NON-LINEAR COMPLEXITY OF THE NAOR–REINGOLD PSEUDO-RANDOM FUNCTION

William D. Banks
Department of Mathematics, University of Missouri
Columbia, MO 65211, USA
bbanks@math.missouri.edu

Frances Griffin
Department of Mathematics, Macquarie University
Sydney, NSW 2109, Australia
fgriffin@ics.mq.edu.au

Daniel Lieman
Department of Mathematics, University of Missouri
Columbia, MO 65211, USA
lieman@math.missouri.edu

Igor E. Shparlinski
Department of Computing, Macquarie University
Sydney, NSW 2109, Australia
igor@mpce.mq.edu.au

Abstract

We obtain an exponential lower bound on the non-linear complexity of the new pseudo-random function introduced recently by M. Naor and O. Reingold. This bound extends and generalizes the lower bound on the linear complexity of this function that has recently been obtained by F. Griffin and I. E. Shparlinski.

1 Introduction

Let p and l be primes with $l|p-1$ and let $n \geq 1$ be an integer.

Denote by $\mathbb{F}_p$ the finite field of p elements which we identify with the set $\{0, \ldots, p-1\}$. Select an element $g \in \mathbb{F}_p^*$ of multiplicative order l, that is,

$$g^i \neq 1, \ 1 \leq i \leq l-1, \qquad g^l = 1.$$

Then for each n-dimensional vector $\mathbf{a} = (a_1, \ldots, a_n) \in (\mathbb{F}_l^*)^n$ one can define the function

$$f_{\mathbf{a}}(X) = g^{a_1^{x_1} \ldots a_n^{x_n}} \in \mathbb{F}_p,$$

where $X = x_1 \ldots x_n$ is the bit representation of an n-bit integer X, $0 \leq X \leq 2^n - 1$, with some extra leading zeros if necessary. Thus, given $\mathbf{a} = (a_1, \ldots, a_n) \in (\mathbb{F}_l^*)^n$, for each $X = 0, \ldots, 2^n - 1$ this function produces a certain element of $\mathbb{F}_p$. After that it can be continued periodically.

For a randomly chosen vector $\mathbf{a} \in (\mathbb{F}_l^*)^n$, M. Naor and O. Reingold [4] have proposed the function $f_{\mathbf{a}}(X)$ as an efficient pseudo-random function (it is assumed in [4] that n is the bit length of p but similar results hold in much more general settings).

It is shown in [4] that the function $f_{\mathbf{a}}(X)$ has some very desirable security properties, provided that certain standard cryptographic assumptions about the hardness of breaking the Diffie-Hellman cryptosystem hold. It is also shown in [4] that this function can be computed in parallel by threshold circuits of bounded depth and polynomial size.

The distribution properties of this function have been studied in [8] and it has been proved that the statistical distribution of $f_{\mathbf{a}}(X)$ is exponentially close to uniform for almost all $\mathbf{a} \in (\mathbb{F}_l^*)^n$.

For the elliptic curve version of this generator similar results have been obtained in [9].

The linear complexity, which is an important cryptographic characteristic of this sequence, has been estimated in [2].

Here we study the more general question of non-linear complexity.

Given an integer $d \geq 1$ and an N-element sequence $W_1, \ldots, W_N$ over a ring $\mathcal{R}$, we define the *degree d complexity*, $L(d)$, as the smallest number L such that there exists a polynomial $F(Z_1, \ldots, Z_L)$ over $\mathcal{R}$ of degree at most d in

L variables such that

$$W_{X+L} = F(W_X, \dots, W_{X+L-1}), \qquad X = 1, \dots, N-L.$$

The case $d = 1$ is closely related to the notion of the *linear complexity*, $\widetilde{L}$, the only distinction being that in the traditional definition of linear complexity only homogeneous linear polynomials are considered. However, this distinction is not very important since one can easily verify that $L(1) \le \widetilde{L} \le L(1) + 1$.

Linear complexity is an essential cryptographic characteristic that has been studied in many works, see [1, 3, 5, 6, 7]. Since non-linear complexity is harder to study, hence much less is known about this characteristic, even though it is of ultimate interest as well, see [1, 5].

In this paper we extend the method of [2] and obtain an exponential lower bound on the degree d complexity, $L_{\mathbf{a}}(d)$, of the sequence $f_{\mathbf{a}}(X)$, $X = 0, \dots, 2^n - 1$, which holds for almost all $\mathbf{a} \in (\mathbb{F}_l^*)^n$.

Throughout the paper, $\log z$ denotes the binary logarithm of z.

2 Preparations

We need some statements about the distribution in $\mathbb{F}_l^*$ of products of the form

$$\mathbf{b}^{\mathbf{z}} = b_1^{z_1} \dots b_m^{z_m}, \qquad \mathbf{z} = (z_1, \dots, z_m) \in \{0, 1\}^m,$$

which are of independent interest.

Denote

$$\mathbf{i} = (1, \dots, 1) \in \{0, 1\}^m,$$

so that $\mathbf{b}^{\mathbf{i}} = b_1 \dots b_m$.

Lemma 1 *For all but at most*

$$N_{m,d} \le \sum_{r=1}^{d} r \binom{2^m + r - 2}{r} (l-1)^{m-1}$$

vectors $\mathbf{b} = (b_1, \dots, b_m) \in (\mathbb{F}_l^*)^m$

$$\mathbf{b}^{\mathbf{z}_1} + \dots + \mathbf{b}^{\mathbf{z}_r} \ne \mathbf{b}^{\mathbf{i}},$$

for any choice of $r \le d$ *vectors* $\mathbf{z}_\nu \in \{0, 1\}^m$ *with* $\mathbf{z}_\nu \ne \mathbf{i}$, $\nu = 1, \dots, r$.

Proof. For all $r = 1, \dots, d$, let $\mathcal{Z}_r$ denote the set of all non-equivalent r-tuples $(\mathbf{z}_1, \dots, \mathbf{z}_r)$ with $\mathbf{z}_\nu \in \{0, 1\}^m$ and $\mathbf{z}_\nu \neq \mathbf{i}$, $\nu = 1, \dots, r$, where two r-tuples are considered to be equivalent if one is a permutation of the other.

The cardinality $\#\mathcal{Z}_r$ of this set is equal to the number of solutions of the equation

$$\sum_{k=1}^{2^m-1} n_k = r$$

in nonnegative integers $n_1, \dots, n_{2^m-1}$. Indeed, if we list the vectors

$$\mathbf{v}_k \in \{0, 1\}^m \backslash \{\mathbf{i}\}, \qquad k = 1, \dots, 2^m - 1,$$

then every r-tuple in $\mathcal{Z}_r$ is uniquely defined by the number of times n_k that the vector $\mathbf{v}_k$ occurs in the r-tuple.

Therefore

$$\#\mathcal{Z}_r = \binom{2^m + r - 2}{r}.$$

For each r-tuple $(\mathbf{z}_1, \dots, \mathbf{z}_r) \in \mathcal{Z}_r$ the number of solutions of the equation

$$\mathbf{b}^{\mathbf{z}_1} + \ldots + \mathbf{b}^{\mathbf{z}_r} = \mathbf{b}^{\mathbf{i}},$$

in $\mathbf{b} \in (\mathbb{F}_l^*)^m$ does not exceed $r(l-1)^{m-1}$. This can easily be proved for all $m \geq 1$ by induction in r.

It is convenient to start the induction with $r = 0$ where the statement is clearly true for all $m \geq 1$ (the equation $\mathbf{b}^{\mathbf{i}} = 0$ has no solutions).

Otherwise we select j such that the vector $\mathbf{z}_r$ has a zero jth component. This is always possible because $\mathbf{z}_r \neq \mathbf{i}$. Then the above equation can be written in the form $A = Bb_j$ where A and B do not depend on b_j. Because of our choice of j, we see that by induction, B vanishes for at most $(r-1)(l-1)^{m-2}$ vectors $(b_1, \dots, b_{j-1}, b_{j+1}, \dots, b_m) \in (\mathbb{F}_l^*)^{m-1}$ and in this case we have at most $l-1$ values for b_j. If $B \neq 0$ then for any vector $(b_1, \dots, b_{j-1}, b_{j+1}, \dots, b_m) \in (\mathbb{F}_l^*)^{m-1}$ the value of b_j is defined uniquely. Therefore the number of solutions does not exceed $(r-1)(l-1)^{m-1} + (l-1)^{m-1} = r(l-1)^{m-1}$. This completes the induction. Accordingly,

$$N_{m,d} \leq \sum_{r=1}^{d} r(l-1)^{m-1} \#\mathcal{Z}_r$$

and the bound follows. $\square$

$\square$

We also need the following Lemma 2 of [2] which shows that for large m, the products $\mathbf{b}^{\mathbf{z}}$ with $\mathbf{z} \in \{0,\, 1\}^m$ are quite dense in $\mathbb{F}_l^*$.

Lemma 2 *Fix an arbitrary $\Delta > 0$. Then for all but at most*

$$M_m \le 2^{-m}\Delta^{-1}(l-1)^{m+2}$$

vectors $\mathbf{b} = (b_1, \dots b_m) \in (\mathbb{F}_l^*)^m$, the 2^m products $\mathbf{b}^{\mathbf{z}}$, $\mathbf{z} \in \{0,\, 1\}^m$ take at least $l-1-\Delta$ values from $\mathbb{F}_l^*$.

3 Lower Bound of the Degree d Complexity

Now we are prepared to prove our main result.

Theorem 3 *Assume that for some $\gamma > 0$*

$$n \ge (1+\gamma)\log l.$$

Then for any integer $d \ge 1$ and any $\delta > 0$ the degree d complexity, $L_{\mathbf{a}}(d)$, of the sequence $f_{\mathbf{a}}(X)$, $X = 0, \dots, 2^n-1$, satisfies

$$L_{\mathbf{a}}(d) \ge \begin{cases} 0.5(l-1)^{1/d-\delta/d}, & \text{if } \gamma \ge 1+1/d; \\ 0.5(l-1)^{\gamma/(d+1)-\delta/d}, & \text{if } \gamma < 1+1/d; \end{cases}$$

for all but at most

$$N \le \left(\frac{d+1}{d!} + o(1)\right)(l-1)^{n-\delta}, \qquad l \to \infty,$$

vectors $\mathbf{a} \in (\mathbb{F}_l^*)^n$

Proof. If $\delta \ge \max\{1, \gamma\}$ then the bound is trivial. Otherwise we put

$$t = \left\lfloor \min\left\{\frac{1-\delta}{d}, \frac{\gamma-\delta}{d+1}\right\}\log(l-1)\right\rfloor, \qquad s = n-t$$

and

$$\Delta = \left\lceil (l-1) \left(\binom{2^t+d}{d} + 1 \right)^{-1} \right\rceil - 1.$$

Therefore

$$2^{-s} = 2^{t-n} \le 2^t l^{-1-\gamma}. \tag{1}$$

¿From the inequality $2^{td} \le (l-1)^{1-\delta}$ we see that

$$\sum_{r=1}^{d} r \binom{2^t+r-2}{r} \le \left(\frac{1}{(d-1)!} + o(1) \right) (l-1)^{1-\delta} \tag{2}$$

We also have

$$2^{t(d+1)} \le (l-1)^{\gamma-\delta} \qquad \text{and} \qquad \Delta \ge (d! + o(1))\,(l-1)2^{-td}. \tag{3}$$

From Lemmas 1 and 2 and the bounds (1), (2) and (3) we derive

$$\begin{aligned} N_{t,d} &\le \sum_{r=1}^{d} r \binom{2^t+r-2}{r} (l-1)^{t-1} \\ &\le \left(\frac{1}{(d-1)!} + o(1) \right) (l-1)^{t-\delta} \end{aligned}$$

and

$$\begin{aligned} M_s &\le 2^{-s} \Delta^{-1} (l-1)^{s+2} \le \left(\frac{1}{d!} + o(1) \right) 2^{t(d+1)} (l-1)^{s-\gamma} \\ &\le \left(\frac{1}{d!} + o(1) \right) (l-1)^{s-\delta}. \end{aligned}$$

Let $\mathcal{A}$ be the set of vectors $\mathbf{a} \in (\mathbb{F}_l^*)^n$ such that simultaneously

$$\# \{ a_1^{y_1} \dots a_s^{y_s} \mid (y_1, \dots, y_s) \in \{0,1\}^s \} \ge l - 1 - \Delta$$

and

$$\sum_{\nu=1}^{d} a_{s+1}^{k_{1,\nu}} \dots a_n^{k_{t,\nu}} \ne a_{s+1} \dots a_n$$

for any $(k_{1,\nu}, \dots, k_{t,\nu}) \in \{0,1\}^t$ with $(k_{1,\nu}, \dots, k_{t,\nu}) \ne (1, \dots, 1)$.

Then, from the above inequalities, we derive

$$\begin{aligned}
\#\mathcal{A} &\geq (l-1)^n - N_{t,d}(l-1)^{n-t} - M_s(l-1)^{n-s} \\
&\geq (l-1)^n - \left(\frac{1}{(d-1)!} + o(1)\right)(l-1)^{n-\delta} \\
&\qquad\qquad\qquad\qquad - \left(\frac{1}{d!} + o(1)\right)(l-1)^{n-\delta} \\
&= (l-1)^n - \left(\frac{d+1}{d!} + o(1)\right)(l-1)^{n-\delta}.
\end{aligned}$$

We show that the lower bound of the theorem holds for any $\mathbf{a} \in \mathcal{A}$, thus from $N \leq (l-1)^n - \#\mathcal{A}$ and the above inequality we obtain the desired upper bound on N.

Let us fix $\mathbf{a} \in \mathcal{A}$. Assume that $L_{\mathbf{a}}(d) \leq 2^t - 1$. Then there exists a polynomial

$$F(Z_1, \ldots, Z_{2^t-1}) \in \mathbb{F}_p\left[Z_1, \ldots, Z_{2^t-1}\right],$$

such that

$$F\left(f_{\mathbf{a}}(X), \ldots, f_{\mathbf{a}}(X + 2^t - 2)\right) = f_{\mathbf{a}}\left(X + 2^t - 1\right)$$

for all $X = 0, \ldots, 2^n - 2^t$.

Now suppose $X = 2^t Y$, where $Y = y_1 \ldots y_s$ is an s-bit integer, and let $K = k_1 \ldots k_t$ be a t-bit integer. We remark that the bits of K form the rightmost bits of the sum $X + K$. Then we have

$$f_{\mathbf{a}}(2^t Y + K) = g^{a_1^{y_1} \ldots a_s^{y_s} e_K}, \qquad Y = 0, \ldots, 2^s - 1,$$

where

$$e_K = a_{s+1}^{k_1} \ldots a_n^{k_t}, \qquad K = 0, \ldots, 2^t - 1,$$

and $K = k_1 \ldots k_t$ is the bit expansion of K.

Denote by $\Phi_{\mathbf{a}}(u)$ the following exponential polynomial

$$\Phi_{\mathbf{a}}(u) = F\left(g_0^u, \ldots, g_{2^t-2}^u\right) - g_{2^t-1}^u, \qquad u \in \mathbb{F}_l,$$

where

$$g_K = g^{e_K}, \qquad K = 0, \ldots, 2^t - 1.$$

Collecting together terms with equal values of exponents and taking into account that, because of the choice of the set $\mathcal{A}$, the value of

$$g_{2^t-1} = g^{a_{s+1}\dots a_n}$$

is unique, we obtain that $\Phi_{\mathbf{a}}(u)$ can be expressed in the form

$$\Phi_{\mathbf{a}}(u) = \sum_{\nu=1}^{R} C_\nu h_\nu^u,$$

where

$$1 \le R \le \binom{2^t+d-1}{d} + 1,$$

with some coefficients $C_\nu \in \mathbb{F}_p^*$ and pairwise distinct $h_\nu \in \mathbb{F}_p^*$, $\nu = 1, \dots, R$. Recalling that $\mathbf{a} \in \mathcal{A}$, we conclude that $\Phi_{\mathbf{a}}(u) \ne 0$ for at most Δ values of $u = 1, \dots, l-1$. On the other hand, from the properties of Vandermonde determinants, it is easy to see that for any $u = 1, \dots, l-1$, $\Phi_{\mathbf{a}}(u+v) \ne 0$ for at least one $v = 0, \dots, R-1$. Therefore, $\Phi_{\mathbf{a}}(u) \ne 0$ for at least

$$(l-1)/R \ge (l-1)\left(\binom{2^t+d-1}{d}+1\right)^{-1} > \Delta$$

values of $u = 1, \dots, l-1$, which is not possible because of the choice of $\mathcal{A}$. The obtained contradiction implies that $L_{\mathbf{a}}(d) \ge 2^t$. □

□

4 Remarks

It is useful to recall that typically the bit length of p and l are of the same order as n. Thus

$$\log p \asymp \log l \asymp n.$$

In the most interesting case n is the bit length of p, that is, $n \sim \log p$. In this case Theorem 3 implies a lower bound on $L_{\mathbf{a}}(d)$ which is exponential in n, if $l \le p^{1-\varepsilon}$ for some $\varepsilon > 0$. On the other hand, it would be interesting to estimate the linear and higher degree complexity for all values of $l \le p$.

It is also an interesting open question to study the linear complexity or higher degree complexity of single bits of $f_{\mathbf{a}}(X)$. For example, one can form the sequence $\beta_{\mathbf{a}}(X)$ of the rightmost bits of $f_{\mathbf{a}}(X)$, $X = 0, \ldots, 2^n - 1$, and study its linear and higher degree complexity (as elements of $\mathbb{F}_2$). Unfortunately we do not see any approaches to this question.

Acknowledgment. This work has been motivated by a question asked during a seminar talk about the results of [2, 8, 9] given by I.S. at the Centre for Applied Cryptographic Research at the University of Waterloo, whose hospitality is gratefully acknowledged.

References

[1] T. W. Cusick, C. Ding and A. Renvall, *Stream Ciphers and Number Theory*, Elsevier, Amsterdam, 1998.

[2] F. Griffin and I. E. Shparlinski, 'On the linear complexity of the Naor-Reingold pseudo-random number generator', *Proc. 2nd Intern. Conf. on Information and Communication Security (ICICS'99), Sydney*, Lect. Notes in Comp. Sci., v.1726, Springer-Verlag, Berlin, 1999.

[3] A. J. Menezes, P. C. van Oorschot and S. A. Vanstone, *Handbook of Cryptography*, CRC Press, Boca Raton, FL, 1996.

[4] M. Naor and O. Reingold, 'Number-theoretic constructions of efficient pseudo-random functions', *Proc. 38th IEEE Symp. on Foundations of Comp. Sci. (FOCS'97), Miami Beach*, IEEE, 1997, 458–467.

[5] H. Niederreiter, 'Some computable complexity measures for binary sequences', *Proc. Intern. Conf. on Sequences and their Applications (SETA'98), Singapore*, C. Ding, T. Helleseth and H. Niederreiter (Eds.), Springer-Verlag, London, 1999, 67–78.

[6] H. Niederreiter and M. Vielhaber, 'Linear complexity profiles: Hausdorff dimension for almost perfect profiles and measures for general profiles', *J. Compl.*, **13** (1996), 353–383.

[7] R. A. Rueppel, 'Stream ciphers', *Contemporary Cryptology: The Science of Information Integrity*, IEEE Press, NY, 1992, 65–134.

[8] I. E. Shparlinski, 'On the uniformity of distribution of the Naor–Reingold pseudo-random function', *Preprint*, 1999, 1–11.

[9] I. E. Shparlinski, 'On the Naor–Reingold pseudo-random number generator from elliptic curves', *Preprint*, 1999, 1–9.