

Divisibility, Smoothness and Cryptographic Applications

DAVID NACCACHE

Équipe de cryptographie

École normale supérieure

45 rue d'Ulm, F-75230 Paris, Cedex 05, France

`david.naccache@ens.fr`

IGOR E. SHPARLINSKI

Department of Computing

Macquarie University

Sydney, NSW 2109, Australia

`igor@comp.mq.edu.au`

October 17, 2008

Abstract

This paper deals with products of moderate-size primes, familiarly known as *smooth numbers*. Smooth numbers play an crucial role in information theory, signal processing and cryptography.

We present various properties of smooth numbers relating to their enumeration, distribution and occurrence in various integer sequences. We then turn our attention to cryptographic applications in which smooth numbers play a pivotal role.

1 Introduction

The goal of this paper is to shed light on the prominent role played by divisibility and smoothness in cryptography and related areas of mathematics. This work intends to survey a wide range of results while steering away from too well-known examples. For doing so, we concentrate on some recently discovered applications of results about the arithmetic structure of integers.

We intend to convey to the reader a general comprehension of the state of the art, allow the devising of correct heuristics when problems cannot be tackled theoretically and help assessing the plausibility of new results.

In Section 3 we overview on a number of number-theoretic results commonly used for studying the multiplicative structure of integers. Most of the elementary results which we use are readily available from [83]; more advanced results can be found, often in much more precise forms, in [44, 80, 85, 92, 144] and in many other standard analytic number theory manuals. Some of them are directly used in this paper, others remain in the background but we illustrate with them the variety of cryptographically useful analytic number theory tools.

We start our exploration of the worlds of divisibility and smoothness by asking a number of natural questions. For instance, given a “typical” integer, what can be said about its largest divisor? Are Euler totient function values $\varphi(n)$ “typical” integers? What are the noteworthy properties of shifted primes $p - 1$? How common are numbers who factor into products of primes which are all smaller than a bound b ? The results listed here are neither exhaustive nor new (we refer the reader to references such number theory books or surveys such as [79, 81, 88] for a more formal and systematic topical treatment).

Then, in Section 9, we use these results to shed light on a number of cryptographic constructions and attacks.

We remark that the specifics of this area is such that many impressive works here may be underrated by non-experts as, at a first glance, they present only very small improvements over previously known results. However these small improvements are often principal steps forward and require the development of new ideas and very refined techniques. Some examples of such breakthrough achievements include:

- the estimate of Ford [55] on the counting function for the number of values of the Euler function, see Section 3.4;
- the estimates of Ford [58] on the counting function of integers with an integer divisor in a given interval, see Section 8.3;
- the very tight estimates of Croot, Granville, Pémantle & Tetali [47] on the stopping time of the Dixon factoring algorithm, see Section 9.1.

Probably the oldest application of smoothness and divisibility is the celebrated *Chinese Remainder Theorem* which allows to accelerate cryptographic functions and basic arithmetic operations using specific integer formats called *residue number systems*, see [118]. Results of this kind certainly deserve an independent treatment and we leave them outside of the framework of this paper.

We also recall that the idea of breaking a complex operation, depending on a parameter n , into a recursion of simpler operations depending on the prime factors of n can also be found in other fields such as signal processing.

The finite Fourier transform of a complex n -dimensional vector y is the n -dimensional vector Y defined as

$$Y_k = \sum_{j=0}^{n-1} \omega_n^{jk} y_j.$$

where $\omega_n = e^{-2\pi i/n}$ is a complex n -th root of unity. Now, assume that $n = 2m$ is even. We see that

$$Y_k = \sum_{\substack{j=0 \\ j \text{ even}}}^{n-1} \omega_n^{jk} y_j + \sum_{\substack{j=0 \\ j \text{ odd}}}^{n-1} \omega_n^{jk} y_j = \sum_{j=0}^{m-1} \omega_m^{jk} y_{2j} + \omega_n^k \sum_{j=0}^{m-1} \omega_m^{jk} y_{2j+1}$$

In other words, the initial finite Fourier transform can be broken into two transforms of length $n/2$ on the projections of y on even and odd dimensions. The same applies to divisibility by any prime or prime power and allows to derive a recursive Fast Fourier Transform algorithm of sub-quadratic complexity when n is smooth, [33, 69]. Applications of this kind are also left out as we restrict ourselves to the cryptographic genre.

2 Conventions

2.1 Notations

Throughout this paper we use Vinogradov's notation ' $f(x) \ll g(x)$ ' which is equivalent to the Landau notation $f(x) = \mathcal{O}(g(x))$, whilst being easier to chain as, for example, $f(x) \ll g(x) = h(x)$.¹ If convenient, we also write $g(x) \gg f(x)$ instead of $f(x) \ll g(x)$. We also write $f(x) \asymp g(x)$ if $f(x) \ll g(x) \ll f(x)$.

The letter p (possibly subscripted) always denotes a prime; ε always stands for a small positive parameter on which implied constants may depend; $\log x$ denotes the natural logarithm of x . Calligraphic letters, for example, $\mathcal{A} = (a_n)$, usually denote sequences of integers.

For a prime power q , we use \mathbb{F}_q to denote the finite field of q elements.

For an integer m , we use \mathbb{Z}_N to denote the residue ring modulo N .

2.2 Arithmetic Functions

We use the following standard notations for the most common arithmetic functions for integers $m \geq 2$:

- $P(m)$, the largest prime divisor of m ,
- $\varphi(m)$, the Euler (totient) function of m ,
- $\omega(m)$, the number of distinct prime divisors of m ,
- $\tau(m)$, the number of positive integer divisors of m .

Recall that $\varphi(m)$ is the number of positive integers $i \leq m$ with $\gcd(i, m) = 1$ and that $\tau(m)$ is sometimes denoted as $\sigma_0(m)$.

We also define $P(1) = \omega(1) = 0$ and $\tau(1) = \varphi(1) = 1$.

Clearly $2^{\omega(m)} \leq \tau(m)$ and the inequality is tight for square-free $m \geq 1$.

Letting $x \geq 0$ be a real number, we denote by:

¹Note that $f(x) = \mathcal{O}(g(x)) = h(x)$ is meaningless and $f(x) = \mathcal{O}(g(x)) = \mathcal{O}(h(x))$ may discard some useful information.

- $\pi(x)$ the number of primes $p \leq x$,
- $\pi(x; q, a)$ the number of primes $p \leq x$ such that $p \equiv a \pmod{q}$.

2.3 Integer Sequences

Besides the sequence of natural numbers \mathbb{N} , we devote in this paper particular attention to the following integer sequences:

- $\mathcal{P}_a = \{p + a : p \text{ prime}\}$,
- $f(\mathbb{N}) = \{f(n) : n = 1, 2, \dots\}$,
- $\varphi(\mathbb{N}) = \{\varphi(n) : n = 1, 2, \dots\}$,
- $\varphi(\mathcal{P}_a) = \{\varphi(p + a) : p \text{ prime}\}$.

In other words, \mathcal{P}_a is the sequence of shifted primes, $f(\mathbb{N})$ is the sequence of polynomial valuations over \mathbb{N} , $\varphi(\mathbb{N})$ is the sequence of Euler function values and $\varphi(\mathcal{P}_a)$ is the sequence of Euler function values of shifted primes.

Amongst the sequence \mathcal{P}_a , the instances $a = \pm 1$ are of special interest in cryptography and thus many papers concentrate only on these values. As results can usually be extended to any $a \neq 0$ at the cost of mere typographical changes, this work usually presents these results in this more general form.

2.4 Smoothness

$n \in \mathbb{N}$ is *smooth* if n has only *small* prime divisors. As the previous sentence does not define what *small* is, we formally define n as *y-smooth* if all prime divisors $p \mid n$ are such that $p \leq y$.

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

Let $\mathcal{A} = (a_n)$ be a sequence. We denote by $\psi(x, y; \mathcal{A})$ the number of *y-smooth* a_n values found amongst the first x elements of \mathcal{A} (that is, for $n \leq x$). The following compact notations are used for the specific sequences defined in Section 2.3:

$$\begin{aligned}
\psi(x, y) &= \psi(x, y; \mathbb{N}), \\
\pi_a(x, y) &= \psi(x, y; \mathcal{P}_a), \\
\psi_f(x, y) &= \psi(x, y; f(\mathbb{N})), \\
\Phi(x, y) &= \psi(x, y; \varphi(\mathbb{N})), \\
\Pi_a(x, y) &= \psi(x, y; \varphi(\mathcal{P}_a)).
\end{aligned}$$

2.5 The Dickman–de Bruijn Function

The *Dickman–de Bruijn function* $\rho(u)$ is probably the most popular smoothness density estimation tool.

$\rho(u)$ is defined recursively by:

$$\rho(u) = \begin{cases} 1 & \text{if } 0 \leq u \leq 1, \\ 1 - \int_n^t \frac{\rho(v-1)}{v} dv & \text{if } u > 1. \end{cases}$$

Note that $\rho(u) = 1 - \log u$ for $1 \leq u \leq 2$. For example, $\rho(\sqrt{e}) = 1/2$, that is, about half of the integers $n \leq x$ has no prime divisors larger than $n^{1/\sqrt{e}} = n^{0.6065\dots}$. This has been used by Vinogradov [152] and by Burgess [30], to estimate the smallest quadratic non-residue modulo a prime.

It is not difficult to show that as $u \rightarrow \infty$:

$$\rho(u) = u^{-u+o(u)} \tag{1}$$

and, more precisely,

$$\rho(u) = \left(\frac{e + o(1)}{u \log u} \right)^u ;$$

even more accurate approximations to $\rho(u)$ are known, see [144, Chapter III.5, Theorem 8].

3 Number Theoretic Facts

3.1 Distribution of Primes

The *Prime Number Theorem* states that for any fixed A :

$$\pi(x) = \text{li}x + \mathcal{O}\left(\frac{x}{(\log x)^A}\right), \quad (2)$$

where

$$\text{li}x = \int_2^x \frac{dt}{\log t}.$$

Alternatively, using a more convenient (yet equivalent) formulation, in terms of the ϑ -function

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

we can write that for any fixed $A > 0$:

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

A commonly committed crime against primes is the assertion that:

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

for any fixed $A > 0$, which is *wrong*, although, of course,

$$\pi(x) \sim \text{li}x \sim \frac{x}{\log x}.$$

An asymptotic estimate of the number of primes in arithmetic progressions is given by the *Siegel–Walfisz theorem* see [44, Theorem 1.4.6] or, in an alternative form, [144, Chapter II.8, Theorem 5], which states that for every fixed $A > 0$ there exists $C > 0$ such that for $x \geq 2$ and for all positive integers $q \leq (\log x)^A$,

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

see also [92, Theorem 5.27].

While for larger values of q , only conditional asymptotic formulae are known, for example, subject to the Generalized Riemann Hypothesis), the *Brun-Titchmarsh theorem*, see [80, Chapter 3, Theorem 3.7], or [92, Theorem 6.6], or [144, Chapter I.4, Theorem 9], gives a tight upper bound on $\pi(x; q, a)$ for all $q \leq x^{1-\varepsilon}$. Namely, we have

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

without any restrictions on x and q .

Clearly for all $q \leq x^{1-\varepsilon}$ we can replace $\log(x/q)$ in the denominator with $\log x$. Furthermore, this is conjectured to hold with just $\log x$ instead of $\log(x/q)$ in a wider range of q (say up to $q \leq x/(\log x)^A$ with some constant $A > 0$).

Finally, although for any given q , the Siegel–Walfisz theorem is the best known result, the *Bombieri–Vinogradov theorem*, see [92, Theorem 17.1], gives a much better estimate of $\pi(x; q, a)$ on average over q . In particular, for every $A > 0$ there exists B such that

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

We conclude with the trivial but helpful remark that the bounds

$$\pi(x; q, a) \leq \pi(x) \quad \text{and} \quad \pi(x; q, a) \leq \frac{x}{q},$$

can also be sufficient sometimes to establish useful results.

3.2 Mertens Formulae

We recall the *Mertens formulae* for the sums over primes

$$\sum_{p \leq x} \frac{1}{p} = \log \log x + A + o(1), \quad \sum_{p \leq x} \frac{\log p}{p} = \log x + B + o(1)$$

and for the product

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right) = \frac{C + o(1)}{\log x}, \tag{3}$$

where $A = 0.2614\dots$, $B = 1.3325\dots$, $C = e^\gamma = 1.7810\dots$ and as before, $\gamma = 0.5772\dots$ is the *Euler-Mascheroni constant*, see [83, Sections 22.7 and 22.8] or [144, Sections I.1.4 and I.1.5]. Vinogradov [151] gives a sharp bound of the error term.

Note that the formula (3) is related to the fact that $\varphi(n)$ is rather large:

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

3.3 Primes and the Zeta-Function

The Riemann Zeta-function $\zeta(s)$ is defined for any $s \in \mathbb{C}$ with $\Re(s) > 1$ by

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

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

The *Riemann Hypothesis* postulates that all the zeros of $\zeta(s)$ with $0 \leq \Re(s) \leq 1$ are such that $\Re(s) = 1/2$. It is important to remind that there are other *trivial* zeros outside of the *critical strip* $0 \leq \Re(s) \leq 1$.

The *Generalized Riemann Hypothesis* asserts that the same property holds for a much wider class of similar functions called *L-functions*.

There are some *explicit* formulae that relate $\pi(x)$ to the zeros of ζ in the critical strip. In particular, the non-vanishing $\zeta(1+it)\zeta(it) \neq 0$ for every $t \in \mathbb{R}$ implies the Prime Number Theorem under the form $\pi(x) \sim \text{li}x$. In fact the more we know about the distribution of the zeros of ζ the better is the bound on $|\pi(x) - \text{li}x|$ we get.

The best known result on the *zero-free region* of ζ is due to Ford [57], who gives a more explicit version of the previous result obtained independently by Korobov [99] and Vinogradov [153], see also [92]. In particular, thanks to these results, the asymptotic formula (2) can be sharpened as

$$\pi(x) - \text{li}x \ll x \exp\left(-C(\log x)^{3/5}(\log \log x)^{-1/5}\right)$$

where $C = 0.2098$. A similar estimate for $\vartheta(x)$ can be obtained as well.

Unfortunately, besides the result of Ford [57] and a few other similar estimates, very little progress has been witnessed in this area over the last decades.

For $\Re(s) > 1$, the *Dirichlet product* is defined as:

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

More generally, letting \mathcal{S} be any set of primes, and letting $\mathcal{N}_{\mathcal{S}}$ be the set of integers obtained by multiplying elements of \mathcal{S} , we have:

$$\prod_{p \in \mathcal{S}} \left(1 - \frac{1}{p^s}\right)^{-1} = \sum_{n \in \mathcal{N}_{\mathcal{S}}} \frac{1}{n^s}. \quad (4)$$

3.4 Euler Function

Here are a few beautiful properties of the Euler function which can be found in many standard number theory manuals (see, for example, [83]) For example, it is easy to see that

$$\sum_{d|m} \varphi(d) = m \quad \text{and} \quad \varphi(m) = m \sum_{d|m} \frac{\mu(d)}{d}, \quad (5)$$

where $\mu(d)$ is the Möbius function. Furthermore, we have the identity

$$\sum_{m=1}^{\infty} \frac{\varphi(m)q^m}{1 - q^m} = \frac{q}{(1 - q)^2}.$$

Using (5) and simple analytic estimates one can derive the following asymptotic formulae:

$$\frac{1}{m^2} \sum_{k=1}^m \varphi(k) = \frac{3}{\pi^2} + \mathcal{O}\left(\frac{\log m}{m}\right)$$

and

$$\frac{1}{m} \sum_{k=1}^m \frac{\varphi(k)}{k} = \frac{6}{\pi^2} + \mathcal{O}\left(\frac{\log m}{m}\right).$$

We also have explicit inequalities such as:

$$\varphi(m) > \frac{m \log \log m}{e^{\gamma} (\log \log m)^2 + 3}$$

for $m \geq 3$, where $\gamma = 0.5772\dots$ is the *Euler-Mascheroni constant*, and

$$\varphi(m) > \sqrt{\frac{m}{2}}$$

for any $m \geq 1$. Finally, for composite m :

$$\varphi(m) \leq m - \sqrt{m}.$$

There are also some much deeper questions about the Euler function. One of the is studying the cardinality

$$F(x) = \#\{\varphi(n) \leq x\}$$

of the set of values of the Euler function up to x , for which Ford [55] obtained a very precise estimate.

Ford [56] has also established the validity of the *Serpiński conjecture* that for any integer $k \geq 2$ there is m such that the equation $\varphi(n) = m$ has exactly k solutions. We recall that by the *Carmichael conjecture* for any m this equation has either at least two solutions or no solutions at all.

4 How Smooth? How Many?

4.1 Empirical Estimates: A Cautionary Note

Empirical estimates abound in cryptography. For examples, many cryptographers readily admit that, in the absence of obvious divisibility conditions, the density of primes in a given integer sequence is identical to the density of primes in \mathbb{N} . This and several similar “postulates” can be frequently found throughout modern cryptographic literature. Let us illustrate the danger of such assumptions by a concrete example.

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

Now, assuming that all primes $p \leq y$ are independent, we may infer that the probability that $p \nmid n$ for all $x \geq p > y$ when $n \leq x$ is chosen at random is close to:

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

by virtue of the Mertens formula, where u is given by

$$u = \frac{\log x}{\log y} \quad \text{or} \quad x = y^u. \quad (6)$$

Here intuition leads to the seemingly elegant asymptotic formula

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

which is . . . completely wrong!

4.2 Estimating Smooth Integer Densities

One of the most popular estimates of $\psi(x, y)$ is:

$$\psi(x, y) = u^{-u+o(u)}x. \quad (7)$$

This formula, due to Canfield, Erdős & Pomerance [31], is applicable in the very large range:

$$u \leq y^{1-\varepsilon} \quad \text{or} \quad y \geq (\log x)^{1+\varepsilon}$$

but the behavior of $\psi(x, y)$ changes for $y < \log x$.

While (7) is not an asymptotic formula (since $o(u)$ is in the exponent), asymptotic formulae for $\psi(x, y)$ exist. In particular, Hildebrand [87] gave the asymptotic formula

$$\psi(x, y) \sim \rho(u)x \quad (8)$$

for

$$u \leq \exp((\log y)^{3/5-\varepsilon}) \quad \text{or} \quad y \geq \exp((\log \log x)^{5/3+\varepsilon}).$$

A precise estimate of the error term in (8) is given by Saias [131].

Note that (7) and (8) imply (1); of course (1) can also be obtained independently.

Unfortunately the validity range of (8) is much narrower than that of (7), and is likely to remain so for quite some time. Indeed, as per another result of Hildebrand [86], the validity of (8) in the range:

$$1 \leq u \leq y^{1/2-\varepsilon} \quad \text{or} \quad y \geq (\log x)^{2+\varepsilon}$$

is equivalent to the Riemann Hypothesis.

5 Estimating $\psi(x, y)$

5.1 Counting Very Smooth Numbers: Lattices

To estimate $\psi(x, y)$ for very small values of y one can resort to a geometric approach introduced by Ennola [52], which has been developed up to its natural limit by Granville [76] (see also [75]):

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

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

Thus our question boils-down to counting integer points in a specific tetrahedron. The number of integer points in any “reasonable” convex body is close to its volume. However, this is correct only if the volume is large with respect to its dimension s .

Thus we may expect that:

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

if y is reasonably small. This approach can yield rigorous estimates, see, for example, [75].

5.2 Upper Bounds: Rankin’s Method

For large values of y , the geometric approach fails to produce useful estimates. If only an upper bound is required, as is the case in many situations, then *Rankin’s method* [128] provides a reliable alternative.

Fix any constant $c > 0$. Then

$$\psi(x, y) = \sum_{\substack{n \leq x \\ p|n \Rightarrow p \leq y}} 1 \leq \sum_{\substack{n \leq x \\ p|n \Rightarrow p \leq y}} \left(\frac{x}{n}\right)^c = \sum_{p|n \Rightarrow p \leq y} \left(\frac{x}{n}\right)^c. \quad (9)$$

The underlying idea is that most of the contribution to $\psi(x, y)$ comes from integers which are close to x , so, although $(x/n)^c$ is larger than one for such integers, it is not much larger. On the other hand, $(x/n)^c$ decreases rapidly to zero when n is much larger than x . So the above two steps do not cause over-counts.

Using the fact that the right hand side of (9) is an infinite series which can be represented as a Dirichlet product (see (4)), we get:

$$\psi(x, y) \leq x^c \sum_{p|n \Rightarrow p \leq y} \frac{1}{n^c} = x^c \prod_{p \leq y} \left(1 - \frac{1}{p^c}\right)^{-1}. \quad (10)$$

Using the Prime Number Theorem (in its best available asymptotic form) we estimate the product on the right hand side of (10) as a function of y and c and minimize over all possible choices of $c > 0$.

This task is technical but feasible and yields the quasi-optimal choice:

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

which, in turn, yields an upper bound of the form (7).

Simplicity (despite a few final technicalities) is the main advantage of this approach. In exchange, it suits only upper bounds and is apparently incapable of producing lower bounds.

5.3 Asymptotic Formula: Buchstab–de Bruijn’s Recurrence

We write each y -smooth $n > 1$, 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:

$$\psi(x, y) = 1 + \sum_{p \leq y} \psi\left(\frac{x}{p}, p\right) \quad (11)$$

(where 1 at the front accounts for $n = 1$), which is called the *Buchstab–de Bruijn’s Recurrence*.

This recurrence formula has been used for both lower and upper bounds and even for deriving asymptotic formulae.

We now use (11) to “prove” (8) for each fixed u .

The “proof” is by induction over N , where $u \in (N, N + 1]$. To ease the comprehension we deliberately ignore error terms and use the sign \approx without specifying its formal meaning. However, we do guarantee to the reader that more careful analysis can re-cast the following formulae into a proper proof.

We start with the 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 $x \geq y \geq \sqrt{x}$), noticing that non- y -smooth numbers have one and only one prime divisor $p \geq y$, we get:

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

Therefore, 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 note that the above step... has not really necessary. It is nonetheless a good warming exercise for the next “induction” step.

Suppose that

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

holds for $0 \leq u \leq N$.

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

Subtracting the Buchstab–de Bruijn relation (11) 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

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

since

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

so the induction hypothesis applies (error terms ignored).

We now recall the definition of the function $\vartheta(z)$ and the Prime Number Theorem (2). Writing $z = x^{1/t}$, by partial summation, we get

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

We confess that a terrible offence has just been committed: instead differentiating $\vartheta(z)$ we have differentiated its approximation z . However, more careful examination shows that the above formulae are still correct.

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”.

Estimating the largest prime divisor is a necessary step in many number-theoretic algorithms. For instance, Bach, von zur Gathen & Lenstra [3] introduce an algorithm for factoring polynomials over finite fields of characteristic p . The complexity of this algorithm depends on the largest prime divisor of the product the k cyclotomic polynomial $\Phi_k(p)$ evaluated at p . A relationship between a number and its largest divisor allows to tune k for every p and optimize complexity. We refer the reader to [130, 133, 156] for more related results.

6 Smoothness Miscellanea

6.1 Evaluating $\psi(x, y)$

To optimize (balance) the complexity of steps in several cryptographic algorithms, one often needs more precise information about $\psi(x, y)$ than current² estimates and asymptotic formulae can provide.

For example, Parsell & Sorenson [123], improving several previous results of Bernstein [15], have shown that for any parameter α , one can estimate $\psi(x, y)$ up to a factor $1 + \mathcal{O}(\alpha^{-1} \log x)$ in time

$$\mathcal{O}\left(\frac{\alpha y^{2/3}}{\log y} + \alpha \log x \log \alpha\right).$$

A number of related results can be found in [90, 136, 141, 142].

6.2 Constructing Constrained Smooth Numbers

Producing smooth numbers is trivial. However, constructing *constrained* smooth numbers appears to be a challenging problem. A natural constraint, stemming from the study of digital signatures, is the requirement that the y -smooth number belongs to a given interval $[x, x + z]$.

Boneh [19], motivated by certain cryptographic problems, has devised a polynomial-time algorithm solving this problem for some x, y, z parameter combinations.

²proven or conjectured

Results about the existence of very smooth numbers with a prescribed bit pattern at a certain position are given in [134], see also [74] which gives an alternative approach (via character sums instead of exponential sums) that may probably be used to further improve the aforementioned result of [134].

More research in this area is certainly very desirable.

6.3 Rough Numbers

An integer n is y -rough if all prime divisors $p \mid n$ are such that $p > y$. We denote by $\Omega(x, y)$ be the number of y -rough integers smaller than x .

Buchstab [29] gives the asymptotic formula:

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

where the *Buchstab* function $\omega(u)$ is defined as follows:

$$\omega(u) = \frac{1}{u} \times \begin{cases} 1, & \text{if } 1 \leq u \leq 2, \\ 1 - \int_1^{u-1} \omega(t) dt, & \text{if } u \geq 2. \end{cases}$$

Rough numbers can be viewed as “approximations” to primes. Rough numbers can be easily found and are proven to exist in various integer sequences of cryptographic interest. For example, Joye, Paillier & Vaudey [94] use rough numbers as “interesting” candidates for primality testing during cryptographic key generation.

6.4 Large Smooth Divisors

It also natural to ask how often integers are expected to have a large smooth divisor; or, from a more quantitative perspective, explore the behavior of:

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

While Θ has been addressed in the classical literature on smooth numbers, see [81, 144, 145], it has not received as much attention as $\psi(x, y)$.

Some asymptotic formulae for Θ have recently been given by Banks & Shparlinski [11] and Tenenbaum [147]. These formulae involve the parameter u appearing in (6) and a parameter v defined as:

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

The formulae also contain an integral involving $\rho(u)$ and its derivative. Part of the motivation in [11] comes from a cryptographic problem discussed by Menezes [113], see as well Section 9.12.

Shifted primes with large smooth divisors are studied by Pomerance & Shparlinski [126].

6.5 Next Largest Prime Divisors

Characterizing the second largest prime divisor is of interest too, as the complexity of factoring an integer n with Lenstra's elliptic curve factorization method [103] (commonly called '*the ECM*'), depends on this prime divisor.

More generally, denoting by $P_j(n)$ the j -th largest prime divisor of n , one may consider the joint distribution

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

The work of Tenenbaum [146] contains the most recent results and further references on this topic.

The case $k = 2$ is especially important. Indeed, using above notation, the ECM algorithm factors n completely in time:

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

where $p = P_2(n)$. This case has also got special attention in [4].

6.6 Other Facts

In this section we present several unrelated results, which while unlikely to have any obvious cryptographic applications, still prove interesting for our exploration of smooth numbers.

Balog and Wooley [8] 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 [8] is based on very nice and elementary explicit constructions.

One can also take $k \rightarrow \infty$ and $\varepsilon \rightarrow 0$ (slowly) when $n \rightarrow \infty$.

Balog [6] proved 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* claiming that all positive even integers $N \geq 4$ can be represented as the sum of two primes.

Finally, various bounds of rational exponential sums

$$S_{a,q}(x, y) = \sum_{\substack{n \leq x \\ n \text{ is } y\text{-smooth}}} \exp\left(2\pi i \frac{an}{q}\right), \quad \text{where } \gcd(a, q) = 1,$$

are given by Fouvry & Tenenbaum [63] and also by de la Bretèche & Tenenbaum [24]. Multiplicative character sums

$$T_{a,q}(x, y) = \sum_{\substack{n \leq x \\ n \text{ is } y\text{-smooth}}} \chi(n - a), \quad \text{where } \gcd(a, q) = 1,$$

with a nonprincipal multiplicative character χ modulo q are estimated [135]. We also note that asymptotic formulae for the sums

$$\sum_{\substack{a < n \leq x \\ n \text{ is } y\text{-smooth}}} \frac{\varphi(n - a)}{n - a} \quad \text{and} \quad \frac{1}{\psi(x, y)} \sum_{\substack{a < n \leq x \\ n \text{ is } y\text{-smooth}}} \varphi(n - a)$$

are given in [109]. On the other hand, obtaining asymptotic formulae (or even good estimates) for the sums

$$\frac{1}{\psi(x, y)} \sum_{\substack{a < n \leq x \\ n \text{ is } y\text{-smooth}}} \tau(n - a) \quad \text{and} \quad \frac{1}{\psi(x, y)} \sum_{\substack{a < n \leq x \\ n \text{ is } y\text{-smooth}}} \omega(n - a),$$

is still an open problem.

7 Smoothness 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 generalization of this question, which is also of cryptographic interest, is the study of smooth numbers with an additional congruence condition.

In particular, we introduce the counting functions

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

and

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

Tenenbaum [143] proved that

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

in a wide range of parameters.

In turn, a family of bounds of the forms

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

(of decreasing strength but in increasingly larger ranges of x , y and q) can be found in [7, 64, 77, 78, 84, 137]. Some of these bounds hold for all a with $\gcd(a, q) = 1$, others hold only for almost all such integers a .

We note that bounds of exponential and character sums $S_{a,q}(x, y)$ and $T_{a,q}(x, y)$, 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

We now turn our attention to the sequence of integers in “short” intervals $[x, x + z)$.

Accordingly,

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

It is natural to expect that:

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

in a wide range of x , y and z .

A series of ingenious results due to Balog [5], Croot [46], Friedlander & Granville [66], Friedlander & Lagarias [66], Harman [84] and Xuan [149] gives various interesting bits of information, but in general the status of this problem is far from being satisfactory.

Croot’s work [46] is particularly interesting as it uses a quite unusual tool: bounds of *bilinear Kloosterman sums* due to Duke, Friedlander & Iwaniec [50].

From both cryptographic and number theoretic perspectives, the main challenge is that of obtaining good lower bounds in $\psi(x, y, 4\sqrt{x})$, which appears to be currently out of reach. This case is of special importance as it is crucial for the rigorous analysis of Lenstra’s elliptic curve factoring algorithm [103]. We note that the result of Croot [46] applies to intervals of similar length but unfortunately for y values which are much larger than these appearing in [103].

Finally, we recall that Lenstra, Pila & Pomerance [104, 105] have found an ingenious way to circumvent this problem by introducing a hyperelliptic factoring algorithm. For this algorithm, smooth numbers in large intervals ought to be studied, which is already a feasible task. This has been achieved at the cost of very delicate arguments and required the developing of new algebraic and analytic tools by the authors.

7.3 Smooth Shifted Primes

Let $\pi_a(x, y)$ be the counting function of smooth shifted primes given in Section 2.4.

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

$$\pi_a(x, y) \sim \rho(u)\pi(x) \tag{12}$$

holds for a wide range of x and y . Unfortunately, results of such breadth seem unreachable using current techniques.

However rather strong upper bounds are known. For example, Pomerance & Shparlinski [126] gave the estimate

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

for

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

In a shorter range

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

the “right” upper bound

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

follows from a result of Fouvry & Tenenbaum [64, Theorem 4].

It is not just the asymptotic formula (12) which is presently out of reach. In fact, even the obtaining of lower bounds on $\pi_a(x, y)$ is an extremely difficult task where progress seems to be very slow.

The best known result, due to Baker & Harman [9] only asserts that there is a positive constant A such that for $a \neq 0$,

$$\pi_a(x, y) \gg \frac{\pi(x)}{(\log x)^A}$$

for $u \leq 3.377\dots$ (where as before, u is defined by (6)), see also [85].

For most applications the logarithmic loss in the density of such primes is not important. However, if this becomes an issue, one can use bound of Friedlander [65]:

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

which, however, is proven only for $u \leq 2\sqrt{e} = 3.2974\dots$

Finally, we recall yet another result of Baker & Harman [9] guarantees that

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

for $u \geq 1.477\dots$

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

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

- there are at least $C\pi(x)/(\log x)^A$ primes $p \leq x$ such that $p + a$ has a prime divisor $q \geq p^{0.6776}$;
- there are at least $C\pi(x)/(\log x)^A$ 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 hold with $A = 0$, with $1 - \varepsilon$ instead of 0.6776 and with ε instead of 0.2962 (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 the deterministic primality test of Agrawal, Kayal & Saxena [1].

7.4 Smooth Values of Polynomials

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

As in the case of shifted primes, rather strong upper bounds on $\psi_f(x, y)$ exist, see for example, the results of Hmyrova [89] and Timofeev [148].

For a squarefree polynomial f , Martin [111] gives an asymptotic formula of the type

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

where d_1, d_2, \dots, d_k are the degrees of irreducible factors of f over $\mathbb{Z}[x]$. This holds only for very large values of y . See as well several related results by Dartyge, Martin & Tenenbaum [48], where smooth values of polynomials at prime valuations (that is, of $f(p)$) are also discussed.

7.5 Smooth Totients

Let $\Phi(x, y)$ and $\Pi_a(x, y)$ be the functions counting smooth values of the Euler function on integers and shifted primes, respectively, see Section 2.4.

Banks, Friedlander, Pomerance & Shparlinski [10] have shown that in the range

$$y \geq (\log \log x)^{1+\varepsilon}$$

we have

$$\Phi(x, y) \leq x \exp(-(1 + o(1)) u \log \log u). \quad (13)$$

There are two interesting things to note about this result:

- the range is wider than that of (7), with $\log \log y$ instead of $\log y$.
- the bound is weaker than that of (7), with $\log \log u$ instead of $\log u$ in the exponent.

Both bullet points reflect the fact that the values of Euler functions tend to be smoother than integers and shifted primes. Furthermore, under a plausible conjecture about smooth shifted primes similar to (12), a matching lower bound on $\Phi(x, y)$ has been obtained in [10]. Under even stronger conjectures, Lamzouri [101] has obtained an asymptotic formula for $\Phi(x, y)$ and similar quantities related to iterations of the Euler function.

To estimate $\Pi_a(x, y)$, we can now use the trivial inequality

$$\Pi_a(x, y) \leq \Phi(x + |a|, y),$$

which, in fact, is quite sufficient for many applications. Furthermore, in a wide range of parameters, unless u is small, the above is equivalent to the expected estimate:

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

which is a full analogue of (13).

However for small values of u , obtaining the above estimate remains an important open problem.

On the other hand, in some ranges, several other bounds for $\Pi_a(x, y)$ have been obtained in [10] using different techniques (such as a sieve method). For example, for $y \geq \exp(\sqrt{\log x \log \log x})$ we have

$$\Pi_a(x, y) \ll \frac{\pi(x)}{u},$$

and for $y \geq \log x$, we have

$$\Pi_a(x, y) \leq \frac{\pi(x)}{\exp((1/2 + o(1))\sqrt{u} \log u)} + \frac{\pi(x) \log \log x}{\exp((1 + o(1))u \log u)}.$$

Note that estimates on the number of smooth values of the Euler function of polynomial sequences with integer and prime arguments, that is, $\varphi(f(n))$ and $\varphi(f(p))$, are given in [10] as well.

7.6 Smooth Cardinalities of Elliptic Curves

One aspect of this question has already been mentioned in Section 7.2 in relation to the elliptic curve factoring algorithm of Lenstra [103].

It is also interesting and important to study the arithmetic structure of cardinalities of the reductions of a given elliptic curve defined over Q modulo distinct primes. More precisely, given an elliptic curve \mathbf{E} over Q , we denote by $N_p = \#\mathbf{E}(\mathbb{F}_p)$ the cardinality of the set of rational points on the reduction of \mathbf{E} modulo p (for a sufficiently large prime p such reduction always leads to an elliptic curve over \mathbb{F}_p).

The number of prime divisors of N_p has been studied by Cojocaru [37], Iwaniec & Jiménez Urroz [91], Jiménez Urroz [95], Liu [106, 107, 108], Miri & Murty [116] and Steuding & A. Weng [139].

Some heuristics about the number of prime values of N_p for $p \leq x$ has been discussed by Galbraith & McKee [68], Koblitz [97, 98] and Weng [154]. An upper bound on this quantity is obtained by Cojocaru, Luca & Shparlinski [38], see also [36].

However, it seems that there are no smoothness results about the numbers N_p , although this issue has been touched in McKee [112]. Probably obtaining an asymptotic formula or even a good lower bound on the number of y -smooth values of N_p for $p \leq x$ is very hard, but perhaps some upper bounds can be established.

Results of this kind are of great importance for the elliptic curve cryptography.

7.7 Smooth Class Numbers

For an integer $d < 0$ we denote by $h(d)$ the class number of the imaginary quadratic field $\mathbb{Q}(\sqrt{d})$. Let \mathcal{D} be the set of *fundamental discriminants*, that is, the set of integers $d < 0$ such that

- either $d \equiv 1 \pmod{4}$ and d is square-free,
- or $d \equiv 0 \pmod{4}$, $d/4 \equiv 2, 3 \pmod{4}$ and $d/4$ is square-free.

Using the so-called *Cohen–Lenstra heuristics* for divisibility of class numbers, see [35], Hamdy & Saidak [82] derived a conditional asymptotic formula for the number of $d \in \mathcal{D}$ with $-d \leq x$ for which $h(d)$ is y -smooth. Unfortunately, this seems to be the only known result in this really exciting direction, see [28] for the relevance of these problems to cryptography. Studying the smoothness of class numbers of other fields is of great interest too but is perhaps a very hard question.

7.8 Smooth Numbers in Sumsets

We recall that de la Bretèche [22] gives a result of surprising generality and strength stating 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

$$\exp((\log x)^{2/3+\varepsilon}) < y \leq x,$$

we have

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

where, as usual, u is given by (6).

Although the authors are unaware of any immediate cryptographic applications of this result, we underline its high potential, given its generality and “condition-free” formulation.

Several other relevant results are given by Croot [45].

7.9 Smooth Polynomials Over Finite Fields

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

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

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

Define

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

(the last expression makes the analogy with formula (6) completely explicit).

The systematic study of $N_q(m, k)$ dates back to the work of Odlyzko [122] who also discovered the relevance of this quantity to the discrete logarithm problem in finite fields.

Several very precise results about $N_q(m, k)$ have recently been given by Bender & Pomerance [14]. For example, by [14, Theorem 2.1] we have

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

as $k \rightarrow \infty$ and $u \rightarrow \infty$, uniformly for $q^k \geq m(\log m)^2$, and by [14, Theorem 2.2] we also have

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

for $k \leq \sqrt{m}$.

8 Distribution of Divisors

8.1 More On Intuition

We have already seen in Section 4.1 that carelessly applied intuition may lead to wrong conclusions. The following is yet another example.

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

Let us relax the relation $n = k^2$ and consider $n = km$ with $k \leq m \leq k^{1.001}$ for $k, m \in N$. Such integers can be called quasi-squares.

It is natural to ask whether the density of quasi-squares is still small. Say, are there only $o(x)$ quasi-squares up to x ? As the below results indicate, such quasi-squares form a set of positive density which perhaps does not match the intuition (too bad for the intuition ...).

8.2 Notations

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

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

As usual, in the case of $\mathcal{A} = \mathbb{N}$ we define

$$H(x, y, z) = H(x, y, z; \mathbb{N}).$$

8.3 Natural Numbers

This case goes back to two old questions of Erdős:

Given an integer N what is the size $M(N)$ of the multiplication table $\{nm : 1 \leq m, n \leq \sqrt{N}\}$?

and

Is it true that almost all integers n have two divisors $d_1 | n$ and $d_2 | n$ with $d_1 < d_2 < 2d_1$?

Hall & Tenenbaum's book [81] contains a very detailed treatment of such questions. For example, for the size of the multiplication table we have from [58, Corollary 3]

$$M(N) \asymp \frac{N}{(\log N)^\delta (\log \log N)^{3/2}}$$

as $N \rightarrow \infty$, where

$$\delta = 1 - \frac{1 + \log \log 2}{\log 2} = 0.008607\dots \tag{14}$$

is the *Erdős number*, see also [144, Theorem 23] for a slightly less precise result.

Ford [58] has recently obtained a series of remarkable improvements of several previously known results. Several related results can also be found in [23, 59, 62, 100, 144]. Unfortunately, exact formulations of precise results lead to rather cluttered technical conditions and estimates, which also depend on the relative sizes of x , y and z as well as on $z - y$ and z/y . Thus we limit our discussion to only a few sample results.

For example, let us define $v > 0$ by the relation

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

Then, by [58, Theorem 1] for any real x , y and z with

$$x \geq \max\{100000, y^2\} \quad \text{and} \quad x \geq z \geq y \geq 100$$

we have

$$\frac{H(x, y, z, \mathbb{N})}{x} \asymp \begin{cases} v^\delta (\log v)^{-3/2} & \text{if } 2y \leq z \leq y^2, \\ 1 & \text{if } z \geq y^2, \end{cases}$$

where δ is the Erdős number.

In particular, we see that for $\varepsilon > 0$ and any sufficiently large y , we have

$$H(x, y, y^{1+\varepsilon}, \mathbb{N}) \gg x, \tag{15}$$

where the implied constant only depends on $\varepsilon > 0$. Thus, there is a positive density of integers $n \leq x$, depending only on $\varepsilon > 0$, which have a divisor $d \mid n$ in the interval $d \in [y, y^{1+\varepsilon}]$.

We now prove (15) in the special case where y is a power of x . That is, we prove that for $0 < \alpha < \beta < 1$:

$$H(x, x^\alpha, x^\beta, \mathbb{N}) \gg x. \tag{16}$$

In our proof we consider only *prime divisors* $p \in [x^\alpha, x^\beta]$ (instead of integer divisors) and make the following two trivial observations:

- there are $x/p + \mathcal{O}(1)$ integers $n \leq x$ divisible by p ;
- each $n \leq x$ may have at most $K = \lceil 1/\alpha \rceil$ of them.

Hence,

$$H(x, x^\alpha, x^\beta, \mathbb{N}) \geq \frac{1}{K} \sum_{x^\alpha \leq p \leq x^\beta} \left(\frac{x}{p} + \mathcal{O}(1) \right)$$

and the sum on the right hand side counts every integer $n \leq x$ with a prime divisor $p \in [x^\alpha, x^\beta]$ at most K times. Therefore

$$H(x, x^\alpha, x^\beta, \mathbb{N}) \geq \frac{x}{K} \sum_{x^\alpha \leq p \leq x^\beta} \frac{1}{p} + \mathcal{O}(x^b).$$

By the Mertens formula, we now obtain

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

and (16) follows.

There are other tell tale signs that integer divisors are densely distributed. For example, for an integer $s \geq 1$ we denote

$$T(n) = \max_{i=1, \dots, \tau(n)-1} \frac{d_{i+1}}{d_i},$$

where $1 = d_1 < \dots < d_{\tau(n)} = n$ are the positive divisors of n . Clearly,

$$T(n) \leq P(n).$$

However, for many integers $T(n)$ is much smaller than $P(n)$. By a result of Saias [132, Theorem 1], we know that for any fixed t and sufficiently large x ,

$$\#\{n \leq x : T(n) \leq t\} \asymp \frac{x \log t}{\log x}.$$

8.4 Shifted Primes

Ford [58] has given upper bounds on $H(x, y, z; \mathcal{P}_a)$ of about the same strength as these applying to $H(x, y, z; \mathcal{N})$, where \mathcal{P}_a is defined in Section 2.3.

The situation with lower bounds $H(x, y, z; \mathcal{P}_a)$ is quite bleak, although heuristically there is little doubt that $H(x, y, z; \mathcal{P}_a)$ should behave similarly to $H(x, y, z; \mathcal{N})$.

One of the very few known lower bounds (yet, with many cryptographic applications) is given in [58, Theorem 7]: for $a \neq 0$ and $0 < \alpha < \beta$:

$$H(x, x^\alpha, x^\beta, \mathcal{P}_a) \gg \pi(x)$$

(where the implied constant depends on a , α and β).

The proof is similar to our proof of (16), but requires some technical analytic number theory tools, namely, the Bombieri–Vinogradov theorem, see Section 3.1, since instead of *integers* $n \leq x$ with $p \mid n$ we need to count *primes* $q \leq x$ with $p \mid (q - a)$.

To implement this approach one also needs the elementary observation that is enough to consider only the case $0 < \alpha < \beta \leq 1/2$ (since if $d \mid n$ then $(n/d) \mid n$).

8.5 Polynomials

Unfortunately there seem to be no results about the distribution of integers divisors of polynomials. Nonetheless, this problem does not look hopeless.

8.6 Totients

Here is 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 $H(x, y, z; \varphi(\mathbb{N}))$ is very different.

Given that typical values of the Euler function

- have more prime divisors, due to a result of Erdős & Pomerance [53],
- have more integer divisors, due to a result of Luca & Pomerance [110],
- are smoother, due to a result of Banks, Friedlander, Pomerance & Shparlinski [10], see also Section 7.5,

than a typical integer, it is also natural to expect that totients have denser divisor sets. This is supported by several recent results of Ford & Hu [60], who in particular show that

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

9 Cryptographic Applications

9.1 Smoothness in Factoring and Discrete Logarithms

Most integer factorization algorithms, such as Dixon’s method, the Quadratic Sieve, index calculus, the Number Field Sieve or Elliptic Curve Factoring have been designed and analyzed (either rigorously or heuristically) using our current *knowledge and understanding* of smooth numbers. The same also applies to many primality tests and algorithms for solving the discrete logarithm problem.

Results about the arithmetic structure of “typical” integers are therefore of high cryptographic relevance. As most results are already well publicized in the community, we illustrate them by one example (next section) and refer the reader to [44] for further information.

Nonetheless, new results and applications keep appearing regularly. The works of Croot, Granville, Pemanle & Tetali [47] and of Agrawal, Kayal & Saxena [1] are typical examples.

In [47] various results about the arithmetic structure of integers are used to give a very precise analysis of Dixon’s factoring algorithm. In [1] results about shifted primes with a large divisor, see Section 7.3, form the core of the algorithm.

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

We start by highlighting the role of smooth numbers in algorithms solving the Discrete Logarithm Problem:

Namely, given two integers a and b and a prime p we consider the problem of computing k (denoted $k = \text{Dlog}_a b$) such that $b \equiv a^k \pmod{p}$ and $0 \leq k \leq p - 2$.

The algorithm is assembled in two steps. We first use a certain (very strong) assumption and then show how to get rid of it.

Initial Assumption: Let us fix some y (to be optimized later) and *assume* that we know the discrete logarithms of *all* primes p_1, \dots, p_s up to y where $s = \pi(y)$.

Under this assumption we perform the following steps:

Step 1: Pick a random integer m and compute

$$c \equiv ba^m \equiv a^{k+m} \pmod{p}, \quad 0 \leq c < p.$$

Note that

$$\text{Dlog}_a c \equiv \text{Dlog}_a b + \text{Dlog}_a a^m \equiv \text{Dlog}_a b + m \pmod{p - 1}.$$

The *cost* of this step is negligible.

Step 2: Try to factor c , assuming that c , treated as an integer, is y -smooth.

Let

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

For doing so use trial division or the elliptic curve factorization algorithm [103].

Note that

$$\text{Dlog}_a c \equiv \alpha_1 \text{Dlog}_a p_1 + \dots + \alpha_s \text{Dlog}_a p_s \pmod{p - 1}.$$

The *cost* of this step is about y operations (less if [103] is used).

Step 3: If the previous step succeeds, output

$$\text{Dlog}_a b \equiv \alpha_1 \text{Dlog}_a p_1 + \dots + \alpha_s \text{Dlog}_a p_s - m \pmod{p-1},$$

otherwise repeat the first step.

The *cost* of this step is about $p/\Psi(p, y) = u_p^{(1+o(1))u_p}$ iterations, where

$$u_p = \frac{\log p}{\log y}$$

(under the assumption that $c < p$ is a random).

Thus the *total cost*, ignoring nonessential factors, is about $yu_p^{u_p}$.

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

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

but... it is premature to celebrate the victory, as we need to get rid of the assumption that the discrete logarithms of all small primes are available.

Removing the Assumption: We apply the same algorithm for each p_i , $i = 1, \dots, s$ as b . Then at Step 3 we get a congruence

$$\text{Dlog}_a p_i \equiv \alpha_{1,i} \text{Dlog}_a p_1 + \dots + \alpha_{s,i} \text{Dlog}_a p_s - m_i \pmod{p-1},$$

for $i = 1, \dots, s$.

We cannot find $\text{Dlog}_a p_i$ immediately but after getting such relations for every p_i , $i = 1, \dots, s$, we have a system of s linear congruences in s variables. If the system is not of full rank we continue to generate a few relations until a full rank system is reached (this overhead is negligible as most “random” matrices are non-singular). Therefore the cost of creating such a system of congruences is about $y^2 u_p^{u_p}$ and the cost of solving it is about y^3 (lesser if fast linear algebra algorithms are used, see, for example, [69]).

Choosing y optimally, we obtain an algorithm of complexity

$$\exp\left(\mathcal{O}\left(\sqrt{\log p \log \log p}\right)\right).$$

The above approach can be improved and optimized in many ways finally yielding a *subexponential* algorithm of asymptotic complexity

$$\exp\left(\sqrt{(2+o(1))\log p \log \log p}\right)$$

that can also be rigorously analyzed; this is done by Pomerance in [125].

We have presented the above example because of its illustrative value although a much faster algorithm exists: the *number field sieve*, see [44], whose complexity is $\mathcal{O}((\log p)^{1/3}(\log \log p)^{2/3})$.

One can note that the above approach uses both the structure of finite fields and the properties of smooth numbers. Thus a prime field is substantial. Over an extension of a field of small characteristic, such as \mathbb{F}_{2^n} , elements can be represented by polynomials and thus smooth polynomials play the role of smooth integers. Hence, the results of Section 7.9 become of great importance.

We note that despite a very common belief that the discrete logarithm problem is solvable in subexponential time, this is *not proved* as we write these lines.

In other words, although over the last decade fast heuristic algorithms for the discrete logarithm problem have been designed to work over any finite field, rigorous subexponential algorithms are known only for very specific fields (such as prime fields \mathbb{F}_p , their quadratic fields \mathbb{F}_{p^2} or fields \mathbb{F}_{p^m} with a fixed p), see [44, Section 6.4] for more details.

It is also clear that the above approach does not apply to the discrete logarithm problem in the elliptic curve settings where smoothness admits no analogous notion.

9.3 Textbook ElGamal Encryption

The ElGamal cryptosystem [51] makes use of two primes p, q with $q \mid p - 1$ and an element $g \in \mathbb{F}_p$ of order q (all of which are public), see also [27, Section 8.6], or [114, Sections 8.4.1 and 8.4.2], or [140, Section 6.1] for further details.

The receiver chooses a random *private key* element $x \in \mathbb{Z}_q$ and computes the *public key* $X = g^x \in \mathbb{Z}_q$.

Encryption: To encrypt a message $\mu \in \mathbb{F}_p$, the sender chooses a random $r \in \mathbb{Z}_q$, computes $R = \mu X^r \in \mathbb{F}_p$, and $Q = g^r \in \mathbb{F}_p$ and sends the pair $(R, Q) = (\mu X^r, g^r)$.

Decryption: The receiver computes (in \mathbb{F}_p)

$$S = Q^x = g^{xr} = X^r \quad \text{and} \quad \frac{R}{S} = \frac{R}{X^r} = \mu.$$

As most public key cryptosystems, the ElGamal protocol is quite slow. It is hence traditionally used to wrap a block-cipher key used for securing the subsequent communication flow.

Doing this in a “textbook fashion”, means that μ is a rather small integer. For example, p can be about 500 bits long to thwart discrete logarithm calculation attempts, but μ can be only 80 bits long to resist the brute force search.

Boneh, Joux & Nguyen [20] have shown that in this case, with a reasonable probability, μ can be recovered significantly faster than by any of the above two attacks.

Let \mathcal{G}_q be the subgroup of \mathbb{F}_p^* of order q generated by g . We note that $R = \mu U$ where $U \in \mathcal{G}_q$.

Let us assume that $1 \leq \mu \leq M$ (where M is much smaller than p). We also choose some bound K which is a parameter of the algorithm (controlling the trade-off between complexity and success probability).

Step 1: Compute $R^q = \mu^q U^q = \mu^q$.

Step 2: For $k = 1, \dots, \lceil K \rceil$ compute, sort and store k^q in a table.

Step 3: For $m = 1, \dots, \lceil M/K \rceil$ compute

$$\frac{R^q}{m^q} = \left(\frac{\mu}{m}\right)^q$$

and check whether this value is present in the table of Step 2.

Step 4: Output $\mu = km$ if there is a *match*.

This algorithm always works with $K = M$ (for example, $m = \mu$, $k = 1$, which is essentially a form of brute force search).

A better choice is $K = M^{1/2+\varepsilon}$. Using (16), we see that the algorithm succeeds for a positive proportion of messages. That is, it works because with a sufficiently high probability a random positive integer $\mu \leq M$ has a representation $\mu = km$ with $1 \leq k, m \leq M^{1/2+\varepsilon}$.

In other words, taking $M = 2^{80}$ (as in the above example as a standard key size for a private key cryptosystem) we see that the attack runs in a little more than 2^{40} steps.

9.4 Affine RSA Padding

The RSA signature scheme [129] makes use of the following parameters: a composite modulus N , a public exponent e and private exponent d which satisfy the congruence:

$$ed \equiv 1 \pmod{\varphi(N)},$$

see also [27, Section 8.3], or [114, Section 8.2], or [140, Section 5.3].

The *signature* $s \in \mathbb{Z}_N$ of a message $m \in \mathbb{Z}_N$ is computed as follows: $s \equiv m^d \pmod{N}$. *Verification* consist in checking that $m \equiv s^e \pmod{N}$.

If this is applied in this “textbook” form, the scheme becomes susceptible to a *chosen message attack* which works as follows.

Assume that the attacker, wishing to sign a target message m , has the ability to ask the legitimate signer to sign seemingly meaningless messages. Then the attacker can:

- choose a random m_1 and compute

$$m_2 \equiv \frac{m}{m_1} \pmod{N};$$

- query the signatures $s_i \equiv m_i^d \pmod{N}$ for $i = 1, 2$ from the legitimate signer;
- and compute $s \equiv s_1 s_2 \pmod{N}$.

This works because

$$s \equiv s_1 s_2 \equiv m_1^d m_2^d \equiv (m_1 m_2)^d \equiv m^d \pmod{N}.$$

In other words, because RSA is *homomorphic* with respect to multiplication, a multiplicative relation between messages shadows a similar relation between the signatures.

A natural defense against this attack is to restrict the signature and the verification algorithms to messages of a prescribed *structure*. For example, if N is n bits long, it is requested that the meaningful message part m is only ℓ bits long to which a fixed $(n - \ell)$ -bit string (called *padding pattern*) is appended. Clearly in the above example m_1 can be chosen to comply with this format but m_2 is unlikely to fulfill this constraint, which thwarts the attack.

In the case of affine padding, signed messages have the following structure:

$$\boxed{\text{fixed } (n - \ell)\text{-bit padding } P \mid \ell\text{-bit message } m}$$

Thus, denoting $R(m) = P + m$ we see that the signature $s(m)$ of an ℓ bit message m is computed as

$$s(m) \equiv R(m)^d \pmod{N}, \quad 1 \leq s(m) \leq N,$$

(that is, $P = 2^\ell \Pi$ where Π is the appended padding pattern).

In a thread of works by Misarsky [117], Girault & Misarsky [71, 72] and Brier, Clavier, Coron & Naccache [25], *existential forgery* attacks on affine-padded RSA signatures have been progressively developed and refined.

Lenstra & Shparlinski [102] have improved [25] by redesigning it as a *selective forgery* attack, where the attacker can sign *any* message.

Let us start by presenting the basic technique introduced in [25].

Our goal is to find four distinct ℓ -bit messages m_1, m_2, m_3, m_4 such that

$$R(m_1) \cdot R(m_2) \equiv R(m_3) \cdot R(m_4) \pmod{N}. \quad (17)$$

In this case we obtain

$$s(m_1) \cdot s(m_2) \equiv s(m_3) \cdot s(m_4) \pmod{N},$$

and hence a signature on m_3 can be computed from signatures on m_1, m_2, m_4 . In [25] this has been applied to the case where all four messages are considered as variable m_1, m_2, m_3, m_4 (which leads to an existential signature forgery), while in [102] the message m_4 is assumed to be fixed (which leads to a selective signature forgery).

One verifies that the congruence (17) is equivalent to

$$P(m_3 + m_4 - m_1 - m_2) \equiv m_1 m_2 - m_3 m_4 \pmod{N}.$$

With

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

this becomes

$$(P + m_4)z \equiv xy \pmod{N}. \tag{18}$$

We note that if m_4 is already chosen, the values of x, y and z define m_1, m_2 and m_3 uniquely.

The congruence (18) is trivial to solve without any restrictions on the variables, but in fact we need “small” x, y and z about ℓ bits long, which is a much harder constraint to deal with.

We show how to solve it when

$$\ell = \left(\frac{1}{3} + \varepsilon\right) n.$$

Before we proceed with the algorithm we note that this choice of ℓ is close to the limit of this approach given that for any fixed $\varepsilon > 0$ a “typical” polynomial congruence in three variables

$$F(x, y, z) \equiv 0 \pmod{N} \tag{19}$$

is unlikely to have a integer solution (x, y, z) with

$$1 \leq x, y, z \leq N^{1/3-\varepsilon}. \tag{20}$$

This is because $F(x, y, z)$ takes only $N^{1-3\varepsilon}$ possible values for such x, y and z , thus (19) is solvable under the condition (20) only with exponentially small “probability” of order $N^{-3\varepsilon}$ (this estimate assumes that F behaves like a random trivariate function and hence must not be taken literally).

Now, to find ℓ -bit solutions to the congruence (18) we first consider the congruence

$$(P + s)z \equiv w \pmod{N}, \quad (21)$$

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

$$w \leq N^{2/3+2\varepsilon} \quad \text{and} \quad |z| \leq N^{1/3}.$$

Let R_i/Q_i denote the i -th continued fraction convergent to $(P + s)/N$, $i = 1, 2, \dots$

Then

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

We now define j by the inequalities $Q_j < N^{1/3} \leq Q_{j+1}$ and set

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

Then

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

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

Now, one can certainly try to apply the above procedure with $s = m_4$ and then check whether w can be factored as $w = xy$ with $1 \leq x, y < 2^\ell$. However, this would unfortunately happen only for a rather sparse sequence of messages m_4 , so we may try to randomize the above idea as follows:

- Pick a random integer r with $0 \leq r < N^\varepsilon/2$ and find

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

with $1 \leq w < N^{2/3}$.

- Let $u = w + r \lfloor N^{1/3} \rfloor z$, thus

$$u \equiv (P + m_3)z \pmod{N}$$

and $1 \leq u < N^{2/3+\varepsilon}$ (provided that N is large enough).

- Try to factor u using the elliptic curve factoring method which requires $\exp((2 + o(1))\sqrt{\log p \log \log p}) (\log N)^{O(1)}$ bit operations, where $p = P_2(u)$. Abort this steps if this takes longer then some selected time bound. Thus we abort this steps if $P_2(u)$ is large.
- Try to find x, y with $u = xy$ and $1 \leq x, y < 2^\ell$.
- If successful, compute m_1, m_2, m_3 , otherwise try another pair (u, z) .

The above works because eventually we hit a reasonably good u of the form $u = P(u)v$ where $p = P(u) \leq 2\sqrt{u}$ and $P(v) = P_2(u)$ is small.

For such a u it is easy to find a representation $u = xy$ with integers x and y of the desired size, $1 \leq x, y \leq 2^\ell$.

The algorithm seems to be very hard to analyze rigorously, but the heuristic analysis given in [102] predicts the runtime as $L_N(1/3, 1)$ which is substantially faster than

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

where

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

for $N \rightarrow \infty$.

Furthermore, Lenstra & Shparlinski [102] give a 1024-bit affine padding forgery example while direct factorization of moduli of this size is currently beyond reach.

A challenging open question is to find a way to use more signatures and thereby extend the range of ℓ which can be attacked. Recent progress on this question, from a rather unexpected direction, can be found in a work by Joux, Naccache & Thomé [93].

Finally, although unrelated, other recreative applications of *ad-hoc* factoring in cryptanalysis can be found in [41] and [42].

9.5 Desmedt-Odlyzko Attack

The attacks that we have just described work because of the inability of the affine padding to eradicate the homomorphic properties of RSA. However, there are other attacks that apply in theory to any type of message padding.

In [49], Desmedt & Odlyzko describe an existential RSA signature forgery scenario. Here, the opponent is allowed to query from the legitimate signer e -th roots (signatures) of validly padded messages of his choosing. Having done so, the opponent crafts a new validly padded signature of his own on a message left unsigned by the legitimate signer. The attack works as follows:

Step 1: Select a bound y and let p_1, \dots, p_s be the primes up to y , that is, $s = \pi(y)$.

Step 2: Find $k + 1$ messages m_i which are y -smooth and factor them

$$m_i = \prod_{j=1}^k p_j^{\alpha_{i,j}}$$

(for example, using elliptic curve factorization [103] or trial division).

Step 3: Solve in $u_1, \dots, u_k \in \{0, \dots, e - 1\}$

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

and write

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

Thus

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

where

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

Step 4: Obtain from the legitimate signer the signatures s_i on m_i for $i = 1, \dots, k$ and forge the signature on m_{k+1} as

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

We see that s is a *valid signature* since:

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

Clearly, in a real-life scenario, instead of generating the y -smooth messages m_1, \dots, m_k on which valid signatures are required, the attacker may also passively monitor the legitimate date exchange and test each signed message for smoothness, putting aside smooth messages together with their signatures.

Using such a strategy, the expected occurrence of a smooth m is about once in

$$\frac{N}{\Psi(N, y)} = u_N^{(1+o(1))u_N}$$

signature rounds, where

$$u_N = \frac{\log N}{\log y}$$

(under the assumption that $m < N$ is random).

Thus, the effort of collecting y messages is $y u_N^{(1+o(1))u_N}$ which for

$$y = \exp\left(\sqrt{0.5 \log N \log \log N}\right)$$

optimizes as $\exp\left(\sqrt{(2 + o(1)) \log N \log \log N}\right)$.

Coppersmith, Coron, Grieru, Halevi, Jutla, Naccache & Stern [40] added a number of improvements and generalizations to this attack and applied it successfully to a number of industry standards.

9.6 Small Prime Based Public-Key Encryption

Products of small primes can also be used for public-key encryption. The idea, due to Naccache & Stern [120], is based on the following problem:

Given a prime p , a positive integer $f < p$ and a set of integers $\{v_1, \dots, v_n\}$, find a binary vector x such that

$$f \equiv \prod_{i=1}^n v_i^{x_i} \pmod{p},$$

if such a vector exists.

It is easy to observe that if the v_1, \dots, v_n are relatively prime and much smaller than p , then the exponent vector x can be found in polynomial time by factoring f . Indeed, instances where

$$p > \prod_{i=1}^n v_i \quad \text{and} \quad \gcd(v_i, v_j) = 1, 1 \leq i < j \leq n,$$

are easy.

Such an easy instance can be hidden by extracting the s -th modular root of each v_i , where s is a secret integer with $\gcd(s, p-1) = 1$.

More formally, let p be a large public prime and denote by n the largest integer such that:

$$p > \prod_{i=1}^n p_i$$

where p_i is the i -th prime.

The secret-key $s < p-1$ is a random integer such that $\gcd(p-1, s) = 1$ and the public-keys are the s -th roots:

$$u_i \equiv p_i^r \equiv p_i^{1/s} \pmod{p}, \quad 0 \leq u_i < p, \quad i = 1, \dots, n,$$

where r satisfies

$$rs \equiv 1 \pmod{p-1}.$$

An n -bit message (m_1, \dots, m_n) is encrypted as

$$c \equiv \prod_{i=1}^n u_i^{m_i} \pmod{p}$$

and recovered by computing

$$f \equiv c^s \pmod{p}, \quad 0 \leq t < p,$$

and then

$$m_i = \begin{cases} 0, & \text{if } p_i \nmid f, \\ 1, & \text{if } p_i \mid f, \end{cases} \quad i = 1, \dots, n.$$

We refer the reader to [120] and [34] for more information on this somewhat unusual public-key encryption scheme, whose encoding idea dates back to 1931, see Section 9.7.

9.7 Gödel Numbers

In his famous work published in 1931, Gödel [73] uses a mapping of mathematical expressions into integers based on divisibility by small prime factors.

Gödel [73] starts by assigning a unique natural number $\tau(\xi)$ to each basic mathematical symbol ξ in the formal language of arithmetic he is dealing with³ (in other words τ is a symbol-to-integer dictionary).

To encode an entire mathematical *expression* Ξ , which is nothing but an ordered sequence of mathematical symbols:

$$\Xi = \langle \xi_1, \dots, \xi_n \rangle.$$

Gödel [73] uses the following system: Each atomic symbol being associated to a positive integer *via* τ , the mathematical expression is mapped into \mathbb{N} as:

$$\tau(\Xi) = \prod_{i=1}^n p_i^{\tau(\xi_i)} \in \mathbb{N}, \quad (22)$$

where p_i stands for the i -th prime.

Given that any number obtained this way can be uniquely factored into prime factors, it is possible to effectively and unambiguously recover any mathematical expression Ξ from its Gödel number $\tau(\Xi)$ (22).

Gödel [73] uses this scheme at two levels: first, to encode sequences of symbols representing formulae, and second, to encode sequences of formulae representing proofs. This has allowed him to show a correspondence between statements about natural numbers and statements about the provability of theorems dealing with natural numbers, which is the cornerstone of the celebrated Gödel Incompleteness Theorem [73].

9.8 Error Correction with Products of Small Primes

Interestingly, Gödel's encoding (22) can also be used for error correction.

Error-correcting codes are used to protect information sent over noisy channels against transmission errors. In [43, 119], Coron & Naccache describe an unusual error-correcting code based on modular arithmetic.

³for example, $\xi \in \{\exists, \forall, \Rightarrow, +, -, \times, \div, 0, 1, 2, \dots\}$

Let m be the n -bit message to encode; we denote by m_i the i -th bit of m . We let p_i be the i -th prime, starting with $p_1 = 2$. Let t be the number of errors which can be corrected. We generate a prime p such that:

$$2p_n^{2t} \leq p < 4p_n^{2t} \quad (23)$$

(which, of course, always exists).

Given m , we generate the following “redundancy”:

$$c(m) \equiv \prod_{i=1}^n p_i^{m_i} \pmod{p}, \quad 0 \leq c(m) < p. \quad (24)$$

The integer $c(m)$ is protected by using an error-correcting code μ resilient to t transmission errors.

The encoded message $E(m)$ is defined as $E(m) = \langle m, \mu(c(m)) \rangle$.

Let $\tilde{E}(m)$ be the received version of $E(m)$ where at most t errors occurred:

$$\tilde{E}(m) = E(m) \oplus e$$

where e is an error vector of Hamming weight at most t , and \oplus stands for bit-wise addition.

Splitting $e = \langle e_m, e_c \rangle$ into parts corresponding to errors in m and to errors in $\mu(c(m))$, we obtain

$$\tilde{E}(m) = \langle \tilde{m}, \tilde{c}(m) \rangle = \langle m \oplus e_m, \mu(c(m)) \oplus e_c \rangle.$$

Since μ can correct t errors, and e_c is only a part of e (whose total Hamming weight is t), $c(m)$ can be safely recovered from $\mu(c(m)) \oplus e_c$.

The receiver computes:

$$s \equiv \frac{c(\tilde{m})}{c(m)} \equiv \frac{c(m \oplus e_m)}{c(m)} \pmod{p}.$$

Using (24) the integer s can be written as:

$$s \equiv a/b \pmod{p},$$

where

$$a = \prod_{\substack{i=1, \dots, n \\ \tilde{m}_i=1 \\ m_i=0}} p_i \quad \text{and} \quad b = \prod_{\substack{i=1, \dots, n \\ \tilde{m}_i=0 \\ m_i=1}} p_i.$$

Since \tilde{m} suffered at most t errors, we have

$$\max\{a, b\} < p_n^t.$$

A result of Stern, Fouque & Wackers [138] shows that given s one can recover a and b efficiently. The algorithm is based on the Gauss reduction algorithm for finding the shortest vector in a two-dimensional lattice [150]. More precisely, let p be a prime with $p > 2AB$ for some $A \in \mathbb{R}$ and $B \in \mathbb{R}$. Let $a, b \in \mathbb{Z}$ be such that $|a| \leq A$ and $0 < b \leq B$. Then given p , A , B and $s \equiv ab^{-1} \pmod{p}$, one can recover a and b in polynomial time. Note that the condition $p > 2AB$ guarantees the uniqueness of a and b . A very similar argument has been used in Section 9.4 to find small solutions to (21).

Taking $A = B = p_n^t - 1$, we have from (23) that $2AB < p$. Moreover, $0 \leq a \leq A$ and $0 < b \leq B$. Therefore, we can recover a and b from s in polynomial time. By testing the divisibility of a and b by the small primes p_1, \dots, p_n , one can recover $e_m = \tilde{m} \oplus m$ and hence $m = \tilde{m} \oplus e_m$.

The process assumes the existence of an error correcting code μ . Note that μ can be nothing but the procedure that we have just presented in miniature. In other words, the described encoding procedure can be iterated to protect $c(m)$ using a new, much smaller, set of primes. In turn, yet another encoding iteration is used at the third level of encoding and so on. Finally, the smallest and last layer can be protected by simple replication ($2t + 1$ times) and decoded using a majority vote.

The proposed code turns out to provide efficient decoding for some specific parameter combinations. For instance, denoting by μ Reed-Muller encoding, and assuming that 5812-bit messages need to be protected against 31 transmission errors, the size of $\mu(m)$ is 8192 bits, whereas the hybrid encoding $\langle m, \mu(c(m)) \rangle$ is only 7860 bits long.

More examples and details can be found in [43, 119]. The full asymptotic analysis of this scheme still remains to be worked out.

9.9 Private Information Retrieval with Products of Small Primes

A *Private Information Retrieval* (PIR) scheme is a combination of encoding and encryption which allows a user to retrieve the k -th bit of an n -bit database, without revealing to the database owner the value of k .

Gentry & Ramzan [70] have used the Chinese Remainder Theorem and properties of products of small primes to design a PIR scheme. The construction of [70] requires a cyclic group \mathcal{G} whose order $t = \#\mathcal{G}$ has a prescribed arithmetic structure; namely a product of a large prime and a very smooth integers. This makes the results of [11, 126, 147] relevant to this problem, see also Section 6.4.

9.10 Zero-Knowledge with Products of Small Primes

A zero-knowledge proof (ZKP) is a protocol allowing Alice to convince Bob that she knows a secret s without revealing to Bob information on s .

The best-known ZKP is probably the protocol of Fiat & Shamir [54] which uses an RSA modulus N and k quadratic residues v_i as public parameters. In its simplest version, Alice uses the k modular square roots s_i such that $s_i^2 \equiv v_i \pmod{N}$ as secret identification keys. The protocol is:

- Alice picks a random $r \in \mathbb{Z}_N$ and sends to Bob $x \equiv r^2 \pmod{N}$.
- Bob picks a random binary vector

$$e = \langle e_0, \dots, e_{k-1} \rangle$$

and sends it to Alice.

- Alice replies to Bob with:

$$y \equiv r \prod_{i=0}^{k-1} s_i^{e_i} \pmod{N}.$$

- Bob verifies that:

$$y^2 \equiv x \prod_{i=0}^{k-1} v_i^{e_i} \pmod{N}.$$

To ease Bob's computational burden, Micali & Shamir [115] suggest to use very small v_i -values. As it turns out, using small primes as v_i -values presents particular security and simplicity advantages.

9.11 The Generalized Diffie-Hellman Problem

Recently, several cryptographic schemes based on the following assumption appeared:

Let g be an element of prime order p of a “generic” Abelian group \mathcal{G} . That is, we assume that \mathcal{G} is a group where only “generic” attacks, such as Shanks’ or Pollard’s algorithms exist and take about \sqrt{p} operations, see [27, Sections 10.3 and 10.4], or [114, Sections 3.6.2 and 3.6.3], or [140, Sections 6.2.1 and 6.2.2]. For example, one may regard \mathcal{G} as the group of points on an elliptic curve over a finite field.

The *traditional Diffie-Hellman problem* is defined as follows:

Given g^x and g^y , compute g^{xy}

solving this problem is believed to be hard.

Due to the identity

$$g^{2xy} = g^{(x+y)^2} g^{-x^2} g^{-y^2}$$

and the fact that computing square roots in groups of prime order is easy the Diffie-Hellman problem can be reformulated in a shorter form:

Given g^x , compute g^{x^2} .

On the other hand, many cryptographic protocols rely on the presumed hardness of the following *generalized Diffie-Hellman problem*:

Given n powers g^x, \dots, g^{x^n} , compute $g^{x^{n+1}}$.

Intuitively it may seem that, despite the fact that more information on x leaks out in the generalized Diffie-Hellman setting, solving it is not easier than solving the traditional Diffie-Hellman problem with the same parameters.

Surprisingly, Brown & Gallant [26] and Cheon [32], have shown this intuition to be wrong.

Here are some results of Cheon [32]:

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

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

This brings up the question of estimating the probability at which primes p are such that $p \pm 1$ has a divisor d of a given size.

More specifically, how rare are primes p such that $p \pm 1$ has a divisor $d \in [n^{1-\varepsilon}, n]$? (which guarantees the asymptotically best advantage if we are given g^x, \dots, g^{x^n} with n which is not too large).

By the result of Ford from Section 8.4 we know that for every $\varepsilon > 0$ this happens for a positive proportion of primes p .

Therefore, we conclude that the attack of [32] can be applied in its asymptotically strongest form with a positive probability. In other words, the generalized Diffie-Hellman problem *is easier* than the traditional Diffie-Hellman problem.

9.12 Large Subgroup Attack

The Digital Signature Algorithm uses two large primes p and q such that $q \mid (p - 1)$, see [27, Section 12.6], or [114, Section 11.5.1], or [140, Section 7.4.2],

Suppose that p and q are generated using the following straightforward method:

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

In [113], Menezes introduces the *Large Subgroup Attack* on some cryptographic protocols, including a version of the HMQV protocol.

The attack can be applied if $n = (p - 1)/(2q)$ has a smooth divisor $s > q$. Some upper bounds on the density of such primes with a large smooth divisor are given by Pomerance & Shparlinski [126].

However this result does not take into account the special structure of p (for example, the presence of a large prime divisor $q \mid (p - 1)$), so it does not (quite) apply.

Furthermore, in the above situation lower bounds become more important. Determining such bounds is unfortunately a much harder question.

On the other hand, using the results of Banks & Shparlinski [11] and Tenenbaum [147] mentioned in Section 6.4 one can get an estimate of the probability $\eta(k, \ell, m)$ that a k -bit integer n has a divisor $s > 2^m$ which is 2^ℓ -smooth. Then, assuming that shifted primes $p - 1$ behave like “random” integers, one can address the original question (at least heuristically).

The most interesting choice of parameters as we write these lines is:

$$k = 863, \quad m = 160, \quad \ell = 80$$

(which produces a 1024-bit prime p).

It has been shown in [11] that for these parameters, the theoretic estimates (together with some heuristic assumptions about the distribution of primes in the sequence $2qn + 1$ for n having a large smooth part) suggest that the attack succeeds with probability

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

over the choices of p and q .

We also note that similar attacks on the ElGamal signature scheme and the Diffie Hellman key exchange protocols, have been outlined by Anderson & Vaudenay [2].

9.13 Smooth Orders

Let $l(n)$ be the multiplicative order of 2 modulo n , $\gcd(2, n) = 1$ (in the following 2 can be replaced by any integer $a \neq 0, \pm 1$).

Motivated by several cryptographic applications, Pomerance & Shparlinski [126] has studied the smoothness of $l(n)$ on integers and on shifted primes $n = p - 1$. This arises from the desire to clarify whether $g = 2$ can safely serve as an exponentiation base in discrete logarithm based cryptosystems⁴. However, in order to avoid the *Pohlig-Hellman attack* $l(n)$ must not be smooth, see [27, Section 10.5], or [114, Section 3.6.4], or [140, Section 6.2.3].

⁴Small values of g allow to significantly speed up square-and-multiply exponentiation.

Also, Boneh & Venkatesan [21] have shown that the Diffie-Hellman protocol with the $g = 2$ has some additional attractive bit security properties which are not known for other g values.

Finally we recall that Pollard's $(p - 1)$ -factorization method works better when $p \mid n$ features a smooth $l(p)$, see [27, Section 9.2], or [44, Section 5.4], or [114, Section 3.2.3], or [140, Section 5.6.1] for details. Some improvements of this algorithm have recently been suggested by Zralek [155].

Let us define the following counting functions:

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

and

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

Pomerance & Shparlinski [126] have shown that for

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

we have

$$L(x, y) \ll u \rho\left(\frac{u}{2}\right) \pi(x).$$

It is also noticed in [126] that it seems quite plausible that in fact the bound also holds with $\rho(u)$ instead of $\rho(u/2)$, which means that the values of $l(p)$ behave as “random” integers.

In fact, this may even happen to be provable under the Generalized Riemann Hypothesis. However this has not been worked out yet and remains an interesting open question.

Furthermore, Banks, Friedlander, Pomerance & Shparlinski [10] proved that for

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

we have

$$N(x, y) \leq x \exp\left(-\left(\frac{1}{2} + o(1)\right) u \log \log u\right).$$

As in the case of $L(x, y)$, one may expect that the same bound should hold with 1 instead of $1/2$ in the exponent, but the appearance of $\log \log u$ instead of $\log u$ seems to be right, see also Section 7.5.

9.14 Smooth-Order Based Public Key Encryption

Smooth orders can also be used constructively to provide public key encryption. Here is one such suggestion due to Naccache & Stern [121]:

Parameter Generation: Let s be an odd, squarefree, y -smooth integer, where y is a certain small parameter and let $N = pq$ be an RSA modulus such that

$$s \mid \varphi(n) \quad \text{and} \quad \gcd\left(s, \frac{\varphi(N)}{s}\right) = 1.$$

Typically, we think of y as being a 10 bit integer and consider N to be at least 768 bits long. Let g be an element whose multiplicative order modulo N is a large multiple of s . Publish N , g and keep p , q and s secret (note that there are very few possibilities for s so its revealing does not give any dramatic advantage to the attacker).

Generation of the modulus appears rather straightforward: pick a family $p_1 < \dots < p_k$ of $2k$ small odd distinct primes and set:

$$u = \prod_{i=1}^k p_{2i-1}, \quad v = \prod_{i=1}^k p_{2i}, \quad s = uv = \prod_{i=1}^k p_i \quad (25)$$

(thus s is p_k -smooth).

Find (using trials and primality testing) two large primes ℓ and r such that both $p = 2\ell u + 1$ and $q = 2rv + 1$ are prime and let $N = pq$.

Note that much faster key generation procedures exist, we refer the reader to [121] for more details.

To generate g , one can choose it at random in \mathbf{Z}_N and check whether it has the possible order $\varphi(N)/4$ or $\varphi(N)/2$ modulo N , Note that for any N multiplicative orders of elements of \mathbf{Z}_N are divisors of the Carmichael function $\lambda(N)$; in the above case $\lambda(N) = \varphi(N)/2$.

The main point is to ensure that g is not a p_i -th power modulo N for each $i = 1, \dots, k$ by testing that

$$g^{\varphi(N)/p_i} \not\equiv 1 \pmod{N}, \quad i = 1, \dots, k.$$

The success probability is:

$$\rho = \prod_{i=1}^k \left(1 - \frac{1}{p_i}\right).$$

If the p_1, \dots, p_k are the first k odd primes, this in turn can be estimated by the Mertens formula as $\rho \sim 1/\log k$. Another method consists in choosing, for each index $i \leq k$, a random g_i until it is not a p_i -th power. With overwhelming probability

$$g = \prod_{i=1}^k g_i^{s/p_i}$$

has the multiplicative order at least $\varphi(n)/4$.

Encryption: A message $m < s$ is encrypted as

$$c \equiv g^m \pmod{N}.$$

Decryption: The algorithm computes the value m_i of the residue of m modulo each prime factor p_i , $i = 1, \dots, k$ of s given by (25), and recovers the message by the Chinese Remainder Theorem, following an idea of Pohlig-Hellman [124], see also [44, 114].

Now for every $i = 1, \dots, k$, to find m_i , given the ciphertext $c \equiv g^m \pmod{N}$, the algorithm computes

$$c_i \equiv c^{\varphi(n)/p_i} \equiv g^{m\varphi(n)/p_i} \equiv g^{m_i\varphi(n)/p_i} \pmod{N},$$

where the congruence $m \equiv m_i \pmod{p_i}$ is used at the last step.

By comparing this result with all possible powers

$$g^{j\varphi(n)/p_i}, \quad j = 0, \dots, p_i - 1,$$

the algorithm finds out the correct value of m_i .

The basic operation used by this (non-optimized) algorithm is a modular exponentiation of complexity $\mathcal{O}((\log N)^3)$, repeated at most :

$$kp_k \ll k^2 \log k \ll (\log N)^2 \log \log N$$

times. Decryption therefore takes $\mathcal{O}((\log N)^5 \log \log N)$ bit operations.

We refer the reader to [121] for more details and optimizations.

9.15 Pratt Trees

Highly critical security applications sometimes require *primality proofs*. Here is a way to provide such proofs, due to Pratt [127]

- Check that the would-be prime p is not a perfect power. This is easy, see, for example, [16, 17].
- Produce a primitive root g modulo p and provide a proof of this. For that sake 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 must also be supplied.

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

The whole algorithm can be viewed as a tree, called the *Pratt Tree*, where each node contains a prime (with p as a root) and with 2 at each leaf.

The algorithm runs in polynomial time and in particular shows that the decision problem PRIMES is in the complexity class NP (which is not so exciting nowadays given that, thanks to [1], we know that PRIMES is actually in P).

Pratt [127] has shown that the number of multiplications required by this algorithm is $\mathcal{O}((\log p)^2)$. On the other hand, Bayless [13] shows that this number is at least $C \log p$ for any fixed $C > 1$ and for almost all primes p .

There are, however, many other interesting questions about this tree, such as estimating its height, number of nodes, number of leaves, and so on, in extreme cases and also for almost all primes.

For example, it is obvious that the Pratt Tree's height $H(p)$ satisfies the inequality

$$H(p) \ll \log p. \tag{26}$$

One can also infer from more general results of Kátai [96] that for some constant $c > 0$ the inequality

$$H(p) \geq c \log \log p \tag{27}$$

holds for almost all primes p .

Ford, Konyagin & Luca [61] have recently given a heuristic argument suggesting that

$$H(p) \gg \frac{\log p}{\log \log p}$$

for infinitely many primes p and also a rigorous proof that

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

holds for almost all p . It is also shown in [61] that (27) holds for almost all primes p with any

$$c < \frac{1}{1 + \log 2}.$$

It seems that the lower bound (27) is of the right order of magnitude and in fact some heuristic arguments, given in [61], lead to the conjecture that

$$H(p) = e \log \log p + \mathcal{O}(\log \log \log p)$$

for almost all primes p .

A number of other challenging open questions and conjectures can be found in [61].

Studying other characteristics of the Pratt Tree is also an interesting and little-researched open question.

For instance, Banks & Shparlinski [12] have shown that the length $L(p)$ of the chain $p \mapsto P(p-1)$ satisfies

$$L(p) \geq (1 + o(1)) \frac{\log \log p}{\log \log \log p} \tag{28}$$

for almost all primes p . This corresponds to a particular path in the Pratt Tree. Furthermore, it may be natural to expect that this should actually be the longest path for almost all primes, so it is possible that

$$L(p) = (1 + o(1))H(p)$$

for almost all primes p . On the other hand, it seems that $L(p) < H(p)$ holds for almost all primes p . Clarifying the matter is an important research challenge. As a first step one may for instance try to use the methods of [61, 96] to improve (28) up to the level of (27).

9.16 Strong Primes

A prime p is called *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, see [114, Section 4.4.2].

To make this definition more formal we say that p is *y-strong* if $p + 1$ has a large prime divisor $q \geq y$, and $p - 1$ has a prime divisor r such that $r - 1$ has a prime divisor $\ell \geq y$.

We note that the combination of [10] and [126] (see Sections 7.3 and 7.5) implies that almost all primes are *y-strong* as $\log x / \log y \rightarrow \infty$.

Indeed, the cardinality of the set of primes $p \leq x$ such that $p + 1$ is *y-smooth* is exactly the function $\pi_1(x, y)$ discussed in Section 7.3.

From the set of remaining primes $p \leq x$ we remove those for which $p - 1$ is divisible by r^2 for a prime $r \geq y$. Since the number of primes $p \leq x$ with $p \equiv 1 \pmod{r^2}$ is at most x/r^2 , the cardinality of this set can be estimated trivially as

$$\sum_{r \geq y} \frac{x}{r^2} = \mathcal{O}(x/y).$$

Hence it is easy to see that if one of the remaining primes is not *y-strong* then $\varphi(p - 1)$ is *y-smooth* and thus the bounds of $\Pi_{-1}(x, y)$ from Section 7.5 can be applied.

9.17 Small Prime Based Hash Functions

The *Very Smooth Hash* function, VSH, recently introduced and studied by Contini, Lenstra & Steinfeld [39], is defined as follows.

Let p_i denote the i -th prime number and let

$$Q_k = \prod_{i=1}^k p_i$$

denote the product of the first k primes.

Assume that integers k and N satisfy

$$Q_k < N \leq Q_{k+1}. \tag{29}$$

Let the message length $\ell < 2^k$ be a positive integer whose k -bit representation (including all leading zeros) is $\ell = \lambda_1 \dots \lambda_k$ that is

$$\ell = \sum_{i=1}^k \lambda_i 2^{i-1}.$$

The VSH takes an ℓ -bit message $m = \mu_1, \dots, \mu_\ell$ and hashes it (in a very efficient way, via a simple iterative procedure) to

$$h_N(m) \equiv \prod_{i=1}^k p_i^{e_i} \pmod{N}, \quad 0 \leq h_N(m) < N,$$

where $L = \lceil \ell/k \rceil$, $\mu_s = 0$, for $\ell < s \leq Lk$, $\mu_{Lk+i} = \lambda_i$, for $1 \leq i \leq k$, and

$$e_i = \sum_{j=0}^L \mu_{jL+i} 2^{L-j}, \quad i = 1, \dots, k.$$

It is demonstrated in [39] that the VSH also admits a rigorous collision-resistance proof based on “natural” number theoretic problems which are presumably hard. As the above problem is related to factoring, it is natural to choose $N = pq$ to be an RSA modulus. The design and the suggested parameter choice are both based on classical facts about the distribution of smooth numbers.

In [18], Blake & Shparlinski harness results about the distribution of smooth numbers to provide rigorous support in favor of the security and the distribution properties of the VSH. In particular, [18] shows that for almost all RSA moduli and any integer a , the probability that for a random ℓ -bit message m we have $h_N(m) \equiv a \pmod{N}$, is negligible for sufficiently large values of ℓ .

This bounds the collision probability and also the probability of finding a second pre-image by brute force.

The above and several other results in [18] are based on the study of the multiplicative subgroup of \mathbf{Z}_N^* generated by p_1, \dots, p_k for integers $N = pq$ where p and q are distinct primes, satisfying the inequality (29).

10 Conclusion

Our goal has been to position this paper at the crossroads of cryptography and number theory. We hope that while reading it cryptographers have enriched their arsenal with a large gamut of little-used, yet powerful, number-theoretic methods and results extending beyond the classical facts used in cryptology. On the other hand, it is our hope that number theorists have enjoyed learning how smooth numbers can be harnessed to provide encryption, private information retrieval, identification, error correction, hashing, primality proofs and other cryptographic functions. Furthermore, final tuning and adjusting already known results and techniques may lead to new advances of intrinsic mathematical interest. Our outline, somewhat sketchy and simplified and also sometimes ignoring subtleties, cannot replace a careful and systematic reading of the original number theoretic and cryptographic literature, such as [44, 80, 85, 92, 144] and [29, 114, 140], respectively.

11 Acknowledgements

The authors would like to thank Kevin Ford for many valuable suggestions.

This second author work was supported in part by ARC grant DP0556431.

References

- [1] M. Agrawal, N. Kayal and N. Saxena, ‘PRIMES is in P’, *Ann. of Math.*, **160** (2004), 781–793.
- [2] R. Anderson and S. Vaudenay, ‘Minding your p ’s and q ’s’, *Proc. Asiacrypt’96*, Lect. Notes in Comp. Sci., vol. 1163, Springer-Verlag, Berlin, 2000, 26–35.
- [3] E. Bach, J. von zur Gathen and H. W. Lenstra, ‘Factoring polynomials over special finite fields’, *Finite Fields Appl.*, **7** (2001), 5–28.
- [4] E. Bach and R. Peralta, ‘Asymptotic semismoothness probabilities’, *Math. Comp.*, **65** (1996), 1701–1715.

- [5] A. Balog, ‘On the distribution of integers having no large prime factors’, *Astérisque*, **147–148** (1987), 27–31.
- [6] A. Balog, ‘On additive representation of integers’, *Acta Math. Hungar.*, **54** (1989), 297–301.
- [7] A. Balog and C. Pomerance, ‘The distribution of smooth numbers in arithmetic progressions’, *Proc. Amer. Math. Soc.*, **115** (1992), 33–43.
- [8] 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.
- [9] R. C. Baker and G. Harman, ‘Shifted primes without large prime factors’, *Acta Arith.*, **83** (1998), 331–361.
- [10] W. Banks, J. B. Friedlander, C. Pomerance and I. E. Shparlinski, ‘Multiplicative structure of values of the Euler function’, *High Primes and Misdemeanours: Lectures in Honour of the 60-th Birthday of Hugh Cowie Williams*, Fields Institute Communications, vol. 41, Amer. Math. Soc., 2004, 29–48.
- [11] W. D. Banks and I. E. Shparlinski, ‘Integers with a large smooth divisor’, *Integers*, **7** (2007), # A17, 1–11.
- [12] W. D. Banks and I. E. Shparlinski, ‘On values taken by the largest prime factor of shifted primes’, *J. Aust. Math. Soc.*, **82** (2007), 133–147.
- [13] J. Bayless, ‘The Lucas-Pratt primality tree’, *Math. Comp.*, **77** (2008), 495–502.
- [14] 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.
- [15] D. J. Bernstein, ‘Bounding smooth integers’, *Proc. 3-rd Algorithmic Number Theory Symp.*, Lect. Notes in Comput. Sci., vol. 1423, Springer-Verlag, Berlin, 1998, 128–130.
- [16] D. J. Bernstein, ‘Detecting perfect powers in essentially linear time’, *Math. Comp.*, **67** (1998), 1253–1283.

- [17] D. J. Bernstein, H. W. Lenstra and J. Pila, ‘Detecting perfect powers by factoring into coprimes’, *Math. Comp.*, **76** (2007), 385–388.
- [18] I. Blake and I. E. Shparlinski, ‘Statistical distribution and collisions of the VSH’, *J. Math. Cryptology*, **1** (2007), 329–349.
- [19] D. Boneh, ‘Finding smooth integers in short intervals using CRT decoding’, *J. Comp. and Syst. Sciences.*, **64** (2002), 768–784.
- [20] D. Boneh, A. Joux and P. Q. Nguyen, ‘Why textbook ElGamal and RSA encryption are insecure’, *Proc. Asiacrypt’00*, Lect. Notes in Comp. Sci., vol. 1976, Springer-Verlag, Berlin, 2000, 30–43.
- [21] D. Boneh and R. Venkatesan, ‘Hardness of computing the most significant bits of secret keys in Diffie–Hellman and related schemes’, *Proc. Crypto’96*, Lect. Notes in Comp. Sci., vol. 1109, Springer-Verlag, Berlin, 1996, 129–142.
- [22] R. de la Bretèche, ‘Sommes sans grand facteur premier’, *Acta Arith.*, **88** (1999), 1–14.
- [23] R. de la Bretèche, ‘Entries ayant exactement r diviseurs dans un intervalle donné’, *Anatomy of Integers*, CRM Proc. and Lect. Notes, vol. 46, Amer. Math. Soc., Providence, RI, 2008, 19–45.
- [24] R. de la Bretèche and G. Tenenbaum, ‘Sommes d’exponentielles friables d’arguments rationnels’, *Funct. Approx. Comment. Math.*, **37** (2007), 31–38.
- [25] 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.
- [26] D. R. L. Brown and R. P. Gallant, ‘The static Diffie-Hellman problem’, *Cryptology ePrint Archive*, Report 2004/306, 2004, (available from <http://eprint.iacr.org/2004/306>).
- [27] J. Buchmann, *Introduction to cryptography*, Springer-Verlag, Berlin, 2004.

- [28] J. Buchmann and S. Hamdy, ‘A survey on IQ cryptography’, *Public-Key Cryptography and Computational Number Theory*, Walter de Gruyter, Berlin, 2001, 1–15.
- [29] 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).
- [30] D. A. Burgess, ‘The distribution of quadratic residues and non-residues’, *Mathematika*, **4** (1957), 106–112.
- [31] E. R. Canfield, P. Erdős and C. Pomerance, ‘On a problem of Oppenheim concerning “*Factorisatio Numerorum*”’, *J. Number Theory*, **17** (1983), 1–28.
- [32] J. Cheon, ‘Security analysis of the strong Diffie-Hellman problem’, *Proc. Eurocrypt’06*, Lect. Notes in Comp. Sci., vol. 4004, Springer-Verlag, Berlin, 2006, 1–11.
- [33] J. Cooley and J. Tukey, ‘An algorithm for the machine calculation of complex Fourier series’, *Math. Comp.*, **19** (1965), 297–301.
- [34] B. Chevallier-Mames, D. Naccache and J. Stern, ‘Linear bandwidth Naccache-Stern encryption’, *Proc. 6-th Conf. on Security and Cryptography for Networks*, Lect. Notes in Comp. Sci., vol. 5229, Springer-Verlag, Berlin, 2008, 327–339.
- [35] H. Cohen and H. W. Lenstra, ‘Heuristics on class groups of number fields’, *Number theory, Noordwijkerhout 1983*, Lect. Notes in Math., vol. 1068, Springer, Berlin, 1984, 33–62.
- [36] A. Cojocaru, ‘Questions about the reductions modulo primes of an elliptic curve’, *Proc. 7th Meeting of the Canadian Number Theory Association (Montreal, 2002)*, CRM Proceedings and Lecture Notes, Vol. 36, Amer. Math. Soc., 2004, 61–79.
- [37] A. Cojocaru, ‘Reductions of an elliptic curve with almost prime orders’, *Acta Arith.*, **119** (2005), 265–289.
- [38] A. Cojocaru, F. Luca and I. E. Shparlinski, ‘Pseudoprime reductions of elliptic curves’, *Math. Proc. Cambr. Phil. Soc.*, (to appear).

- [39] S. Contini, A. K. Lenstra and R. Steinfeld, ‘VSH, an efficient and provable collision-resistant hash function’, *Proc. Eurocrypt’06*, Lect. Notes in Comp. Sci., vol. 4004, Springer-Verlag, Berlin, 2006, 165–182.
- [40] D. Coppersmith, J. S. Coron, F. Grieru, S. Halevi, C. Jutla, D. Naccache, J. P. Stern, ‘Cryptanalysis of ISO/IEC 9796-1’, *J. Cryptology*, **12** (2008), 27–51.
- [41] J.-S. Coron, M. Joye, D. Naccache and P. Paillier, ‘New attacks on PKCS#1 v1.5 encryption’, *Proc. Eurocrypt’00*, Lect. Notes in Comp. Sci., vol. 1807, Springer-Verlag, Berlin, 2000, 369–381.
- [42] J.-S. Coron and D. Naccache, ‘Security analysis of the Gennaro-Halevi-Rabin signature scheme’, *Proc. Eurocrypt’00*, Lect. Notes in Comp. Sci., vol. 1807, Springer-Verlag, Berlin, 2000, 91–101.
- [43] J.-S. Coron and D. Naccache, ‘A new error-correcting code based on modular arithmetic’, *Preprint*, 2004.
- [44] R. Crandall and C. Pomerance, *Prime numbers: A computational perspective*, 2-nd edition, Springer-Verlag, New York, 2005.
- [45] E. Croot, ‘On a combinatorial method for counting smooth numbers in sets of integers’, *J. Number Theory*, **126** (2007), 237–253.
- [46] E. Croot, ‘Smooth numbers in short intervals’, *Int. J. Number Theory*, **3** (2007), 159–169.
- [47] E. Croot, A. Granville, R. Pemanle and P. Tetali, ‘Running time predictions for factoring algorithms’, *Proc. 8-th Algorithmic Number Theory Symp.*, Lect. Notes in Comput. Sci., vol. 5011, Springer-Verlag, Berlin, 2008, 1–36.
- [48] C. Dartyge, G. Martin and G. Tenenbaum, ‘Polynomial values free of large prime factors’, *Periodica Math. Hungar.*, **43** (2001), 111–119.
- [49] Y. Desmedt and A. Odlyzko, ‘A chosen text attack on the RSA cryptosystem and some discrete logarithm schemes’, *Proc. Eurocrypt’85*, Lect. Notes in Comput. Sci., vol. 218, Springer-Verlag, Berlin, 1985, 516–522.

- [50] W. Duke, J. B. Friedlander and H. Iwaniec, ‘Bilinear forms with Kloosterman fractions’, *Invent. Math.*, **128** (1997), 23–43.
- [51] T. El-Gamal, ‘A public-key cryptosystem and a signature scheme based on discrete logarithms’, *Proc. Crypto’84*, Lect. Notes in Comput. Sci., vol. 196, Springer-Verlag, Berlin, 1985, 10–18.
- [52] V. Ennola, ‘On numbers with small prime divisors’, *Ann. Acad. Sci. Fenn., Ser. AI*, **440** (1969), 1–16.
- [53] P. Erdős and C. Pomerance, ‘On the normal number of prime factors of $\varphi(n)$ ’, *Rocky Mountain J. Math.*, **15** (1985), 343–352.
- [54] A. Fiat and A. Shamir, ‘How to prove yourself: Practical solutions to identification and signature problems’, *Proc. Crypto’86*, Lect. Notes in Comput. Sci., vol. 263, Springer-Verlag, Berlin, 1987, 186–194.
- [55] K. Ford, ‘The distribution of totients’, *The Ramanujan J.*, **2** (1998), 67–151.
- [56] K. Ford, ‘The number of solutions of $\varphi(x) = m$ ’, *Annals of Math.*, **150** (1999), 283–311.
- [57] K. Ford, ‘Vinogradov’s integral and bounds for the Riemann zeta function’, *Proc. London Math. Soc.*, **85** (2002), 565–633.
- [58] K. Ford, ‘The distribution of integers with a divisor in a given interval’, *Annals Math.*, **168** (2008), 367–433.
- [59] K. Ford, ‘Integers with a divisor in $(y, 2y]$ ’, *Anatomy of Integers*, CRM Proc. and Lect. Notes, vol. 46, Amer. Math. Soc., Providence, RI, 2008, 65–80.
- [60] K. Ford and Y. Hu, ‘Divisors of the Euler and Carmichael functions’, *Acta Arith.*, **133** (2008), 199–208.
- [61] K. Ford, S. Konyagin and F. Luca, ‘Prime chains and Pratt trees’, *Preprint*, 2008.
- [62] K. Ford and G. Tenenbaum, ‘The distribution of integers with at least two divisors in a short interval’, *Quart. J. Math.*, **58** (2007), 187–201.

- [63] É. Fouvry and G. Tenenbaum, ‘Entiers sans grand facteur premier en progressions arithmétiques’, *Proc. London Math. Soc.*, **63** (1991), 449–494.
- [64] 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.
- [65] J. B. Friedlander, ‘Shifted primes without large prime factors’, *Number Theory and Applications*, Kluwer Acad. Publ., Dordrecht, 1989, 393–401.
- [66] J. B. Friedlander and A. Granville, ‘Smoothing ‘smooth’ numbers’, *Philos. Trans. Roy. Soc. London, Ser. A*, **345** (1993), 339–347.
- [67] 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.
- [68] S. D. Galbraith and J. McKee, ‘The probability that the number of points on an elliptic curve over a finite field is prime’, *J. London Math. Soc.*, **62** (2000), 671–684.
- [69] J. von zur Gathen and J. Gerhard, *Modern computer algebra*, Cambridge University Press, Cambridge, 2003.
- [70] C. Gentry and Z. Ramzan, ‘Single-database private information retrieval with constant communication rate’, *Proc. 32nd Intern. Coll. Automata, Languages and Programming*, Lect. Notes in Comput. Sci., vol. 3580, Springer-Verlag, Berlin, 2005, 803–815.
- [71] 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.
- [72] M. Girault and J.-F. Misarsky, ‘Cryptoanalysis of countermeasures proposed for repairing ISO 9796’, *Proc. Eurocrypt’00*, Lect. Notes in Comp. Sci., vol. 1807, Springer-Verlag, Berlin, 2000, 81–90.
- [73] K. Gödel, ‘Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme’, *I. Monatshefte fr Mathematik und Physik*, **38** (1931), 173–198.

- [74] S. W. Graham and I. E. Shparlinski, ‘On RSA moduli with almost half of the bits prescribed’, *Disc. Appl. Math.*, **156** (2008), 3150–3154.
- [75] A. Granville, ‘On positive integers $\leq x$ with prime factors $\leq t \log x$ ’, *Number Theory and Applications* Kluwer, 1989, 403–422.
- [76] A. Granville, ‘The lattice points of an n -dimensional tetrahedron’, *Aequationes Math.*, **41** (1991), 234–241.
- [77] A. Granville, ‘Integers, without large prime factors, in arithmetic progressions I’, *Acta Math.*, **170** (1993), 255–273.
- [78] A. Granville, ‘Integers, without large prime factors, in arithmetic progressions II’, *Philos. Trans. Roy. Soc. London, Ser. A*, **345** (1993), 349–362.
- [79] A. Granville, ‘Smooth numbers: Computational number theory and beyond’, *Proc. MSRI Conf. Algorithmic Number Theory: Lattices, Number Fields, Curves, and Cryptography, Berkeley 2000*, Cambridge University Press, (to appear).
- [80] H. Halberstam and H.–E. Richert, *Sieve methods*, Academic Press, London, 1974.
- [81] R. Hall and G. Tenenbaum, *Divisors*, Cambridge Tracts in Mathematics, vol. 90, Cambridge University Press, 1988.
- [82] S. Hamdy and F. Saidak, ‘Arithmetic properties of class numbers of imaginary quadratic fields’, *J. Algebra Number Theory Appl.*, **6** (2006), 129–148.
- [83] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Oxford Univ. Press, Oxford, 1979.
- [84] G. Harman, ‘Integers without large prime factors in short intervals and arithmetic progressions’, *Acta Arith.*, **91** (1999), 279–289.
- [85] G. Harman, *Prime-detecting sieves*, Princeton Univ. Press, Princeton, NJ, 2007.
- [86] A. Hildebrand, ‘Integers free of large prime factors and the Riemann Hypothesis’, *Mathematika*, **31** (1984), 258–271.

- [87] A. Hildebrand, ‘On the number of positive integers $\leq x$ and free of prime factors $\leq y$ ’, *J. Number Theory*, **22** (1986), 289–307.
- [88] A. Hildebrand and G. Tenenbaum, ‘Integers without large prime factors’, *J. de Théorie des Nombres de Bordeaux*, **5** (1993), 411–484.
- [89] N. A. Hmyrova, ‘On polynomials with small prime divisors, II’, *Izv. Akad. Nauk SSSR Ser. Mat.*, **30** (1966), 1367–1372 (in Russian).
- [90] S. Hunter and J. P. Sorenson, ‘Approximating the number of integers free of large prime factors’, *Mathem. Comp.*, **66** (1997), 1729–1741.
- [91] H. Iwaniec and J. Jiménez Urroz, ‘Orders of CM elliptic curves modulo p with at most two primes’, *Preprint*, 2006.
- [92] H. Iwaniec and E. Kowalski, *Analytic number theory*, Amer. Math. Soc., Providence, RI, 2004.
- [93] A. Joux, D. Naccache and E. Thomé, ‘When e -th roots become easier than factoring’, *Proc. Asiacrypt’07*, Lect. Notes in Comp. Sci., vol. 4833, Springer-Verlag, Berlin, 2007, 13–28.
- [94] M. Joye, P. Paillier and S. Vaudenay, ‘Efficient generation of prime numbers’, *Proc. Cryptographic Hardware and Embedded Systems (CHES’00)*, Lect. Notes in Comp. Sci., vol. 1965, Springer-Verlag, Berlin, 2000, 340–354.
- [95] J. Jiménez Urroz, ‘Almost prime orders of CM elliptic curves modulo p ’, *Proc. 4-th Algorithmic Number Theory Symp.*, Lect. Notes in Comp. Sci., vol. 5011, Springer-Verlag, Berlin, 2008, 74–87.
- [96] I. Kátai, ‘On the iteration of multiplicative functions’, *Publ. Math. Debrecen*, **36** (1989), 129–134.
- [97] N. Koblitz, ‘Primality of the number of points on an elliptic curve over a finite field’, *Pacific J. Math.*, **131** (1988), 157–166.
- [98] N. Koblitz, ‘Almost primality of group orders of elliptic curves defined over small finite fields’, *Experiment. Math.*, **10** (2001), 553–558.
- [99] N. M. Korobov, ‘Estimates of trigonometric sums and their applications’, *Uspehi Mat. Nauk*, **13** (1958), 185–192, (Russian).

- [100] D. Koukoulopoulos, ‘Localized factorizations of integers’, *Preprint*, 2008, (available from <http://arxiv.org/abs/0809.1072>).
- [101] Y. Lamzouri, ‘Smooth values of the iterates of the Euler φ -function’, *Canadian J. Math.*, **59** (2007) 127–147.
- [102] A. K. Lenstra and I. E. Shparlinski, ‘Selective forgery of RSA signatures with fixed-pattern padding’, *Proc. Intern. Workshop on Practice and Theory in Public Key Cryptography*, Lect. Notes in Comp. Sci., vol. 2274, Springer-Verlag, Berlin, 2002, 228–236.
- [103] H. W. Lenstra, ‘Factoring integers with elliptic curves’, *Annals of Math.*, **126** (1987), 649–673.
- [104] 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.
- [105] H. W. Lenstra, J. Pila and C. Pomerance, ‘A hyperelliptic smoothness test, II’, *Proc. London Math. Soc.*, **84** (2002), 105–146.
- [106] Y.-R. Liu, ‘Prime divisors of the number of rational points on elliptic curves with complex multiplication’, *Bull. London Math. Soc.*, **37** (2005), 658–664.
- [107] Y.-R. Liu, ‘A prime analogue to Erdős-Pomerance’s conjecture for elliptic curves’, *Comment. Math. Helv.*, **80** (2005), 755–769.
- [108] Y.-R. Liu, ‘Prime analogues of the ErdősKac theorem for elliptic curves’, *J. Number Theory*, **119** (2006), 155–170.
- [109] S. S. Loiperdinger and I. E. Shparlinski, ‘On the distribution of the Euler function of shifted smooth numbers’, *Preprint*, 2008, (available from <http://arxiv.org/abs/0810.1093>).
- [110] F. Luca and C. Pomerance, ‘On the average number of divisors of the Euler function’, *Publ. Math. Debrecen*, **70** (2007), 125–148.
- [111] G. Martin, ‘An asymptotic formula for the number of smooth values of a polynomial’, *J. Number Theory*, **93** (2002), 108–182.

- [112] J. McKee, ‘Subtleties in the distribution of the numbers of points on elliptic curves over a finite prime field’, *J. London Math. Soc.*, **59** (1999), 448–460.
- [113] A. J. Menezes, ‘Another look at HMQV’, *J. Math. Cryptology*, **1** (2007), 47–64.
- [114] A. J. Menezes, P. C. van Oorschot and S. A. Vanstone, *Handbook of applied cryptography*, CRC Press, Boca Raton, FL, 1996.
- [115] S. Micali and A. Shamir, ‘An improvement of the Fiat Shamir identification and signature scheme’, *Proc. Crypto’87*, Lect. Notes in Comp. Sci., vol. 403, Springer-Verlag, Berlin, 1990, 244–247.
- [116] S. A. Miri and V. K. Murty, ‘An application of sieve methods to elliptic curves’, *Proc. Indocrypt’01*, Lect. Notes in Comp. Sci., vol. 2247, Springer-Verlag, Berlin, 2001, 91–98.
- [117] J.-F. Misarsky, ‘A multiplicative attack using LLL algorithm on RSA signatures with redundancy’, *Proc. Crypto’97*, Lect. Notes in Comp. Sci., vol. 1294, Springer-Verlag, Berlin, 1997, 221–234.
- [118] A. Mohan, *Residue number systems: Algorithms and architectures*, Springer-Verlag, Berlin, 2002.
- [119] D. Naccache, ‘Sécurité, cryptographie : théorie et pratique’, *Mémoire d’habilitation à diriger des recherches*, Université Paris VII - Denis Diderot, 2004.
- [120] D. Naccache and J. Stern, ‘A new public-key cryptosystem’, *Proc. Eurocrypt’97*, Lect. Notes in Comput. Sci., vol. 1233, Springer-Verlag, Berlin, 1997, 27–36.
- [121] D. Naccache and J. Stern, ‘A new public-key cryptosystem based on higher residues’, *Proc. 5-th ACM Conference on Computer and Communications Security*, ACM Press, 1998, 59–66.
- [122] A. M. Odlyzko, ‘Discrete logarithms in finite fields and their cryptographic significance’, *Proc. Eurocrypt’84*, Lect. Notes in Comput. Sci., vol. 209, Springer-Verlag, Berlin, 1985, 224–314.

- [123] S. T. Parsell and J. P. Sorenson, ‘Fast bounds on the distribution of smooth numbers’, *Proc. 7-th Algorithmic Number Theory Symp.*, Lect. Notes in Comput. Sci., vol. 4076, Springer-Verlag, Berlin, 2006, 168–181.
- [124] S. C. Pohlig and M. E. Hellman, ‘An improved algorithm for computing logarithms over $GF(p)$ and its cryptographic significance’, *IEEE Trans. on Inform. Theory*, **24** (1978), 106–110.
- [125] C. Pomerance, ‘Fast, rigorous factorization and discrete logarithm algorithms’, *Discrete Algorithms and Complexity*, Academic Press, 1987, 119–143.
- [126] C. Pomerance and I. E. Shparlinski, ‘Smooth orders and cryptographic applications’, *Proc. 5-th Algorithmic Number Theory Symp.*, Lect. Notes in Comput. Sci., vol. 2369, Springer-Verlag, Berlin, 2002, 338–348.
- [127] V. Pratt, ‘Every prime has a succinct certificate’, *SIAM J. Comput.*, **4** (1975), 214–220.
- [128] R. A. Rankin, ‘The difference between consecutive prime numbers’, *J. London Math. Soc.*, **13** (1938), 242–247.
- [129] R. Rivest, A. Shamir, and L. Adleman, ‘A method for obtaining digital signatures and public key cryptosystems’, *Commun. ACM*, **21** (1978), 120–126.
- [130] L. Rónyai, ‘Factoring polynomials modulo special primes’, *Combinatorica*, **9** (1989), 199–206.
- [131] É. Saias, ‘Sur le nombre des entiers sans grand facteur premier,’ *J. Number Theory*, **32** (1989), 78–99.
- [132] É. Saias, ‘Entiers à Diviseurs Denses 1’, *J. Number Theory*, **62** (1997), 163–191.
- [133] V. Shoup, ‘Smoothness and factoring polynomials over finite fields,’ *Inform. Process. Lett.*, **38** (1991), 39–42.
- [134] I. E. Shparlinski, ‘On RSA moduli with prescribed bit patterns’, *Designs, Codes and Cryptography*, **39** (2006), 113–122.

- [135] I. E. Shparlinski, ‘Character sums over shifted smooth numbers’, *Proc. Amer. Math. Soc.*, **135** (2007), 2699–2705.
- [136] J. P. Sorenson, ‘A fast algorithm for approximately counting smooth numbers’, *Proc. 4-th Algorithmic Number Theory Symp.*, Lect. Notes in Comput. Sci., vol. 1838, Springer-Verlag, Berlin, 2000, 539–549.
- [137] K. Soundararajan, ‘The distribution of smooth numbers in arithmetic progressions’, *Anatomy of Integers*, CRM Proc. and Lect. Notes, vol. 46, Amer. Math. Soc., Providence, RI, 2008, 115–128.
- [138] J. Stern, P.-A. Fouque and G.-J. Wackers, ‘CryptoComputing with rationals’, *Proc. Financial Cryptography’2002*, Lect. Notes in Comput. Sci., vol. 2357, Springer-Verlag, Berlin, 2002, 136–146.
- [139] J. Steuding and A. Weng, ‘On the number of prime divisors of the order of elliptic curves modulo p ’, *Acta Arith.*, **117** (2005), 341–352; ‘Erratum’, *Acta Arith.*, **119** (2005), 407–408.
- [140] D. R. Stinson, *Cryptography: Theory and practice*, CRC Press, Boca Raton, FL, 2006.
- [141] K. Suzuki, ‘An estimate for the number of integers without large prime factors’, *Mathem. Comp.*, **73** (2004), 1013–1022.
- [142] K. Suzuki, ‘Approximating the number of integers without large prime factors’, *Mathem. Comp.*, **75** (2006), 1015–1024.
- [143] G. Tenenbaum, ‘Crible les entiers sans grand facteur premier’, *Philos. Trans. Roy. Soc. London, Ser. A*, **345** (1993), 377–384.
- [144] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge University Press, 1995.
- [145] G. Tenenbaum, ‘Crible d’Ératosthène et modèle de Kubilius’, *Number theory in progress (Zakopane-Kościelisko, 1997)*, vol. 2, Walter de Gruyter, Berlin, 1999, 1099–1129.
- [146] G. Tenenbaum, ‘A rate estimate in Billingsley’s theorem for the size distribution of large prime factors’, *Quart. J. Math.*, **51** (2000), 385–403.

- [147] G. Tenenbaum, ‘Integers with a large friable component’, *Acta Arith.*, **124** (2006), 287–291.
- [148] N. M. Timofeev, ‘Polynomials with small prime divisors’, *Taškent. Gos. Univ., Naučn. Trudy No. 548, Voprosy Mat.*, Taškent, 1977, 87–91 (Russian).
- [149] T. Z. Xuan, ‘On smooth integers in short intervals under the Riemann hypothesis.’, *Acta Arith.*, **88** (1999), 327–332.
- [150] B. Vallée, ‘Gauss’ algorithm revisited’. *J. Algorithms*, **12** (1991), 556–572.
- [151] A. I. Vinogradov, ‘On the remainder in Merten’s formula,’ *Dokl. Akad. Nauk SSSR*, **148** (1963), 262–263, (Russian).
- [152] I. M. Vinogradov, ‘On a bound for the smallest non-residue of n -th power’, *Izv. Akad. Nauk SSSR, Ser. Mat.*, **20** (1926), 47–58, (Russian).
- [153] I. M. Vinogradov, ‘A new estimate for $\zeta(1+it)$ ’, *Izv. Akad. Nauk SSSR, Ser. Mat.*, **22** (1958), 161–164, (Russian).
- [154] A. Weng, ‘On group orders of rational points of elliptic curves’, *Quaest. Math.*, **25** (2002), 513–525.
- [155] B. Zralek, ‘A deterministic version of Pollard’s $p - 1$ algorithm’, *Preprint*, 2007, (available from <http://arxiv.org/abs/0707.4102>).
- [156] B. Zralek, ‘Using the smoothness of $p-1$ for computing roots modulo p ’, *Preprint*, 2008, (available from <http://arxiv.org/abs/0803.0471>).