On the Linear Complexity of the Naor-Reingold Sequence

Paula Bustillo, Domingo Gómez, Jaime Gutierrez, and Álvar Ibeas

1 Introduction

In this paper we provide a bound for the linear complexity of the so-called Naor-Reingold sequence of elements in a finite prime field. This sequence is presented in [4] as a primitive for crytptographic protocols.

For a prime p, we denote by $\mathbb{F}_p$ the field with p elements and identify them with the integers in the range $\{0, \ldots, p-1\}$. We write $\mathbb{F}_p^*$ for the group of units in this field.

Let p, t be prime numbers, n a positive integer, and $g \in \mathbb{F}_p^*$ an element of multiplicative order t. Then, a vector $\mathbf{a} = (a_0, \ldots, a_{n-1}) \in (\mathbb{F}_t^*)^n$ defines the following finite sequence in the subgroup $\langle g \rangle$:

$$f_{\mathbf{a}}(x) := g^{\varphi_{\mathbf{a}}(x)},$$

where $\varphi_{\mathbf{a}}(x) := a_0^{x_0} \cdots a_{n-1}^{x_{n-1}} \in \mathbb{F}_t^*$ and $x = \sum_{i=0}^{n-1} x_i 2^i$ is the binary representation of x. Note that we have required that the order of g is prime, which is not necessary for the definition of the sequence. It would be interesting to deal with this general case as well.

It has been shown that, if the decisional Diffie-Hellman assumption holds, then in general the index x is not enough to compute in polynomial time $f_{\mathbf{a}}(x)$, even if an attacker performs polynomially many oracle calls (see [4, Theorem 4.1]). A bound on the discrepancy of the Naor-Reingold sequence is given in [6] and the article [3] investigates its period.

We recall that the *linear complexity* of an N-element sequence over a ring:

$$f(x), \qquad x = 0, \ldots, N-1$$

is the order L of the shortest linear recurrence

$$f(x+L) = a_{L-1} f(x+L-1) + \cdots + a_1 f(x+1) + a_0 f(x), \qquad x = 0, \ldots, N-L-1.$$

The linear complexity of the Naor-Reingold sequence has been studied in [2, 5]. Those articles provide lower bounds for it, assuming that the dimension n of the parameter $\mathbf{a}$ is bigger than the logarithm of t. In particular, those bounds do not apply in the case $n \sim \log t$, for the set of excluded parameters covers almost the whole range. Note that this is a natural assumption, for those two parameters define the representation size of the sequence. In this article we modify the proof of [5] and obtain a lower bound for the linear complexity that is nontrivial even when $n \sim \log t$.

2 Preliminaries

Throughout the paper the implied constants in the symbols 'O' and '$\gg$' are absolute, and log denotes the binary logarithm. We state now an immediate consequence of the proof of Lemma 2 from [1]:

Lemma 1 *Let $\mathcal{K} \subseteq \mathbb{F}_t^*$ be a set of cardinality $\#\mathcal{K} = K$, and write*

$$L_r(\mathcal{K}, h) = \#\{(k, y) \in \mathcal{K} \times \mathbb{F}_t^* \mid rk = y,\ 0 \le y < h\}.$$

There exists $r \in \mathbb{F}_t^*$ such that $L_r(\mathcal{K}, h) \ge Kh/t$.

We will also use Lemma 2 from [5].

Lemma 2 *Consider a finite sequence $(f(x))_{x=0}^{N-1}$ in a field $\mathbb{K}$, with linear complexity L. Then, for any integers $M \ge 1,\ h \ge 1$, and $0 \le e_0, \ldots, e_L \le h$ there are some elements $c_0, \ldots, c_L \in \mathbb{K}$ (not all zero) such that*

$$\sum_{j=0}^{L} c_j f(Mb + e_j) = 0$$

for any integer b with $0 \le Mk + h \le N - 1$.

We prove now a result about the number of elements in a generic Naor-Reingold sequence. A more general result can be found in [7].

Proposition 3 *For any integers $n \ge j > 0$ and for all except at most $(3^j - 1)(t-1)^{n-1}/2$ vectors $\mathbf{a} \in (\mathbb{F}_t^*)^n$, the Naor-Reingold sequence contains at least 2^j distinct elements.*

Proof. If the sequence $f_{\mathbf{a}}(x),\ x = 0, \ldots,\ 2^n - 1$ contains fewer than 2^j values, there must be one repetition among the first 2^j. Suppose that $f_{\mathbf{a}}(x) = f_{\mathbf{a}}(y)$, with

$$x = \sum_{i=0}^{j-1} x_i 2^i, \qquad y = \sum_{i=0}^{j-1} y_i 2^i.$$

Then,

$$a_0^{x_0} \cdots a_j^{x_j} = a_0^{y_0} \cdots a_j^{y_j}. \tag{1}$$

Let i be the most significant position such that $x_i \ne y_i$. Without loss of generality, we can suppose that $x_i = 1,\ y_i = 0$. Equation (1) gives

$$a_i = a_0^{y_0 - x_0} \cdots a_{i-1}^{y_{i-1} - x_{i-1}}.$$

Once fixed the values $a_0 \ldots, a_{i-1}$ and the exponents $y_0 - x_0, \ldots,\ y_{i-1} - x_{i-1}$; then a_i is determined as well. Therefore, there are at most $3^i(t-1)^{n-1}$ possibilities for the parameter vector $\mathbf{a}$. Summing up all these values from the possible indices i, we obtain the result. ■

3 Linear complexity bound

We are ready to prove the main result of the article. The combination of the technique developed in [5, 7] with Lemma 1 yields a nontrivial result even in the case $n \sim \log t$. Furthermore, this bound improves the one provided in [5] when $n \ll (1 + \log_3 2) \log t$.

Theorem 4 *Let $\gamma > 0$ and $0 < \varepsilon < 1$ such that*

$$n > \log t + \gamma - 4.$$

The linear complexity $L_{\mathbf{a}}$ of the sequence $(f_{\mathbf{a}}(x))_{x=0}^{2^n-1}$ satisfies:

$$L_{\mathbf{a}} \geq \min(2^\gamma, t^{(1-\varepsilon)/\log 3})$$

for all but at most $O(t^{n-\varepsilon})$ vectors $\mathbf{a} \in (\mathbb{F}_t^*)^n$.

Proof. Take $\mathbf{a} = (a_0, \ldots, a_{n-1})$ and split it into the following two:

$$\begin{aligned} \mathbf{a}^- &= (a_0, \ldots, a_{s-1}), \\ \mathbf{a}^+ &= (a_s, \ldots, a_{n-1}), \end{aligned}$$

where

$$s = \min\left(\lfloor n/2 \rfloor, \left\lfloor \frac{1-\varepsilon}{\log 3} \log t \right\rfloor\right).$$

Let $\mathcal{A}$ be the set of vectors such that each of the Naor-Reingold sequences defined by vectors $\mathbf{a}^-$ and $\mathbf{a}^+$ generate at least 2^s distinct elements. Using Proposition 3, we have that $\#\mathcal{A} = (t-1)^n + O(t^{n-\varepsilon})$.

We show that the bound holds for any vector $\mathbf{a} \in \mathcal{A}$. Let us consider the set

$$\mathcal{K} := \{a_0^{x_0} \ldots a_{s-1}^{x_{s-1}} \mid x_0, \ldots, x_{s-1} \in \{0,1\}\} = \varphi_a[0, 2^s - 1].$$

The cardinality of this set is at least 2^s, for $\mathbf{a} \in \mathcal{A}$. By Lemma 1, there exists $r \in \mathbb{F}_t^*$ such that $L_r(\mathcal{K}, 2^{s-1}) \geq 2^{2s-1}/t$.

If $L_r(\mathcal{K}, 2^{s-1}) > L_{\mathbf{a}}$, we choose $d_1, \ldots, d_{L_{\mathbf{a}}} \in \mathcal{K}$ and such that $0 \leq y_i := d_i r < 2^{s-1}$. Let $0 \leq e_0 < \cdots < e_{L_{\mathbf{a}}} < 2^s$ be the integers such that $\varphi_{\mathbf{a}}(e_i) = d_i$. Using Lemma 2, we derive

$$\sum_{i=0}^{L_{\mathbf{a}}} c_i f_{\mathbf{a}}(2^s b + e_i) = 0$$

for $0 \leq b < 2^{n-s} - 1$. Let $m := r^{-1} \in \mathbb{F}_t^*$. We have that

$$f_{\mathbf{a}}(2^s b + e_i) = g^{m\varphi_a(2^s b) r \varphi_a(e_i)} = \left(g^{m\varphi_a(2^s b)}\right)^{y_i},$$

where $b = \sum_{j=0}^{n-s-1} b_j 2^j$. Now, as $\mathbf{a} \in \mathcal{A}$, the polynomial

$$F(X) = c_0 X^{y_0} + c_1 X^{y_1} + \ldots + c_{L_\mathbf{a}} X^{y_{L_\mathbf{a}}}$$

has at least $2^s - 1$ roots. This is impossible because $\deg F \leq 2^{s-1}$. Therefore, $L_\mathbf{a} \geq L_r(\mathcal{K}, 2^{s-1})$ and the result follows. ■

Acknowlegdments This article was partialled supported by the Spanish Ministry of Science, project MTM2007-67088.

References

[1] John Friedlander, Jan Hansen, and Igor Shparlinski. On character sums with exponential functions. *Mathematika*, 47:75–85, 2000.

[2] Frances Griffin and Igor Shparlinski. On the linear complexity of the naor-reingold pseudo-random function. *2nd Internatinal Conference on Information and Communication Security*, pages 301–308, 1999.

[3] Álvar Ibeas. On the period of the Naor-Reingold sequence. *Information Processing Letters*, 108(5):304–307, 2008.

[4] Moni Naor and Omer Reingold. Number-theoretic constructions of efficient pseudo-random functions. *J. ACM*, 51(2):231–262 (electronic), 2004.

[5] Igor Shparlinski. Linear complexity of the naor-reingold pseudo-random function. *Information Processing Letters*, 76:95–99, 2000.

[6] Igor Shparlinski. On the uniformity of distribution of the naor reingold pseudo-random function. *Finite Field and their Applications*, 7(2):318–326, 2001.

[7] John Silverman and Igor Shparlinski. On the linear complexity of the naor reingold pseudo-random function from elliptic curves. *Designs, Codes and Cryptography*, 24(3):279–289, 2001.