ON A FAMILY OF PREIMAGE-RESISTANT FUNCTIONS

ATTILA BÉRCZES, JÁNOS FOLLÁTH, AND ATTILA PETHŐ

1. Introduction

One of the most basic notion for cryptographic applications is the one-way function. These functions are important building blocks for most of the protocols and play a fundamental role in verifying passwords and creating digital signatures. Their use is important for constructing cryptographically secure pseudo-random-number generators. There is an extensive literature on one-way functions and their applications. We refer here only to two fundamental books on cryptography [21] and [30].

A one-way function is a function which is "easy" to compute but "hard" to invert. Complexity theoretical point of view this means, that a one-way function can be computed in polynomial time, but all of its inverses only in superpolynomial time. If a function belongs to the **P** (polynomial) class, then its inverses belong to the **NP** class and there can exist a one-way function in the above sense only if $\mathbf{P} \neq \mathbf{NP}$, (see e.g. [23]).

Despite the lack of the safe theoretical background, there appeared in the literature several suggestions for the construction of one-way functions. The papers [32], [22], [12] and [16] show how to construct a candidate one-way function. O. Goldreich, L. Levin and N. Nisan [14] make a one-to-one candidate one-way function based on the hardness of inverting RSA and the discrete log problem.

2000 Mathematics Subject Classification: Primary 94A60; Secondary 68P25, 11G25, 11D79.

Keywords and Phrases: hash function, collision, polynomials.

The research was supported in part by the Hungarian Academy of Sciences (A.B.,A.P.), and by grants T67580 (A.B., A.P.) and T75566 (A.B.) of the Hungarian National Foundation for Scientific Research, the János Bolyai Research Scholarship (A.B.).

J. Buchmann and S. Paulus [9] use results from algebraic number theory to construct a candidate one-way function. It is based on the hardness of the discrete logarithm problem with respect to the ideal class group of algebraic number fields.

The lattice-based one-way function candidate, introduced by Ajtai and Dwork [1] is the most promising from theoretical point of view. It is based on the computation of shortest vector in a lattice and its average case complexity is the same as its worst case complexity.

Important property of one-way functions is the *collision resistance*. Informally this means that "it is computationally infeasible to find any two distinct inputs x, x' which hash to the same output, i.e., such that $h(x) = h(x')$." (c.f. [25]). A weaker form of collision resistance is the *preimage-resistance*. It means that "it is computationally infeasible to find [...] any preimage x' such that $h(x') = y$ when given any y for which a corresponding input is not known".

Bérczes, Ködmön and Pethő [7] constructed a family of preimage-resistant functions based on norm functions, well studied in the theory of diophantine equations. Bérczes and Járási [8] extended this result to a family based on index forms. In both cases the functions were reduced modulo m, where m is the product of two large primes. For security reasons m should have at least 1024 binary digits. The first construction was implemented by the company Crypto Ltd under the name CODEFISH. J.-P. Aumasson [4] pointed out some vulnerability of the implemented algorithm.

The aim of this paper is to continue the investigations of [7] and [8] on the preimage-resistance of functions defined over finite rings and improve their results in two directions. First, we are working on finite fields $\mathbb{F}_q$ and not on finite rings $\mathbb{Z}_m$, where m is the product of two primes. For the security of the construction of [7, 8] m has to be hard to factorize, i.e., it must be at least 1024 bit long. In contrast the length of q can be considerably shorter, e.g., 256 or 512 bits. Second, we are able to handle functions over finite fields of characteristic two, which makes the implementation of the proposed algorithms much more efficient. The main difference of the new construction with respect to the previous ones is that our functions are inhomogenous polynomials.

We mention, that the above mentioned vulnerability of CODEFISH was caused because it was possible to compute the value of the hash function using circulant matrices. Aumasson used the properties of such matrices to prove the vulnerability of the function. Since the present construction has no connection to circulant matrices, the vulnerabilities pointed out by Aumasson do not occur in the case of this construction.

In Theorem 2.1 we define a large family of polynomials $\mathcal{F}$. It is proved that under mild and easily decidable conditions the members of this family are nearly permutational polynomials. In Section 5 we define a subfamily $\mathcal{F}_1$ such that its members are easy to evaluate. For $f \in \mathcal{F}$ the preimage-resistance means that for any $\gamma \in \mathbb{F}_q$ it is infeasible to find $\mathbf{x} \in \mathbb{F}_q^n$ such that $f(\mathbf{x}) = \gamma$. Our result implies that if q is large enough then the solution of this equation by chance is computationally infeasible.

There are algorithms for the root finding problem over finite fields. The best known algorithm is due to Berlekamp [6], which is exponential in the characteristic of $\mathbb{F}_q$, but its probabilistic version is polynomial in this parameter [11]. Both versions are polynomial in k, the degree of the polynomial. However, if k is as large as q, then the root finding problem becomes intractable. Shparlinski [31] provides arguments for the hardness of the discrete logarithm problem by proving that the discrete logarithm function cannot be represented by a low degree polynomial. The members of the family $\mathcal{F}_1$ are large degree multivariate sparse polynomials. Kaltofen and Koiran [17] claimed that the "the complexity of root finding of supersparse polynomials over finite fields is open". Moreover, they proved that a Monte Carlo polynomial time irreducibility test for supersparse polynomials in $\mathbb{F}_{2^m}[X, Y]$ would imply a Las Vegas polynomial-time factorization algorithm for integers.

Of course zero is a trivial root of a univariate polynomial with zero constant term. To find such a specialization for members of $\mathcal{F}_1$ means the solution of the root finding problem for polynomials with number of unknowns one less than the original one. For this there does not exist efficient algorithm and by Theorem 2.1 the random choice does not work as well. By this reasons we believe that the members of $\mathcal{F}_1$ are good candidates to be preimage-resistance functions.

2. Main results

Let p be a prime and let $q = p^f$ with $f \geq 1$ an integer. Denote by $\mathbb{F}_q$ the finite field with q elements. For any polynomial $P(X_1, \ldots, X_m) \in K[X_1, \ldots, X_m]$ denote by $\deg P$ the total degree of P and by $\deg_{X_i} P(\mathbf{X})$ the partial degree of P with respect to the variable X_i.

Our main result is the following theorem.

Theorem 2.1. *Let $f(\boldsymbol{X}) \in \mathbb{F}_q[X_1, \ldots, X_m]$ be a polynomial such that*

$$f(\boldsymbol{X}) := b(X_1, \ldots, X_m) + a(X_1, \ldots, X_m)$$

with homogeneous polynomials $a(\boldsymbol{X}), b(\boldsymbol{X})$ satisfying $k = \deg a(\boldsymbol{X}) < \deg b(\boldsymbol{X}) = n$, $\deg_{X_i} b(\boldsymbol{X}) = n$ for $1 \leq i \leq m$. Further, suppose that there exist indices $1 \leq j_1 < j_2 \leq n$ such that the binary form

$$b_0(X_{j_1}, X_{j_2}) := b(0, \ldots, 0, X_{j_1}, 0, \ldots, 0, X_{j_2}, 0, \ldots, 0) \tag{1}$$

has no multiple zero.

Let $N(f, \gamma, q)$ denote the number of solutions of the equation $f(x_1, \ldots, x_m) = \gamma$ in $x_1, \ldots, x_m \in \mathbb{F}_q$. Then

$$|N(f, \gamma, q) - q^{m-1}| \leq (n-1)(n-2)q^{m-3/2} + 5n^{13/3}q^{m-2}. \tag{2}$$

Moreover, if $q > 15n^{13/3}$, then

$$|N(f, \gamma, q) - q^{m-1}| \leq (n-1)(n-2)q^{m-3/2} + (5n^2 + n + 1)q^{m-2}. \tag{3}$$

As a simple corollary we get that the functions defined in Theorem 2.1 are preimage-resistant. More precisely we have:

Corollary 2.2. *Assume that the polynomial $f(\boldsymbol{X}) \in \mathbb{F}_q[X_1, \ldots, X_m]$ satisfies the requirements of Theorem 2.1. Denote by $P_{coll}(f, \gamma)$ the probability that $f(\boldsymbol{x})$ assumes the value $\gamma \in \mathbb{F}_q^*$, when $\boldsymbol{x}$ runs uniformly through the elements of $\mathbb{F}_q^m$. Then*

$$P_{coll}(f, \gamma) \leq \frac{1}{q} + \frac{(n-1)(n-2)}{q^{3/2}} + \frac{5n^{13/3}}{q^2}.$$

Moreover, if $q > 5n^{13/3}$, then

$$P_{coll}(f, \gamma) \leq \frac{3}{q}.$$

For practical application we propose to choose $q = p$, with a prime $p > 2^{256}$ or $q = 2^f$ with $f = 256$ or 512. The polynomial $b(\mathbf{X})$ can be chosen in the first case a norm function, like in [7] or a diagonal form $\alpha_1 X_1^n + \cdots + \alpha_m X_m^n$ with $\alpha_1, \ldots, \alpha_m$ as well as $\beta_1, \ldots, \beta_m$ random elements of $\mathbb{F}_q^*$. The choice of m, n we discuss in Section 5.

3. Auxiliary results

First we cite a general result of Cafure and Matera [10] (c.f. also [18, 29]) about the number of $\mathbb{F}_q$ points lying on a hypersurface defined over $\mathbb{F}_q$.

Theorem 3.1. *For an absolutely irreducible $\mathbb{F}_q$ -hypersurface H of $\mathbb{A}^n$ of degree δ the following estimate holds:*

$$\left||H \cap \mathbb{F}_q^n| - q^{n-1}\right| \leq (\delta - 1)(\delta - 2)q^{n-3/2} + 5\delta^{13/3}q^{n-2}.$$

In the theorem $\mathbb{A}^n$ denotes the affine space of dimension n over $\mathbb{F}_q$. Under a regularity condition, i.e., if q is large enough, much better remainder term was proved by Cafure and Matera [10].

Theorem 3.2. *Let $q > 15\delta^{13/3}$ and let $H \subseteq \mathbb{A}^n$ be an absolutely irreducible $\mathbb{F}_q$ - hypersurface of degree δ. Then the following estimate holds:*

$$\left||H \cap \mathbb{F}_q^n| - q^{n-1}\right| \leq (\delta - 1)(\delta - 2)q^{n-3/2} + (5\delta^2 + \delta + 1)q^{n-2}.$$

The next lemma shows that under certain condition lacunary polynomials are absolutely irreducible.

Lemma 3.1. *Let K be any field, and fix an algebraic closure $\overline{K}$ of K. Let $n \geq 4$ be an integer and let*

$$G(X, Y) = Y^n + A(X)Y^{n-1} + B(X) \in K[X, Y]$$

be a polynomial with the properties $A(X), B(X) \in K[X]$, $B(X)$ has no multiple zeros and $\deg A(X) \neq \deg B(X) \geq 1$. Then $G(X, Y)$ is irreducible over $\overline{K}$, i.e., it is absolutely irreducible.

Proof. Suppose indirectly that $G(X, Y)$ is reducible, i.e., $G(X, Y) = U(X, Y)V(X, Y)$ with

$$U(X, Y) = Y^k + a_{k-1}(X)Y^{k-1} + \cdots + a_1(X)Y + a_0(X) \in \overline{K}[X, Y]$$
$$V(X, Y) = Y^{n-k} + b_{n-k-1}(X)Y^{n-k-1} + \cdots + b_1(X)Y + b_0(X) \in \overline{K}[X, Y],$$

where $1 \leq k \leq n-1$, $a_i(X), b_j(X) \in \overline{K}[X]$ for $i, j \in Z_{\geq 0}$, $a_k(X) = 1$, $b_{n-k}(X) = 1$ and $a_i(X) = 0$, for $i > k$ and $b_j(X) = 0$ for $j > n-k$ are constant polynomials.

Case I. First we suppose that $min(k, n-k) \geq 2$. We have

$$G(X,Y) = U(X,Y)V(X,Y) = \sum_{i=0}^{n} c_i(X)Y^i, \tag{4}$$

with

$$c_i(X) = \sum_{j=0}^{i} a_j(X)b_{i-j}(X). \tag{5}$$

Since $\deg B(X) \geq 1$, without loss of generality we may suppose that $\deg a_0(X) \geq 1$. Then there exists an $\alpha \in \overline{K}$ with $a_0(\alpha) = 0$. Since $B(X) = a_0(X)b_0(X)$, and $B(X)$ has no multiple zero, we get $b_0(\alpha) \neq 0$. By comparing (4) with $G(X,Y) = Y^n + A(X)Y^{n-1} + B(X)$ we get that $c_i(X) = 0$ is the constant 0 polynomial for $i = 1, \ldots, n-2$. Thus we have $c_i(\alpha) = 0$ for $i = 1, \ldots, n-2$, which together with (5) leads to

$$\sum_{j=0}^{i} a_j(\alpha)b_{i-j}(\alpha) = 0 \qquad \text{for } \ i = 1, \ldots, n-2. \tag{6}$$

Now (6) for $i = 1$, together with $a_0(\alpha) = 0$ and $b_0(\alpha) \neq 0$ proves $a_1(\alpha) = 0$. Similarly, (6) for $i = l$, together with $a_0(\alpha) = 0, \ldots, a_{l-1}(\alpha) = 0$ and $b_0(\alpha) \neq 0$ proves $a_l(\alpha) = 0$ for any $l = 1, \ldots, n-2$. Thus we conclude with $a_i(\alpha) = 0$ for $i = 0, \ldots, n-2$. Since $\min(k, n-k) \geq 2$ we have $U(\alpha, Y) = Y^k$ and

$$\begin{aligned} Y^n + A(\alpha)Y^{n-1} + B(\alpha) &= U(\alpha,Y)V(\alpha,Y) \\ &= Y^n + b_{n-k-1}(\alpha)Y^{n-1} + \cdots + b_0(\alpha)Y^k \end{aligned}$$

is clearly a contradiction.

Case II. Suppose now that $min(k, n-k) = 1$. Without loss of generality we may suppose that $k = 1$. Then

$$\begin{aligned} &U(X,Y) = Y + a_0(X) \\ &V(X,Y) = Y^{n-1} + b_{n-2}(X)Y^{n-2} + \cdots + b_1(X)Y + b_0(X). \end{aligned} \tag{7}$$

Then $Y^n + A(X)Y^{n-1} + B(X) = U(X,Y)V(X,Y)$ combined with (7) leads to the relations $B(X) = a_0(X)b_0(X)$, $a_0(X)b_l(X) - b_{l-1}(X) = 0$ for $l = 1, \ldots, n-2$, and $a_0(X) + b_{n-2}(X) = A(X)$. The first two relations show that $B(x)$ is divisible by $a_0(X)^2$, which together with the condition that $B(X)$ has no multiple zero, leads to the conclusion that $a_0(X) = a$ is a constant.

Now the above relations mean that $b_{n-k-2}(X) = (-a)^k b_{n-2}(X)$ for $k = 1, \ldots, n-2$. This shows that

$$\begin{aligned} Y^n + A(X)Y^{n-1} + B(X) &= U(X,Y)V(X,Y) \\ &= Y^n + (a + b_{n-2}(X))Y^{n-1} + a(-a)^{n-2}b_{n-2}(X). \end{aligned}$$

However, this means that $\deg A(X) = \deg B(X)$, which contradicts the assumption of the lemma.

This concludes the proof of Lemma 3.1. □

Note. In the above proof the fact that $G(X,Y)$ cannot have a factor with $\deg_Y > 2$ follows as a simple consequence of a much deeper result of Schinzel [26] (see also [27], [28]), however, since we have to deal with the case of quadratic factors anyway, we have proved Case I in general without using the result of Schinzel.

Lemma 3.2. *Let K be any field. Let $f(\boldsymbol{X}) \in K[X_1, \ldots, X_m]$ be a polynomial such that*

$$f(\boldsymbol{X}) := b(X_1, \ldots, X_m) + a(X_1, \ldots, X_m)$$

with homogeneous polynomials $a(\boldsymbol{X}), b(\boldsymbol{X})$ satisfying $k = \deg a(\boldsymbol{X}) < \deg b(\boldsymbol{X}) = n$, $\deg_{X_i} b(\boldsymbol{X}) = n$ for $1 \le i \le m$. Further, suppose that there exist indices $1 \le j_1 < j_2 \le m$ such that the binary form

$$b_0(X_{j_1}, X_{j_2}) := b(0, \ldots, 0, X_{j_1}, 0, \ldots, 0, X_{j_2}, 0, \ldots, 0) \tag{8}$$

has no multiple zero. Then the polynomial $f(\boldsymbol{X}) + \gamma$ is absolutely irreducible for every $0 \neq \gamma \in K$.

Proof. Put $g(\mathbf{X}) := f(\mathbf{X}) + \gamma$, $f_0(\mathbf{X}) := b_0(\mathbf{X}) + a_0(\mathbf{X})$ and $g_0(\mathbf{X}) := f_0(\mathbf{X}) + \gamma$, where $a_0(\mathbf{X}) := a(0, \ldots, 0, X_{j_1}, 0, \ldots, 0, X_{j_2}, 0, \ldots, 0)$.

Suppose indirectly, that $g(\mathbf{X})$ is reducible, i.e., $g(\mathbf{X}) = U(\mathbf{X})V(\mathbf{X})$, with $\deg U(\mathbf{X}) \geq 1$ and $\deg V(\mathbf{X}) \geq 1$. Thus there exists $i \in \{1, \dots, m\}$ such that $\deg_{X_i} U(\mathbf{X}) \geq 1$. Now using that $\deg_{X_j} g(\mathbf{X}) = n$ for each $j \in \{1, \dots, m\}$, we see that $\deg_{X_j} V(\mathbf{X}) < n$ and thus $\deg_{X_j} U(\mathbf{X}) > 0$ for each $j \in \{1, \dots, m\}$. Similarly, we infer that $\deg_{X_j} U(\mathbf{X}) < n$ and thus $\deg_{X_j} V(\mathbf{X}) > 0$ for each $j \in \{1, \dots, m\}$. Altogether, this means that we have

(9)
$$1 \leq \deg_{X_j} U(\mathbf{X}) \leq n-1 \text{ and } 1 \leq \deg_{X_j} V(\mathbf{X}) \leq n-1 \text{ for each } j \in \{1, \dots, m\}.$$

Now put

$$U_0(X_{j_1}, X_{j_2}) := U(0, \dots, 0, X_{j_1}, 0, \dots, 0, X_{j_2}, 0, \dots, 0)$$

and

$$V_0(X_{j_1}, X_{j_2}) := V(0, \dots, 0, X_{j_1}, 0, \dots, 0, X_{j_2}, 0, \dots, 0).$$

By (9) we see that $g_0(X_{j_1}, X_{j_2}) = U_0(X_{j_1}, X_{j_2})V_0(X_{j_1}, X_{j_2})$ is a non-trivial factorization of g_0.

However, since

$$g_0(X_{j_1}, X_{j_2}) = b_0(X_{j_1}, X_{j_2}) + a_0(X_{j_1}, X_{j_2}) + \gamma = X_{j_2}^n \left[b_0\left(\frac{X_{j_1}}{X_{j_2}}, 1\right) + \frac{1}{X_{j_2}^{n-k}} a_0\left(\frac{X_{j_1}}{X_{j_2}}, 1\right) + \gamma \frac{1}{X_{j_2}^n} \right] \tag{10}$$

the above non-trivial factorization of g_0 leads to a non-trivial factorization of the polynomial

$$Y^n + A(X)Y^{n-k} + B(X),$$

where $X := \frac{X_{j_1}}{X_{j_2}}, Y := \frac{1}{X_{j_2}}$, $A(X) := \frac{1}{\gamma} a_0(X, 1)$ and $B(X) := \frac{1}{\gamma} b_0(X, 1)$. However, this is impossible by Lemma 3.1. So we get a contradiction, which proves Lemma 3.2 □

4. Proof of Theorem 2.1 and its Corollary

Proof of Theorem 2.1 It follows from Lemma 3.2 that the polynomial $f - \gamma$ is absolutely irreducible over $\mathbb{F}_q$.

Thus by Theorems 3.1 and 3.2 the result follows.□

Proof of Corollary 2.2 Obviously, $\mathbb{F}_q^m$ has q^m elements and $P_{coll}(f,\gamma) = \frac{N(f,\gamma,q)}{|\mathbb{F}_q^m|}$, which together with Theorem 2.1 implies the first statement immediately.

If $q > 5n^{13/3}$ then $q^{1/2} > (n-1)(n-2)$ and we get the second statement from the first one at once. □

5. Practical considerations

In this section we are dealing with practical aspects of the proposed family of collision-free functions. The implementation was based on this analysis.

There are two typical ways for the choice of the finite field; either q is a prime, or q is a power of 2. To avoid brute force attack the binary length of q must be at least 128. The computation time depends very much on m, we decided to choose $m = 4$.

We decided to choose $f(\mathbf{X}) := b(\mathbf{X}) + a(\mathbf{X})$ such that $b(\mathbf{X})$ and $a(\mathbf{X})$ are in diagonal form, i.e., $b(\mathbf{X}) = \beta_1 X_1^r + \cdots + \beta_m X_m^r$ and $a(\mathbf{X}) = \alpha_1 X_1^s + \ldots \alpha_m X_m^s$ with $0 < s < r < q$ and $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m \neq 0$. With this choice all assumptions of Theorem 2.1 automatically hold except that the polynomial $b_0(X_i, X_j) = \beta_i X_i^r + \beta_j X_j^r$ has no multiple roots.

The polynomial $b_0(X_i, X_j)$ has multiple roots in $\bar{\mathbb{F}}_q$ if and only if $c(\mathbf{X}) = X^r + \gamma$ with $\mathbf{X} = \mathbf{X}_i/X_j$ and $\gamma = \beta_j/\beta_i$ has multiple roots in $\bar{\mathbb{F}}_q$. It is well-known that the multiple roots of $c(\mathbf{X})$ are roots of $\gcd(c(\mathbf{X}), c'(\mathbf{X}))$. Since $c'(\mathbf{X}) = r\mathbf{X}^{r-1}$, it is non-zero if r and the characteristic of $\mathbb{F}_q$ are coprime. This holds for all r, if q is a prime, and for all odd r, if $q = 2^f$. Further, if $c'(\mathbf{X}) \neq 0$, then its only root is 0, which is a zero of $c(\mathbf{X})$ if and only if $\gamma = 0$, but this is excluded by the choice of the β's. Thus we proved the following assertion.

Proposition 5.1. *Let $f(\boldsymbol{X}) := b(\boldsymbol{X}) + a(\boldsymbol{X})$ such that $b(\boldsymbol{X}) = \beta_1 X_1^r + \cdots + \beta_m X_m^r$, $a(\boldsymbol{X}) = \alpha_1 X_1^s + \cdots + \alpha_m X_m^s$ and $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m \neq 0$. If $0 < s < r < q$ and r is odd if $q = 2^f$, then $f(\boldsymbol{X})$ satisfies all assumptions of Theorem 2.1.*

The coefficients of $b(\mathbf{X})$ and $a(\mathbf{X})$ should be chosen distinct random elements of $\mathbb{F}_q$. Further, if we choose r such that $q > 5r^{13/3}$, then by Corollary

2.2 the probability that $f(\mathbf{X})$ takes a fixed element of $\mathbb{F}_q$ is at most $3/q$. If q is a prime, then we put $s = 1$, i.e., $a(\mathbf{X})$ a linear polynomial.

If $q = 2^f$, then choose a normal basis (see e.g. [19]) of $\mathbb{F}_q$ and represent the elements in this basis. Then squaring means a periodic left shift, while multiplication with a fixed element means mixing of the coordinates. Thus a good choice of r is, if its binary representation has at least 7 non-zero digits. Since the highest and the lowest digits are one (r must be odd), the remaining five ones should be distributed among the remaining positions. To be more specific, let $f = 128$. Then, for example $r = 2^{28} + 2^{24} + 2^{20} + 2^{15} + 2^{10} + 2^5 + 1 = 286295073$ satisfies all requirements. We propose to choose the exponent $s < r$ on the same principles as r.

In our implementation, we used the following iteration to hash messages of arbitrary length:

$$h_0 = f(w_0, \ldots, w_{m-1}), \quad h_{i+1} = f(h_i, w_{(m-1)i+1}, \ldots, w_{(m-1)(i+1)}),$$

where the w_j is the field element represented by the $\log q$ bits, beginning with the $j \log q + 1$th bit of the input message. We did a C language implementation of $f(\mathbf{X})$ with several choices of the parameters. The result of the computational time is displayed in the following table [1].

Hash length	Characteristic	Kilobyte/second
256	odd	338
512	odd	121
254	even	8
509	even	6

Although the efficiency of the odd characteristic version is near the efficiency of the VSH [13], the Fast VSH is 3-4 times faster.

6. Implementation

In this section we will consider the implementation issues regarding the above described families of hash functions. By the implementation of the

[1]The results were obtained on a 2GHz Intel(R) Core(TM)2 Duo CPU for several megabytes of messages

proposed functions choosing the field has a great impact on the performance. Beyond the obvious importance of the field size the choice of the characteristic has the greatest significance. It depends on the characteristic whether we can use simple modular arithmetic (which is advantageous on general purpose processors) or the (with hardware fast implementable) even characteristic arithmetic can be applied.

6.1. **Prime field arithmetic.** The (odd) prime field arithmetic (which means simple modular arithmetic in practice) is better suited for general purpose processors. On general purpose processors a single difficulty arises in conjunction with the proposed algorithm: the size of the operands. The parameters suggested in the previous section imply that the representation of the field elements is significantly longer than the nowadays widespread general purpose processors word size. There exists many implementation of arbitrary precision arithmetic for various programming languages. They are well tested and optimized, accelerated by assembly language fragments. Some of them also take a staged approach to the multiplication and squaring algorithms. Namely, they implement multiple algorithms and the fastest is used by a given operand length. Still, arbitrary precision arithmetic is significantly slower than word-level arithmetic. Consequently the performance of the proposed hash algorithm will be only comparable to those algorithms that also have to use arbitrary precision arithmetic (like for example [13]). We used the GNU Multiprecision Library in our implementation.

6.2. **Even characteristic arithmetic.** The primary strength of the proposed construction lies in the hardware implementation. If we define the function over an even characteristic field and use a normal basis representation, the squaring can be done with a simple cyclic shift which is extremely fast. The normal basis multiplication is also well studied and multiple fast architectures and implementations were proposed ([2],[3],[15],[20],[33]). Consequently the even characteristic version is a viable practical hash function when a hardware solution is needed.

The implementations of even characteristic multiplication on general purpose processors are usually slower than the prime field arithmetic, but since

the fast normal basis squaring, it may be also worth of consideration. Normal basis multiplication algorithms require many bit level operations, and thats why the implementations cannot make use of the full data path of the processor. The algorithms proposed by Reyhani-Masoleh and Hasan [24] avoid this disadvantage, making the multiplication much faster. In our implementation we used Algorithm 2 of [24]. This algorithm can only be applied by fields having a Gaussian normal base of even type. Gaussian normal basis are special, low complexity normal bases. A Gaussian normal base (GNB) exists for $GF(2^k)$ if k is not divisible by 8. GNB type t exists if and only if $p = tk + 1$ is prime and $\gcd(\frac{tk}{l}, k) = 1$), where l is the multiplicative order of 2 modulo p. By the tests we used GNB type 2 in both cases. Our implementation is only a pure C reference implementation of the construction, it can be significantly accelerated by further optimization, Algorithm 3 of [24], assembly language fragments and by reducing the Hamming weight of the exponents (notice that the algebraic complexity does not change with this modification). According to our estimates, with the above improvements, the performance of the even characteristic version can reach the odd characteristic one's. Further improvements can reach by using ideas described in the very recent paper of Bernstein and Lange [5].

Acknowledgement The authors are grateful for the referees for their thorough report on the paper, and also on the important modifications and corrections they suggested. The authors would like to thank Lajos Rónyai for his valuable remarks about the complexity of the root finding problem.

References

[1] M. Ajtai and C. Dwork, *A public-key cryptosystem with worst-case/average-case equivalence*, STOC '97 (El Paso, TX), 284–293 (electronic), ACM, New York, 1999.

[2] G.B. Agnew, R.C. Mullin, I.M. Onyszchuk, and S.A. Vanstone, *An Implementation for a Fast Public-Key Cryptosystem*, J. Cryptology, vol. 3, pp. 63-79, (1991)

[3] G.B. Agnew, R.C. Mullin, and S.A. Vanstone, *An Implementation of Elliptic Curve Cryptosystems over F2155*, IEEE J. Selected Areas in Comm., vol. 11, no. 5, pp. 804-813, June (1993)

[4] J.-P. Aumasson, *Cryptanalysis of a hash function based on norm form equations*, Cryptologia, **33** (2009), 1-4.

[5] D.J. Bernstein and T. Lange, *Type-II Optimal Polynomial Bases*, http://eprint.iacr.org/2010/069

[6] E.R. Berlekamp, *Factoring polynomials over large finite fields*, Math. Comp. **24** (1970), 713-715.

[7] A. Bérczes, J. Ködmön and A. Pethő, *A one-way function based on norm form equations*, Periodica Mathematica Hungarica, **49** (2004), 1-13.

[8] A. Bérczes and I. Járási, *An application of index forms in cryptography*, Periodica Math. Hungar., **58** (2008), 35-45.

[9] J. Buchmann, S. Paulus, *A one-way function based on ideal arithmetic in number fields*, Lect. Notes Comput. Sci. 1294 (1997), 385-394.

[10] A. Cafure and G. Matera, *Improved explicit estimates on the number of solutions of equations over a finite field*, Finite Fields Appl. **12** (2006), 155-185.

[11] D.G. Cantor and H. Zassenhaus, *A new algorithm for factoring polynomials over finite fields*, Math. Comp. **36** (1981), 587–592.

[12] L.R. Chao, Y.C. Lin, *Associative one-way function and its significances to cryptographics*, In. J. Inf. Manage. Sci. 5 (1994), 53-59.

[13] S. Contini, A.K. Lenstra and R. Steinfeld, *VSH, an efficient and provable collision-resistant hash function*, Advances in cryptology—EUROCRYPT 2006, 165–182, Lecture Notes in Comput. Sci., 4004, Springer, Berlin, (2006)

[14] O. Goldreich, L. Levin, N. Nisan, *On constructing 1-1 one-way functions*, Electronic colloquium on computational complexity, TR-95-029, 6/25/95, 1995.

[15] M.A. Hasan, M.Z. Wang, and V.K. Bhargava, *A Modified Massey-Omura Parallel Multiplier for a Class of Finite Fields*, IEEE Trans. Computers, vol. 42, no. 10, pp. 1278-1280, Oct. (1993)

[16] L.A. Hemaspaandra, J. Rothe, *Creating strong, total, commutative, associative one-way functions from any one-way function in complexity theory*, J. Comput. Syst. Sci. 58 (1999), 648-659.

[17] E. Kaltofen and P. Koiran, *On the complexity of factoring bivariate supersparse (lacunary) polynomials*. ISSAC'05, 208–215, ACM, New York, 2005.

[18] S. Lang and A. Weil, *The number of points of varieties in finite fields*, Amer. J. Math. **76** (1954) 819-827.

[19] R. Lidl and H. Niederreiter, *Finite fields. With a foreword by P. M. Cohn.* Second edition. Encyclopedia of Mathematics and its Applications, 20. Cambridge University Press, Cambridge, 1997.

[20] J.L. Massey and J.K. Omura, *Computational Method and Apparatus for Finite Field Arithmetic*, US Patent No. 4,587,627, (1986)

[21] A.J. Menezes, P.C.van Oorschot, S. Vanstone, *Handbook of Applied Cryptography*, CRC Press, 1997.

[22] R.C. Merkle, *A fast software one-way hash function*, J. Cryptology 3 (1990), 43-58.

[23] C.H. Papadimitriou, *Computational complexity*, Addison-Wesley Publishing Company, Reading, MA, 1994.

[24] A. Reyhani-Masoleh and M.A. Hasan, *Fast Normal Basis Multiplication Using General Purpose Processors*, IEEE Trans. Computers, vol. 52, no. 11, pp. 1379-1390, Nov.

[25] P. Rogaway and T. Shrimpton, *Cryptographic hash-function basics: definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance.* In B. Roy and W. Meier, editors, FSE 2004, volume 3017 of LNCS, page 371388, Berlin, 2004, Springer-Verlag.

[26] A. Schinzel, *On reducible trinomials*, Dissert. Math., **329** (1993), errata, Acta Arith. **73** (1995) 399-400.

[27] A. Schinzel, *On reducible trinomials. II*, Publ. Math. Debrecen, **56** (2000), 575–608.

[28] A. Schinzel, *On reducible trinomials. III*, Period. Math. Hungar., **43** (2001), 43–69.

[29] W.M. Schmidt, *A lower bound for the number of solutions of equations over finite fields*, J. Number Theory **6** (1974) 448-480.

[30] B. Schneier, *Applied Cryptography*, John Wiley & Sons,1996.

[31] I. Shparlinski, *Number theoretic methods in cryptography*, Complexity lower bounds. Progress in Computer Science and Applied Logic, vol. 17. Birkhäuser Verlag, Basel, 1999.

[32] Q. Sun, *A kind of trap-door one-way function over algebraic integers*, J. Sichuan Univ., Nat. Sci. Ed. No. 2 (1986), 22-27.

[33] B. Sunar and C.K. Koc, *An Efficient Optimal Normal Basis Type II Multiplier*, IEEE Trans. Computers, vol. 50, no. 1, pp. 83-88, Jan. (2001)

A. Bérczes
Institute of Mathematics, University of Debrecen
Number Theory Research Group, Hungarian Academy of Sciences and University of Debrecen
H-4010 Debrecen, P.O. Box 12, Hungary
E-mail address: berczesa@math.klte.hu

J. Folláth
Faculty of Informatics, University of Debrecen
H-4010 Debrecen, P.O. Box 12, Hungary
E-mail address: follathj@inf.unideb.hu

A. Pethő
Faculty of Informatics, University of Debrecen
Number Theory Research Group, Hungarian Academy of Sciences and University of Debrecen
H-4010 Debrecen, P.O. Box 12, Hungary
E-mail address: pethoe@inf.unideb.hu