Indice

Elenco delle figure

Elenco delle tabelle

1 Introduzione e obiettivi

-sicurezza -perchè rsa -perchè problema della fattorizzazione -perchè quest'algoritmo -Obiettivo implementazione

-fattorizzazione -conseguenze del problema (rsa) -sicurezza -Rompere rsa (implementazione parallela qs rsa-challenge) -ottimizzazione dell'algoritmo (cluster // uso linux-mpi)

Per il teorema fondamentale dell'Arimetica dato un numero N non primo, il quale possiede quindi dei divisori non banali, esiste ed è unica, prescindendo dall'ordine dei fattori, la sua fattorizzazione esprimibile come prodotto di numeri primi elevati ad opportune potenze. Il problema della fattorizzazione di numeri interi viene affrontato sin dalle elementari per trovare relazioni tra due numeri quali il massimo comun divisore o il minimo comune multiplo, ognuno di noi ha quindi presente di cosa si tratti, per lo meno ad un livello intuitivo. Quando ci si addentra nell'ostico compito di fattorizzare numeri che contengono un numero di cifre nell'ordine delle centinaia il problema, però, si dimostra essere molto più difficile di quanto si potesse pensare analizzandolo in maniera intuitiva. Sotto l'aspetto della teoria della complessità computazionale il problema risulta essere esponenziale, subesponenziale nel caso di alcuni algoritmi particolari. Sulla difficoltà di questo problema si basa un famosissimo algoritmo di cifratura: RSA.

RSA fa parte di quella branca della crittografia moderna che prende il nome di crittografia asimmetrica. A differenza della crittografia classica o simmetrica, quella asimmetrica prevede l'esistenza di due tipi di chiavi, una pubblica ed una privata. Questo approccio permette situazioni del seguente tipo:

- Alice vuole spedire un messaggio a Bob di modo che solo Bob possa leggerlo, userà quindi la chiave pubblica di Bob per cifrare il messaggio sapendo che solo Bob con la sua chiave privata potrà decifrarlo.

- Alice vuole spedire un messaggio a Bob di modo che, non solo Bob sia l'unico a poterlo leggere, ma che esso abbia la certezza che il mittente del messaggio sia proprio Alice. Alice quindi cifra il messaggio con la propria chiave privata e successivamente con quella pubblica di Bob. In questo modo all'atto della ricezione, Bob potrà applicare la propria chiave privata e la chiave pubblica di Alice per poter decifrare il messaggio essendo certo della provenienza dello stesso.

Il funzionamento di RSA non è particolarmente complesso. Supponiamo che Alice e Bob stiano avendo un dialogo segreto, ossia non vogliono che un eventuale haker possa intercettare la loro comunicazione e comprenderne il significato. Sia quindi M il messaggio che si vogliono scambiare. Ognuno di loro sceglie due numeri primi p e q li moltiplica tra di loro ottenendo N=$p.q$. In seguito calcolano $\varphi(N) = (p-1)(q-1)$. Scelgono infine un numero e coprimo con $\varphi(N)$ e minore dello stesso e calcolano d tale per cui $e.d \equiv 1 mod(N)$. Ora (N, e) è la chiave pubblica, mentre quella privata è (N, d). Il messaggio visibile sulla rete è il seguente: $\exp(M, 1) = \exp(M, e) mod(N)$ che verrá poi decifrato applicando una semplice esponenziazione di esponente d, elemento della chiave pubblica del mittente.

La sicurezza di quest'algoritmo risiede nella difficoltà computazionale di fattorizzare il numero N nei suoi fattori primi e quindi nel trovare la funzione $\varphi(N)$ che permetterebbe di trovare l'inverso moltiplicativo di e e quindi rompere i sistema.

Nel 1991 la RSA Laboratories propose come sfida la fattorizzazione di 54 semiprimi (prodotti di due primi) con un numero di cifre compreso tra 100 e 617. Ad oggi solo i 12 piú piccoli sono stati fattorizzati e, nonostante il 2007 vide la chiusura dell'RSA Challenge in molti ancora si dilettano nel tentativo di fattorizzarli.

2 Crivello Quadratico

Il Crivello quadratico è assieme al Crivello coi Campi di Numeri l'algoritmo di fattorizzazione piú veloce ad oggi conosciuto. Il costo computazionale risulta essere asintoticamente subesponenziale nell'ordine di $O(exp(sqrt(log(N))log(log(N))))$.
Per comprendere l'algoritmo occorre presentare almeno in parte la tecnica di fattorizzazione di Fermat. Fermat osservó che per trovare una fattorizzazione di un numero N si possono trovare due numeri X e Y tali per cui $X^2 - Y^2 = N$ trovando quindi $(X + Y)(X - Y)$. Qualora $X + Y$ o $X - Y$ non risultassero fattori banali, ossia 1 o N stesso, avremmo trovato una fattorizzazione completa per N.
Il Crivello Quadratico parte da questa semplice idea sviluppandola per ottenere un algoritmo alquanto efficiente e piuttosto articolato.

algoritmo

Dato un numero N intero, il crivello, a differenza dell'algoritmo di Fermat, cerca dei valori X e Y tali per cui valga la relazione $X \equiv Y \mod N$, successivamente ricerca il massimo comun divisore tra $(X - Y, N)$. Iniziamo calcolando una base di fattori primi FB di dimensione $k = k(N)$. Questo calcolo avviene mediante la scrematura di una base data dal crivello di eratostene. La discriminante per i fattori primi p della base è che essi abbiano simbolo di Legendre $(N|P) = 1$, ossia siano tali per cui N sia

un residuo quadratico modulo p Sia ora $s = \sqrt[2]{N}$ impostiamo il seguente polinomio: $Q(A) = (A + s)^2 - N$. Siamo certi che $Q(A) \equiv N$ sia un quadrato perfetto, il lavoro ora consiste nel trovare dei valori di A tali per qui $Q(A)$ si fattorizzi completamente sulla base di fattori precedentemente calcolata. Quando uno di questi polinomi si fattorizza completamente sulla base FB creiamo un vettore $v = (\alpha_1, \alpha_2, .., \alpha_n)$ dove ogni α_i rappresenta l'esponente dell'i-esimo numero primo nella fattorizzazione di $Q(A)$. Calcoliamo quindi un altro vettore $v_2 = (\alpha_1, \alpha_2, .., \alpha_n)_2$ ossia il vettore degli esponenti in base binaria. Nel caso banale in cui v_i fosse identicamente nullo, allora ogni primo avrebbe un esponente pari, in questo caso sarebbe un quadrato perfetto e, quindi, avremmo trovato una congruenza del tipo $X^2 \equiv Y^2 \mod N$. Anche se questo accadesse potremmo aver ottenuto una fattorizzazione banale, si prosegue quindi con la parte dell'algoritmo che riguarda prettamente l'algebra lineare. I vettori $v_2 i$ vengono inseriti tutti in una grossa matrice. Dall'algebra sappiamo che, data una matrice di k colonne necessita di almeno k+1 righe per ottenere almeno una dipendenza lineare. Nel nostro caso occorrerà quindi trovare almeno $k + m$ con $m \geq 1$ per poter ottenere un numero di righe tale da permettere di trovare almeno una dipendenza lineare.
Consideriamo le seguenti matrici:

v_1	3	4	2	7
v_2	3	2	2	2
v_3	5	2	3	1
v_4	6	1	3	3
v_5	2	2	2	3

v_1^2	1	0	0	1
v_2^2	1	0	0	0
v_3^2	1	0	1	1
v_4^2	0	1	1	1
v_5^2	0	0	0	1

Notiamo che dalla combinazione lineare di $v1_2$ $v2_2$ e $v4_2$ otteniamo un vettore nullo. Andiamo quindi a considerare ora i relativi vettori non modulati $v1v2ev5$ e sommiamoli tra loro, ottenendo $v_125 = (8, 8, 6, 12)$. Se ora consideriamo la congruenza

$$Q(A_1)Q(A_2)Q(A_5) \equiv 2^8 * 3^8 * 5^6 * 7^12 \mod N$$

Considerando il membro di sinistra come la X e quello di destra come la Y della nostra relazione iniziale, possiamo ricercare la congruenza desiderata e tentare di fattorizzare N.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at,

mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

3 Architettura del sistema di calcolo distribuito

3.1 MPI

MPI, acronimo per *Message Passing Interface*, è un'interfaccia che permette lo scambio di dati tramite il paradigma di scambio di messaggi fra processi. Lo scambio può avvenire sia fra processi sulla stessa macchina (tipicamente tramite memoria condivisa), che fra processi su macchine differenti (tramite qualche protocollo di rete come TCP/IP). Il principale vantaggio è l'indipendenza del codice dalla configurazione utilizzata. In altre parole, il programmatore che utilizza questa interfaccia non ha bisogno di sapere quanti processori avrà a disposizione e su quante macchine essi siano distribuiti, si dovrà semplicemente occupare di scambiare i dati con le primitive offerte dall'interfaccia, e il mezzo di comunicazione verrà poi determinato automaticamente a seconda dell'implementazione di MPI e della disponibilità di risorse.

4 Tecnologie utilizzate

4.1 OpenMPI

OpenMPI è un'implementazione open source di MPI (3.1). È in grado di gestire la comunicazione fra processi con una moltitudine di tecnologie, fra cui TCP per la comunicazione fra processi in esecuzione su macchine facenti parte di una rete di tipo classico, e memoria condivisa per la comunicazione (molto più rapida) fra processi in esecuzione sulla stessa macchina. L'esecuzione su macchine multiple è gestita tramite il protocollo SSH: la macchina su cui viene lanciato il processo contatterà le altre tramite protocollo SSH, e lancerà opportunamente il programma richiesto. La libreria funziona assumendo che le macchine abbiano un utente con lo stesso nome e con lo stesso contenuto della home (tenterà infatti di contattare tramite SSH la macchina con lo stesso nome utente da cui è stato lanciato il processo, e cercherà l'eseguibile nello stesso path della macchina di orgine). Al fine di evitare possibili problemi in questo senso è opportuno utilizzare uno dei tanti sistemi in grado di sincronizzare il contenuto dei dischi, o se possibile montare la home su un file system di rete.

Di seguito verranno esposte le primitive più importanti della libreria.

- MPI_Init: Inizializza la libreria, va chiamata prima di utilizzare qualunque altra funzione di MPI.
- MPI_Comm_size: Ritorna il numero di processi all'interno del comunicatore specificato.
- MPI_Comm_rank: Ritorna il rank del processo all'interno del comunicatore specificato.
- MPI_Abort: Termina tutti i processi nel comunicatore specificato.
- MPI_Get_processor: Ritorna il nome del processore che sta eseguendo il processo.

- MPI_Finalize: Finalizza la libreria, dopo questa chiamata non è più possibile utilizzare funzioni di libreria.
- MPI_Send: Send bloccante classica, la funzione ritorna quando il buffer contenente i dati da spedire è riutilizzabile.
- MPI_Recv: Receive bloccante classica, la funzione ritorna quando il buffer contiene i dati ricevuti ed è utilizzabile.
- MPI_Isend: Send non bloccante, ritorna un id della richiesta, utilizzabile per verificare lo stato dell'operazione.
- MPI_Irecv: Receive non bloccante, ritorna un id della richiesta, utilizzabile per verificare lo stato dell'operazione.
- MPI_Test: Controlla se l'operazione richiesta è terminata o meno.
- MPI_Wait: Attende il termine dell'operazione richiesta.
- MPI_Pack e MPI_Unpack: Impacchettano e spacchettano dati da spedire.

5 Strategia di parallelizzazione dell'algoritmo

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

6 Algoritmo parallelizzato

L'algoritmo ha una particolaritá molto utile sotto l'aspetto della complessita temporale: la sezione del codice deputata alla valutazione del polinomio $Q(A)$ è fortemente parallelizzabile. Per ottenere le congruenze necessarie dobbiamo scomporre un numero notevole di polinomi valutati sulla base di fattori FB. Questa parte del lavoro puó essere fortemente parallelizzata in quanto aumentando il lavoro demandato ai singoli processori è essenzialmente quello di valutare e scomporre i polinomi su intervalli diversi; questo permette di demandare un notevole calcolo ai nodi paralleli e quindi uno speed-up notevole all'aumentare del numero di nodi. per quanto la parte appena successiva, quella di eliminazione gaussiana è intrinsecamente seriale.

7 Risultati

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu

libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

8 Sviluppi futuri

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.