Anatomy of Integers and Cryptography

IGOR SHPARLINSKI
Department of Computing
Macquarie University
Sydney, NSW 2109, Australia
igor@comp.mq.edu.au

May 16, 2008

Abstract

It is well-known that heuristic and rigorous analysis of many integer factorisation and discrete logarithm algorithms depends on our various results about the distribution of smooth numbers. Here we give a survey of some other important cryptographic algorithms which rely on our knowledge and understanding of the multiplicative structure of "typical" integers and also "typical" terms of various sequences such as shifted primes, polynomials, totients and so on.

Part I

Introduction

1 Outline of the contents

1.1 Motivation

It is a common knowledge that number theory provides an enabling background for public key cryptography which used to design and analyze a variety of algorithms. Certainly algorithms for the integer factorisation and discrete logarithm problems are classical examples of the great importance.

However it seems that there are still many powerful results and techniques of analytic number theory which are now known and used widely enough by the cryptographers.

In Section 3 ee also provide a very brief and incomplete summary of some important number theoretic results which are commonly utilized in studying the multiplicative structure of integer numbers. These results can be found, in much more precise forms, in [17, 44, 60] and in a many other standard manuals. Some of them are directly used in this paper, some remain in the background.

We do not aim to replace the systematic study of analytic number theory, or even provide a "tool-box" of relevant results. Our purpose is much more modest. We intent to survey various presently available results and methods. This may assist in developing some taste and feeling of what kind of results can be obtained within our present techniques and which ones have a chance to be true at all. As well, it may help to develop correct heuristics for problems that are too complicated to be tackled theoretically.

On the other hand, number theorists may wish to learn about new important areas of applications of their skills and knowledge. Furthermore, final tuning and adjusting already knows results and techniques may lead to new advances of intrinsic mathematical interest. However our outline, somewhat sketchy and simplified and also sometimes ignoring subtleties, cannot replace a careful and systematic reading of the original cryptographic literature.

1.2 Number theory background

In the first part we give a very brief outline of some important facts about the multiplicative structure of integer numbers. For example, we address the following questions: Given a "typical" integer n what can we say about

- the largest prime divisor of n?
- the distibution of integer divisors of n?

We also discuss whether the answers to similar questions are much different for "typical" integers of cryptographic interest, such as

- shifted primes $p-1$;

- polynomial values $f(n)$;

- values of the Euler function.

We note that we do not attempt to give any systematic knowledge or even a complete survey. Rather we intent to give some taste of this area and provide a guide to the literature. In particular, many of the presented here results can be found with complete proofs in [17, 36, 37, 60] and many other manuals, we also recommend the surveys [35, 41].

1.3 Cryptographic applications

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

- Naive ElGamal protocol for private key exchange;

- Fix-padded RSA;

- Generalised Diffie-Hellman protocol;

- Pratt primality certificate;

- Using small exponentiation base;

- Strong primes for RSA.

As in the case of the number theoretic part, we only explain the meaning and importance of theses results and sketch the main underlying ideas.

2 Notation

2.1 General conventions

We recall that $A \ll B$ and $B \gg A$ are both equaivalent to $A = O(B)$. However the symbols '$\ll$' and '$\gg$'are more convenient to use as they are more compact and admits more informative chains like $A \ll B = C$ (while $A = O(B) = C$ is meaningless and $A = O(B) = O(C)$ may discard some useful information).

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

We use ε to denote a small positive parameter always allow all implied constants to depend on it.

We use $\log x$ to denote the natural logarithm and when we write $\log x$, $\log\log x$ and so on, we always assume that the argument is large enough

2.2 Arithmetic functions

As usual, for an integer $m \geq 2$, we use $P(m)$, $\omega(m)$, $\tau(m)$ and $\varphi(m)$ to denote the largest prime divisor, the number of distinct prime divisors, the number of positive integer divisors and the Euler function of m. We also put $P(1) = \omega(1) = 0$ and $\tau(1) = \varphi(1) = 1$.

For a real $x \geq 0$ we denote by $\pi(x)$ the number of primes $p \leq x$ and by $\pi(x; q, a)$ the number of primes $p \leq x$ with $p \equiv a \pmod q$.

2.3 Sequences of interest

The following sequences $\mathcal{A} = (a_n)$ of integer numbers 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 $f \in \mathbb{Z}[X]$, polynomial sequences
- $\mathcal{A} = \mathit{\Phi} = \{\varphi(n) \ : \ n = 1, 2 \ldots\}$, values of the Euler function

2.4 Smooth numbers

Colloquially, an integer n is *smooth* if it has only small prime divisors.

In a more precise quantitative form, we say that n is *y-smooth* if all prime divisors $p \mid n$ satisfy $p \le y$.

Alternatively, n is y-smooth if and only if $P(n) \le y$.

For a sequence $\mathcal{A} = (a_n)$ of integer numbers we use $\psi(x, y; \mathcal{A})$ to denote the number $n \le x$ for which a_n is y-smooth.

We also denote for brevity

$$\begin{aligned} \psi(x,y) &= \psi(x,y;\mathcal{N}), \\ \pi_a(x,y) &= \psi(x,y;\mathcal{P}_a), \\ \psi_f(x,y) &= \psi(x,y;\mathcal{F}_f), \\ \Phi(x,y) &= \psi(x,y;\mathit{\Phi}). \end{aligned}$$

2.5 The Dickman–de Bruijn function

The following function $\rho(u)$, known as the *he Dickman–de Bruijn function* plays a prominent role in investigating smooth numbers.

It is defined by recursively by the relations:

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

and

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

Here we summarise some properties of $\rho(u)$

We recall that

$$\rho(u) = u^{-u+o(u)}, \qquad u \to \infty \tag{1}$$

and more precisely

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

We also have

$$\rho(u) = 1 - \log u, \qquad 1 \le u \le 2.$$

For example, $\rho(e^{1/2}) = 1/2$, that is, $\sim$ %50 of integers n have all prime divisors up to $n^{1/e^{1/2}}$. This has been used by I. M. Vinogradov, and then by D. A. Burgess, to estimate the smallest quadratic non-residue modulo a prime p.

3 Some Fasic Number Theory Facts

3.1 Distribution of prime numbers

We recall that by the *prime number theorem* we have

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

for any fixed K, where

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

or, in an more common Nowadays it is more commonly formulated in the following fully equivalent form

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

(again for any fixed K) in terms of the function

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

A very commonly committed crime against primes is the assertion that

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

which is *wrong*, although, of course,

$$\pi(x) \sim \mathrm{li}x \sim x/\log x.$$

An asymptotic formula for the number of primes in arithmetic progression is given by the *Siegel–Walfisz theorem* . For every fixed $A > 0$, there is $B > 0$ such that for all $x \ge 2$ and all positive integers $q \le \log^A x$,

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

For larger values of q, only conditional asymptotic formulas are known (for example under the Generalised Riemman Hypotheis).

However the *Brun-Titchmarsh theorem* gives an upper bound of the right order of magnitude in for all $q \leq x^{1-\varepsilon}$. For all $x \geq 2$ and all positive integers q,

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

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

Finally, although "individually" for every q, the Siegel–Walfisz theorem is the best know result; on "average" over q, the *Bombieri-Vinogradov theorem* gives a much better estimate. For every fixed $A > 0$, there is $B > 0$ such that

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

3.2 Sums and products over prime numbers

We recall the *Mertens formulas*

$$\begin{aligned} \sum_{p \leq x} \frac{1}{p} &= \log\log x + A + o(1), \quad A = 0.2614\ldots, \\ \sum_{p \leq x} \frac{\log p}{p} &= \log x + B + o(1), \quad B = 1.3325\ldots, \\ \prod_{p \leq x} \left(1 - \frac{1}{p}\right) &= \frac{C + o(1)}{\log x}, \quad C = e^{\gamma} = 1.7810\ldots, \end{aligned}$$

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

In particular the last formula show that the Euler function is "large":

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

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 $s \in \mathbb{C}$.

The *Riemman Hypotheis* asserts that all zeros of $\zeta(s)$ with $0 \le \Re s \le 1$ have $\Re s = 1/2$. It is important to remember that there are other *trivial* zeros outside of the *critical strip* $0 \le \Re s \le 1$.

The *Generalised Riemman Hypotheis* predicts that the same is true for a much wider class of similar functions called L-functions (and even more general functions).

There are some *explicit* formulas which related $\pi(x)$ with the zeros of $\zeta(s)$ in the critical strip. In particular, the nonvanishing $\zeta(1+it)\zeta(it) \ne 0$ for every $t \in \mathrm{I\!R}$ implies that the Prime Number theorem in the form $\pi(x) \sim \mathrm{li} x$. In fact the more we know about the distribution of the zeros of $\zeta(s)$ the better we bound on $|\pi(x) - \mathrm{li} x|$ we can get.

The best known result on the *zero-free region* of $\zeta(s)$ are due to the results of N. M. Korobov and I. M. Vinogradov, who independently obtained them in 1953, see [44]. Unfortunately over the last decades very little progress has been achieved in this area.

For $\Re s > 1$ we the *Dirichlet product*:

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

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 $\mathcal{S}$:

$$\prod_{p \in \mathcal{S}} \left(1 - \frac{1}{p^s}\right)^{-1} = \sum_{n \in \mathcal{N}_{\mathcal{S}}} \frac{1}{n^s}. \qquad (2)$$

Part II

Arithmetic Structure of Integers

4 Rough Introduction to Smoothness

4.1 Counting smooth numbers: Intuition

It is widely accepted that intuition plays a very important tool in cryptography. For examples, many cryptographic papers readily accept that unless there are some obvious divisibility conditions, the density of prime values among the element of a given sequence is the same as for the set of natural numbers. This and several similar "postulates" can be found throughout the modern cryptographic literature. So let us where our intuition can take us in the case of smooth numbers and see what it suggest for the behavious of $\psi(x, y)$ (defined in Section 2.4).

It is natural to approximate the probability that $p \nmid n$ when $n \le x$ is chosen at random by $1 - 1/p$.

Now assuming that all prime $p \le y$ are independent, we can suggest that the probability that $p \nmid n$ for all $x \ge p > y$ when $n \le x$ is chosen at random is close to

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

by the Mertens formula, where u is given by

$$u = \frac{\log x}{\log y} \qquad \text{or} \qquad x = y^u. \tag{3}$$

Thus our intuation leads us the a very nice asymptotic formula

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

This formula is nice and easy to use, but annoyingly enough has a minor disadvanrage, it is completely wrong.

This failure should certainly be taken as a precaution against uncritical use of intuition, especially when the intuition is wrong.

4.2 Counting smooth numbers: Theoretic estimates

We start with an observation that although the intuitive approach of Section 4.1 failed, the parameter u, given by (3), is indeed very important and plays a very significant role in computational number theory and cryptography.

Probably one of the most convenient and readily available for applications results is the estimate of

$$\psi(x,y) = u^{-u+o(u)}x \tag{4}$$

due to E. R. Canfield, P. Erdős and C. Pomerance [15] which holds in the very large range:

$$u \le y^{1-\varepsilon} \qquad \text{or} \qquad y \ge (\log x)^{1+\varepsilon}.$$

The range in which (4) holds is close to best possible as the behaviour of $\psi(x,y)$ changes for $y < \log x$.

We note that (4) is not an asymptotic formula (since $o(u)$ in the exponent). However an asymptotic formula for $\psi(x,y)$ is also possible. In particular, A. Hildebrand [40] has given the asymptotic formula

$$\psi(x,y) \sim \rho(u)x \tag{5}$$

for

$$u \le \exp\left((\log y)^{3/5-\varepsilon}\right) \qquad \text{or} \qquad y \ge \exp\left((\log\log x)^{5/3+\varepsilon}\right).$$

A precise estimate on the error term in (5) is given by É. Saias [54].

We note that (4) and (5) put together imply (1), which of course can be obtained independently.

Unfortunately the range of the validity of (5) is much narrowed that that of (4), and will probably remain this way for quite some time as by a result of A. Hildebrand [39] the validity of (5) in the range

$$1 \le u \le y^{1/2-\varepsilon} \qquad \text{or} \qquad y \ge (\log x)^{2+\varepsilon}$$

is equivalent to the Riemann Hypothesis.

5 Methods for Estimation of $\psi(x, y)$

5.1 Counting very smooth numbers: Lattices

To estimate $\psi(x, y)$ for rather small values of y one can use the following geometric approach, which has been studied in depth by A. Granville [32]. Here are some ideas behind this approach.

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

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

Thus our question is reduced to counting integer points in a certain tetrahedron.

The number of integer points in any "reasonable" convex body is close to its volume. However this is correct only is the volume is large compared to the dimension s. Thus we may expect that

$$\psi(x, y) \approx \frac{\log^s x}{s! \prod_{i=1}^{s} \log p_i}.$$

provided that y is reasonably small. This approach can and has been transformed in rigorous estimate, see [32].

5.2 Upper bounds: Rankin's method

For large values of y the above method fails to produce any useful estimate. If only an upper bound is required, as is the case in may situations, the so-called *Rankin's method* provides a reliable and easy to use alternative.

Fix any constant $c > 0$. Then

$$\psi(x, y) = \sum_{\substack{n \le x \\ p|n \Rightarrow p \le y}} 1 \le \sum_{\substack{n \le x \\ p|n \Rightarrow p \le y}} \left(\frac{x}{n}\right)^c = \sum_{p|n \Rightarrow p \le y} \left(\frac{x}{n}\right)^c. \qquad (6)$$

The underlying “philosophy” is that most of the contribution to $\psi(x, y)$ comes from integers which are close to x, so, although $\left(\frac{x}{n}\right)^c$ is large than 1 for such integers, it is not much larger. On the other hand, this function $\left(\frac{x}{n}\right)^c$ decreases rapidly to 0 n much large than x. So the above two steps do not introduce too many extras to our counting. We now take advantage of the fact that the right hand side of (6) is an infinite series which can be represented as a Dirichlet product, see (2). Hence

$$\psi(x, y) \leq x^c \sum_{p|n \Rightarrow p \leq y} \frac{1}{n^c} = x^c \prod_{p \leq y} \left(1 - \frac{1}{p^c}\right)^{-1}. \tag{7}$$

Using the Prime Number Theorem (in the best available asymptotic form) we estimate the product on the right hand side of (7) as a function of y and c and minimize over all possible choices of $c > 0$. This task is technically non-trivial but is quite feasible and leads to the quasi-optimal choice

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

which in turn produces a upper bound of the form (4).

Certainly the ease of application (despite some technical complications at the end) is a main advantage of this approach. On the other hand it suits only for upper bounds and is not able to produce a lower bound.

5.3 Asymptotic formula: Buchstab–de Bruijn recurrent relation

We write each y-smooth n with $n > 1$, as $n = pm$ where $p = P(n)$ is the largest prime factor of n. We note that $m \leq x/p$ and is p-smooth.

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

$$\psi(x, y) = 1 + \sum_{p \leq y} \psi\left(\frac{x}{p}, p\right) \tag{8}$$

(where 1 as the front accounts for $n = 1$).

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

We now use it to "prove" the asymptotic formula (5) for each fixed u.

The "proof" is by induction over N, where $u \in (N, N+1]$ and during our argument we completely ignore error terms and many other things. In particular we use the sign $\approx$ without even specifying its precise meaning. However, with a little bit of careful analysis, it can be re-casted into a proper proof.

We start with an observation that for $0 < u \leq 1$ we trivially have $\psi(x, x^{1/u}) = \lfloor x \rfloor$.

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

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

Now, by the Mertens formula,

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

We now remarls that in fact the above step has not been necessary but is good warming exercise for the next "induction" step.

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

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

Subtracting the Buchstab–de Bruijn relation (8) with $y = x^{1/N}$:

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

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

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

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 (certainly the presence of error terms is totally ignored here).

We now recall the definition of the function $\vartheta(z)$ and the prime number theorem, see Section 3.1. Writing $z = x^{1/t}$, by partial summation, we get

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

Therefore

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

which concludes our “proof”.

6 Variations

6.1 Evaluation of $\psi(x, y)$

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

For example, S. T. Parsell and J. P. Sorenson [52], improving several results of D. Berstein [9], have shown that for any parameter α, one can estimate $\psi(x, y)$ up to a factor $1+O(\alpha^{-1} \log x)$ in time $O\left(\alpha y^{2/3}/ \log y + \alpha \log x \log \alpha\right)$. A number of other results can be found in [43, 56, 57, 58].

6.2 Constructing smooth numbers

It is certainly trivial to produce a smooth number, for example, 2^k is as smooth as it gets.

The problem becomes much harder and thus more interesting if one need to find a y-smooth number in a given interval $[x, x + z]$. This question gives an example of the referese influance of cryptographic techniques on number theory.

Namely, D. Boneh [10] has used some ideas and algorithms originated from cryptographic applications to design a polynomial algorithm to for this problem for some relations between x, y and z.

Some results about the existence of very smooth numbers with a prescribed bit pattern at certain position are given in [55], see also [31] whose approach (via character sums instead of exponential sums) may probably be used to improve that result of [55].

Certainly more research in this area would be very desirable.

6.3 Rough numbers

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

The question has been addressed by A. A. Buchstab [14] who give the asymptotic formula

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

where the *Buchstab* function $\omega(u)$ is defined as follows: $\omega(u) = 1/u$ for $1 \le u \le 2$ and

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

We note that such numbers can be considered as “approximations” to prime numbers but maybe easily found and also proven to exist in various sequences of cryptographic interest (when proving the existence of primes among its terms is out of reach).

6.4 Integers with a large smooth divisor

It also natural to ask how often an integers has a large smooth divisor. In a more quantitative form, one may ask about the behaviour of

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

This question has been studuied in the classical literature on smooth numbers, see [37, 60, 61], however it has not received as much attention as the question of estimating $\psi(x,y)$.

More recently it has been addressed independently in [7, 63] where asymptotic formulas for $\Theta(x,y,z)$ are given. This formulas a given in terms of the same parameter u in the case of $\psi(x,y)$ (see (3)) and also in terms of another parameter

$$v = \frac{\log z}{\log y}$$

and also involve some integral expression with the de Bruijn function $\rho(u)$ and its derivative. We note that part of the motivation of [7] has come from a concrete cryptographic problem discussed by A. J. Menezes [49], see also Section [**?**].

6.5 Other prime divisors

The notion of smoothness addresses the distribution of the largest prime divisor $P(n)$ of integers. However studying the second largest prime divisor is of ultimate interest two as the complexity of obtaining full integers factorisation of n via the elliptic curve factorisation algorithm of H. W. Lenstra [45] depends on this prime divisor.

More generally, using $P_j(n)$ to denote the j-th largest prime divisor of n, one can ask about the joint distribution

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

We refer to [62] for the most recent results on this topic and further references. The case of $k = 2$ is especially important:

In the above notation the the elliptic curve factorisation algorithm [45] factors an integer n completely in time $\exp\left((2 + o(1))\sqrt{\log p \log\log p}\right) n^{O(1)}$ where $p = P_2(n)$.

6.6 Miscellaneous

In this section we present several disconnected results, which however unlikely to have any cryptographic applications may still contribute to developing some understanding smooth numbers and also indicate what kind of problems one may hope to successfully tackle.

A. Balog and T. D. Wooley [5] have considered k-tuples of consecutive smooth integers and proved that for any k and $\varepsilon > 0$ there are infinitely many n such that $n + i$ is n^{ε}-smooth for $i = 1, \dots, k$. In fact the proof in [5] yields a very nice and elementary explicit constructions. One can also take $k \to \infty$ and $\varepsilon \to 0$ (slowly) when $n \to \infty$.

A. Balog [3] has proven that each sufficiently large integer N can be written as $N = n_1 + n_2$ where n_1, n_2 are N^{α}-smooth, where

$$\alpha = \frac{4}{9\sqrt{e}} = 0.2695\dots .$$

Results of this type may be considered as dual to the *binary Goldbach conjecture* that all positive even integers $N \geq 4$ an be represented as the sum of two primes.

Finally, we remark that various bounds of rational exponential sums

$$S_{a,q}(x, y) = \sum_{\substack{n \leq x \\ n \text{ is } y\text{-smooth}}} \exp(2\pi i a n/q)$$

where $\gcd(a, q) = 1$, are given by É. Fouvry and G. Tenenbaum [24] and also by R. de la Bretéche and G. Tenenbaum [12].

7 Smooth Elements in Integers Sequences

7.1 Smooth Numbers in arithmetic progressions

So far we considered the distribution of smooth values in the set of all natural numbers. A very natural generalisation of this question, which also comes up in some applications, is to study smooth numbers with an additional conguence condition. In particular, we introduce the counting functions

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

and

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

G. Tenenbaum [59] has proved that

$$\psi_q(x, y) \sim \frac{\varphi(q)}{q} \psi(x, y)$$

in a wide range of parameters.

In turn, a scope of bounds of the forms

$$\begin{array}{rcl} \psi(x, y; a, q) & \sim & \dfrac{1}{\varphi(q)} \psi_q(x, y), \\ \psi(x, y; a, q) & \asymp & \dfrac{1}{\varphi(q)} \psi_q(x, y), \\ \psi(x, y; a, q) & \gg & \dfrac{1}{\varphi(q)} \psi_q(x, y), \end{array}$$

(of descreasing strength but in incresingly larger ragnes of x, y and q) can be found in [4, 33, 34, 25, 38]. Some of this bounds hold for for all a with $\gcd(a, q) = 1$, some of them hold only for almost all such integers a.

We note that bounds of exponential sums $S_{a,q}(x, y)$, see Section 6.6 can also be interpreted as results about the uniformity of distribution of smooth numbers in arithmetic progressions "on average".

7.2 Smooth Numbers in Small Intervals

The next sequence we consider is the sequence of integers in "short" intervals $[x, x+z)$. Accordingly, we put

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

It is natural to expect that

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

in a wide range of x, y and z.

There is a series of results due to Balog [2], E. Croot [19], J. B. Friedlander and A. Granville [27], J. B. Friedlander and J. C. Lagarias [27], G. Harman[38], T. Z. Xuan [65] which give various results in this direction, but in general the situation is far from satisfactory here.

It is especially interesting the work of E. Croot [19] uses an a very unusual for this area tool: bounds of so-called *bilinear Kloosterman sums* due to W. Duke, J. B. Friedlander and H. Iwaniec [23].

For example, from the point of view of cryptography and computational number theory, the main challenge in this area is obtaining good lower bounds in $\psi(x, y, 4x^{1/2})$, which appears to be out of reach nowadays. This case is of special importance of as it is crucial for the rigorous analysis of the elliptic curve factoring algorithms of H. W. Lenstra [45]. We note that, the result of E. Croot [19] applies to intervals of similar length, but for much large values of y that those appearing in [45].

We remark, that H. W. Lenstra, J. Pila and C. Pomerance [46, 47] found an igeneous way to circumvent this problem by introducing a hyperelliptic factoring algorithms. For this algorithm smooth numbers in large intervals ought to be studied which is a feasible task. Of course this has been achieved at the cost of very delicate arguments and required the the authors to develop new algebraic and analytic tools.

7.3 Smooth Shifted Primes

Recall, the definiton of the counting function $\pi_a(x, y)$ of smooth shifted primes given in Section 2.4. The values of $a = \pm 1$ are certainly of special interest for cryptography.

It is strongly believed that that for any fixed $a \neq 0$ the asymptotic formula

$$\pi_a(x, y) \sim \rho(u)\pi(x) \tag{9}$$

holds for a wide range of x and y. Unfortunately results of such strength seem to be unavailable within the present techniques.

However rather strong upper bounds are known. For example, C. Pomerance and I. E. Shparlinski [53] have given the estimate

$$\pi_a(x, y) \ll u\rho(u)\pi(x)$$

for

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

In a shorter range

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

the “right” upper bound

$$\pi_a(x, y) \ll \rho(u)\pi(x)$$

follows from a result of É. Fouvry and G. Tenenbaum [25, Theorem 4].

It is not just the asymptotic formula (9) is presently our of reach. In fact even obtaining lower bounds on $\pi_a(x, y)$ is a an extremely difficult task with very slowly progressing results.

The best known result, which is due to R. C. Baker and G. Harman [6] only asserts that there is a positive constant A such that for $a \neq 0$,

$$\pi_a(x, y) \gg \pi(x)/(\log x)^A$$

for $u \le 3.377\ldots$ (where as before, u is defined by (3)). In most of the applications the logarithmic loss in the density of such primes is not important. However, if this becomes an issue, one can use the bound of J. B. Friedlander [26]

$$\pi_a(x, y) \gg \pi(x)$$

which, however, is proven only for $u \le 2\sqrt{e} = 3.2974\ldots$.

We finally recall yet another result, implied by the work of R. C. Baker and G. Harman [6], which says that

$$\pi(x) - \pi_a(x, y) \gg \pi(x)$$

for $u \geq 1.477\ldots$.

The above results can be reformulated in the following equivalent forms which are usually better known and in which they are more frequently used.

For some absolute constants $A, C > 0$ such that for any $a \neq 0$:

- there are at least $C\pi(x)/\log^A x$ primes $p \leq x$ such that $p + a$ has a prime divisor $q \geq p^{0.6776}$;
- there are at least $C\pi(x)/\log^A x$ primes $p \leq x$ such that all prime divisors q of $p + a$ satisfy $q \leq p^{0.2962}$.

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 $\varepsilon > 0$).

It is interesting to recall, that results about shifted primes $p - 1$ having a large prime divisor play a central role in deterministic primality test of M. Agrawal, N. Kayal and N. Saxena [1].

7.4 Smooth Values of Polynomials

Let $f(X) \in \mathbb{Z}[X]$ and let $\psi_f(x, y)$ is defined as in Section 2.4.

As in the case of shifted primes, here are rather strong upper bounds on $\psi_f(x, y)$, see for example, the results of N. A. Hmyrova [42] and N. M. Timofeev [64].

For a squarefree polynomial f, G. Martin [48] gives an asymptotic formula of the type

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

where $d_1, d_2, \ldots, d_k$ be the degrees of irreducible factors of f over $\mathbb{Z}[x]$, however only for very large values of y. Several related results can also be found in the work of C. Dartyge, G. Martin and G. Tenenbaum [21], where also smooth values of polynomials at prime arguments (that is, of $f(p)$) are discussed.

7.5 Smooth numbers in sumsets

R. de la Bretéche [11] gives a result of surprising generality and strength which claims that under some conditions the proportion of smooth numbers among the sums $a+b$, where $a \in \mathcal{A}$ and $b \in \mathcal{B}$ is close to the expected value for a wide class of sets $\mathcal{A}, \mathcal{B} \subseteq \mathbb{Z}$.

For example, let $\mathcal{A}$ and $\mathcal{B}$ be two sets of integers in the interval $[1, x]$. Then, for any fixed $\varepsilon > 0$ and uniformly for $\exp\left((\log x)^{2/3+\varepsilon}\right) < y \le x$, we have

$$\begin{aligned} &\#\{(a,b) \in \mathcal{A} \times \mathcal{B} \ : \ a+b \text{ is } y\text{-smooth}\} \\ &\qquad\qquad = \rho(u) \cdot \#\mathcal{A}\#\mathcal{B}\left(1 + O\left(\frac{x \log(u+1)}{(\#\mathcal{A}\#\mathcal{B})^{1/2}\log y}\right)\right), \end{aligned}$$

where, as usual, u is given by (3).

Although the author is unaware of any cryptographic existing applications of this result, it seems to have a great potential due to its generality and essentially “condition-free” formulation.

Several more relevant results are also given by E. Croot (**??**).

7.6 Smooth polynomials over finite fields

In full analogie with the case of integer numbers, we say that polynomial $F \in \mathbb{K}[x]$ over a field $\mathbb{K}$ sis k-smooth if all irreducible divisors $f \mid F$ satisfy $\deg f \le k$.

For a finite field $\mathbb{F}_q$ of q-elements we denote

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

Define

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

(the last expression makes the analogy with the formula (3) completely explicit).

The systematic study of $N_q(m,k)$ dates back to the work of A. M. Odlyzko [51] who also discovered the relevance of this quatity for the disrete logarithm problem in finite fields.

More recently a series of very precise results about $N_q(m,k)$ have been given be R. L. Bender and C. Pomerance [8]. For example, by [8, Theorem 2.1] we have

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

as $k \to \infty$ and $u \to \infty$, uniformly for $q^k \geq m \log^2 m$, and by [8, Theorem 2.2] we also have

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

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

8 Distribution of Divisors

8.1 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.

8.2 Notation

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

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

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 $f \in \mathbb{Z}[X]$, polynomial sequences
- $\mathcal{A} = \Phi = \{\varphi(n) \ : \ n = 1, 2 \ldots\}$, values of the Euler function

8.3 Natural Numbers

This case goes back to an 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

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

is the *Erdős number*.

Special case: $c_1(\varepsilon)x \le H(x, y, y^{1+\varepsilon}, \mathbb{N}) \le c_2(\varepsilon)$ where the implied constants depend on $\varepsilon > 0$.

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

Let's prove something . . .

Special-special case:

For $0 < \alpha < \beta$:

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

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

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

- There are $\geq x/p$ integers $n \leq x$ divisible by p
- Each $n \leq x$ may have at most $K = \lceil \alpha^{-1} \rceil$ of them.

The sum

$$\sum_{x^\alpha \leq p \leq x^\beta} \frac{x}{p}$$

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

$$\Downarrow$$

$$H(x, x^\alpha, x^\beta, \mathbb{N}) \geq \frac{1}{K} \sum_{x^\alpha \leq p \leq x^\beta} \frac{x}{p} = \frac{x}{K} \sum_{x^\alpha \leq p \leq x^\beta} \frac{1}{p}$$

By the Mertens formula

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

8.4 Shifted Primes

Ford, **2007**:

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

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

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 $0 < \alpha < \beta$:

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

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** $n \leq x$ with $p \mid n$ we need to count **primes** $q \leq x$ with $p \mid q - a$.

8.5 Polynomials

I wish I could say something here ...

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

8.6 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 $H(x, y, z; \mathcal{N})$.

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 $[K, 2K]$, $1 \le K \le n$.

Part III

Applications

8.7 Primality, Factorisation, Dlog

E. Croot, A. Granville, R. Pemantle and P. Tetali [20]

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)

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

Initial Assumption

Let us fix some y (to be optimised later) and **assume** that we know discrete logarithms 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

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

  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

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

We cannot find Dlogp_i immediately

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

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

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

8.9 Text-book ElGamal

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

ElGamal Scheme Primes p, q with $q \mid p-1$

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

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

Public Key: $X = g^x$

Encryption of a Message μ:

For a random $r \in \mathbb{Z}_q$, 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

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.

8.10 Desmedt-Odlyzko Attack

RSA Singature Scheme:

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

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

Y. Desmedt and A. Odlyzko [22] have given an existential forgery attack on this scheme. In this scenario we are allowed to ask for signatures on some "allowed" message (for example, padded in a prescribed way), and them we must produce a signature on one more "allowed" message.

- Select a bound y and let $p_1, ..., p_k$ be the primes up to y, i.e. $k = \pi(y)$.
- Take $k+1$ messages m_i which are y-smooth and factor them $m_i = \prod_{j=1}^{k} p_j^{\alpha_{i,j}}$
- Express $\prod_{i=1}^{k+1} m_i^{u_i} = 1$ as a multiplicative combination of $m_1, \ldots, m_k$, by solving

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

  Thus

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

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

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

D. Coppersmith, J. S. Coron, F. Grieu, S. Halevi, C. Jutla, D. Naccache, J. P. Stern [16] have recently introduced a number of improvements and generalisations in this attack and also given some applications to concrete protocols.

8.11 Generalised Diffie-Hellman Problem

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 $\mathcal{G}$.

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 $g^x, \ldots g^{x^d}$ 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.

8.12 Fix-Padded RSA

N = n-bit RSA modulus.

"Text-book" RSA signature scheme:

Message $m \quad \Longrightarrow \quad$ Singature $s \equiv m^d \bmod N$

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

8.13 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:

$\boxed{\text{fixed } n-\ell\text{-bit padding } P \;\; | \;\; \ell\text{-bit message } m}$

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

Some existing standards still use this scheme.

[13, 29, 30, 50]

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

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

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

Let

$$w = |(P + s)Q_j - NP_j|.$$

Then

$$0 < w \leq N/Q_{j+1} < N^{2/3} \quad \text{and} \quad (P + s)z \equiv w \bmod N$$

for some z with $|z| < N^{1/3}$, namely $z = \pm Q_j$.

- Choose a "random" r with $0 \leq r < 0.5N^{\varepsilon}$ and find

  $$w \equiv \left(P + m_3 - r \left\lfloor N^{1/3} \right\rfloor\right) z \bmod N$$

  with $w < N^{2/3}$

- 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.

8.14 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 (HMQV).

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

8.15 Smooth Orders

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

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

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

and

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

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

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

8.16 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

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

  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**: $O((\log p)^2)$.
authBayless2007 $C\log p$ for any $C > 1$ and almost all primes

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

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

for almost all primes p.

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

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

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.

8.17 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 smooth;
2. or $p-1$ is divisible by a r^2 for a large prime r;
3. or $\varphi(p-1)$ is smooth

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

For 2:

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

For 3: ???

Let

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

and

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

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.

How tight are they?
Under some plausible conjecture, matching lower bound.

We can now use $\Pi(x,y) \le \Phi(x,y)$ to take care of 3.

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

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

Dream Result:

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

References

[1] M. Agrawal, N. Kayal and N. Saxena, ‘PRIMES is in **P**’, *Ann. of Math.*, **160** (2004), 781–793.

[2] A. Balog, 'On the distribution of integers having no large prime factors', *Astérisque*, **147–148** (1987), 27–31.

[3] A. Balog, 'On additive representation of integers', *Acta Math. Hungar.*, **54** (1989), 297–301.

[4] A. Balog and C. Pomerance, 'The distribution of smooth numbers in arithmetic progressions', *Proc. Amer. Math. Soc.*, **115** (1992), 33–43.

[5] A. Balog and T. D. Wooley, 'On strings of consecutive integers with no large prime factors', *J. Austral. Math. Soc., Ser. A*, **64** (1998), 266–276.

[6] R. C. Baker and G. Harman, 'Shifted primes without large prime factors,' *Acta Arith.*, **83** (1998), 331–361.

[7] W. D. Banks and I. E. Shparlinski, 'Integers with a large smooth divisor', *Integers*, **7** (2007), # A17, 1-11.

[8] R. L. Bender and C. Pomerance, 'Rigorous discrete logarithm computations in finite fields via smooth polynomials', *Computational Perspectives on Number Theory*, Amer. Math. Soc., Providence, RI, 1998, 221–232.

[9] D. J. Bernstein, 'Bounding smooth integers', *Proc. 3rd Algorithmic Number Theory Symp.*, Lecture Notes in Comput. Sci., vol. 1423, Springer-Verlag, Berlin, 1998, 128–130.

[10] D. Boneh, 'Finding smooth integers in short intervals using CRT decoding', *J. Comp. and Syst. Sciences.*, **64** (2002), 768–784.

[11] R. de la Bretéche, 'Sommes sans grand facteur premier', *Acta Arith.*, **88** (1999), 1–14.

[12] R. de la Bretéche and G. Tenenbaum, 'Sommes d'exponentielles friables d'arguments rationnels', *Funct. Approx. Comment. Math.* , **37** (2007), 31–38.

[13] E. Brier, C. Clavier, J.-S. Coron and D. Naccache, 'Cryptanalysis of RSA signatures with fixed-pattern padding', *Proc. Crypto'01*, Lect. Notes in Comp. Sci., vol. 2139, Springer-Verlag, Berlin, 2001, 433–439.

[14] A. A. Buchstab, 'On those numbers in an arithmetic progression all prime factors of which are small in magnitude', *Dokl. Akad. Nauk SSSR*, **67** (1949), 5–8 (in Russian).

[15] E. R. Canfield, P. Erdős and C. Pomerance, 'On a problem of Oppenheim concerning "Factorisatio Numerorum"', *J. Number Theory*, **17** (1983), 1–28.

[16] D. Coppersmith, J. S. Coron, F. Grieu, S. Halevi, C. Jutla, D. Naccache, J. P. Stern, 'Cryptanalysis of ISO/IEC 9796-1', *J. Cryptology*, **12** (2008), 27–51.

[17] R. Crandall and C. Pomerance, *Prime numbers: A computational perspective*, 2nd edition, Springer-Verlag, New York, 2005.

[18] E. Croot, 'On a combinatorial method for counting smooth numbers in sets of integers', *J. Number Theory*, **126** (2007), 237–253.

[19] E. Croot, 'Smooth numbers in short intervals', *Int. J. Number Theory*, **3** (2007), 159–169.

[20] E. Croot, A. Granville, R. Pemantle and P. Tetali, 'Running time predictions for factoring algorithms', *Proc. 8th Algorithmic Number Theory Symp.*, Lecture Notes in Comput. Sci., vol. 5011, Springer-Verlag, Berlin, 2008, 1–36.

[21] C. Dartyge, G. Martin and G. Tenenbaum, 'Polynomial values free of large prime factors', *Periodica Math. Hungar.*, **43** (2001), 111–119.

[22] Y. Desmedt and A. Odlyzko, 'A chosen text attack on the RSA cryptosystem and some discrete logarithm schemes', *Proc. Eurocrypr'85*, Lecture Notes in Comput. Sci., vol. 218, Springer-Verlag, Berlin, 1985, 516–522.

[23] W. Duke, J. B. Friedlander and H. Iwaniec, 'Bilinear forms with Kloosterman fractions', *Invent. Math.*, **128** (1997), 23–43.

[24] É. Fouvry and G. Tenenbaum, 'Entiers sans grand facteur premier en progressions arithmetiques', *Proc. London Math. Soc.*, **63** (1991), 449–494.

[25] É. Fouvry and G. Tenenbaum, 'Répartition statistique des entiers sans grand facteur premier dans les progressions arithmétiques', *Proc. London Math. Soc.*, **72** (1996), 481–514.

[26] J. B. Friedlander, 'Shifted primes without large prime factors', *Number Theory and Applications*, Kluwer Acad. Publ., Dordrecht, 1989, 393–401.

[27] J. B. Friedlander and A. Granville, 'Smoothing 'smooth' numbers', *Philos. Trans. Roy. Soc. London, Ser. A*, **345** (1993), 339–347.

[28] J. B. Friedlander and J. C. Lagarias, 'On the distribution in short intervals of integers having no large prime factor', *J. Number Theory*, **25** (1987), 249–273.

[29] M. Girault and J.-F. Misarsky, 'Selective forgery of RSA signatures using redundancy', *Proc. Eurocrypt'97*, Lect. Notes in Comp. Sci., vol. 1233, Springer-Verlag, Berlin, 1997, 495–507.

[30] M. Girault and J.-F. Misarsky, 'Cryptoanalysis of countermeasures proposed for repairing ISO 9796', *Proc. Eurocrypt'00, Bruges*, Lect. Notes in Comp. Sci., vol. 1807, Springer-Verlag, Berlin, 2000, 81–90.

[31] S. W. Graham and I. E. Shparlinski, 'On RSA moduli with almost half of the bits prescribed', *Disc. Appl. Math.*, (to appear).

[32] A. Granville, 'On positive integers $\leq x$ with prime factors $\leq t \log x$', *Number Theory and Applications* Kluwer, 1989, 403–422.

[33] A. Granville, 'Integers, without large prime factors, in arithmetic progressions I', *Acta Math.*, **170** (1993), 255–273.

[34] A. Granville, 'Integers, without large prime factors, in arithmetic progressions II', *Philos. Trans. Roy. Soc. London, Ser. A*, **345** (1993), 349–362.

[35] 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).

[36] H. Halberstam and H.–E. Richert, *Sieve methods*, Academic Press, London, 1974.

[37] R. Hall and G. Tenenbaum, *Divisors*, Cambridge Tracts in Mathematics, vol. 90, Cambridge Univ. Press, 1988.

[38] G. Harman, 'Integers without large prime factors in short intervals and arithmetic progressions ', *Acta Arith.*, **91** (1999), 279–289.

[39] A. Hildebrand, 'Integers free of large prime factors and the Riemann Hypothesis', *Mathematika*, **31** (1984), 258–271.

[40] A. Hildebrand, 'On the number of positive integers $\leq x$ and free of prime factors $\leq y$', *J. Number Theory*, **22** (1986), 289–307.

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

[42] N. A. Hmyrova, 'On polynomials with small prime divisors, II', *Izv. Akad. Nauk SSSR Ser. Mat.*, **30** (1966), 1367–1372 (in Russian).

[43] S. Hunter and J. P. Sorenson, 'Approximating the number of integers free of large prime factors', *Mathem. Comp.*, **66** (1997), 1729–1741.

[44] H. Iwaniec and E. Kowalski, *Analytic number theory*, Amer. Math. Soc., Providence, RI, 2004.

[45] H. W. Lenstra, Jr., 'Factoring integers with elliptic curves', *Annals of Math.*, **126** (1987), 649–673.

[46] H. W. Lenstra, J. Pila and C. Pomerance, 'A hyperelliptic smoothness test, I', *Phil. Trans. of the Royal Society of London, Ser. A.*, **345** (1993), 397–408.

[47] H. W. Lenstra, J. Pila and C. Pomerance, 'A hyperelliptic smoothness test, II', *Proc. London Math. Soc*, **84** (2002), 105–146.

[48] G. Martin, 'An asymptotic formula for the number of smooth values of a polynomial', *J. Number Theory*, **93** (2002), 108–182.

[49] A. J. Menezes, 'Another look at HMQV', *J. Math. Cryptology*, **1** (2007), 47-64.

[50] J.-F. Misarsky, 'A multiplicative attack using LLL algorithm on RSA signatures with redundancy, ', *Proc. Crypto'97, Santa Barbara*, Lect. Notes in Comp. Sci., vol. 1294, Springer-Verlag, Berlin, 1997, 221–234.

[51] A. M. Odlyzko, 'Discrete logarithms in finite fields and their cryptographic significance', *Proc. Eurocrypr'84*, Lecture Notes in Comput. Sci., vol. 209, Springer-Verlag, Berlin, 1985, 224–314.

[52] S. T. Parsell and J. P. Sorenson, 'Fast bounds on the distribution of smooth numbers', *Proc. 7th Algorithmic Number Theory Symp.*, Lecture Notes in Comput. Sci., vol. 4076, Springer-Verlag, Berlin, 2006, 168–181.

[53] C. Pomerance and I. E. Shparlinski, 'Smooth orders and cryptographic applications', *Proc. 5th Algorithmic Number Theory Symp.*, Lecture Notes in Comput. Sci., vol. 2369, Springer-Verlag, Berlin, 2002, 338–348.

[54] É. Saias, 'Sur le nombre des entiers sans grand facteur premier,' *J. Number Theory*, **32** (1989), 78–99.

[55] I. E. Shparlinski, 'On RSA moduli with prescribed bit patterns', *Designs, Codes and Cryptography*, **39** (2006), 113–122.

[56] J. P. Sorenson, 'A fast algorithm for approximately counting smooth numbers', *Proc. 4th Algorithmic Number Theory Symp.*, Lecture Notes in Comput. Sci., vol. 1838, Springer-Verlag, Berlin, 2000, 539–549.

[57] K. Suzuki, 'An estimate for the number of integers without large prime factors', *Mathem. Comp.*, **73** (2004), 1013–1022.

[58] K. Suzuki, 'Approximating the number of integers without large prime factors', *Mathem. Comp.*, **75** (2006), 1015–1024.

[59] G. Tenenbaum, 'Cribler les entiers sans grand facteur premier', *Philos. Trans. Roy. Soc. London, Ser. A*, **345** (1993), 377–384.

[60] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge U. Press, 1995.

[61] G. Tenenbaum, 'Crible d'Ératosthéne et modéle de Kubilius', *Number theory in progress (Zakopane-Kościelisko, 1997)*, vol. 2, de Gruyter, Berlin, 1999, 1099–1129.

[62] G. Tenenbaum, 'A rate estimate in Billingsleys theorem for the size distribution of large prime factors', *Quart. J. Math.*, **51** (2000), 385–403.

[63] G. Tenenbaum, 'Integers with a large friable component', *Acta Arith.*, **124** (2006), 287–291.

[64] N. M. Timofeev, 'Polynomials with small prime divisors', *Taškent. Gos. Univ., Naučn. Trudy No. 548, Voprosy Mat.*, Taškent, 1977, 87–91 (Russian).

[65] T. Z. Xuan, 'On smooth integers in short intervals under the Riemann hypothesis.', *Acta Arith.*, **88** (1999), 327–332.