Anatomy of Integers and Cryptography

Igor E. Shparlinski

Centre for Advanced Computing:
Algorithms and Cryptography

Macquarie University

igor@comp.mq.edu.au

Outline

Part I: Integers and Their Divisors

We discuss some *not-so-well-known* facts about the arithmetic structure of integer numbers

- Given a "typical" integer n what can we say about
  - the largest prime divisor of n?
  - the distibution of integer divisors of n?
- Are the answers much different for "typical" cryptographic integers, such as
  - shifted primes $p-1$?
  - polynomial values $f(n)$?
  - values of the Euler function?
  - cardinalities of elliptic curves over $\mathbb{F}_q$?

Part II: Cryptography

We apply this knowledge to analysis of several *not-so-well-known* attacks on various cryptographic protocols and algorithms:

- Naive ElGamal protocol for key exchange;
- Fix-padded RSA;
- Generalised Diffie-Hellman protocol;
- Pratt primality certificate;
- Using small exponentiation base;
- Strong primes for RSA.

Rough Introduction to Smoothness

What are Smooth Numbers?

An integer n is *smooth* if it has only small prime divisors.

In a quantitative form: n is y-*smooth* if all prime divisors $p|n$ satisfy $p \leq y$.

$P(n)$ = the largest prime divisor of n.

$$\boxed{n \text{ is } y\text{-smooth} \quad \Longleftrightarrow \quad P(n) \leq y}$$

How Many are There?

Let

$$\psi(x, y) = \#\{n \leq x \mid n \text{ is } y\text{-smooth}\}.$$

Important parameter:

$$u = \frac{\log x}{\log y} \;\Leftrightarrow\; x = y^u$$

Intuition

Question: *What is the probability that $p \not| n$ when $n \leq x$ is chosen at random?*

$$1 - \frac{1}{p}$$

Question: *What is the probability that $p \not| n$ for all $x \geq p > y$ when $n \leq x$ is chosen at random?*

$$\prod_{x \geq p > y} \left(1 - \frac{1}{p}\right) = \prod_{p \leq x} \left(1 - \frac{1}{p}\right) \prod_{p \leq y} \left(1 - \frac{1}{p}\right)^{-1}$$
$$\sim \frac{\log y}{\log x} = \frac{1}{u}$$

by the Mertens formula.

So one can certainly predict a very nice bound

$$\psi(x, y) \sim \frac{1}{u} x$$

Too bad that this is completely wrong ...

Fundamental Theorem of Cryptography:

> *If we have no clue about something then we can safely assume that it behaves as an idependent random variable*

. . . is unfortunately not always correct.

Sometimes we need to work with very slowly changing functions, where numerical experiments are useless and only deep theoretic understanding may help us:

> $\log\log\log n$ *has been proved to go to infinity with* n, *but it has never been observed doing so . . .*

Carl Pomerance

Approximate Answer:

Canfield, Erdős, and Pomerance, **1983**:

$$\psi(x,y) = u^{-u+o(u)}x$$

for

$$1 \le u \le y^{1-\varepsilon} \qquad \forall \varepsilon > 0,$$

or, equaivalently,

$$y \ge (\log x)^{1+\varepsilon} \qquad \forall \varepsilon > 0,$$

More Precise Answer:

Hildebrand, **1986**:

$$\psi(x,y) \sim \rho(u)x$$

for

$$u \le \exp\left((\log y)^{3/5-\varepsilon}\right) \qquad \forall \varepsilon > 0,$$

or, equaivalently,

$$y > \exp\left((\log \log x)^{5/3+\varepsilon}\right) \qquad \forall \varepsilon > 0,$$

$u \le \exp\left((\log y)^{3/5-\varepsilon}\right) \;\Leftrightarrow\; y > \exp\left((\log \log x)^{5/3+\varepsilon}\right)$,
where $\rho(u)$ is the *Dickman–de Bruijn* function

Dickman–de Bruijn function $\rho(u)$

$$\rho(u) \; = \; 1, \qquad 0 \le u \le 1,$$

and

$$\rho(u) \; = \; 1 - \int_1^u \frac{\rho(v-1)}{v}\, dv, \qquad u > 1.$$

Some properties of the $\rho(u)$

We recall that

$$\rho(u) = u^{-u+o(u)}, \qquad u \to \infty$$

(this can be obtained independently).

More precisely

$$\rho(u) = \left(\frac{e + o(1)}{u \log u}\right)^u, \qquad u \to \infty$$

We also have $\rho(u) = 1 - \log u$ for $1 \le u \le 2$.

E.g. $\rho(e^{1/2}) = 1/2$, that is $\sim 50\%$ of integers n have all prime divisors $\le n^{1/e^{1/2}}$.

This has been used by I. M. Vinogradov, and then by D. A. Burgess, to estimate the smallest quadratic non-residue.

Conditional Answer:

Hildebrand, **1984**:

$$\psi(x,y) \sim \rho(u)x$$

for a wider range

$$1 \leq u \leq y^{1/2-\varepsilon} \qquad \forall \varepsilon > 0,$$

or, equaivalently,

$$y \geq (\log x)^{2+\varepsilon} \qquad \forall \varepsilon > 0,$$

iff the Riemann Hypothesis is true.

Very Smooth Numbers

Granville, **1993**:

$$\psi(x, \log^A x) = x^{1-1/A+o(1)}, \quad \text{for any} \quad A > 1.$$

Important Special Case

$$\psi(x, L(x)^c) = x/L(x)^{1/2c+o(1)},$$

where

$$L(x) := \exp(\sqrt{\log x \log\log x}).$$

How Do We Count Them?

Counting very smooth numbers

Lattices:

Let $2 = p_1 < \ldots < p_s \leq y$ be all $s = \pi(y)$ primes up to y.

$$\begin{aligned}
\psi(x,y) &= \#\left\{(\alpha_1,\ldots,\alpha_s) \mid \prod_{i=1}^{s} p_i^{\alpha_i} \leq x\right\} \\
&= \#\left\{(\alpha_1,\ldots,\alpha_s) \mid \sum_{i=1}^{s} \alpha_i \log p_i \leq \log x\right\}
\end{aligned}$$

— counting integer points in a tetrahedron.

The number of integer points is close to its volume

$$V = \frac{\log^s x}{s! \prod_{i=1}^{s} \log p_i}$$

if V is large compared to the dimension s

$$\Downarrow$$

s (and thus p_s) must be reasonably small.

Counting not so smooth numbers

Rankin's method

Fix any constant $c > 0$. Then

$$\begin{aligned}\psi(x,y) &= \sum_{\substack{n\le x\\ p\mid n\Rightarrow p\le y}} 1 \le \sum_{\substack{n\le x\\ p\mid n\Rightarrow p\le y}} \left(\frac{x}{n}\right)^c\\ &\le \sum_{p\mid n\Rightarrow p\le y} \left(\frac{x}{n}\right)^c = x^c \sum_{p\mid n\Rightarrow p\le y} \frac{1}{n^c}\\ &= x^c \prod_{p\le y} \sum_{i=0}^{\infty} \frac{1}{p^{ic}} = x^c \prod_{p\le y} \left(1-\frac{1}{p^c}\right)^{-1}.\end{aligned}$$

Using the prime number theorem, we estimate the product as a function of y and c (non-trivial!) and minimize over all choices of c.

The (quasi-) optimal value is

$$c = 1 - \frac{u \log u}{\log y}.$$

Buchstab-de Bruijn Recurrent relation

Write each y-smooth n with $n > 1$, as $n = pm$ where $p = P(n)$ is the largest prime factor of n.

Collecting together integers n with $P(n) = p$ we get

$$\psi(x,y) = 1 + \sum_{p \le y} \psi\left(\frac{x}{p}, p\right),$$

(since $P(m) \le P(n) = p$ and $m = n/p \le x/p$).

This identity has been used for both lower and upper bounds and even for asymptotic formulas.

We now use it to "prove" that for each fixed u

$$\psi(x, x^{1/u}) \sim x\rho(u).$$

The "proof" is by induction over N, where $u \in (N, N+1]$.

For $0 < u \leq 1$ we trivially have $\psi(x, x^{1/u}) = \lfloor x \rfloor$.

For $1 < u \leq 2$ (i.e. $x \geq y \geq x^{1/2}$), noticing that non-y-smooth numbers have one and only one prime divisor $p > y$, we get

$$\begin{aligned}\psi(x,y) &= \lfloor x \rfloor - \sum_{y<p\leq x} \#\{m : m \leq x/p\} \\ &= \lfloor x \rfloor - \sum_{y<p\leq x} \left\lfloor \frac{x}{p} \right\rfloor \approx x - x \sum_{y<p\leq x} \frac{1}{p} \\ &= x \left(1 - \sum_{2\leq p\leq x} \frac{1}{p} + \sum_{2\leq p\leq y} \frac{1}{p} \right)\end{aligned}$$

Now, by the Mertens formula,

$$\begin{aligned}\psi(x,y) &\approx x(1 - (\log\log x - \log\log y)) \\ &\approx x\left(1 - \log \frac{\log x}{\log y}\right) = x(1 - \log u) = x\rho(u)\end{aligned}$$

This step wasn't necessary but is good warming up for the proof

Suppose $\psi(x, x^{1/u}) \sim x\rho(u)$ holds for $0 \le u \le N$.

Consider values of $u \in (N, N+1]$.

Subtracting the Buchstab-de Bruijn relation with $y = x^{1/N}$:

$$\psi(x, x^{1/N}) = 1 + \sum_{p \le x^{1/N}} \psi\left(\frac{x}{p}, p\right)$$

from the same equation with $y = x^{1/u}$:

$$\psi(x, x^{1/u}) = 1 + \sum_{p \le x^{1/u}} \psi\left(\frac{x}{p}, p\right),$$

we obtain

$$\begin{aligned} \psi(x, x^{1/u}) &= \psi(x, x^{1/N}) - \sum_{x^{1/u} < p \le x^{1/N}} \psi\left(\frac{x}{p}, p\right) \\ &\approx x\left(\rho(N) - \sum_{x^{1/u} < p \le x^{1/N}} \frac{1}{p}\rho\left(\frac{\log(x/p)}{\log p}\right)\right). \end{aligned}$$

since

$$\frac{\log(x/p)}{\log p} = \frac{\log x}{\log p} - 1 < \frac{\log x}{\log(x^{1/u})} - 1 = u - 1 \le N,$$

so the induction hypothesis applies.

Let

$$\vartheta(z) = \sum_{p \le z} \log p.$$

By the prime number theorem

$$\vartheta(z) = z + O(z/(\log z)^A)$$

for any fixed A.

Writing $z = x^{1/t}$, by partial summation, we get

$$\begin{aligned} \sum_{x^{1/u} < p \le x^{1/N}} \frac{1}{p} \rho\left(\frac{\log(x/p)}{\log p}\right) &= \sum_{x^{1/u} < p \le x^{1/N}} \frac{\vartheta(p) - \vartheta(p-1)}{p \log p} \rho\left(\frac{\log(x/p)}{\log p}\right) \\ &= \int_{x^{1/u}}^{x^{1/N}} \rho\left(\frac{\log x}{\log z} - 1\right) \frac{d\vartheta(z)}{z \log z} \\ &\approx \int_{x^{1/u}}^{x^{1/N}} \rho\left(\frac{\log x}{\log z} - 1\right) \frac{dz}{z \log z} \\ &= \int_N^u \rho(t-1) \frac{dt}{t}, \end{aligned}$$

Therefore

$$\psi(x, x^{1/u}) \approx x \left(\rho(N) - \int_N^u \rho(t-1) \frac{dt}{t}\right) = \rho(u) x$$

Sieve method

This is how **not** to count:

A general purpose sieve method gives an upper bound

$$\psi(x,y) = O(x/u)$$

which is very weak.

How Can We Evaluate $\psi(x,y)$?

To optimise and balance many cryptographic algorithms, need more precise information about $\psi(x,y)$ than estimates and asymptotic formulas provide.

Hunter and Sorenson, **1997,2000**:

One can approximate $\psi(x,y)$ up to a factor

$$1 + O\left(\frac{1}{u} + \frac{\log y}{y}\right)$$

in time

$$O\left(\frac{y \log\log x}{\log y} + \frac{y}{\log\log y}\right).$$

Several more results: *Sorenson,* **2000**:
Bernstein, **2001**:
Suzuki, **2004,2006**:
Parsell and Sorenson, **2006**:
to be continued . . .

Constructing Smooth Numbers

It is easy to produce a smooth number, e.g. 2^k is such.

What if one needs a smooth number close to the target value x?

Boneh, **2001**:

Efficient **constructions** of y-smooth numbers in short intervals $[x, x+z]$ for some relations between x, y and z.

More research here would be very welcome . . .

For Those Who Want it Rough

Let $\Theta(x,y)$ be the number of $n \leq x$ which are y-*rough*, that is, all prime divisors $p|n$ satisfy $p > y$.

Buchstab, **1949**:

$$\Theta(x,y) \sim \omega(u)\frac{x}{\log y},$$

where $\omega(u) = 1/u$ for $1 \leq u \leq 2$ and

$$u\omega(u) = 1 + \int_1^{u-1} \omega(t)dt \quad \text{for all} \quad u \geq 2.$$

Note that:

- $(u\omega(u))' = \omega(u-1)$;

- $\lim_{u\to\infty} \omega(u) = e^{-\gamma}$.

Large Smooth Divisors

Banks and Shparlinski, **2007**:
Tenenbaum, **2007**:

Asymptotic formula for

$\Theta(x,y,z) = \#\{n \le x \mid \exists d \mid n,\ d > z,\ d$ is y-smooth$\}$

in a wide range of parameters x, y, z.

Other Prime Divisors

Let $P_k(n)$ be the k-th largest prime divisor of n.

Billingsley, **1972**:
Tenenbaum, **2000**:

Joint distribution

$$\psi(x, y_1, \ldots, y_k) = \#\{n \le x \mid P_j(n) \le y_j,\ j = 1, \ldots, k\}.$$

The case of $k = 2$ is especially important:

Lenstra, **1987**:
The elliptic curve factorisation algorithm which factors an integer n in time $\exp\left(2\sqrt{\log p \log\log p}\right)$ where $p = P_2(n)$.

Variations and Open Questions

Smooth Numbers in Arithmetic Progressions:

Let

$$\psi(x,y;a,q) = \#\{n \le x \mid n \text{ is } y\text{-smooth, } n \equiv a \bmod q\}$$

and

$$\psi_q(x,y) = \#\{n \le x \mid n \text{ is } y\text{-smooth, } \gcd(n,q) = 1\}$$

Balog and Pomerance, **1992**:
Granville, **1993**:
Fouvry and Tenenbaum, **1996**:
Harman, **2001**:

In many cases, the situation is at about the same level as for $\psi(x,y)$. For $\gcd(a,q)=1$,

$$\psi(x,y;a,q) \sim \frac{1}{\varphi(q)}\psi_q(x,y) \quad \text{and} \quad \psi_q(x,y) \sim \frac{\varphi(q)}{q}\psi(x,y)$$

or a little weaker

$$\psi(x,y;a,q) \asymp \frac{1}{\varphi(q)}\psi_q(x,y) \quad \text{and} \quad \psi_q(x,y) \asymp \frac{\varphi(q)}{q}\psi(x,y)$$

are known in wide ranges of x, y and q.

Smooth Numbers in Small Intervals

Define $\psi(x,y,z)$ as the number of y-smooth $n \in [x, x+z]$:

$$\psi(x,y,z) = \psi(x+z,y) - \psi(x,y)$$

It is expected that

$$\psi(x,y,z) \sim \rho(u)z$$

in a wide range. It is known in some ranges but not in general.

Hildebrad, **1986**:
Balog, **1987**:
Friedlander and Lagarias, **1987**:
Harman, **1991**:
Friedlander and Granville, **1993**:

Some results . . . but the main challenge —

$$\psi(x, \exp\left(2\sqrt{\log x \log\log x}\right), 2x^{1/2})$$

is out of reach. This case is crucial for analysis of the **elliptic curve factoring**.

Lenstra, Pila and Pomerance, **1993**:
The current knowledge is enough to analyse rigorously the hyperelliptic smoothness test (larger intervals: $\psi(x,y,Cx^{3/4})$ is already doable for rather "small" y!).

Smooth k-Tuples

Balog and Wooley, **1998**: *Konyagin,* **2000**:

For any k and $\varepsilon > 0$ there are infinitely many n such that $n+i$ is n^{ε}-smooth for $i = 1, \ldots, k$.

Very nice and elementary explicit constructions. One can take $k \to \infty$ and $\varepsilon \to 0$ (slowly) when $n \to \infty$.

Smooth Partitions

Balog, **1989**:

Each sufficiently large integer N can be written as $N = n_1 + n_2$ where n_1, n_2 are $N^{0.2695}$-smooth.

Smooth Shifted Primes

Let

$$\pi(x,y) = \#\{p \le x \mid p-1 \text{ is } y\text{-smooth}\}.$$

It is strongly believed that that

$$\pi(x,y) \sim \rho(u)\pi(x)$$

(out of reach).

Pomerance and Shparlinski, **2002**:

$$\pi(x,y) = O\left(u\rho(u)\pi(x)\right)$$

for

$$\exp\left(\sqrt{\log x \log\log x}\right) \le y \le x$$

In a shorter range we have the "right" upper bound

Fouvry and Tenenbaum, **1996**:

$$\pi(x,y) = O\left(\rho(u)\pi(x)\right)$$

for

$$\exp\left((\log x)^{2/3+\varepsilon}\right) \le y \le x$$

Friedlander, **1989**:

$$\pi(x,y) \ge C\pi(x)/\log x$$

for $u \le 2\sqrt{e} = 3.2974\ldots.$

Baker and Harman, **1998**:

- $$\pi(x,y) \ge C\pi(x)/\log^A x$$

  for $u \le 3.377\ldots.$

- $$\pi(x) - \pi(x,y) \ge C\pi(x)/\log^A x$$

  for $u \ge 1.477\ldots.$

Equivalent Form

There are $C\pi(x)/\log^A x$

- primes $p \le x$ such that $p-1$ has a prime divisor $q \ge p^{0.6776}$

  Results of these type play a central role in deterministic primality test of *Agrawal, Kayal, Saxena,* **2004**:

- primes $p \le x$ such that all prime divisors q of $p-1$ satisfy $q \le p^{0.2962}$.

The above two statements are expected to be true with $A = 0$ and with $1-\varepsilon$ instead of 0.6776 and ε instead of 0.2962, respectively (for any $\varepsilon > 0$).

Smooth Values of the Euler Function

Let

$$\Pi(x, y) = \#\{p \le x \mid \varphi(p-1) \text{ is } y\text{-smooth}\}.$$

and

$$\Phi(x, y) = \#\{n \le x \mid \varphi(n) \text{ is } y\text{-smooth}\}.$$

Banks, Friedlander, Pomerance and Shparlinski, **2003**:
For $(\log\log x)^{1+\varepsilon} \le y \le x$, we have

$$\Phi(x, y) \le x \exp(-(1 + o(1))\, u \log\log u)$$

Remark: Mind $\log\log u$ rather than $\log u$ in the exponent and mind the very wide range. Recall that $\rho(u) = \exp(-(1 + o(1))\, u \log u)$ and the “typical” range for $\psi(x, y)$ starts with $y \ge (\log x)^{1+\varepsilon}$.

How tight is this?
Lamzouri, **2007**: Under some plausible conjecture, there is a matching lower bound.

We can now simply use the trivial inequality $\Pi(x,y) \le \Phi(x,y)$.

Other bounds:

For $\exp\left(\sqrt{\log x \log\log x}\right) \le y \le x$ we have

$$\Pi(x,y) \ll u^{-1}\pi(x).$$

For $\log x \le y \le x$, we have

$$\begin{aligned}\Pi(x,y) \;\; &\le \;\; \frac{\pi(x)}{\exp((\frac{1}{2}+o(1))u^{1/2}\log u)} \\ &\qquad +\frac{\pi(x)\log\log x}{\exp((1+o(1))u\log u)}\end{aligned}$$

Dream Result:

$$\Pi(x,y) \ll \pi(x)\exp(-(1+o(1))\,u\log\log u) \quad ???$$

Smooth Values of Polynomials

$$\psi_f(x,y) = \#\{n \le x \mid |f(n)| \text{ is } y\text{-smooth}\}.$$

Martin, **2001**:

Conjecture: Let f be squarefree and let $d_1, d_2, \ldots, d_k$ be the degrees of irreducible factors of f over $\mathbf{Z}[x]$,

$$\psi_f(x,y) \sim \rho(d_1 u)\rho(d_2 u)\ldots\rho(d_k u)x$$

(out of reach).

Some rigorous **upper** bounds are known (and some **lower** bounds for very small $u = O(1)$).

Smooth Polynomials

A polynomial $F \in \mathbb{F}_q[x]$ is k-smooth if all irreducible divisors $f|F$ satisfy $\deg f \leq k$.

$$N_q(m,k) = \#\{f \in \mathbb{F}_q[x] \mid \deg f \leq m, \\ f \text{ is monic and } k\text{-smooth }\}.$$

Define

$$u = \frac{m}{k} = \frac{\log q^m}{\log q^k}$$

Odlyzko, **1985**:
Bender and Pomerance, **1998**:

$$N_q(m,k) = \rho(u) q^m \exp\left(O\left(\frac{m \log k}{k^2}\right)\right)$$

(if $k \geq m^{1/2} \log m$ it is an asymptotic formula).

$$N_q(m,k) = u^{-u+o(u)} q^m$$

for $q^k \geq m \log^2 m$.

$$N_q(m,k) \geq m^{-u} q^m$$

for $k \leq m^{1/2}$.

Distribution of Divisors

More About Intuition

It is obvious that the density of perfect squares $n = d^2$ is extremely small as there are only about $\sim x^{1/2}$ perfect squares up to x.

Let's relax the relation $n = k^2$ and consider $n = km$ with $k \leq m \leq k^{1.001}$. Such integers can be called "almost" squares.

Question: Is the density of "almost" squares small? Are there only $o(x)$ of "almost" squares up to x?

Answer: **NO!**

"Almost" squares occur with positive density.

Notation

Given a sequence of integers $\mathcal{A} = (a_n)$ we denote

$$H(x, y, z; \mathcal{A}) = \#\{n \le x \ : \ \exists\, d | a_n \text{ with } y < d \le z\}.$$

The following sequences $\mathcal{A}$ are of our primal interest:

- $\mathcal{A} = \mathbb{N}$, natural numbers
- $\mathcal{A} = \mathcal{P}_a = \{p + a \ : \ p \text{ prime}\}$, shifted primes
- $\mathcal{A} = \mathcal{F}_f = \{f(n) \ : \ n = 1, 2 \ldots\}$, where $f \in \mathbf{Z}[X]$, polynomial sequences
- $\mathcal{A} = \Phi = \{\varphi(n) \ : \ n = 1, 2 \ldots\}$, values of the Euler function

Natural Numbers

This case goes back to two old questions of Erdős:

> Given an integer N what is the size of the *multiplication table* $\{nm \ : \ 1 \le m, n \le N\}$.

> Show that almost all n have two divisors $d_1 < d_2 < 2d_1$.

Erdős, Ford, Hall, Hooley, Maier, Saias, Tenenbaum . . . , **1980 − ???**:

Many various results, upper and lower bounds on $H(x, y, z, \mathbb{N})$, depending on relative sizes of x, y, z as well as of $z - y$ and z/y.

A sample result (will be used later)

Define $v > 0$ by the relation

$$z = y^{1+1/v}$$

Then, if

$$2y \le z \le \min\{y^{3/2}, x^{1/2}\}$$

then

$$\exp(-c\sqrt{\log v \log\log v}) \le \frac{H(x,y,z,\mathbb{N})}{xv^{-\delta}} \le \frac{\log\log v}{\sqrt{\log v}}$$

where $c > 0$ is an absolute constant and

$$\delta = 1 - \frac{1 + \log\log 2}{\log 2} = 0.008607\ldots$$

is the *Erdős number*.

Special case: For any $\varepsilon > 0$,

$$c_1(\varepsilon)x \le H(x,y,y^{1+\varepsilon},\mathbb{N}) \le c_2(\varepsilon)x$$

where $O < c_1(\varepsilon) < c_2(\varepsilon) < 1$

Equivalent form: The set of integers $n \le x$, which have a divisor $d \in [y, y^{1+\epsilon}]$, is of positive density (depending only on $\varepsilon > 0$).

Let's prove something . . .

Special-special case:

For $0 \le \alpha < \beta \le 1$:

$$x \ll H(x, x^{\alpha}, x^{\beta}, \mathbb{N}) \ll x$$

It is enough to consider $0 < \alpha < \beta < 1/2$ (since if $d|n$ then $(n/d) \mid n$).

Consider only **prime divisors** $p \in [x^{\alpha}, x^{\beta}]$.

- There are $x/p + O(1)$ integers $n \le x$ divisible by p
- Each $n \le x$ may have at most $K = \left\lceil \alpha^{-1} \right\rceil$ of prime divisors $p \ge x^{\alpha}$.

The sum

$$\sum_{x^\alpha \le p \le x^\beta} \left(\frac{x}{p} + O(1) \right)$$

count every integer $n \le x$ with a prime divisor $p \in [x^\alpha, x^\beta]$ at most K times.

$$\Downarrow$$

$$\begin{aligned} H(x, x^\alpha, x^\beta, \mathbb{N}) &\ge \frac{1}{K} \sum_{x^\alpha \le p \le x^\beta} \frac{x}{p} + O(x^\beta) \\ &= \frac{x}{K} \sum_{x^\alpha \le p \le x^\beta} \frac{1}{p} + O(x^\beta) \end{aligned}$$

By the Mertens formula

$$\begin{aligned} H(x, x^\alpha, x^\beta, \mathbb{N}) &\ge \frac{x}{K} \Big(\log\log(x^\beta) - \log\log(x^\alpha) + o(1) \Big) \\ &= \frac{x}{K} \left(\log \frac{\log(x^\beta)}{\log(x^\alpha)} + o(1) \right) \\ &\sim \frac{\log(\beta/\alpha)}{K} x \end{aligned}$$

Shifted Primes

Ford, **2007**:

Upper bounds on $H(x,y,z;\mathcal{P}_a)$ of the same strength as for $H(x,y,z;\mathbb{N})$.

Lower bounds are much weaker although heuristically there is little doubt that $H(x,y,z;\mathcal{P}_a)$ behaves similarly to $H(x,y,z;\mathbb{N})$.

One of the very few known lower bounds (yet, with many important applications to cryptography) is due to *Ford,* **2007**:
For $a \neq 0$ and $0 < \alpha < \beta$:

$$c_1(\varepsilon)\pi(x) \ll H(x,x^\alpha,x^\beta,\mathcal{P}_a) \le c_2(\varepsilon)\pi(x)$$

The proof follows the same path as our previous proof, but needs rather deep tools from the analytic number theory, the *Bombieri–Vinogradov* theorem:

Instead of **integers** $n \le x$ with $p \mid n$ we need to count **primes** $q \le x$ with $p \mid q - a$.

Polynomials

I wish I could say something here ...

However, it is not hopeless. It is just needs more attention, and fully deserves it!

Euler Function

Here is just yet another confirmation that **totients** are not typical integers.

As we have mentioned, $H(x, y, z; \mathcal{P}_a)$ is expected to behave similarly to $H(x, y, z; \mathbb{N})$.

However the behaviour of $H(x, y, z; \Phi)$ is very different! Totients have larger/denser divisor sets.

Ford and Hu, **2007**:

- Uniformly over $1 \le y \le x/2$, we have $H(x, y, 2y; \Phi) \gg x$.
- For $y = x^{o(1)}$, we have $H(x, y, 2y; \Phi) \sim x$.
- For a positive proportion of integers n, there is a divisor $d \mid \varphi(n)$ in every interval of the form $[K, 2K]$, $1 \le K \le n$.

Applications

Primality, Factorisation, Dlog

90% of applications are in these areas.

90% of this talk is about other applications.

Examples:

- Dixon's Method
- Quadratic Sieve
- Number Field Sieve
- Index Calculus
- Elliptic Curve Factoring

Some are rigorously analysed, some are heuristic (but based on our *understanding* (???) of smooth numbers)

Index Calculus in $\mathbb{F}_p^*$

Initial Assumption

Let us fix some y (to be optimised later) and **assume** that we know **Dlog**'s of **all** primes $p_1, \ldots, p_s$ up to y.

To compute k from $b \equiv a^k \bmod p$ we

- take a random integer m and compute

  $$c \equiv ba^m \equiv a^{k+m} \bmod p$$

  Note that

  $$\mathrm{Dlog}_a\, c = \mathrm{Dlog}_a\, b + \mathrm{Dlog}_a\, a^m = \mathrm{Dlog}_a\, b + m$$

  Cost: negligible

- Try to factor c, assuming that c, treated as an integer, is y-smooth and try to factor c as

  $$c = p_1^{\alpha_1} \ldots p_s^{\alpha_s}$$

  factors, by using the brute force trial division.

  Note that

  $$\mathrm{Dlog}_a\, c = \alpha_1 \mathrm{Dlog}_a\, p_1 + \ldots + \alpha_s \mathrm{Dlog}_a\, p_s$$

  Cost: About y operations

- If the previous step succeeds, output

  $$\mathsf{Dlog}_a b = \alpha_1 \mathsf{Dlog}_a p_1 + \ldots + \alpha_s \mathsf{Dlog}_a p_s - m,$$

  otherwise repeat the first step.

  Cost: About $u_p^{u_p}$ repetitions, where $u_p = \frac{\log p}{\log y}$ (under the assumption that c is a random integer up to p).

Total Cost: $y u_p^{u_p}$

Taking $y = \exp\left(\sqrt{\log p \log\log p}\right)$ we get an algorithm of complexity about

$$\exp\left(2\sqrt{\log p \log\log p}\right)$$

. . . but it is too early to celebrate yet.

Removing the Assumption

We apply the same algorithm for each p_i as b. Then at the 3rd step we get an equation

$$\mathrm{Dlog}_a p_i = \alpha_{1,i}\mathrm{Dlog}_a p_1 + \ldots + \alpha_{s,i}\mathrm{Dlog}_a p_s - m_i$$

We cannot find $\mathrm{Dlog} p_i$ immediately

...but after we have this relations for every p_i we have a system of s linear equations with s variables!!

Cost: About $s^3 \leq y^3$ or even less — still subexponential!!

This algorithm (due to Andrew Odlyzko, AT&T, 1967) has the overall **subexponential** complexity about

$$\exp\left(c\sqrt{\log p \log\log p}\right)$$

for some constant c.

Nowadays there is an algorithm, **Number Field Sieve**, of complexity

$$\exp\left(c(\log p)^{1/3}(\log\log p)^{2/3}\right)$$

Comments

The above approach belongs to the generic class of so-called **index calculus** algorithms.

However it does not work for elliptic curves:

- What are primes points?
- What are small points?
- What is the prime number factorizations?

Igor Semaev, Pierrick Gaudry, Claus Diem etc., **2005–??**: some attempts to address these questions (and some real achievements!!).

In fact even extension to general finite field is not easy (and is **not** known in full generality).

For fields of small characteristic such as $\mathbb{F}_{2^n} = \mathbb{F}_2[X]/f(X)$, where $f \in \mathbb{F}_2[X]$ is irreducible of degree n, irreducible polynomials of small degree play roles of small primes. Thus counting smooth polynomials in finite fields becomes very important.

“Text-book” ElGamal

Boneh, Joux and Nguyen, **2000**:

<u>ElGamal Scheme</u> Primes p, q with $q|p-1$

$g \in \mathbb{F}_p$ of order q.

Private Key: $x \in \mathbf{Z}_q$

Public Key: $X = g^x$

Encryption of a Message μ:
For a random $r \in \mathbf{Z}_q$, compute $R = \mu X^r$ and $Q = g^r$, send $C = (R, Q) = (\mu X^r, g^r)$

Decryption:
Compute $S = Q^x = g^{xr} = X^r$ and $R/S = R/X^r = \mu$

$$\boxed{\text{Assume that } \mu \text{ is small}}$$

E.g. μ is a key for a private key cryptosystem (e.g. p is 500 bits long, μ is 80 bits long).

Attack

We have $R = \mu U$ where $U \in \mathcal{G}_q$, the subgroup of $\mathbb{F}_p^*$ of order q.

Let $1 \le \mu \le M$.

- Compute $R^q = \mu^q U^q = \mu^q$;
- Choose some bound B and for $m = 1, \ldots, B$ compute, sort and store m^q;
- For $k = 1, \ldots, M/B$ compute $R^q/k^q = (\mu/k)^q$ and check whether they are in the table;
- Output $\mu = km$ if there is a **match**.

The Algorithm works with:

- $B = M$ for all messages (trivial; e.g., $m = \mu$, $k = 1$)

- $B = M^{1/2+\varepsilon}$ for a positive proportion of messages (nontrivial; it works because with a positive probability a random integer μ has a representation $\mu = km$ with $1 \le k \le m \le \mu^{1/2+\varepsilon}$).

Example: $M = 2^{80}$ (standard key size for a private key cryptosystem). The attack runs in a little more than 2^{40} steps.

Desmedt-Odlyzko Attack

RSA Signature Scheme:

N = RSA modulus
e = public exponent
d = private exponent; $ed \equiv 1 \pmod{\varphi(N)}$

Message: m Signature: $s \equiv m^d \pmod{N}$
The pair (m, s) is sent

Verification: : $m \equiv s^e \pmod{N}$

Desmedt and Odlyzko, **1985**:
Existential Forgery Attack: we are allowed to ask for signatures on some "allowed" message (e.g. padded in a prescribed way), and them we must produce a signature on one more "allowed" message.

- Select a bound y and let $p_1, \ldots, p_k$ be the primes up to y, i.e. $k = \pi(y)$.

- Take $k + 1$ messages m_i which are y-smooth and factor them $m_i = \prod_{j=1}^{k} p_j^{\alpha_{i,j}}$

- Solve in $u_1, \ldots, u_k \in \{0, \ldots, e - 1\}$

$$\sum_{i=1}^{k} \alpha_{i,j} u_i \equiv \alpha_{k+1} \pmod{e}, \quad j = 1, \ldots, k,$$

  and the write

$$\sum_{i=1}^{k} \alpha_{i,j} u_i + \gamma_i e = \alpha_{k+1}, \quad j = 1, \ldots, k.$$

  Thus

$$m_{k+1} \equiv r^e \prod_{i=1}^{k} m_i^{u_i} \pmod{N}$$

  where

$$r = \prod_{j=1}^{k} p_j^{\gamma_j}.$$

- Ask for the signatures s_i on m_i for $i = 1, \ldots, k$ and forge the signature on m_{k+1} as

$$s \equiv r \prod_{i=1}^{k} s_i^{u_i} \equiv r \prod_{i=1}^{k} m_i^{du_i}$$

It is a **valid signature** since:

$$\begin{aligned} s^e &\equiv r^e \prod_{i=1}^{k} m_i^{edu_i} \\ &\equiv r^e \prod_{i=1}^{k} m_i^{u_i} \equiv m_{k+1}^d \pmod{N} \end{aligned}$$

Coppersmith, Coron, Grieu, Halevi, Jutla, Naccache and Stern, **2007**:
Improvements, generalisations, concrete applications

Generalised Diffie-Hellman Problem

DH Assumption: Given g^x with some "hidden" integer x, it is hard to compute g^{x^2}.

Recently, several cryptographic schemes have appeared which base their security on the following assumption:

Let g be an element g of prime order p of a "generic" Abelian group $\mathcal{G}$.

Generalised DH Assumption: Given n powers $g^x, \ldots g^{x^n}$ with some "hidden" integer x, it is hard to compute $g^{x^{n+1}}$.

A "generic" attack (e.g. Shanks or Pollard algorithms) take about $p^{1/2}$ operations.

Brown, Gallant, **2006**:
and, in more detail, *Cheon,* **2006**:

- Given g^x and g^{x^d} for some $d \mid p-1$, one can find x in time about $(p/d)^{1/2} + d^{1/2}$ (which is $O(p^{1/4})$ for $d \sim p^{1/2}$).

- Given $g^x, \dots, g^{x^d}$ for some $d \mid p+1$, one can find x in time about $(p/d)^{1/2} + d$ (which is $O(p^{1/3})$ for $d \sim p^{1/3}$).

Question: How often primes p are such that $p \pm 1$ has a divisor d of a give size?

More specifically:

Question: How often primes p are such that $p \pm 1$ has a divisor $d \in [n^{1-\varepsilon}, n]$ (which will guarantee the maximal advantage if we are given $g^x, \dots, g^{x^n}$).

Ford, **2006**:

For every $\varepsilon > 0$ this happens for a positive proportion of primes p.

Moral: The conditions for this attack are satisfied with a positive probability!! The new problem is weaker than the traditional Diffie-Hellam problem.

Fix-Padded RSA

N = n-bit RSA modulus.

"Text-book" RSA signature scheme:

Message m $\Longrightarrow$ Signature $s \equiv m^d \bmod N$

Verification: $s^e \equiv m \bmod N$ — ???

Chosen Message Attack

Assume that the attacker wants to sign an important message m and has an ability to ask a *demo version* to decrypt some innocent messages.

The attacker:

- chooses a random m_1 and computes m_2 from $m_1 m_2 \equiv m \bmod N$ (and gets to (meaningless) messages m_1 and m_2).
- asks the demo version to sign $s_i \equiv m_i^d \bmod N$
- computes $s \equiv s_1 s_2 \bmod N$

This works because

$$s \equiv s_1 s_2 \equiv m_1^d m_2^d \equiv (m_1 m_2)^d \equiv m^d \bmod N$$

RSA is *homogeneous*:
A relation between messages implies a relation between signatures.

Defence:
Allow the signature/verfication algorithms to work only for messages of special structure, e.g., ending with some function of the message itself or say with 100 binary digits of π:
m_1 and m_2 are not likely to be of this type $\Longrightarrow$ the attack fails.

Fixed-pattern padding scheme:

fixed $n-\ell$-bit padding P	ℓ-bit message m

$$m \to P + m = R(m), \quad s(m) \equiv R(m)^d \bmod N$$

Some existing standards still use this scheme.

Misarsky, **1997**:
Girault and Misarsky, **1997**:
Brier, Clavier, Coron and Naccache, **2001**:

$\boxed{\text{Existential forgery}}$

that is, the attacker can sign **some** message.

Lenstra and Shparlinski, **2002**:

$\boxed{\text{Selective forgery}}$

that is, the attacker can sign **any** message.

Idea of the Forgery

Find four distinct ℓ-bit messages $m_1, \ldots, m_4$ such that

$$R(m_1) \cdot R(m_2) \equiv R(m_3) \cdot R(m_4) \bmod N.$$

Then

$$s(m_1) \cdot s(m_2) \equiv s(m_3) \cdot s(m_4) \bmod N.$$

$\Longrightarrow$ signature on m_3 can be computed from signatures on m_1, m_2, m_4.

The above congruence is equivalent to

$$P(m_3 + m_4 - m_1 - m_2) \equiv m_1 m_2 - m_3 m_4 \bmod N.$$

With

$x = m_1 - m_3, \quad y = m_2 - m_3, \quad z = m_3 + m_4 - m_1 - m_2$

this becomes

$$(P + m_3)z \equiv xy \bmod N.$$

This congruence would be trivial to solve by we need "small" x, y and z about ℓ bits long

Let $\ell = (1/3 + \varepsilon)n$.

We start with the congruence

$$(P + s)z \equiv w \bmod N.$$

where $|s| \leq N^{1/3+\varepsilon}$ is given and the variables w and s satisfy where $|z| \leq N^{1/3}$ and $w \leq N^{2/3+2\varepsilon}$

Let R_i/Q_i denote the i-th continued fraction convergent to $(P + s)/N$. Then

$$\left|\frac{P+s}{N} - \frac{R_i}{Q_i}\right| \leq \frac{1}{Q_i Q_{i+1}}.$$

Define j by $Q_j < N^{1/3} \leq Q_{j+1}$.

Let $w = |(P + s)Q_j - NP_j|$ and $z = \pm Q_j$

$$w = NQ_j \left|\frac{P+s}{N} - \frac{R_j}{Q_j}\right| \leq N/Q_{j+1} < N^{2/3}.$$

and

$$|z| < N^{1/3}$$

For at least one choice of the sign ± we have

$$(P + s)z \equiv w \bmod N$$

- Choose a “random” r with $0 \le r < 0.5N^{\varepsilon}$ and find
$$w \equiv \left(P + m_3 - r \left\lfloor N^{1/3} \right\rfloor\right) z \bmod N$$
with $w < N^{2/3}$ (i.e., use $s = m_3 - r \left\lfloor N^{1/3} \right\rfloor$)

- Put $u = w + r \left\lfloor N^{1/3} \right\rfloor z$, thus
$$u \equiv (P + m_3) z \bmod N$$
and $u < N^{2/3+\varepsilon}$

- Try to use elliptic curve factorisation to factor u which runs in time $\exp\left(2\sqrt{\log p \log\log p}\right)$ where $p = P(u/P(u)) = P_2(u)$ (but terminate this steps if it takes too long).

- Try to find x, y with $u = xy$ and $x, y < N^{1/3+\varepsilon}$

- If successful, compute m_1, m_2, m_4, otherwise try another pair z, u

Why does it work?

Eventually we hit a reasonably good u:

- u is of the form $u = P(u)v$ where $P(v) = P_2(u)$ is small .

- u has a divisor $x \in [N^{1/3+\varepsilon/2}, N^{1/3+\varepsilon}]$

Heuristic run-time: $L_N(1/3, 1)$ which is substantially faster than

$$L_N(1/3, (128/27)^{1/3}) \approx L_N(1/3, 1.68),$$

where as usual

$L_N(\alpha, \gamma) = \exp((\gamma + o(1))(\log N)^{\alpha}(\log\log N)^{1-\alpha}).$

for $M \to \infty$.

Lenstra and Shparlinski, **2002**:
Selective forgery for 1024 RSA modulus.

Question: Find a way to use more signatures and thus extend the range of ℓ which can be attacked this way.

Note: The congruence

$$(P + m_3)z \equiv xy \bmod N$$

as any other congruence

$$F(x, y, z) \equiv 0 \bmod N$$

with a "generic" polynomial F is **not** likely to have a solution with

$$1 \le x, y, z \le N^{\alpha}$$

for $\alpha < 1/3$.

$$\Downarrow$$

$$\boxed{\text{One needs a relation involving more signatures}}$$

Large Subgroup Attack

Digital Signature Algorithm (**DSA**), uses two large primes p and q with $q \mid p-1$.

Suppose that p and q are selected for **DSA** using the following standard method:

- Select a random m-bit prime q;
- Randomly generate k-bit integers n until a prime $p = 2nq + 1$ is reached.

Menezes, **2007**:
The *Large subgroup attack* on some cryptographic protocols (e.g. HMQV) which contain **DSA** as their part. These attacks lead to the following

Question: What is the probability $\eta(k, \ell, m)$ that

$$n = \frac{p-1}{2q}$$

has a divisor $s > q$ which is 2^ℓ-smooth?

Banks and Shparlinski, **2007**:
(heuristically, assuming that shifted primes $p-1$ behave like “random” integers):

In the most interesting choice of parameters at the present time is $k = 863$, $\ell = 80$, and $m = 160$ (which produces a 1024-bit prime p), for which one expects that the attack succeeds with probality

$$\eta(863, 80, 160) \approx 0.09576 > 9.5\%$$

over the choices of p and q.

Smooth Orders

Let $l(n)$ be the order of 2 modulo n, gcd$(2, n) = 1$ (change 2 with your favourite integer $a \geq 2$). That is, $l(n)$ is the smallest integer k with

$$2^k \equiv 1 \pmod{n}.$$

Question: *Can we use $g = 2$ as the base for Diffie-Hellman, ElGamal and other exponentiation based cryptoschemes modulo n?*

Yes, but only if $l(n)$ is not smooth – mind **Pohlig-Hellman!**

Question: *Why would we want $g = 2$?*

Boneh and Venkatesan, **1996**:
Nice bit security properties
(and a little easier to compute).

Also remember **Pollard's** $p-1$ factorisation method: if $p|n$ with $l(n)$ smooth, n can be easily factored.

Let

$$L(x,y) = \#\{p \le x \mid l(p) \text{ is } y\text{-smooth}\}.$$

and

$$N(x,y) = \#\{n \le x \mid l(n) \text{ is } y\text{-smooth}\}.$$

Pomerance and Shparlinski, **2002**:
For $\exp\left(\sqrt{\log x \log\log x}\right) \le y \le x$, we have

$$L(x,y) \ll u\rho(u/2)\pi(x),$$

Banks, Friedlander, Pomerance and Shparlinski, **2003**:
For $\exp\left(\sqrt{\log x \log\log x}\right) \le y \le x$, we have

$$N(x,y) \le x \exp(-(1/2 + o(1))\, u \log\log u)$$

Remark: Mind $\log\log u$ rather than $\log u$ in the exponent. Recall that $\rho(u) = \exp(-(1+o(1))\, u \log u)$.

How tight are they?
Probably quite tight (but 1/2 should be 1 in both cases).

Pratt Tree

Assume that somebody wants to “sell” a large prime p, but the buyer requests a proof that p is prime indeed.

Here is a way to do this.
Pratt, **1975**:

- Ask the buyer to check that p is not a perfect power (easy!!).

- Produce a primitive root g modulo p and ask the buyer to check this. It is enough to verify that

  $$g^{p-1} \not\equiv 1 \pmod{p} \quad \text{and} \quad g^{(p-1)/q} \not\equiv 1 \pmod{p}$$

  for all prime divisors $q \mid p-1$, so the list of these primes q also must be supplied.

- Give a proof that each q on the above list is prime by iterating the above procedure.

The algorithm runs in polynomial time and in particular shows that PRIMES $\in$ NP (not so exciting nowdays as we know that PRIMES $\in$ P).

The whole algorithm can be viewed as a tree where each node contains a prime (with p as a root), with 2 at each leave.

The number of multiplication required by this algorithm is:

Pratt, **1975**: $O((\log p)^2)$. *Bayless,* **2007**: At least $C \log p$ for any $C > 1$ and almost all primes p.

This tree is called the **Pratt Tree**.

Question: *What is the size of this tree, e.g. the height, the number of nodes, the number of leaves, etc.*

Banks, Shparlinski, **2007**: The length $L(p)$ of the chain $p \mapsto P(p-1)$ is at least

$$(1+o(1))\frac{\log\log p}{\log\log\log p}$$

for almost all primes p.

Ford, Konyagin, Luca, **2008 (?)**: The height $H(p)$ of the Pratt Tree is at least

$$(\log p)^{0.9622} \gg H(p) \gg \log\log p$$

for almost all primes p.

Ford, Konyagin, Luca, **2008 (?)**: Heuristically

$$H(p) = e\log\log p + O(\log\log\log p)$$

for almost all primes.

Strong Primes

A prime p is *strong* if $p-1$ and $p+1$ have a large prime divisor, and $p-1$ has a prime divisor r such that $r-1$ has a large prime divisor.

If p is **not** strong then

1. either $p-1$ or $p+1$ are y-smooth;
2. (extra condition) or $p-1$ is divisible by a r^2 for a prime $r \geq y$;
3. or $\varphi(p-1)$ is y-smooth

For 1: Bounds on $\pi(x,y)$

For 2:

$$\sum_{r \geq y} \sum_{p \leq x, p \equiv 1 \text{ mod } r^2} 1 \leq \sum_{r \geq y} \frac{x}{r^2} = O(x/y)$$

For 3: Bounds on $\pi(x,y)$

What to Read?

General Theory

A. Granville, ‘Smooth numbers: Computational number theory and beyond’, *Proc. MSRI Conf. Algorithmic Number Theory: Lattices, Number Fields, Curves, and Cryptography, Berkeley 2000*, Cambridge Univ. Press, (to appear).

A. Hildebrand and G. Tenenbaum, ‘Integers without large prime factors’, *J. de Théorie des Nombres de Bordeaux*, **5** (1993), 411–484.

G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge Univ. Press, 1995.

Primality, Factorisation, Dlog

R. Crandall and C. Pomerance, *Prime numbers: A Computational perspective*, Springer-Verlag, Berlin, 2005.

Other Cryptographic Applications

Original papers
. . . or these notes

Postmortem

What Was This All About?

The goal was **not** to make you experts in analytic number theory but rather:

1. give you a glimpse of a beautiful and diverse world of integers;

2. give you a glimpse of a no less beautiful and diverse world of cryptanalysis;

3. give you some feelings on what can and what cannot be true;

4. give you some ideas of underlying methods and ideas;

5. show you that the intuition and heuristic guessing should be based on knowledge, not on the lack of it;

6. help you to be able to formulate your number theory questions, in a coherent form, using standard terminology;

7. give you a literature guide, your questions may have already been addressed in the literature and either have been answered (or known to be out of reach);

8. give you some names of the most distinguished experts in the area who might be asked for help.

If you

- ask your question in a coherent form;
- give some evidences that you have checked standard sources and did not find answers;
- ask a right person;

this may lead to a very interesting and useful collaboration and extend you network of co-drinkers

Numbers and You

Numbers are your friends:

- be relaxed dealing with them
- do not worry about making mistakes, they will excuse you, as friends always do;
- learn from your mistakes, you make the same mistake twice and they may not be your friends anymore. . .

Tutorial Problems

General Comments

This problems are designed to show you different aspects of the material which has been discussed. Some of them are easy, some of them require more efforts, some may potentially lead to more serious research projects.

Sore are of experimental nature, some are more theoretic, some are designed to teach you to develop heuristic prediction skill and avoid standard traps.

Some of them can be too difficult to answer, but even thinking about possible way to tackle them.

Feel free to email for help, clarifications of further problems anytime:

`igor@ics.mq.edu.au` .

Arithmetic Structure of Integers

1. Given a prime p, give a heuristic guess on the "probability" that $n^2 \equiv 1 \pmod{p}$.

2. Prove it!

3. Given a prime p, give a heuristic guess on the "probability" that $n^2 \equiv -1 \pmod{p}$.

4. Prove it!

5. Given a real y, give a heuristic prediction for the number of $n \leq x$ such that

   - $n^2 \equiv 1 \pmod{p}$ for all primes $p \leq y$,
   - $n^2 \equiv 1 \pmod{p}$ for at least one prime $p \leq y$.

6. Prove them for as large as possible values of y compared to x.

7. Given a real y, give a heuristic prediction for the number of $n \le x$ such that

- $n^2 \equiv -1 \pmod{p}$ for all primes $p \le y$,
- $n^2 \equiv -1 \pmod{p}$ for at least one prime $p \le y$.

8. Prove them for as large as possible values of y compared to x.

9. Is this true that if $\gcd(a, q) > 1$ then $\pi(x; q, a) = 0$?

10. We usually say that the "probability" that a "random" integer n is prime is about $1/\log n$. What is the expect number of Mersenne numbers $2^p - 1$ with prime exponents $p \le x$, which are prime?

11. What is the expect number of Fermat numbers $2^{2^n} + 1$ with integer exponents $n \le x$, which are prime?

Euler function

We say that m is a *totient* if there exists some integer n with $m = \varphi(n)$, where

$$\varphi(n) = \prod_{p^{\alpha_p} \| n} p^{\alpha_p - 1}(p-1) = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$$

is the Euler function ($p^{\alpha_p} \| n =$ "exact divisibility").

1. Find all odd totients.

2. Let $F(x)$ be the number of totients $m \leq x$, show that $x/2 + 1 \geq F(x) \geq (1 + o(1))x/\log x$.

3. Prove that $\varphi(n) \geq n^{1/2}$ for $n \geq 2$ and use it to desine a brute force algorithm to check whether a given m is a totient.

4. Search the internet, MathSciNet and/or number theory literature in order to find sharp and fully explicit lower bounds on $\varphi(n)$. Use them to improve your algorithm.

Let $N(m)$ be the number of n with $m = \varphi(n)$

5. Try to find m with $N(m) = 1$

Hint: Do not spend too much time on this . . .

6. Try to find m for which $N(m)$ is large (the larger the better).

7. Analyze the structure of m for which $N(m)$ is large. Try to predict what m are likely to lead to large values of $N(m)$ and verify your guess by constructing such champions m.

8. By Turán-Kubilius, a "typical" integer n has about $\log\log n$ distinct prime factors. Explain why it is reasonable to expect that "typically" $\varphi(n)$ has many more prime factors.

9. Estimate the probability $\vartheta(x, z)$ that for two randomly chosen primes $p, q \le x$ we have

$$\gcd(p - 1, q - 1) \ge z.$$

Hint: Write

$$\begin{aligned}
\vartheta(x,z) &= \frac{1}{\pi(x)^2} \sum_{\substack{p,q\le x\\ \gcd(p-1,q-1)\ge z}} 1 \\
&= \frac{1}{\pi(x)^2} \sum_{d\ge z} \sum_{\substack{p,q\le x\\ \gcd(p-1,q-1)=d}} 1 \\
&\le \frac{1}{\pi(x)^2} \sum_{d\ge z} \sum_{\substack{p,q\le x\\ p\equiv q\equiv 1 \pmod d}} 1 \\
&= \frac{1}{\pi(x)^2} \sum_{d\ge z} \pi(x;d,1)^2
\end{aligned}$$

Now choose some parameter Z and use the Brun–Titchmarsh theorem for $d \le Z$ and the trivial estimate $\pi(x;d,1) \le x/d$ for $d > Z$. Optimize Z.

Note that even taking $Z = z$ (that is, using the trivial estimate for all d) already gives a nontrivial result).

10. Use the results of the previous two problems to give a motivated guess what should be the number of distinct prime factors of $\varphi(n)$ for a "typical" integer n.

11. Verify your guess numerically, re-assess and adjust it, if necessary. Verify the new guess again.

12. Use inclusion-exclusion principle (expressed in terms of the Möbius function) and try to get an asymptotic formula for $\vartheta(x, z)$ (for the values of z as large as possible).

13. Use inclusion-exclusion principle (expressed in terms of the Möbius function) and try to get an asymptotic formula for $\vartheta(x, z)$ (for the values of z as large as possible).

Hint:

(a) The condition gcd$(a, b) = d$ can be expressed as $a = a_0d$, $b = b_0d$ where gcd$(a_0, b_0) = 1$. The last condition is equivalent to that gcd(a_0, b_0) is not divisible by any prime p.

(b) The result can now be expresses via some sums involving $\pi(x; d, 1)$ and the Möbius function (via the inclusion-exclusion principle).

(c) Use the Siegel–Walfisz theorem for $d \le Z$ and the Brun–Titchmarsh theorem for $d > Z$ where Z is a parameter to be optimised.

(d) Now try to improve the result by using the Bombieri–Vinogradov theorem instead of the Siegel–Walfisz theorem.

14. Check what the Extended Riemann Hypothesis gives for the previous problem.

Counting Smooth Numbers

1. Evaluate $\psi(100, 2)$. Give an exact formula for $\psi(100, 2)$.

2. Evaluate $\psi(100, 10)$. Try different approaches and discuss which one seems to be more efficient. Now evaluate $\psi(100, 9)$, $\psi(100, 8)$, $\psi(100, 7)$.

3. Using your results and conclusions from the previous problem, evaluate $\psi(100, 6)$.

4. Count the number of primes $p \leq x$ for which $p-1$ is 2-smooth, and the the number of primes $p \leq x$ for which $p+1$ is 2-smooth. Such primes are named after two distinguished mathematicians. What are these names?

Duality of Divisors

1. We have claimed that in order to prove that for $0 \le \alpha < \beta \le 1$ we have

$$x \ll H(x, x^\alpha, x^\beta, \mathbb{N}) \ll x$$

it is enough to consider $0 < \alpha < \beta < 1/2$.

Prove this!

2. Does our proof show that for any $0 \le \alpha < \beta \le 1$ a positive proportion of integers $n \le x$, there is a prime divisor $p \mid n$ with $x^\alpha \le p \le x^\beta$?

3. Prove that if

$$\sum_{d|m} d = \sum_{f|n} f$$

and

$$\sum_{d|m} \frac{1}{d} = \sum_{f|n} \frac{1}{f}$$

then $m = n$.

Algorithms and Smooth Numbers

H. W. Lenstra's elliptic curve factorisation algorithm of completely factors an integer n in time

$$\exp\left(2\sqrt{\log p \log\log p}\right)(\log n)^{O(1)}$$

where $p = P(n/P(n))$.

1. Considering the products of the shape $n = qs$ where $q \le Q$ and is prime and $s \le S$ is prime, obtain a lower bound on $F(T)$ integers $n \le N$ which can be factored in time T. Optimize Q and S over all choices with $QS \approx N$.

2. Work out the complexity of the Desmedt-Odlyzko Attack when all messages satisfy $m \le M$ (optimise the choice of y).

3. Design an algorithm to verify whether n is perfect power (that is, $n = m^k$ for some integers m and $k \ge 2$).

Number Theory Background

Notation

- $A \ll B$ or $B \gg A$ (I. M. Vinogradov)

$$\Updownarrow$$

$$A = O(B) \quad \text{(E. Landau)}$$

$\ll$ is more compact and easier to use and it admits more informative chains like

$$A \ll B = C$$

while

$$A = O(B) = C$$

is nonsense.

We also recall that

$$A \sim B \qquad \Leftrightarrow \qquad A \ll B \ll A$$

- p (with or without a subscript) always denotes a prime number

- $\varphi(n)$ is the Euler function;

  $\varphi(n) = \#\{1 \leq k < n \mid \text{gcd}(k,n) = 1\} = \#(\mathbf{Z}/n\mathbf{Z})^*$

- $\pi(x)$ is the number of primes $p \leq x$

- $\log x$ is the natural logarithm

- When we write $\log x$, $\log\log x$, etc. we always assume that the argument is large enough

Some Fasic Facts

- Prime Number Theorem

$$\mathbf{1.}\qquad \pi(x) = \mathrm{li}x + O\left(\frac{x}{(\log x)^K}\right)$$

for any fixed K, where

$$\mathrm{li}x = \int_2^x \frac{1}{\log t} d\,t$$

or

$$\mathbf{2.}\qquad \sum_{p \le x} \log p = x + O\left(\frac{x}{(\log x)^K}\right)$$

for any fixed K.

Warning: $\mathrm{li}x \sim x/\log x$ but

$$\pi(x) = \frac{x}{\log x} + O\left(\frac{x}{(\log x)^K}\right)$$

is **wrong!**

Remark: The best know result is due to N. M. Korobov and I. M. Vingogradov (and dates back to 1953), correct by H.-E. Richert and improved by K. Ford:

$$\pi(x) = \mathrm{li}x + O\left(x \exp\left(-0.2098\frac{(\log x)^{3/5}}{(\log\log x)^{1/5}}\right)\right)$$

- $\pi(x;q,a)$ = the number of primes $p \le x$ with $p \equiv a \pmod{q}$ (only $\gcd(a,q)=1$ should be considered).

- Siegel–Walfisz theorem

  For every fixed $A > 0$, there is $B > 0$ such that for all $x \ge 2$ and all positive integers $q \le \log^A x$,

$$\max_{\gcd(a,q)=1} \left| \pi(x;q,a) - \frac{\mathrm{li}x}{\varphi(q)} \right| \ll x \exp\left(-B\sqrt{\log x}\right).$$

- Brun–Titchmarsh theorem

  For all $x \ge 2$ and all positive integers q,

$$\pi(x;q,a) \ll \frac{x}{\varphi(q)\log(x/q)}$$

  (it is expected to hold with just $\log x$ instead of $\log(x/q)$).

- Bombieri–Vinogradov theorem:

  For every fixed $A > 0$, there is $B > 0$ such that

$$\sum_{q \le x^{1/2}(\log x)^{-B}} \max_{y \le x} \max_{\gcd(a,q)=1} \left| \pi(y;q,a) - \frac{\mathrm{li}y}{\varphi(q)} \right| \ll x(\log x)^{-A}.$$

- Euler function is "large":

$$n \geq \varphi(n) \gg \frac{n}{\log\log n}$$

- Mertens Formulas

1. $$\sum_{p \leq x} \frac{1}{p} = \log\log x + A + o(1), \qquad A = 0.2614\ldots$$

2. $$\sum_{p \leq x} \frac{\log p}{p} = \log x + B + o(1), \qquad B = 1.3325\ldots$$

3. $$\prod_{p \leq x} \left(1 - \frac{1}{p}\right) = \frac{C + o(1)}{\log x}, \quad C = e^{\gamma} = 1.7810\ldots$$

where $\gamma = 0.5772\ldots$ is the *Euler-Mascheroni* constant.

- For any complex number s with $\Re s > 1$ the Riemann Zeta-function is given by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

then it is analytically continued to all $s \in \mathbb{C}$.

- Riemann Hypotheis: All zeros of $\zeta(s)$ with $0 \le \Re s \le 1$ have $\Re s = 1/2$.

  Warning There are other *trivial* zeros outside of the *critical strip* $0 \le \Re s \le 1$.

- $\zeta(1+it)\zeta(it) \ne 0$ for every $t \in \mathbb{R}$

$$\Updownarrow$$

  Prime Number Theorem

- The best known result is due to Korobov and Vinogradov (independently, 1953)

- Generalised Riemann Hypotheis (**GRH**): the same is true for a much wider class of similar functions called L-functions (and even more general functions).

- The Riemann Hypothesis implies that

$$\pi(x) = \mathrm{li}x + O\left(x^{1/2}\log x\right).$$

- The **GRH** implies that

$$\pi(x; q, a) = \frac{\mathrm{li}x}{\varphi(q)} + O\left(x^{1/2}\log x\right), \qquad \gcd(q, a) = 1.$$

Remarks

- The Bombieri–Vinogradov theorem gives “on average” over q a result comparable with what the **GRH** implies.

- The Brun–Titchmarsh is **stronger** than the **GRH** for, say $q > x^{1/2}$.

- Euler Product:

$$\begin{aligned}\prod_p \left(1-\frac{1}{p^s}\right)^{-1} &= \prod_p \left(1+\frac{1}{p^s}+\frac{1}{p^{2s}}+\frac{1}{p^{3s}}+\dots\right)\\ &= \sum_{n=1}^{\infty}\frac{1}{n^s}=\zeta(s).\end{aligned}$$

- More generally, let $\mathcal{S}$ be any set of primes, and let $\mathcal{N}_{\mathcal{S}}$ be the set of integers whose all prime factors are from $\mathcal{S}$:

$$\prod_{p\in\mathcal{S}}\left(1-\frac{1}{p^s}\right)^{-1}=\sum_{n\in\mathcal{N}_{\mathcal{S}}}\frac{1}{n^s}$$