Introduzione alla crittografia

Maurizio Monge
maurizio.monge@sns.it
http://poisson.phc.unipi.it/~monge/

Scuola Normale Superiore di Pisa

Seminario

26 luglio 2011

Cenni storici (1/3)

La Scitala lacedemonica è un sistema crittografico usato nell'antica Grecia.

- Gli antichi spartani (400 a.C.) usavano un cilindro sul quale è possibile arrotolare una striscia di pelle per codificare in messaggi,
- Plutarco, nella "Vita di Lisandro", racconta come il metodo fosse stato utilizzato da Lisandro nel 404 a.C. in un episodio risolutivo della Guerra del Peloponneso,
- la chiave crittografica consiste nella forma e dimensioni del cilindro di legno.

Cenni storici (2/3)

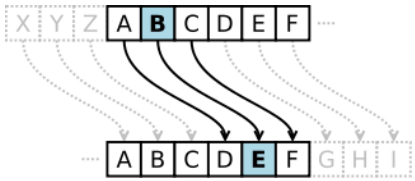
Gli antichi ebrei utilizzavano alcuni sistemi di cifratura

- il metodo detto Atbash consisteva nel riflettere le lettere dell'alfabeto, nel nostro alfabeto di tratta di rimpiazzare la A con la Z, la B con la Y, e così via,
- similmente i sistemi Albam e Atbah consistevano rispettivamente nel ruotare l'alfabeto di 13 lettere, e riflettere alcuni intervalli dell'alfabeto.

Anche l'antico testo indiano Kama Sutra raccomanda agli amanti l'uso della crittografia come sistema per comunicare senza essere scoperti (Parte I, Cap. III).

Cenni storici (3/3)

Il cifrario di Cesare consisteva nello ruotare le lettere dell'alfabeto di un numero fissato di posizioni.



- Come per gli altri sistemi crittografici usati nell'antichità, la sicurezza si basava sulla speranza che l'avversario non fosse a conoscenza del metodo di crittazione usato.
- A Cesare andava bene, visto che la maggior parte dei suoi avversari non era neppure in grado di leggere un testo in chiaro.
- Meno bene andava per il boss della mafia Bernardo Provenzano, che ai nostri giorni usava ancora il cifrario di Cesare per impartire ordini.

Rottura dei crittosistemi a sostituzione (1/3)

- sisuoyo eruur bmralripr:
  iru iziz yrkzuz u smsxz su-joieo bl sldzmr eru fsizykmoddz “yluus erkobmspozir ero fryyscco kmoddsdo” kgr r ou tmofz dmsddsdz izdz eo kmoddsisuoyo. ou drydz kzidorir lis erykmopozir eru frdzez eruu sisuoyo eruur bmralripr.
- iruur uoiclr isdlmsuo krmdr urddrmr kzftsoziz kzi lis bmralrips fsccozmr eo sudmr, r us kmoddzcmsbos kzi yzyodlpozir kziyrmvs ur bmralripr kzi ur alsuo kzftsoziz ur urddrmr.

Rottura dei crittosistemi a sostituzione (2/3)

```
sisuoyo eruur bmralripr: iru iziz yrkzuz u smsxz su-joieo bl sldzmr eru fsizykmoddz
A_ALI_I DELLE __E__E__E: _EL ____ _E__L_ L A_A__ AL-_I_DI __ A____E DEL _A_____I___
yluus erkobmspozir ero fryyscco kmoddsdo kgr r ou tmofz dmsddsdz izdz eo
__LLA DE_I__A_I__E DEI _E__A__I __I__A_I __E E IL __I__ __A__A__ ____ DI
kmoddsisuoyo. ou drydz kzidorir lis erykmopozir eru frdzez eruu sisuoyo eruur bmralripr.
__I__A_ALI_I. IL _E___ ____IE_E __A DE___I_I__E DEL _E__D_ DELL A_ALI_I DELLE __E__E__E.
iruur uoiclr isdlmsuo krmdr urddrmr kzftsoziz kzi lis bmralrips fsccozmr eo sudmr, r us
_ELLE LI___E _A___ALI _E__E LE__E_E ____AI___ ___ __A __E__E__A _A__I__E DI AL__E, E LA
kmoddzcmsbos kzi yzyodlpozir kziyrmvs ur bmralripr kzi ur alsuo kzftsoziz ur urddrmr.
__I_____A_IA ___ ___I___I__E ____E__A LE __E__E__E ___ LE __ALI ____AI___ LE LE__E_E.
```

- Le frequenze con cui compaiono le lettere dell'alfabeto italiano è:

E	A	I	O	T	N	R	L	S	C	D	P	U
11.7%	11.7%	11.3%	8.7%	7.2%	7.1%	7.0%	6.2%	5.1%	4.1%	3.9%	2.8%	2.7%
M	G	V	F	B	Z	H	Q	K	J	Y	W	X
2.5%	2.0%	1.8%	1.1%	1.1%	0.9%	0.9%	0.4%	0.0%	0.0%	0.0%	0.0%	0.0%

- Nel testo invece sono:

r	o	z	s	u	i	d	m	k	l	y	e	b
14.8%	9.6%	9.1%	8.8%	8.8%	8.2%	7.6%	6.2%	4.5%	3.7%	3.7%	3.4%	1.9%
f	p	c	a	t	g	j	v	x				
1.9%	1.9%	1.7%	1.4%	0.8%	0.2%	0.2%	0.2%	0.2%				

- Proviamo con $r = E$?
- La u si trova in fine di parola, potrebbe essere la L? E e la D?
- Probabilmente anche $s = A$ e $o = I$. E la prima parola sembra "ANALISI", da cui $i = N$, $y = S$.
- Aggiungiamo ancora $k = C$ e $z = O$. E completiamo la decrittazione.

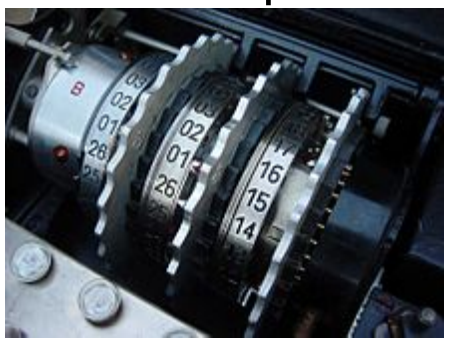
Rottura dei crittosistemi a sostituzione (3/3)

- Analisi delle frequenze
  Nel nono secolo l'arabo al-Kindi
  fu autore del manoscritto "Sulla decifrazione
  dei messaggi crittati" che è il primo trattato noto
  di crittanalisi. Il testo contiene una descrizione
  del metodo dell'analisi delle frequenze.
- Nelle lingue naturali certe
  lettere compaiono con una frequenza maggiore di
  altre, e la crittografia con sostituzione conserva
  le frequenze con le quali compaiono le lettere.

Sistemi polialfabetici

I sistemi a sostituzione fissa smisero di essere sicuri nel nono secolo d.C.

- Blaise de Vignère nel 1586 propose un sistema a sostituzione variabile, con rotazione di intervalli diversi in dipendenza di una chiave ripetuta.
- I tedeschi usarono nella seconda guerra mondiale le macchine Enigma, erano formate da rulli scorrevoli che cambiavano la sostituzione delle lettere con una certo livello di variabilità.



- Nonostante il sistema usato al meglio potesse essere molto difficile da decifrare, le chiavi erano spesso prevedibili, e gli Alleati furono in grado di leggere alcuni messaggi codificati.
- Gran parte dei messaggi era crittografato usando come chiave “Hitler”.

Crittosistemi rotti di recente

- Il sistema di hash crittografica LM (LAN Manager) è il sistema usato da Windows NT per verificare le password. Il realtà la funzione non era realmente a senso unico e la password veniva spezzettata in parti più piccole, ed è possibile recuperare la password dalla hash in poche ore.
- Il sistema CSS (Content Scramble System) usato per crittare i DVD è stato forzato nel 1999, anche se le chiavi sono di 40 bit la loro lunghezza effettiva è di 16 bit, e possono essere recuperate in pochi minuti.
- La crittografia WEP (Wired Equivalent Privacy) usata per proteggere le reti wireless è stata rotta nel 2001, ascoltando un numero sufficiente di dati da una rete wireless l'attaccante ottiene informazione su quale possa essere la chiave.

Fondamenti della crittografia moderna

La crittografia moderna cerca di ottenere sistemi che soddisfino alcuni requisiti fondamentali:

- Il sistema deve essere sicuro anche supponendo che il sistema crittografico sia perfettamente noto ad un avversario (Principio di Kerckhoff, riassunto nella massima di Shannon "Il nemico conosce il sistema").
- È necessario fare i conti con la potenza di calcolo fornita dal computer, che permette a un avversario di fare un gran numero di tentativi durante un attacco.
- L'avversario potrebbe conoscere un grosso pezzo del testo in chiaro.
- Si considera compromesso un sistema crittografico in cui esista un qualsiasi attacco migliore di quello di provare tutte le possibili chiavi crittografiche (anche se spesso questo non è sufficiente per decrittare un messaggio nel mondo reale).

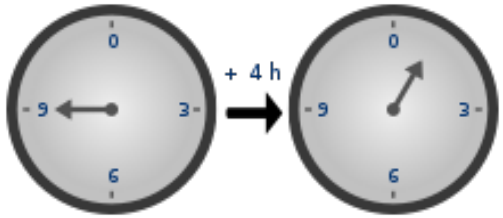
Estensione delle funzionalità

- Crittografia a chiave pubblica, esiste un sistema che permetta a tutti di scrivere messaggi che solo uno potrà leggere?
- Firma digitale: è possibile creare messaggi che tutti possono leggere, anche se uno solo è in grado di scriverli?
- Funzioni di hash crittografiche: è possibile associare a un insieme di dati un numero di controllo, o "hash", che permetta di verificare che i dati sono quelli giusti e sia difficile riprodurre con dati contraffatti?
- Crittografia a chiave simmetrica: per normali utilizzi di cifratura si cercano sistemi in cui la procedura per decifrare sia (quasi) identica a quella usata per cifrare.
- Condivisione di segreti: tecniche per suddividere un messaggio fra n persone in modo che k qualsiasi di esse possano conoscere il messaggio, ma $k-1$ non abbiano alcuna informazione su di esso.

Aritmetica modulare (1/2)

Posso rimpiazzare ogni lettera con una cifra, in un opportuno sistema di numerazione in base 26 ad esempio. Quindi un messaggio (o un pezzo di un messaggio) può essere visto come un numero molto grande.

- Fissiamo n, l'aritmetica modulo n è un sistema in cui i numeri ricominciano da zero quando sono arrivato ad n.



- Due numeri interi sono considerati lo stesso se mi lasciano resti uguali dividendo per n. O la loro differenza è un multiplo intero di n.
- Quando lavoro modulo n ho a disposizione le operazioni $+, -, \times$:

$$\begin{aligned} 4 \times (5 - 9) + 8 &\equiv 4 \times 7 + 8 \pmod{11} \\ &\equiv 6 + 8 \pmod{11} \\ &\equiv 3 \pmod{11} \end{aligned}$$

Aritmetica modulare (2/2)

- Modulo n posso anche dividere per qualsiasi numero primo con n. Ad esempio se consideriamo

$$3x \equiv 1 \pmod{14}$$

  l'equazione ha soluzione $x = 5$. Cioè $5 \equiv 1/3$.
- Sia $\phi(n)$ la funzione di Eulero, che conta i numeri primi con n nell'intervallo $0, 1, 2, \ldots, (n-1)$.
- Ad esempio $\phi(12) = 4$, dato che i resti primi con 12 sono $1, 5, 7, 11$.
- Più in generale se $n = p_1^{a_1} p_2^{a_2} \ldots p_k^{a_k}$, allora

$$\phi(n) = n \cdot \frac{(p_1 - 1)}{p_1} \cdot \frac{(p_2 - 1)}{p_2} \ldots \frac{(p_k - 1)}{p_k}.$$

Teorema di Eulero

- La funzione $\phi(n)$ soddisfa il seguente fatto fondamentale: se a è primo con n, allora
$$a^{\phi(n)} \equiv 1 \pmod{n}.$$
- La dimostrazione si ha facilmente usando il fatto che i possibili resti primi con n sono esattamente $\phi(n)$ e formano un gruppo rispetto alla moltiplicazione. Se non sapete cos'è un gruppo e non avete mai visto questo teorema fidatevi! ♪♫
- Se conosco la fattorizzazione di n, è facile calcolare $\phi(n)$.
- Usando le proprietà della ϕ è possibile costruire un sistema a chiave pubblica, l'RSA.

Crittosistema a chiave pubblica RSA (1/2)

- Poniamo mi venga detto un numero b tale che $b = a^e \pmod{n}$, e che conosca l'esponente e. Allora posso risalire ad a, se so $\phi(n)$.
- Sia in infatti d tale che $d = 1/e \mod \phi(n)$, o equivalentemente

$$de = 1 + k\phi(n)$$

per qualche k. Conoscendo $\phi(n)$ si calcola rapidamente d.

- Allora

$$\begin{aligned} b^d &\equiv a^{de} \\ &\equiv a^{1+k\phi(n)} \\ &\equiv a \cdot (a^{\phi(n)})^k \equiv a \pmod{n} \end{aligned}$$

grazie al teorema di Eulero.

- Elevare a d è come estrarre la radice e-esima!

Crittosistema a chiave pubblica RSA (2/2)

Il crittosistema RSA funziona nel seguente modo:

- Alice cerca due numeri primi grandi p, q, e calcola $n = pq$ e $\phi(n)$.
- Cerca anche un e primo con $\phi(n)$, e calcola $d = 1/e \mod \phi(n)$.
- Rende pubblici n ed e, e tiene segreto d (e anche $\phi(n)$ e i fattori p e q)
- e è la chiave pubblica, che serve a Bob per cifrare

$$a \to a^e \mod n$$

- d è la chiave privata, che serve ad Alice per decifrare

$$a^e \to (a^e)^d \equiv a \mod n$$

Sicurezza dell'RSA (1/2)

- Se conosco due numeri primi p, q calcolare $n = pq$ è questione di un attimo.
- Se conosco solo n è invece molto difficile trovare i fattori.
- Se n è un numero di 100 cifre, e provo tutti i possibili fattori fino a $\sqrt{n} \approx 10^{50}$, e faccio un mille miliardi di tentativi al secondo, ci metto $\approx 10^{38}$ secondi, $\approx 10^{30}$ anni. L'universo avrà smesso di esistere prima.
- Esistono algoritmi migliori che impiegano un tempo che è circa $\exp(c^{1/3})$, dove c è il numero di cifre. Ma il problema per numeri di > 250 cifre è comunque intrattabile.

Sicurezza dell'RSA (2/2)

D'altra parte se $n = pq$ allora $\phi(n) = (p-1)(q-1)$

- Se conosco n e $\phi(n)$, allora posso trovare p, q risolvendo il sistema

$$\begin{cases} n & = pq \\ \phi(n) & = (p-1)(q-1) \end{cases}$$

- Quindi ci si aspetta che trovare $\phi(n)$ a partire da n sia un problema difficile, altrimenti sarebbe anche facile fattorizzare n.
- Ci si aspetta anche che se un attaccante conosca sia a che a^e sia difficile risalire a e (problema del logaritmo discreto).
- Attenzione: gli a devono essere interi grandi che rappresentano un grosso pezzo del messaggio! Se ogni a che cifro è una sola lettera dell'alfabeto sostituire $a \to a^e$ è un cifrario con sostituzione!

Firma digitale

- Come abbiamo visto, Alice è l'unica in grado di estrarre la radice e-esima di a, calcolando a^d.
- Tutti conoscono la chiave pubblica e, e possono verificare che $(a^d)^e$ è proprio a. Anche senza essere in grado di calcolare a^d da a, perché non conoscono d.
- Questa osservazione fornisce un sistema di firma digitale, perché tutti possono verificare che il messaggio è stato scritto realmente da Alice, mentre lei è l'unica a poter scrivere.
- Comunicando con Bob, Alice firmerà il suo messaggio, e lo invierà a Bob cifrandolo con la chiave pubblica di Bob, in modo che solo Bob possa leggere.

Condivisione di segreti (1/4)

È possibile dare un messaggio a cinque persone, in modo che tre qualsiasi fra esse possano recuperare il messaggio, ma se sono solo in due non abbiano alcuna informazione?

- Consideriamo l'aritmetica modulo un primo p. Tale struttura è un campo! In matematica si indica in generale con $\mathbb{Z}/p\mathbb{Z}$, o $\mathbb{F}_p$.
- Ovvero, abbiamo le quattro operazioni $+, -, \times, /$. Possiamo dividere per qualsiasi elemento diverso da 0, e quindi risolvere sistemi lineari.
- Se $f(x)$ è un polinomio di grado $\leq k$, allora $f(x)$ è univocamente determinato dai suoi valori $f(x_i)$ su $k+1$ elementi distinti x_i, con $1 \leq i \leq k+1$.

Condivisione di segreti (2/4)

- Infatti, facciamo un esempio con $f(x)$ di grado 2

$$f(x) = f_2 x^2 + f_1 x + f_0,$$

  e supponiamo per elementi a, b, c noi conosciamo $f(a) = A$, $f(b) = B$, e $f(c) = C$.

- Allora abbiamo

$$\begin{cases} f_2 a^2 + f_1 a + f_0 &= A \\ f_2 b^2 + f_1 b + f_0 &= B \\ f_2 c^2 + f_1 c + f_0 &= C \end{cases}$$

- Consideriamo gli f_i come incognite. Abbiamo tre equazioni per tre incognite!
- Bisognerebbe dimostrare che il sistema è non degenere e quindi le incognite sono univocamente determinate, la dimostrazione usa le matrici (Cramer), è un po' tecnica ma non difficile.

Condivisione di segreti (3/4)

- Di conseguenza possiamo condividere un segreto s fra cinque persone prendendo un polinomio di secondo grado

$$f(x) = f_2 x^2 + f_1 x + f_0,$$

  con $f_0 = s$, e f_1, f_2 scelti a caso.
- Prendiamo cinque numeri distinti x_i (noti a tutti), e diciamo un valore $f(x_i)$ a ciascuna delle cinque persone.
- Tre qualsiasi fra esse sanno ricostruire il segreto!

Condivisione di segreti (4/4)

- Ma vediamo con due. Dovrebbero provare a trovare f_0 nel sistema

$$\begin{cases} f_2 a^2 + f_1 a + f_0 &= A \\ f_2 b^2 + f_1 b + f_0 &= B \end{cases}$$

- Ma tale sistema ha precisamente una soluzione nelle incognite f_2, f_1 qualsiasi potesse essere il segreto f_0! (due equazioni e due incognite).
- Di conseguenza due sole persone non hanno alcuna informazione su quale potesse essere il segreto f_0.

Bibliografia

Grazie per l'ascolto! Domande?

[1] H. Davenport: "*Aritmetica superiore. Un'introduzione alla teoria dei numeri*". Zanichelli, 1994.

[2] N.Koblitz: "*A Course in Number Theory and Cryptography*". Graduate texts in Mathematics **114**, Springer, New York, 1987.

[3] S.Singh: "*The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography*", Anchor Books, Londra, 1999. Traduzione italiana: "*Codici & Segreti*", Rizzoli, Milano, 1999.