Fall October 18, 2011

Constructing Large Numbers with Cheap Computers

Andrew Shallue, *Illinois Wesleyan University*
Steven Hayman, *Illinois Wesleyan University*

Constructing Large Numbers with Cheap Computers

Steven Hayman '12 and Andrew Shallue

Department of Mathematics/Computer Science
Illinois Wesleyan University
Bloomington, IL

October 18, 2011

Algorithmic/computational number theory

Number theory is the study of integers

An algorithm is a sequence of steps that terminates

Algorithmic number theory: development and analysis of algorithms related to number theory

Computational number theory: implementing algorithms and generating data in order to break records, show off new techniques, or confirm conjectures from number theory.

Holy Grail problems: primality, factoring

Hyperion

Built with the help of Mark Liffiton for $\sim$ \$3000
6 nodes, 24 processors, $\sim$ 70 Gigs RAM (2 nodes with 16 Gigs),
3.4 GHz

Comparison: Blue Waters, UIUC, proposed $\sim$ 500,000 processors,
$\sim$ \$2.5 million

Ideal: one node with 1 Terabyte RAM

Carmichael numbers

A Carmichael number is a pseudoprime, i.e. a composite number that looks prime in a certain technical fashion.

More precisely, a Carmichael number n is a composite that passes the base a Fermat primality test, for all a coprime to n.

Theorem

n is Carmichael if it is squarefree (no prime divides n twice) and for all p dividing n, $p-1$ divides $n-1$.

A bit of history

First studied by Carmichael and Korselt around 1900

Alford, Granville, Pomerance 1994: infinitely many Carmichael numbers

Recent records:

Löh and Niebuhr (1996): $1,101,518$ prime factors

Alford and Grantham (2003): $19,565,300$ prime factors

Hayman and Shallue (Oct 2011): $1,021,449,091$ prime factors

Hayman and Shallue (Nov 2011): $10,333,229,505$ prime factors

Motivation

Carmichael numbers are interesting.

Primality testing is fundamental to many algorithms, in particular the cryptography that protects the internet.

The development of new algorithms is always important.

330864320785720230954728761814534076584544218400403366619873
292561819896001746711768353878020418262491027974945395961615
466588983533976932029571190787250078231501124849761112783479
985132502101869918441335155919295368738804318490685494258145
173639135198946560355699527376560127749310175740677407352750
474009428951345538930734728858675392754139131676191487097954
780882212819797722231199355394929924457129954925804652705120
856990960927213977704135331685033761631553330666978209286852
830118363644834308017408730103369454389939688179873051679337
037310386887345146884024348044379512056754691373012915026506
847475287331677569524867262619520556073845364610955579117766
270988105148909093048193517093264565117315594092619480358259
700563219010405446473590473934779257068172465082629245535790
949244441749878596866139423918080252106258667952454071719228
289472078352487944473654288386151875672167637156530881100001
957863316536663442093878357090462085661305498985255093231275
081024023695363156023129401777459200001

How large?

We have constructed a Carmichael number with $10,333,229,505$ prime factors.

Number of digits is roughly $30 \cdot 10^{10} = 3 \cdot 10^{11}$.

At 1000 digits per slide, that's $\sim 3 \cdot 10^8 = 300,000,000$ slides.

Modular arithmetic

Definition

Let a, b be integers and m a positive integer. As an operation, $a \mod m$ is the remainder after dividing a by m. If m evenly divides $b - a$ we say that a is equivalent to b modulo m, denoted $a \equiv b \mod m$.

$32 \mod 5 = 2$ i.e. $32 \equiv 2 \mod 5$

$18 \mod 6 = 0$ i.e. $18 \equiv 0 \mod 6$

$17 \mod 16 = 1$ i.e. $17 \equiv 1 \mod 16$

$5 \mod 101 = 5$

$16 \equiv 22 \mod 3$

Modular exponentiation

"$\equiv$" acts as an equal sign should, and operations have all the standard properties (commutative, distributive, and so on)

Theorem

Let a, b be integers and m a positive integer. Then

$$(a+b) \mod m = [(a \mod m) + (b \mod m)] \mod m$$
$$(a \times b) \mod m = [(a \mod m) \times (b \mod m)] \mod m$$

$(2^4 \mod 5) = (16 \mod 5) = 1$

or

$(2^3 \mod 5) = (2 \cdot 4 \mod 5) = (8 \mod 5) = 3$ so
$(2^4 \mod 5) = (3 \cdot 2 \mod 5) = (6 \mod 5) = 1$

Theorem (Pierre de Fermat, 1640)

Let p be prime and a be an integer not divisible by p. Then $a^{p-1} \equiv 1 \mod p$.

$$
\begin{aligned}
1^{12} &\mod 13 = 1 \\
2^{12} &\mod 13 = 1 \\
3^{12} &\mod 13 = 1 \\
4^{12} &\mod 13 = 1 \\
5^{12} &\mod 13 = 1 \\
6^{12} &\mod 13 = 1 \\
7^{12} &\mod 13 = 1 \\
8^{12} &\mod 13 = 1 \\
9^{12} &\mod 13 = 1 \\
10^{12} &\mod 13 = 1 \\
11^{12} &\mod 13 = 1 \\
12^{12} &\mod 13 = 1
\end{aligned}
$$

$$
\begin{aligned}
1^{14} &\mod 15 = 1 \\
2^{14} &\mod 15 = 4 \\
3^{14} &\mod 15 = 9 \\
4^{14} &\mod 15 = 1 \\
5^{14} &\mod 15 = 10 \\
6^{14} &\mod 15 = 6 \\
7^{14} &\mod 15 = 4 \\
8^{14} &\mod 15 = 4 \\
9^{14} &\mod 15 = 6 \\
10^{14} &\mod 15 = 10 \\
11^{14} &\mod 15 = 1 \\
12^{14} &\mod 15 = 9 \\
13^{14} &\mod 15 = 4 \\
14^{14} &\mod 15 = 1
\end{aligned}
$$

Basic algorithm

1. Choose Λ
2. Find primes p with $p-1$ dividing Λ (and $p \nmid \Lambda$)
3. Find subset of these primes that product to 1 modulo Λ

e.g.

$\Lambda =$

$2^{15} \cdot 3^8 \cdot 5^5 \cdot 7^4 \cdot 11^3 \cdot 13^2 \cdot 17^2 \cdot 19^2 \cdot 23^2 \cdot 29 \cdot 31 \cdot 37 \cdot 41 \cdot 43 \cdot 47 \cdot 53 \cdot 59 \cdot 61 \cdot 67 \cdot 71 \cdot 73 \cdot 79$

$= 288828494392627542423975683172283292832395366400000$

How many operations?

Goal reached: construct a Carmichael number with a billion prime factors.

Attempt 1: check all $2^{1,000,000,000}$ subsets

Attempt 2: pick random subset, hope it is 1 modulo Λ.
Expected number of trials is $|\Lambda| \sim 2^{160}$.

New algorithm: among the many possible solutions, some have special properties that make them easier to find.

Theoretical result

Theorem

Let $G = (\mathbb{Z}/\Lambda\mathbb{Z})^\times$ *and* $\mathcal{P}$ *be defined as in the Erdös construction. Assume that the elements of* $\mathcal{P}$ *are distributed symmetrically in* G*, that they are* 1 *modulo* Q_i *with probability at least* $1/(h_i + 1)$*, and that* $N = |\mathcal{P}| = K(\Lambda)$*. Then there is an algorithm that with high probability finds a subset of* $\mathcal{P}$ *that products to* b *in* G *and requires time and space*

$$2^{O(\sqrt[4]{\log N})}.$$