A RIGOROUS PROOF OF THE WATERLOO ALGORITHM FOR THE DISCRETE LOGARITHM PROBLEM

MICHAEL DRMOTA, DANIEL PANARIO

Abstract. In this paper we are concerned with the Waterloo variant of the index calculus method for the discrete logarithm problem in $\mathbb{F}_{2^n}$. We provide a rigorous proof for the heuristic arguments for the running time of the Waterloo algorithm. This implies in studying the behavior of pairs of coprime smooth polynomials over finite fields. Our proof involves a double saddle point method, and it is in nature similar to the one of Odlyzko for the rigorous analysis of the basic index calculus.

1. Introduction

Given a group G and elements g and y, the *discrete logarithm problem* for G is to find the smallest integer x such that $y = g^x$. This is a fundamental problem not only from a mathematical standpoint but also from a practical point of view. Indeed, many public-key cryptosystems depend their security on the difficulty of computing discrete logarithms fast [5, 8, 9], at least for certain groups.

In this paper we are concerned with the discrete logarithm problem in the multiplicative group of the finite field $\mathbb{F}_{2^n}$. However our proofs are general and apply to other finite fields.

The most important algorithm for the discrete logarithm problem in finite fields is the *index calculus method*. This method for computing discrete logarithms was created by Western and J.C.P. Miller [18], and was rigorously analized by Adleman [1] and Odlyzko [15]. This paper deals with a variant of the basic index calculus method proposed by Blake et al. [3, 4], known as the Waterloo algorithm. This variant improves the running time of the method by introducing a heuristic argument that makes its analysis not rigorous. A classical problem in computational number theory and theoretical cryptography is to provide rigorous proofs for heuristic arguments. We give such a proof for the Waterloo algorithm.

The index calculus algorithm consists of two stages. In the first stage, the logarithms of all irreducible polynomials of degree at most m are computed, where m is a parameter of the process. In the second stage, individual logarithms are computed. For a detailed explanation of the index calculus method, the Waterloo variant, and several improvements, see Odlyzko's survey [15].

Date: May 11, 2001.

The basic idea behind the improvement of the Waterloo variant over the basic index calculus method is that it is more likely that two polynomials of degree at most $(n-1)/2$ have all their irreducible factors of degree smaller than or equal to m that one polynomial of degree at most $n-1$ have all its irreducible factors of degree smaller than or equal to m. Polynomials that factor into irreducibles of degree smaller or equal to m are called *m-smooth* polynomials. The problem for a rigorous analysis of this algorithm is that not much is known about the distribution of pairs of coprime m-smooth polynomials. The focus of this paper is on that problem.

We should point out that the fastest variant of the index calculus method in $\mathbb{F}_{2^n}$ is due to Coppersmith [6]. His algorithm is also based on heuristic arguments. We return to this algorithm on the last section of the paper.

We conclude this section commenting on the structure of the paper. Section 2 provides an introduction to the index calculus method and the Waterloo variant, and to the estimates needed for a rigorous proof of its running time. Section 3 presents those generating functions required in this paper, whereas the main results of the paper are in Section 4. We establish an asymptotic approximation for the number of coprimes m-smooth polynomials of degree n_1 and n_2 in terms of the product of the number of m-smooth polynomials of degree n_1 by the number of m-smooth polynomials of degree n_2. This implies the rigorous analysis of the running time for the Waterloo algorithm; see Section 5. Finally, Section 6 indicates further directions of research.

2. The Waterloo Algorithm

In this section we review the Waterloo variant of the index calculus method. We concentrate on a key change introduced by Blake et al. [3, 4]. We exemplify the change using the second stage of the algorithm, that is, the computation of individual logarithms once we have computed the database of logarithms of irreducible polynomials of degree smaller or equal than m. A similar argument can be used for the first stage, see [3, 4, 15].

The algorithm deals with the computation of the logarithm of an element in $\mathbb{F}_{2^n}$. The elements in $\mathbb{F}_{2^n}$ are represented by polynomials over $\mathbb{F}_2$ of degree smaller than n by means of the isomorphism $\mathbb{F}_{2^n} \cong \mathbb{F}_2[x]/(f)$, where f is a monic irreducible polynomial over $\mathbb{F}_2$ of degree n. To compute any arithmetic operation in $\mathbb{F}_{2^n}$ we simply perform the arithmetic operation with the associated polynomials modulo f.

We are given the polynomial f over $\mathbb{F}_2$ of degree n that defines the extension $\mathbb{F}_{2^n}$, a generator g of the multiplicative group of $\mathbb{F}_{2^n}$, a parameter m, and a polynomial $h(x) \in \mathbb{F}_2[x]$ whose discrete logarithm we want to compute.

Input: polynomials f, g, h as defined above and an integer m
Output: $\log_g h(x)$

0) Set A to 0.
1) If $\deg h(x) \leq m$ and $h(x) = \prod_i h_i(x)^{e_i}$, then $\log_g h(x) \equiv \sum_i e_i \log_g h_i(x) \bmod (q-1)$. Stop.
2) Generate a random integer a; set A to $A + a$ and $h(x)$ to $h(x)g(x)^a$. Apply the extended Euclidean algorithm to $h(x)$ and $f(x)$ to obtain polynomials

$$t(x)h(x) \equiv r(x) \pmod{f(x)},$$

with $\deg r(x), \deg t(x) \leq (n-1)/2$ if n is odd (respectively, $\deg r(x) \leq n/2$ and $\deg t(x) \leq (n-2)/2$ if n is even).
3) Factor $t(x) = \prod_i p_i(x)^{d_i}$ and $r(x) = \prod_j p_j(x)^{d_j}$. If $\deg p_i(x) \leq m$ and $\deg p_j(x) \leq m$ for all i, j, then compute the required discrete logarithm as

$$\log_g h(x) = \sum_j d_j \log_g p_j(x) - \sum_i d_i \log_g p_i(x) - A,$$

and stop. Otherwise, return to 2.

The essential step of this method is step 2 (see McEliece [14] and Blake et al. [3, p. 278] for the usage of the extended Euclidean algorithm to show the correctness of the method). It should be further mentioned that these polynomials $r(x), t(x)$ are coprime and unique if we, for example, assume that $r(x)$ is monic. More precisely, there is a bijection between all non-zero polynomials $h(x)$ of degree $\leq n$ and all pairs of polynomials $r(x), t(x)$ of degees $\leq (n-1)/2$, if n is odd (and of degrees $\leq n/2$ and $\leq (n-2)/2$, if n is even), which are coprime and for which $r(x)$ is monic. Thus, if a is chosen at random (in step 2), then we get the same (uniform) probability distribution on all possible polynomials $h(x)$ as choosing two polynomials $r(x), t(x)$ independently (under the above restrictions). This means that we can analyze step 2 by counting arguments.

It seems that this has not been observed before. For example, Blake et al. approximated that probability $P(m; (n-1)/2, (n-1)/2)$ (that a pair of polynomials of degrees $\leq (n-1)/2$ is coprime and m-smooth, i.e. all irreducible factors have degree $\leq m$) by $P^2(m; (n-1)/2)$ (where $P(m; (n-1)/2)$ is the probability that a polynomial of degree $\leq (n-1)/2$ is m-smooth). This entails the heuristic justification of the running time in their algorithm. As a corollary of our main result (Theorem 1), we show that this approximation is correct in asymptotic terms, and we provide a precise estimate for this relation.

We remark that the optimal value of m for the Waterloo variant is $m = k\sqrt{n \log n}$, for a computable constant k (see [15]). Our proof is for a large range of values of m including this optimal value.

3. Combinatorial Steps

The first part of our proof is the derivation of appropriate generating functions for counting the number of polynomials of interest in this work. Before that, we briefly comment on generating functions for polynomials over finite fields.

3.1. **Generating functions for polynomials over finite fields.**

Generating functions are a classical tool to count polynomials over finite fields with particular characteristics (see Chapter 3 of Berlekamp's book [2], and the exercise section 4.6.2 of Knuth's book [12]). The reader is pointed out to [10] for a more detailed introduction to the usage of generating functions to analize algorithms that deal with polynomials over finite fields.

The number of monic polynomials over $\mathbb{F}_q$ of degree n is q^n. Thus, the generating function of polynomials over $\mathbb{F}_q$ is

$$P(z) = \sum_{k\ge 0} q^k z^k = \frac{1}{1-qz}.$$

Another way of obtaining this generating function is by considering an enumerator of a fixed irreducible factor of degree k

$$1 + z^k + z^{2k} + \cdots .$$

This enumerator counts the number of times, $0, 1, 2, \ldots$, that this particular irreducible factor appears in a polynomial. Let I_k be the number of monic irreducible polynomials over $\mathbb{F}_q$ of degree k. Then, considering the I_k irreducible factors of degree k, and varying on k, we obtain (by distributivity law) the generating function of polynomials over $\mathbb{F}_q$

$$P(z) = \prod_{k\ge 1} \left(1 + z^k + z^{2k} + \cdots\right)^{I_k} = \prod_{k\ge 1} \left(\frac{1}{1-z^k}\right)^{I_k}.$$

We conclude that

$$P(z) = \prod_{k\ge 1} \left(\frac{1}{1-z^k}\right)^{I_k} = \frac{1}{1-qz}. \tag{1}$$

In this paper, we need to count pairs of polynomials over $\mathbb{F}_q$ with particular characteristics. This involves bivariate generating functions. As an introduction to the bivariate generating functions required here, we now provide the probability that two polynomials over $\mathbb{F}_q$ are relatively prime.

Proposition 1. *Let f, g be two monic polynomials over $\mathbb{F}_q$ of degree $n_1 > 0$ and $n_2 > 0$. Then, the probability that f and g are relatively prime is $1 - 1/q$.*

Proof. Let us consider an enumerator of a fixed irreducible factor of degree k that counts the presence of the factor in one of the polynomials f or g but not in both

$$1 + z^k + z^{2k} + \cdots + w^k + w^{2k} + \cdots .$$

Now, we consider the I_k irreducible factors of degree k, and varying on k, we obtain the generating function of pairs of coprime polynomials over $\mathbb{F}_q$

$$
\begin{aligned}
P(z,w) &= \prod_{k\geq 1}\left(1+z^k+z^{2k}+\cdots+w^k+w^{2k}+\cdots\right)^{I_k}\\
&= \prod_{k\geq 1}\left(\frac{1}{1-z^k}+\frac{1}{1-w^k}-1\right)^{I_k} = \prod_{k\geq 1}\left(\frac{1-z^k w^k}{(1-z^k)(1-w^k)}\right)^{I_k}\\
&= \frac{P(z)P(w)}{P(zw)} = \frac{1-qzw}{(1-qz)(1-qw)}.
\end{aligned}
$$

As usual, $[z^{n_1}w^{n_2}]P(z,w)$ denotes the coefficient of $z^{n_1}w^{n_2}$ in the generating function $P(z,w)$. Then, the number of pairs of coprime polynomials over $\mathbb{F}_q$ of degree $n_1 > 0$ and $n_2 > 0$, respectively, is given by

$$
\begin{aligned}
[z^{n_1}w^{n_2}]P(z,w) &= [z^{n_1}w^{n_2}]\frac{1-qzw}{(1-qz)(1-qw)}\\
&= [z^{n_1}w^{n_2}]\frac{1}{(1-qz)(1-qw)} - q[z^{n_1-1}w^{n_2-1}]\frac{1}{(1-qz)(1-qw)}\\
&= q^{n_1+n_2} - q\,q^{n_1-1+n_2-1} = q^{n_1+n_2}\left(1-\frac{1}{q}\right).
\end{aligned}
$$

The total number of pairs of polynomials over $\mathbb{F}_q$ of degree n_1 and n_2, respectively, is $q^{n_1+n_2}$. Thus, the probability that two polynomials are coprime is

$$
\frac{q^{n_1+n_2}\left(1-\frac{1}{q}\right)}{q^{n_1+n_2}} = 1-\frac{1}{q}.
$$

■

Remarks. In the previous result we excluded the case when either $n_1 = 0$ or $n_2 = 0$. In these cases, the probability that f and g are relatively prime is 1.

Corteel et al. [7] prove that the probability that ℓ polynomials over $\mathbb{F}_q$ of equal degree n are relatively prime is $1-1/q^{\ell-1}$ (see also [16]). This provides an extension to [7] in the case $\ell = 2$. More important, our proof serves as an introduction to the next section when we add the "smooth" condition to the pairs of coprime polynomials.

3.2. **The number of coprime pairs of m-smooth polynomials.** Let $N_q(m;n)$ denote the number of monic polynomials over $\mathbb{F}_q$ of degree n which are m-smooth, i.e. all irreducible factors have degree smaller or equal to m. Using a similar argument as above, one can show (see [15]) that the generating function of these numbers

$$
S_m(z) = \sum_{n\geq 0} N_q(m;n)\,z^n
$$

is given by

$$S_m(z) = \prod_{k=1}^{m} \left(\frac{1}{1-z^k} \right)^{I_k}.$$

We use the well known approximation $I_k = \frac{1}{k}\left(q^k + O(q^{k/2})\right)$.

Odlyzko [15] provides an asymptotic estimate for these numbers as $n \to \infty$ (and uniformly for m in the range $n^{1/100} \le m \le n^{99/100}$). Indeed, his results hold for $n^{\delta} \le m \le n^{1-\delta}$ for a constant $\delta > 0$. The main tool of his proof is the saddle point method. Our approach is based on Odlyzko's work, and thus we review his proof when we derive our main asymptotic results (Theorems 1 and 2).

Similarly we can consider the numbers

$$\tilde{N}_q(m;n) = \sum_{n' \le n} N_q(m;n')$$

of monic polynomials over $\mathbb{F}_q$ of degree $\le n$ which are m-smooth. Their generating function $\tilde{S}_m(z)$ obviously satisfies

$$\tilde{S}_m(z) = \frac{1}{1-z} S_m(z) = \frac{1}{1-z} \prod_{k=1}^{m} \left(\frac{1}{1-z^k} \right)^{I_k}.$$

As we commented when we presented the Waterloo algorithm, we need an estimate for the number of coprime pairs of m-smooth polynomials over finite fields. Thus, let $N_q(m;n_1,n_2)$ denote the number of coprime pairs of monic polynomials over $\mathbb{F}_q$ of degrees n_1 and n_2, respectively, which are m-smooth. Let $F_m(z,w)$ be their generating function, that is,

$$F_m(z,w) = \sum_{n_1,n_2 \ge 0} N_q(m;n_1,n_2)\, z^{n_1} w^{n_2}.$$

By a similar argument as the one in Proposition 1 but not allowing irreducibles of degree greater than m, we have that this generating function is given by

$$\begin{aligned} F_m(z,w) &= \prod_{k=1}^{m} \left(1 + z^k + z^{2k} + \cdots + w^k + w^{2k} + \cdots\right)^{I_k} \\ &= \prod_{k=1}^{m} \left(\frac{1}{1-z^k} + \frac{1}{1-w^k} - 1 \right)^{I_k} \\ &= \prod_{k=1}^{m} \left(\frac{1-z^k w^k}{(1-z^k)(1-w^k)} \right)^{I_k} = \frac{S_m(z) S_m(w)}{S_m(zw)}. \qquad (2) \end{aligned}$$

It is clear that $N_q(m;n_1,n_2) \le N_q(m;n_1) N_q(m;n_2)$. Theorem 1 shows that this inequality is sharp despite of a constant.

For Theorem 2, we will also consider the numbers

$$\tilde{N}_q(m;n_1,n_2) = \sum_{n' \le n_1} \sum_{n'' \le n_2} N_q(m;n',n'')$$

of coprime pairs of monic polynomials over $\mathbb{F}_q$ of degrees $\le n_1$ and $\le n_2$, respectively, which are m-smooth. As above it is easy to express their generating function

$$\tilde{F}_m(z,w) = \frac{1}{1-z}\frac{1}{1-w}F_m(z,w) = \frac{1}{1-z}\frac{1}{1-w}\frac{S_m(z)S_m(w)}{S_m(zw)} = \frac{\tilde{S}_m(z)\tilde{S}_m(w)}{S_m(zw)}.$$

4. Asymptotic Expansion

In this section we prove the following theorems.

Theorem 1. *Let $\delta > 0$ be given. Then we have, uniformly for $m, n_1, n_2 \to \infty$ with $n_1^\delta \le m \le n_1^{1-\delta}$ and $n_2^\delta \le m \le n_2^{1-\delta}$,*

$$N_q(m;n_1,n_2) \sim \left(1-\frac{1}{q}\right) N_q(m;n_1)N_q(m;n_2).$$

Theorem 2. *Let $\delta > 0$ be given. Then we have, uniformly for $m, n_1, n_2 \to \infty$ with $n_1^\delta \le m \le n_1^{1-\delta}$ and $n_2^\delta \le m \le n_2^{1-\delta}$,*

$$\tilde{N}_q(m;n_1,n_2) \sim \left(1-\frac{1}{q}\right) \tilde{N}_q(m;n_1)\tilde{N}_q(m;n_2).$$

The proof of Theorem 1 follows from three lemmas. Before we give these lemmas, we informally introduce the ideas behind the proofs.

It is well known (see [13]) that in the range $\sqrt{n}\log n \ll m \le n$

$$N_q(m;n) \sim q^n \rho\left(\frac{n}{m}\right),$$

where $\rho(u)$ denotes the Dickman function. Estimates for the Dickman function are known in the wider range $\log n \ll m \le n$ (see [17]), and from this we can conclude that

$$N_q(m;n) = q^n e^{-(1+o(1))\, n/m\, \log(n/m)}. \tag{3}$$

Therefore, as a conclusion of our theorem, we also get uniform asymptotic expansions for $N_q(m;n_1,n_2)$. It should be noted that with some extra effort it is possible to extend the range for m in Theorem 1 to $\log n_i \ll m \le n_i$ $(i = 1,2)$. However, in order to keep the presentation short, we consider only the range which is most important for our application.

The proof relies on Cauchy's formula

$$\begin{aligned} N(m;n_1,n_2) &= \frac{1}{(2\pi i)^2}\int_{|z|=r_1}\int_{|w|=r_2} F_m(z,w)z^{-n_1-1}w^{-n_2-1}\,dz\,dw \\ &= \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} F_m\left(r_1e^{i\theta_1}, r_2e^{i\theta_2}\right) r_1^{-n_1}r_2^{-n_2}e^{-i(n_1\theta_1+n_2\theta_2)}\,d\theta_1\,d\theta_2, \end{aligned}$$

where $0 < r_1, r_2 < 1$ are real numbers. In fact, we will make use of Odlyzko's proof [15, Appendix A], who used this approach to extract

$$\begin{aligned} N(m;n) &= \frac{1}{2\pi i}\int_{|z|=r} S_m(z) z^{-n-1}\, dz \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_m\left(re^{i\theta}\right) r^{-n} e^{-in\theta}\, d\theta. \end{aligned}$$

He showed that it is possible to calculate this integral asymptotically via a usual application of the saddle point method. Let $r_0 = r_0(m,n)$ be the unique solution of the equation

$$\frac{rS_m'(r)}{S_m(r)} = n,$$

then we have, as $n \to \infty$ (and uniformly for m in the range $n^\delta \le m \le n^{1-\delta}$)

$$N(m;n) \sim \frac{S_m(r_0) r_0^{-n}}{\sqrt{2\pi b(r_0)}}, \tag{4}$$

where

$$b(r) = \left(\frac{S_m'(r)}{S_m(r)}\right)'.$$

Odlyzko also showed that $r_0 = r_0(m,n)$ is asymptotically given by

$$r_0 = \frac{1}{q}\exp\left(\frac{\log(n/m) + \log\log(n/m) + o(1)}{m}\right), \tag{5}$$

so that in the given range

$$r_0 - \frac{1}{q} = O\left(\frac{\log n}{n^\delta}\right). \tag{6}$$

Furthermore, he showed that

$$b(r_0) \sim 4mn \tag{7}$$

and that (for $n^\delta \le m \le n^{1-\delta}$ and $0 \le k \le 2m$)

$$\frac{N(m;n+k)}{N(m;n)} \sim q^k \left(\frac{n}{m}\log\frac{m}{n}\right)^{k/m}. \tag{8}$$

We use these results by Odlyzko in our proof.

The most important observation for the proof of Theorem 1 is that if $z \sim \frac{1}{q}$ and $w \sim \frac{1}{q}$, then the factor $1/S_m(zw)$ is almost constant in

$$F_m(z,w) = \frac{S_m(z)S_m(w)}{S_m(zw)}.$$

Asymptotically, we get an additional factor $1/(1-1/q)$. Therefore, one expects that, as proposed,

$$\begin{aligned} [z^{n_1}w^{n_2}]F_m(z,w) &\sim \frac{1}{S_m(q^{-2})}[z^{n_1}w^{n_2}]S_m(z)S_m(w) \\ &\sim \left(1-\frac{1}{q}\right) N(m;n_1)N(m;n_2). \end{aligned}$$

In what follows we show that these observations can be made rigorous.

Lemma 1. *Using the previous notation, we have*

$$\left|\frac{1}{S_m(z)} - \frac{1}{P(z)}\right| \ll \frac{|qz|^m}{m}$$

uniformly for z with $|z| \le q^{-1} - \eta$ (for any fixed $\eta < q^{-1}$).

PROOF. For $|z| \le q^{-1} - \eta$, we get

$$|\log P(z) - \log S_m(z)| = \left|\sum_{k>m} I_k \log\frac{1}{1-z^k}\right| \ll \sum_{k>m}\frac{q^k|z|^k}{k} \ll \frac{|qz|^m}{m},$$

where $P(z)$ is given in Equation (1). Furthermore, since $\log P(z)$ is bounded (for $|z| \le q^{-1} - \eta$), we finally obtain

$$\left|\frac{1}{S_m(z)} - \frac{1}{P(z)}\right| = \frac{1}{|P(z)|}\cdot\left|1 - \frac{P(z)}{S_m(z)}\right| \ll \frac{|qz|^m}{m}.$$

■

Lemma 2. *We have, uniformly for $m, n_1, n_2 \to \infty$ with $n_i^\delta \le m \le n_i^{1-\delta}$ $(i = 1, 2)$*

$$[z^{n_1}w^{n_2}]F_m(z,w) - [z^{n_1}w^{n_2}]\frac{S_m(z)S_m(w)}{P(zw)} = o\left(N_q(m;n_1)N_q(m;n_2)\right).$$

PROOF. We use Cauchy's formula for $|r_i - q^{-2}| \le \eta$ $(i = 1, 2)$, where η is any fixed constant smaller than $1/q$,

$$\begin{aligned}&[z^{n_1}w^{n_2}]F_m(z,w) - [z^{n_1}w^{n_2}]\frac{S_m(z)S_m(w)}{P(zw)}\\ &= \frac{1}{(2\pi i)^2}\int\limits_{|z|=r_1}\int\limits_{|w|=r_2}\left(\frac{S_m(z)S_m(w)}{S_m(zw)} - \frac{S_m(z)S_m(w)}{P(zw)}\right)z^{-n_1-1}w^{-n_2-1}\,dz\,dw.\end{aligned}$$

Using Lemma 1, we get the estimate

$$\left|[z^{n_1}w^{n_2}]F_m(z,w) - [z^{n_1}w^{n_2}]\frac{S_m(z)S_m(w)}{P(zw)}\right| \ll (qr_1r_2)^m S_m(r_1)S_m(r_2)r_1^{-n_1}r_2^{-n_2}.$$

It is now convenient to use Equation (4), the saddle points $r_1 = r_0(m,n_1) \sim 1/q$ and $r_2 = r_0(m,n_2) \sim 1/q$ given in Equation (6), and the asymptotic relations $b(r_i) \sim 4mn_i$ given in Equation (7), to observe that

$$(qr_1r_2)^m S_m(r_1)S_m(r_2)r_1^{-n_1}r_2^{-n_2} = o(N_q(m;n_1)N_q(m;n_2)).$$

This completes the proof of the lemma. ■

Lemma 3. *We have*

$$[z^{n_1}w^{n_2}]\frac{S_m(z)S_m(w)}{P(zw)} \sim \left(1 - \frac{1}{q}\right)N_q(m;n_1)N_q(m;n_2).$$

Remark: It is also quite easy to obtain the more general relation

$$[z^{n_1}w^{n_2}]G(z,w)S_m(z)S_m(w) \sim G(1/q,1/q)\,N_q(m;n_1)N_q(m;n_2) \tag{9}$$

for functions $G(z,w)$ which are analytic for $|z|,|w| \le 1/q+\eta$. However, for this purpose we have to use a proof which is a little bit more involved than the following one. Equation (9) follows from an adapted version of Odlyzko's proof [15] based on saddle point approximations.

PROOF. We note that $1/P(zw) = 1 - qzw$. Hence, we have

$$[z^{n_1}w^{n_2}]\frac{S_m(z)S_m(w)}{P(zw)} = N_q(m;n_1)N_q(m;n_2) - qN_q(m;n_1-1)N_q(m;n_2-1).$$

By Equation (8) we have

$$N_q(m;n-1) \sim \frac{1}{q}N_q(m;n),$$

so the result follows. ∎

The proof of Theorem 1 is a direct consequence of Lemma 2 and Lemma 3.

In the same way we can prove Theorem 2. We only have to observe that the factor $1/S_m(zw)$ is almost constant in the representation

$$\tilde{F}_m(z,w) = \frac{\tilde{S}_m(z)\tilde{S}_m(w)}{S_m(z,w)}.$$

We note that

$$\tilde{N}_q(m;n) \sim \frac{N_q(m;n)}{1-\frac{1}{q}}. \tag{10}$$

5. A rigorous proof for the Waterloo algorithm

We now return to the Waterloo algorithm. The probabilities associated with $N_q(m;n_1,n_2)$ and $N_q(m;n)$ can be computed dividing these numbers by proper powers of q. In other words, the probability $P(m;n_1,n_2)$ that a pair of polynomials of degrees n_1 and n_2 is m-smooth and coprime is

$$P(m;n_1,n_2) = \frac{N_q(m;n_1,n_2)}{q^{n_1+n_2}}.$$

In the same way, the probability that a polynomial of degree n_i, $i=1,2$, is m-smooth is given by

$$P(m;n_i) = \frac{N_q(m;n_i)}{q^{n_i}}.$$

Similarly we define the probability

$$\tilde{P}(m;n_1,n_2) = \frac{\tilde{N}_q(m;n_1,n_2)}{q^{n_1+n_2}},$$

and for $i=1,2$,

$$\tilde{P}(m;n_i) = \frac{\tilde{N}_q(m;n_i)}{q^{n_i}}$$

An immediate consequence of our theorems is the following result.

Corollary 1. *Let $\delta > 0$ be given. Then we have, uniformly for $m, n_1, n_2 \to \infty$ with $n_1^\delta \le m \le n_1^{1-\delta}$ and $n_2^\delta \le m \le n_2^{1-\delta}$,*

$$P(m; n_1, n_2) \sim \left(1 - \frac{1}{q}\right) P(m; n_1) P(m; n_2).$$

and

$$\tilde{P}(m; n_1, n_2) \sim \left(1 - \frac{1}{q}\right) \tilde{P}(m; n_1) \tilde{P}(m; n_2).$$

The probability that two random monic independent polynomials of degree at most $n/2$ factor into irreducibles of degree at most m is

$$e^{-2(1+o(1))\frac{n}{2m}\log\frac{n}{2m}} = e^{-(1+o(1))\frac{n}{m}\log\frac{n}{2m}}, \tag{11}$$

where we used Corollary 1, Theorem 2, Equation (10) and finally Equation (3).

The size of the factor base remains the same as in the basic version (see [15], pp. 238-243), therefore an upper bound on the running time for the first stage is

$$e^{(1+o(1))\left(m\log 2 + \frac{n}{m}\log\frac{n}{2m}\right)} + e^{(2+o(1))m\log 2}. \tag{12}$$

We let $m = k\sqrt{n\log n}$, and Equation (12) becomes

$$e^{(1+o(1))\left(k\log 2 + \frac{1}{2k}\right)\sqrt{n\log n}} + e^{(1+o(1))(2k\log 2)\sqrt{n\log n}}.$$

We use the notation $L(n) = e^{(1+o(1))\sqrt{n\log n}}$. The above expression is minimized for $k = (2\log 2)^{-1/2}$, and the running time of the first stage is bounded by $L(n)^{\sqrt{2\log 2}}$. The second stage has running time bounded by $L(n)^{\sqrt{\log 2/2}}$. Therefore, putting all pieces together, we have rigorously proven that the running time of the Waterloo variant is bounded by

$$L(n)^{\sqrt{2\log 2}}.$$

6. Conclusions

In this paper we study the distribution of pairs of coprime smooth polynomials over finite fields. Our results provide rigorous estimates for the running time of the index calculus variant proposed by Blake et al.

It would be interesting to extend the range of applicability of our Theorems 1 and 2. These extensions will not cover the full range of the parameter m. When m is very small or very large with respect to n, other techniques are needed.

The fastest known method for computing discrete logarithms in $\mathbb{F}_{2^n}$ is due to Coppersmith; see [6, 11, 15]. It is a heuristic method, like the Waterloo variant. Providing a rigorous proof for the Coppersmith's method seems out of reach at the moment. However, the methods introduced in this paper may shed some light on how to obtain a rigorous proof for the Coppersmith's variant.

References

[1] Adleman, L. A subexponential algorithm for the discrete logarithm problem with applications. In *Proc. 20th IEEE Symp. Foundations of Computer Science*, pages 55–60, 1979.

[2] Berlekamp, E.R. *Algebraic coding theory.* McGraw Hill, New York NY, 1968.

[3] Blake, I.F., Fuji-Hara, R., Mullin, R.C., and Vanstone, S.A. Computing discrete logarithms in finite fields of characteristic two. *SIAM J. Alg. Disc. Meth.*, 5:276–285, 1984.

[4] Blake, I.F., Mullin, R.C., and Vanstone, S.A. Computing logarithms in GF(2^n). In *Advances in cryptology*, volume 196 of *Lecture Notes in Computer Science*, pages 73–82. Springer-Verlag, 1985.

[5] Blum, M., and Micali, S. How to generate cryptographically strong sequences of pseudorandom bits. *SIAM J. Comput.*, 13:850–864, 1984.

[6] Coppersmith, D. Fast evaluation of logarithms in fields of characteristic two. *IEEE Trans. Info. Theory*, 30:587–594, 1984.

[7] Corteel, S., Savage, C., Wilf, H., and Zeilberger, D. A pentagonal number sieve. *Journal of Combinatorial Theory,* Series A, 82:186–192, 1998.

[8] Diffie, W., and Hellman, M. New directions in cryptography. *IEEE Trans. Inform. Theory*, 22:644–654, 1976.

[9] ElGamal, T. A public key cryptosystem and a signature scheme based on discrete logarithms. *IEEE Trans. Info. Theory*, 31:469–472, 1985.

[10] Flajolet, P., Gourdon, X., and Panario, D. The complete analysis of a polynomial factorization algorithm over finite fields. To appear in J. of Algorithms. [Extended abstract in *Proc. 23rd ICALP Symp.*, Lecture Notes in Computer Science, vol. 1099, p. 232–243, 1996.].

[11] Garefalakis, T., and Panario, D. The index calculus method using non-smooth polynomials. *Mathematics of Computation*, 70:1253–1264, 2001.

[12] Knuth, D.E. *The art of computer programming, vol.2: seminumerical algorithms.* Addison-Wesley, Reading MA, 3 edition, 1997.

[13] Manstavičius, E. Semigroup elements free of large prime factors. In *New trends in Probability and Statistics*, pages 135–153. VSP/TEV, 1992.

[14] McEliece, R.J. *The Theory of Information and Coding: a Mathematical Framework for Communications.* Addison-Wesley Publishing, Reading, MA, 1977.

[15] Odlyzko, A. Discrete logarithms and their cryptographic significance. In *Advances in Cryptology, Proceedings of Eurocrypt 1984*, volume 209 of *Lecture Notes in Computer Science*, pages 224–314. Springer-Verlag, 1985.

[16] Reifegerste, A. On an involution concerning pairs of polynomials over $\mathbb{F}_2$. *Journal of Combinatorial Theory,* Series A, 90:216–220, 2000.

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

[18] Western, A.E., and Miller, J.C.P. Tables of indices and primitive roots. In *Royal Society Mathematical Tables*, volume 9. Cambridge University Press, 1968.

Institut für Geometrie, TU Wien, Wien, A-1040, Austria, michael.drmota@tuwien.ac.at

School of Mathematics and Statistics, Carleton University, K1S 5B6, Ottawa, Canada, daniel@math.carleton.ca