Numeri Primi e Applicazioni crittografiche

Andrea Previtali

Dipartimento di Matematica e Fisica
Università dell'Insubria-Como

http://www.unico.it/matematica/Previtali
andrea.previtali@uninsubria.it

Corsi di Orientamento Universitario 2004

Costruire primi grandi

Grazie ai tre informatici indiani Agrawal, Kayal e Saxena (2002) sappiamo stabilire celermente, ossia in tempo polinomiale, se un intero è primo.

Colla tecnologia attuale è difficile fattorizzare interi i cui fattori primi hanno piú di 50 cifre.

Scelgo un intero $p \approx 10^{50}$. Qual è la probabilità che sia primo?

Basta saper contare i numeri primi minori di un dato valore reale.

Contare i numeri primi

La risposta a questo problema è stata congetturata da Gauss e Legendre alla fine del Settecento e dimostrata 100 anni piú tardi da Hadamard e de La Vallée-Poussin. Sia $\pi(x)$ la cardinalità dell'insieme

$$\{p \in \mathbb{N} \,:\, p \leq x,\, p \quad \text{primo}\}.$$

Teorema 1 (Teorema dei Numeri Primi). *La funzione* $\pi(x)$ *vale circa* $\frac{x}{\ln x}$.

Qui $\ln$ indica il logaritmo in base $E \approx 2.718281828$ (Notate qualcosa?).

Quindi dopo $\ln(10^{50})$ tentativi ho ottime possibilità di trovare un primo.

Ora $\ln(10^{50}) = 50\ln(10) \approx 116$. Usando Mathematica & Maple ottengo due primi p e q vicini a 10^{50} e $2 \cdot 10^{50}$ dopo 150 e 308 tentativi (a essere precisi ne bastano meno della metà).

La regina delle scienze

Sia $n = pq$, allora n ha circa 100 cifre. Anche gli algoritmi piú sofisticati richiedono mesi o anni per fattorizzare n.

A cosa servono queste considerazioni?

Un (famoso) matematico inglese G.H. Hardy (da non confondere col comico americano!) sosteneva che nessuno dei suoi teoremi avrebbe mai migliorato la qualità della vita o trovato applicazioni pratiche. Ma proprio mentre Hardy scriveva la sua autobiografia, un manipolo di Matematici, Fisici, Logici, Linguisti e appassionati di Enigmistica si sforzava a decifrare i messaggi di guerra scambiati dai Tedeschi (vedi Harris, "Enigma"). Questo sforzi hanno stimolato la costruzione dei primi calcolatori, nonché lo studio della Logica moderna.

Il codice di Cesare

L'idea di cifrare o crittare messaggi (ossia di nasconderne il contenuto) era già nota agli egizi. Tra i metodi di cifratura resta noto quello attribuito a Cesare (si proprio lui il vincitore dei Galli).

Vediamo come funziona. Pensate al giovane Cesare costretto ad ascoltare il suo insegnante di Matematica poco prima di pranzo.

Vuole comunicare la sua noia ad una compagna (esistevano già classi miste?) ma ha paura che il suo messaggio venga intercettato dal severo precettore. Allora si accorda anzitempo con la sua compagna su una **CHIAVE**, un intero k, ad esempio $k = 5$.

La chiave

Che ruolo gioca questa chiave? Semplice, se il messaggio M è ”vorrei andare a casa”, Cesare sposta ogni lettera in avanti di 5 (dovete pensare alle 26 lettere del nostro alfabeto disposte su una circonferenza). Quindi ”v” diventa ”a”, ”o” diventa ”t” e così via.

Usando Mathematica & Maple ho scritto un codice che converte il Messaggio M in un messaggio Cifrato C (ho escluso gli spazi per semplicità). Nell’esempio M=”vorreiandareacasa” diventa C=”atwwjnfsifwjfhfxf”. Difficile venir puniti per questo.

Funzioni di cifratura simmetriche

Come fa il compagno di Cesare a capirci qualcosa?

Rispetto al precettore ha due vantaggi

- Conosce la procedura usata da Cesare per cifrare;
- CONOSCE la chiave k.

Gli basta ripercorrere la circonferenza all'indietro di 5, ossia invertire la cifratura di Cesare ed ecco ricomparire il messaggio M.

In formule ho una funzione f che alla posizione della lettera sostituisce la posizione aumentata della chiave k

$$f : i \mapsto i + k \mod 26$$

Per decifrare mi basta osservare che esiste f^{-1} ed è facile da esprimere

$$f^{-1} : i \mapsto i - k \mod 26$$

Anche se il precettore sospettasse quale procedura usa Cesare per cifrare i suoi messaggi, senza la chiave k, non potrebbe decifrarli.

Crittanalisi e analisi delle frequenze

Attenzione però, oggi i criptoanalisti conosco metodi molto efficienti per risalire alla chiave anche se non la conoscono.

Ossia il metodo di Cesare non è inattaccabile. Basta infatti contare quante occorrenze ha ogni lettera nel messaggio cifrato e confrontarle colla frequenza in cui compaiono le varie lettere nella lingua italiana. Paragonando quelle di maggior frequenza si ottiene facilmente la chiave.

Questo sistema viene detto a chiave privata o simmetrica poiché presuppone che Cesare e il suo compagno si siano accordati sulla chiave in precedenza.

Chiavi pubbliche e chiavi private

Supponiamo vogliate comprare l'ultimo videogioco della Nintendo. Cosa fate? Mandare i soldi in una busta è troppo rischioso e lento. Forse posso usare Internet e inviare il mio numero di carta di credito. Ma chi mi assicura che qualcuno non sia in grado di leggerlo e usarlo per i suoi acquisti?

Bisogna cifrare tale numero, ma con quale chiave?

Negli anni '70 Internet stava nascendo e molte persone erano interessate a commercializzare questo strumento.

Nel 76 due Informatici e un Matematico, Rivest, Shamir e Adleman, idearono un sistema oggi noto come RSA che risolve il vostro problema.

RSA con carta e matita

Come funziona? Supponiamo che $M = 7$ sia il numero della vostra carta di credito. Sia $n = 15$ ed $e = 3$, allora lo cifratura avviene nel seguente modo $C = M^e \mod n$. Al contrario del metodo di Cesare dove sommavo qualcosa al mio messaggio, qui il messaggio viene elevato per un certo esponente. $\mod n$ significa che se supero n devo sostituire quanto ottengo col resto della divisione mediante n.

Comincio a calcolare $M^2 = 49 = 15 \cdot 3 + 4$. Per cui nella mia strana aritmetica M^2 equivale a 4. Allora M^3 equivale a $4 \cdot 7 = 28 = 15 + 13$. Quindi $C = 13$.

Questo metodo chiamato ”divide et impera” consente di effettuare elevamenti a potenza in modo molto veloce, ossia in $\lg_2(e)$ passi.

A questo punto inviamo al nostro fornitore $C = 13$. Come fa a recuperare il numero corretto della nostra carta? Il primo passo richiede di saper fattorizzare n nel prodotto di primi. Con Mathematica & Maple ottengo $n = pq$, ove $p = 3$ e $q = 5$. Il secondo problema consiste nel trovare un intero d tale che $ed = 1 \mod (p-1)(q-1)$. Qui interviene l'algoritmo euclideo a fornirmi la seguente equazione $3 \cdot 3 - 8 = 1$. Quindi $d = 3$. A cosa serve d? Proviamo a calcolare $C^d = 13^3 \mod 15$. Contravvenendo all'opinione di Pascal che chiunque creda all'esistenza dei numeri negativi è pazzo, osserviamo che $13 = -2 + 15$, quindi C^d equivale a $(-2)^3 = -8$ che equivale a sua volta a 7. Ma 7 è proprio il numero della nostra carta di credito!

Aritmetica modulare

Analizziamo i vari passaggi. Mi serve

- un intero n che è prodotto di due primi molto grandi ($\approx 10^{50}$) p e q;
- un esponente e coprimo con $(p-1)(q-1)$;
- l'inverso d di e modulo $(p-1)(q-1)$, ossia $ed-1$ deve essere un multiplo di $(p-1)(q-1)$.

Per ogni intero n sia $\phi(n)$ il numero di interi inferiori a n e coprimi con n. Ad esempio $\phi(p) = p-1$ se p è primo e $\phi(pq) = (p-1)(q-1)$ se p, q sono primi distinti. Allora vale il seguente risultato.

Teorema 2 (Eulero 1740). *Siano M ed n due interi coprimi, allora* $M^{\phi(n)} = 1 \mod n$.

Naturalmente M sarà il numero della nostra carta di credito e $n = pq$ un intero che ci viene fornito dal nostro venditore (di solito viene inviato direttamente al nostro computer senza che noi ci accorgiamo di niente, l'applicazione della Matematica passa sotto i nostri occhi e non la vediamo).

Sui sistemi Microsoft-Windows e era uguale a 17 (porta sfortuna anche se non siete superstiziosi!).

Alla luce di questo Teorema calcoliamo C^d. Allora

$$C^d = M^{ed} = M^{k\phi(n)+1} = (M^{\phi(n)})^k M = M \mod \phi(n),$$

purché M risulti coprimo con n. Matematicamente questo NON avviene una volta su 10^{50}. In realtà non avviene MAI (quante cifre ha una carta di credito?).

Il costo del RSA

Quanto tempo ci vuole?

Allora il costo per calcolare $M^e \mod n$ richiede in teoria e moltiplicazioni. Usando il trucco di elevare al quadrato e memorizzare i risultati riduce le moltiplicazioni a $\log_2 e$. Ad esempio $M^{15} = M^8 M^4 M^2 M$ e M^4 si calcola mediante due prodotti M^2 e $(M^2)^2$.

L'intero d si ottiene mediante l'Algoritmo Euclideo applicato a e e $\phi(n) = (p-1)(q-1)$ e questo, come abbiamo visto, necessita di circa $2\log_2 n$ operazioni.

Il tutto richiede pochi centesimi di secondi se avete un buon computer e un buon programma di manipolazione simbolica. Allora perché questo processo è sicuro?

Fattorizzare

La difficoltà consiste nel conoscere p e q a partire da $n = pq$. Il migliore algoritmo noto per fattorizzare un intero n è dovuto a Pomerance e Dixon e richiede

$$\exp(\sqrt{2 \ln n \ln \ln n}) = n^{\sqrt{2 \ln \ln n / \ln n}}$$

operazioni. Per $n \approx 10^{100}$ si tratta di $6 \cdot 10^{21}$. Un computer con clock di 3 GHz compie $3 \cdot 10^{12}$ operazioni al secondo, quindi servono circa 10^8 secondi, ossia circa 3 anni. La vostra carta di credito nel frattempo è già scaduta in barba al pirata.