condizione facilmente verificabile) è possibile violare il codice utilizzando le proprietà statistiche dei linguaggi naturali oppure la presenza di parole probabili
- L'operazione può essere fatta piuttosto velocemente perchè non si esplorano tutte le possibilità ma solo una piccola parte

Cenni storici

- Il metodo fu scoperto dal filosofo arabo Abu Yusuf ibn Ishaq al-Kindi del IX secolo d.C. applicando considerazioni di statistica, fonetica e sistassi
- La stessa tecnica si sviluppò in Occidente (forse in modo indipendente) nel XV secolo.

Esempio

- In italiano la i è la lettera più comune, seguita da *e, a, o, t, n, ...*
- I digrammi sono *er, es, on, re, el, en , ...*
- Per violare il codice:
  - si trovano le lettere più frequenti
  - si ipotizza che corrispondano alle frequenze normali e si sostituiscono
  - si fanno ulteriori ipotesi sui digrammi e trigrammi fino ad arrivare alla soluzione

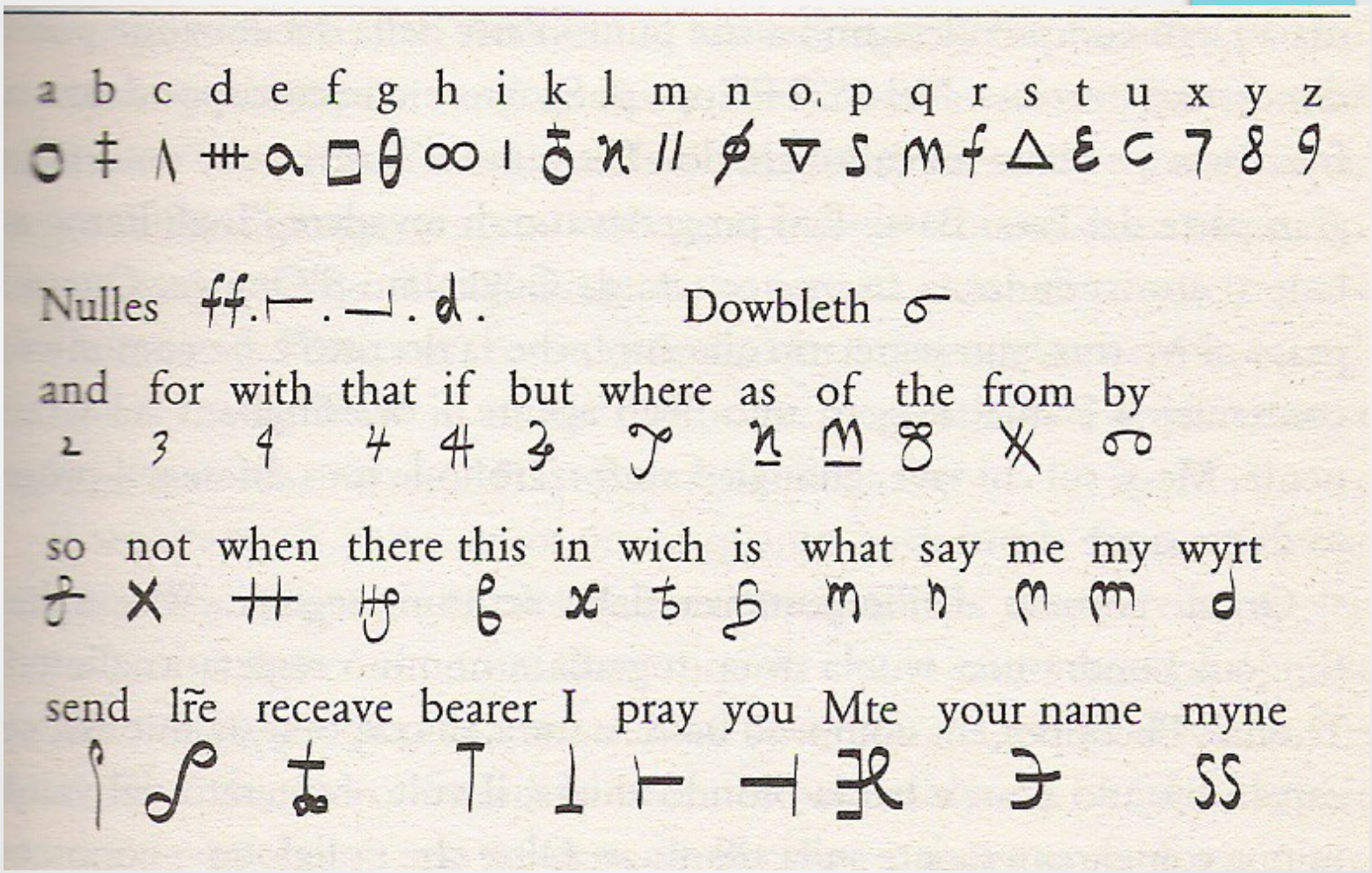
Miglioramenti

- Inserimento di nulle: elementi senza significato, ignorati dal mittente, ma in grado di confondere il crittoanalista
- Introduzione di frequenti errori ortografici
  - esemmpo di errrorri orttogrficci
- Nessuno di questi espedienti assicura grossi miglioramenti in termini di sicurezza

Cifratura con nomenclatore

- Oltre alla chiave monoalfabetica per il messaggio si utilizzano altri simboli ad ognuno dei quali viene associata una parola
- Un esempio famoso è il nomenclatore di Maria Stuarda
- In realtà non aggiunge molta sicurezza alla semplice codifica monoalfabetica

Nomenclatore di Maria Stuarda



Cifratura omofona

- Metodo che serve ad “appiattire” le proprietà statistiche dei linguaggi naturali
- Ogni lettera dell’alfabeto in chiaro può avere più simboli corrispondenti nell’alfabeto cifrante
- Più una lettera è comune più saranno i simboli con i quali può essere sostituita, in modo che tutte le frequenze siano uguali

Tavola per la cifratura omofona

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
16	03	12	04	13	01	23	00	17	–	–	05	18	08	15	09	21	07	11	14	06	10	–	–	–	02
26		20	32	35		49		31	–	–	41	28	29	42	81		48	30	19	24	25				
27		22	43	59				37	–	–	44	45	33	51	92		66	55	36	65					
34		47	53	63				50	–	–	54		40	57			82	72	39						
38				67				60	–	–	61		56	68			94	86	46						
52				73				64	–	–	75		62	77			98		69						
58				79				70	–	–	96		87	78											
71				83				74	–	–				80											
76				89				84	–	–				88											
95				91				85	–	–				97											
99				93				90	–	–															

Limiti della cifratura omofona

- Normalmente nei linguaggi naturali le lettere hanno una propria "personalità"
- In italiano un esempio evidente è la **q**, che è sempre seguita da **u**. Essendo la **q** rara verrà rappresentata con un numero, mentre la **u** con 3 numeri. Se si trova un numero che è sempre seguito dagli stessi tre numeri quello probabilmente è una **q** e di conseguenza i tre numeri sono una **u**

La “chiffrè indèchiffrable”

- Nel 1500 ormai era chiaro che la cifratura monoalfabetica non era più sicura e quindi si era alla ricerca di un metodo alternativo
- Il diplomatico francese Blaise de Vigenère inventò un metodo che da lui prese il nome e che è il capostipite dei metodi a sostituzione polialfabetica.

La cifratura di Vigenère

- Viene scelta una chiave, ad esempio la parola **MONTE** (potrebbe anche essere una parola inventata).
- La prima lettera del messaggio viene crittata con un cifrario di Cesare spostato di 13 (la M), la seconda con un cifrario spostato di 15 (la O) ecc. Dalla sesta lettera si incomincia da capo e così via

Tavola di Vigenère

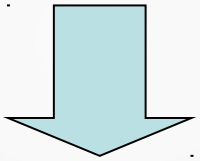
Chiaro	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
1	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
2	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
3	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
4	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
5	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
6	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
7	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
8	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
9	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
10	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
11	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
12	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
13	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
14	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
15	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
16	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
17	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
18	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
19	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
20	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
21	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
22	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
23	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
24	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
25	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
26	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

Esempio

Il testo

spostaretruppesucimaest

diventa



edblxmfrmvgdcxwgqvfeqgg

Forza della cifratura

- Rende inefficace l'analisi delle frequenze: nel caso precedente la lettera più comune nel testo cifrato, la **G**, corrispondente non a una ma a 3 lettere del testo in chiaro (la **u**, la **s** e la **t**)
- Allo stesso modo la stessa lettera del testo in chiaro viene codificata con lettere diverse (la doppia **p** di truppe diventa **dc**)

Problemi

- Nonostante l'apparente robustezza e la mancanza di punti deboli non venne usata per altri due secoli. Perchè?
- Era difficile da usare, molto meno pratica della cifratura monoalfabetica
- In generale un metodo crittografico per avere successo deve essere robusto ma anche pratico da utilizzare

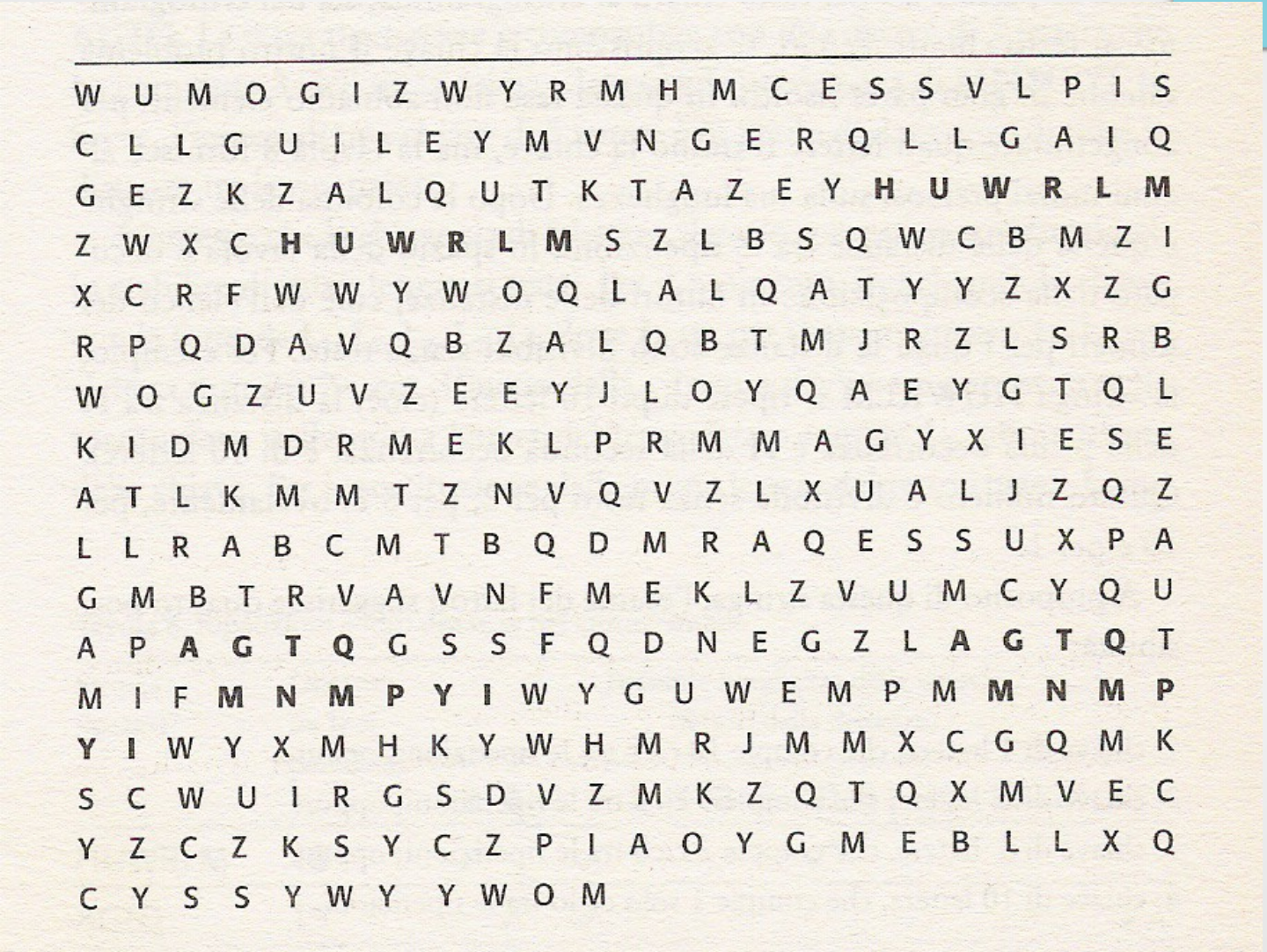
Attacchi al cifrario di Vigenère

- Charles Babbage a metà dell'Ottocento scoprì un sistema per attaccare questo metodo.
- Gli alfabeti cifranti sono più di uno, ma si continuano a ripetere, quindi una parola (o pezzo di parola) può essere cifrata solo in pochi modi diversi (se la chiave è lunga 4 esistono solo 4 modi possibili)

Metodo di Babbage

- Se la parola chiave fosse **SOLE**, la parola "**non**" del testo in chiaro potrebbe essere **FCY**, **BZR**, **YSF**, **RGB** e non altro.
- Se la parola "**non**" si ripete nel testo più volte è probabile trovare qualche ripetizione (se si ripete più di 4 volte è certo). Potrebbero anche altre parole che danno la stessa codifica, ma è improbabile.

Esempio



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

Ricerca della lunghezza della chiave

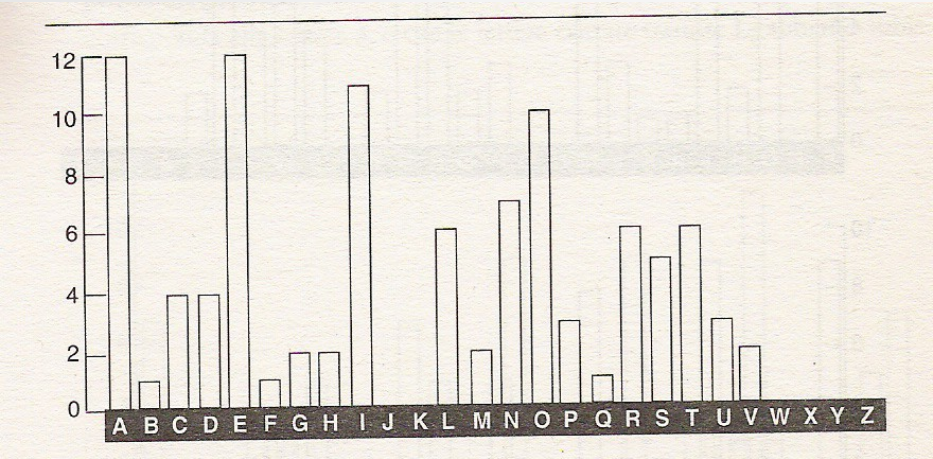
- Si può utilizzare la seguente tabella, sfruttando la distanza delle parole ripetute

Stringa ripetuta:	Distanza tra le ripetizioni:	Possibile lunghezza della chiave: (fattori della distanza)													
		2	3	4	5	6	7	8	9	10	11	12	13	14	15
HUWRLM	10	✓			✓					✓					
AGTQ	15		✓		✓										✓
MNMPYI	15		✓		✓										✓

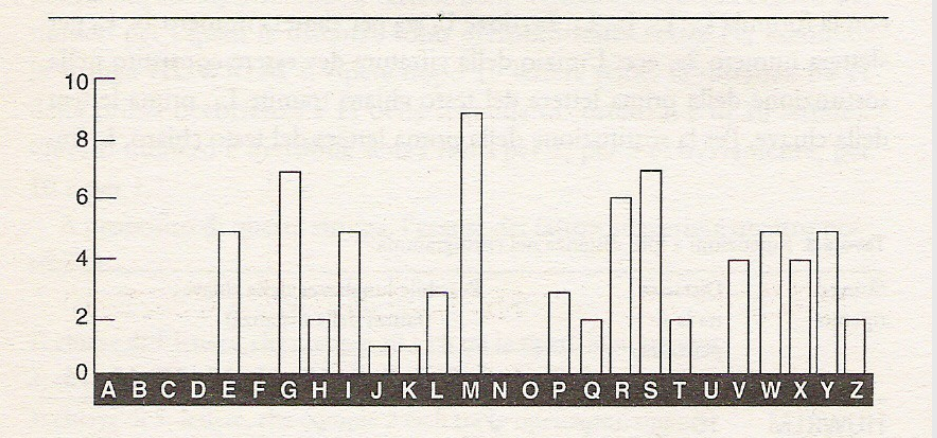
Scomposizione degli alfabeti cifranti

- L'ipotesi più probabile è che la chiave sia lunga 5 e a questo punto è come se avessi 5 messaggi con cifratura di Cesare
- Il primo messaggio è composto dalla prima, dalla sesta, dall'undicesima ecc... lettera, il secondo dalla seconda, dalla settima ecc.... lettera e così via.
- Posso utilizzare l'analisi delle frequenze per trovare la chiave

Soluzione



Distribuzione lingua italiana



Distribuzione "primo messaggio"

Riapplicando il procedimento altre 4 volte si trova che la parola chiave è

EMILY

Codifica a book cypher

- Viene scelto un testo come chiave del messaggio.
- Ogni parola viene numerata e la lettera iniziale viene associata al numero della parola
- Il testo in chiaro viene cifrato utilizzando i numeri così prodotti
- Il tesoro di Beale

Linguaggio come codice segreto

- Se il linguaggio è sconosciuto può servire da codice segreto
- Nella seconda guerra mondiale indiani navajo furono utilizzati come marconisti delle truppe americane nel Pacifico
- Il navajo era una lingua sconosciuta e di difficile comprensione
- I termini tecnici furono tradotti ad hoc

Metodo one-time pad

- Cifrario non violabile a patto di avere una chiave sufficientemente lunga
- Si converte un testo in una sequenza di bit
- Si esegue un OR ESCLUSIVO tra la chiave (una sequenza casuale di bit) e la stringa che rappresenta il testo in chiaro
- Non fornisce alcuna informazione per un attacco

Esempio

Messaggio 1 0 0 1 0 1 0 0 0 1 0 0 1 0 0 1 0 0 1 0 .....

XOR

Chiave 0 1 0 1 1 0 1 0 1 0 0 0 1 0 0 1 0 0 0 1 .....

Msg cifrato 1 1 0 0 1 1 1 0 1 1 0 0 0 0 0 0 0 0 1 1 .....

One Time Pad: svantaggi

- La quantità totale di dati che si possono trasmettere è limitata dalla lunghezza della chiave

$$\textbf{Lunghezza}_{\textbf{MAX}}\ \textbf{messaggio} \leq \textbf{Lunghezza chiave}$$

- Sensibilità alla perdita o inserzione di caratteri

One Time Pad: svantaggi

- Difficoltà a generare una chiave realmente casuale (bisogna utilizzare un fenomeno fisico casuale, ad esempio il decadimento radioattivo)
- Difficoltà di distribuire chiavi enormi
- In generale viene usata solo in casi estremamente particolari (ad esempio la linea rossa)

Riepilogo


STEGANOGRAFIA
(occultamento del messaggio)
SCRITTURE SEGRETE
CODICE
(sostituzione di parole)
SOSTITUZIONE
CRITTOGRAFIA
(alterazione del messaggio)
CIFRATURA
(sostituzione di lettere)
TRASPOSIZIONE
ONE-TIME PAD