On the fractal nature of the set of all binary sequences with almost perfect linear complexity profile

H. Niederreiter and M. Vielhaber
Institute for Information Processing
Austrian Academy of Sciences
Sonnenfelsgasse 19, A–1010 Vienna, Austria
E-mail: nied@qiinfo.oeaw.ac.at

Abstract

Stream ciphers usually employ some sort of pseudo-randomly generated bit strings to be added to the plaintext. The cryptographic properties of such binary sequences can be stated in terms of the so-called linear complexity profile. This paper shows that the set of all sequences with an almost perfect linear complexity profile maps onto a fractal subset of $[0,1]$.

The space $\mathbb{F}_2^\infty$ of all infinite binary sequences can be mapped onto $[0,1]$ by $\iota : (a_i)_{i=1}^\infty \mapsto \sum_{i=1}^\infty a_i 2^{-i}$. Any such sequence admits a linear complexity profile (l.c.p.) $(L_i)_{i=1}^\infty$, stating for each n that the initial string $(a_1, \ldots, a_n)$ can be produced by an LFSR of length L_n (but not $L_n - 1$). Usually $L_n \approx n/2$, and so $m(n) := 2 \cdot L_n - n$ should vary around zero.

Let $\mathcal{A}_d$ be the set of those sequences from $\mathbb{F}_2^\infty$ whose l.c.p. is almost perfect in the sense of $|m(n)| \leq d$, $\forall n$ (Niederreiter, 1988a). The subset of $[0,1]$ obtained as $\iota(\mathcal{A}_d)$ is fractal and its Hausdorff dimension is bounded from above by

$$D_H(\iota(\mathcal{A}_d)) \leq \frac{1 + \log_2 \varphi_d}{2},$$

where φ_d is the positive real root of $x^d = \sum_{i=0}^{d-1} x^i$, e.g. $\varphi_1 = 1, \varphi_2 = 1.618\ldots$ (Fibonacci's golden ratio). Thus, although all the $\mathcal{A}_d$ have Haar measure zero in $\mathbb{F}_2^\infty$, a sharper distinction can be made by looking at their Hausdorff dimension. As a by-product the paper gives explicit formulae for the number of sequences of length n in $\mathcal{A}_d$, for all n and d.

Keywords

Linear complexity, Hausdorff dimension

R. Posch (ed.), *Communications and Multimedia Security*

1 INTRODUCTION

In the space $\mathbb{F}_2^\infty$ of all infinite binary sequences, to each sequence can be assigned a linear complexity profile (Rueppel, 1986), $\mathbb{F}_2^\infty \ni (a_i)_{i=1}^\infty \mapsto (L_i)_{i=1}^\infty \in \mathbb{N}_0{}^\infty$.

The number L_n is the length of a shortest linear feedback shift register that produces the initial string $(a_1, \dots, a_n)$. Generally $0 \le L_n \le n$ and $L_n \le L_{n+1}, \forall n$. As typically L_n is close to $n/2$, it has merits to introduce the following concept.

DEFINITION 1. Let $\underline{a} = (a_i)_{i=1}^N$, $N \in \mathbb{N} \cup \{\infty\}$, be a given binary sequence, $(L_i)_{i=1}^N$ its linear complexity profile (l.c.p.), then the *linear complexity deviation* of $\underline{a}$ at n is defined as

$$m_{\underline{a}}(n) := 2 \cdot L_n - n.$$

The l.c.p. can be computed by the Berlekamp-Massey algorithm (Lidl and Niederreiter, 1994). The following result describes the dynamic behaviour of L_n and $m_{\underline{a}}(n)$.

PROPOSITION 1.

1. *If* $L_n > n/2$, *then* $L_{n+1} = L_n$.
2. *If* $L_n \le n/2$, *then*
$$\exists_1 a \in \mathbb{F}_2 : \quad L_{n+1}(a_1, \dots, a_n, a) = L_n,$$
$$\forall b \neq a : \quad L_{n+1}(a_1, \dots, a_n, b) = n + 1 - L_n.$$
3. *If* $m_{\underline{a}}(n) > 0$, *then* $m_{\underline{a}}(n+1) = m_{\underline{a}}(n) - 1$.
4. *If* $m_{\underline{a}}(n) \le 0$, *then*
$$\exists_1 a \in \mathbb{F}_2 : \quad m_{(a_1,\dots,a_n,a)}(n+1) = m_{(a_1,\dots,a_n)}(n) - 1,$$
$$\forall b \neq a : \quad m_{(a_1,\dots,a_n,b)}(n+1) = 1 - m_{(a_1,\dots,a_n)}(n).$$

PROOF.

1., 2. See Rueppel (1986, p.34).

3. $m_{\underline{a}}(n+1) = 2 \cdot L_{n+1} - n - 1 = (2 \cdot L_n - n) - 1 = m_{\underline{a}}(n) - 1$ by 1.

4. $\exists_1 a$: see 3.
$$\forall b \neq a : m_{\underline{a}}(n+1) = 2 \cdot L_{n+1} - n - 1 = 2 \cdot (n + 1 - L_n) - n - 1$$
$$= 1 - m_{\underline{a}}(n) \text{ by 2.}$$
□

Niederreiter (1988a) and Dai (1989) have shown the intimate connection between the l.c.p. of $(a_i)_{i=1}^\infty$ and the continued fraction expansion of $\sum_{i=1}^\infty a_i x^{-i}$ in the field of formal Laurent series. Thus, a jump by k in the l.c.p. is equivalent to a partial quotient of degree k in the continued fraction expansion.

Rueppel (1986, p.45) introduced the notion of a perfect linear complexity profile, given when the l.c.p. always jumps by 1 only or, stated in continued fraction terms, when the partial quotients all have degree 1. Niederreiter extended this to almost perfect linear complexity profiles: given a fixed number $d \in \mathbb{N}$, every jump must have height $\leq d$. He showed in (Niederreiter, 1988b) that for any d the set of sequences with partial quotients whose degrees do not exceed d has Haar measure 0.

DEFINITION 2. Let $\mathcal{A}_d \subset \mathbb{F}_2^\infty$ be the set of all sequences $\underline{a}$ with $|m_{\underline{a}}(n)| \leq d$ for all n.

2 TRANSLATION THEOREM

As a simple consequence of Proposition 1 we obtain the following translation theorem.

THEOREM 1. *Let $\underline{\alpha} = (\alpha_1, \ldots, \alpha_k)$ and $\underline{\beta} = (\beta_1, \ldots, \beta_l)$ be given binary strings with $m_{\underline{\alpha}}(k) = m_{\underline{\beta}}(l)$. For any length t and deviation d, we have*

$$\#\{\underline{a} \in \mathbb{F}_2^{k+t} \mid a_i = \alpha_i,\ i \leq k,\ m_{\underline{a}}(k+t) = d\} =$$
$$\#\{\underline{b} \in \mathbb{F}_2^{l+t} \mid b_i = \beta_i,\ i \leq l,\ m_{\underline{b}}(l+t) = d\}.$$

In other words: the distribution of l.c. deviations m on all suffixes of a given finite initial string depends only on m at the end of that string, but not on the length or the elements of the initial string.

PROOF. Induction on t starts for $t = 0$ with both cardinalities being 1 for $d = m_{\underline{\alpha}}(k)$ and 0 otherwise by assumption. The step $t \to t+1$ follows by Proposition 1(3,4). □

3 SOME COUNTING FORMULAE

DEFINITION 3. For $t \in \mathbb{N}_0, d \in \mathbb{N}, m \in \mathbb{Z}$ define $A^{(t)}_{m|d}$ as the number of sequences $\underline{a}$ of length t with $m_{\underline{a}}(t) = m$ and $|m_{\underline{a}}(\tau)| \leq d$ for $1 \leq \tau \leq t$. For $t = 0$ set $A^{(0)}_{0|d} = 1$ (the empty sequence ε) and $A^{(0)}_{m|d} = 0$ for $m \neq 0$.

THEOREM 2.

1. $A^{(t+1)}_{m|d} = A^{(t)}_{m+1|d}$ *for* $-d \le m < 0$.
2. $A^{(t+1)}_{0|d} = 2 \cdot A^{(t)}_{1|d}$.
3. $A^{(t+1)}_{m|d} = 2 \cdot A^{(t)}_{m+1|d} + A^{(t)}_{-m+1|d}$ *for* $0 < m \le d-1$.
4. $A^{(t+1)}_{d|d} = A^{(t)}_{-d+1|d}$.
5. $A^{(t)}_{m|d} = 0$ *for* $|m| > d$.
6. $A^{(t)}_{m|d} = 0$ *for* $m \not\equiv t(2)$.

PROOF.

1., 2. and 3. follow by Proposition 1.
4. By Proposition 1, considering $A^{(t)}_{d+1|d} = 0$.
5. By the definition of $A^{(t)}_{m|d}$.
6. By the definition of $m_{\underline{a}}(t)$. □

THEOREM 3. *Every* $A^{(t)}_{m|d}$ *can be expressed in terms of* $A^{(t-\tau)}_{0|d}$ *as follows:*

1. $A^{(t)}_{m|d} = A^{(t+m)}_{0|d}$ *for* $-d \le m \le 0$.
2. $A^{(t)}_{d|d} = A^{(t-d)}_{0|d}$.
3. $A^{(t)}_{m|d} = \sum_{k=0}^{d-m} 2^k \cdot A^{(t-m-2k)}_{0|d}$ *for* $1 \le m \le d-1$.

PROOF.

1. This follows by induction from Theorem 2(1).
2. From 1. and $A^{(t)}_{d|d} = A^{(t)}_{-d|d}, \forall t$ (by Theorem 2(1,4)).
3. $A^{(t)}_{m|d} = 2 \cdot A^{(t-1)}_{m+1|d} + A^{(t-1)}_{-m+1|d}$, $1 \le m \le d-1$, by Theorem 2(3),

$= 2^k \cdot A^{(t-k)}_{m+k|d} + \sum_{i=1}^{k} 2^{i-1} \cdot A^{(t-i)}_{-m-i+2|d}$ holds for $k = 1, \dots, d-m$ by induction and Theorem 2(3),

$= 2^{d-m} \cdot A^{(t-d+m)}_{d|d} + \sum_{i=0}^{d-m-1} 2^i \cdot A^{(t-m-2i)}_{0|d}$ by part 1,

$= \sum_{i=0}^{d-m} 2^i \cdot A^{(t-m-2i)}_{0|d}$ by part 2. □

DEFINITION 4. For $d \in \mathbb{N}$ and $t \in \mathbf{Z}$ we define generalized Fibonacci numbers by

$$\mathrm{Fib}_d(t) = \begin{cases} 0, & t < 0, \\ 1, & t = 0, \\ \sum_{k=1}^{d} \mathrm{Fib}_d(t-k), & t > 0. \end{cases}$$

DEFINITION 5. Let $O^{(t)}_d := 2 \cdot A^{(t-1)}_{-d|d} = 2 \cdot A^{(t-d-1)}_{0|d}$ be the number of sequences leaving the bound $|m| \le d$ at time t by leading to $m(t) = d+1$ or $m(t) = -d-1$.

THEOREM 4.

1. $A^{(t)}_{0|d} = \sum_{i=1}^{d} 2^i \cdot A^{(t-2i)}_{0|d}$ *for* $t \geq 2d$.

2. $A^{(2t)}_{0|d} = 2^t \cdot \mathrm{Fib}_d(t)$.

3. $O^{(t)}_d = \begin{cases} 0, & t \equiv d(2), \\ 2^{\frac{t-d+1}{2}} \cdot \mathrm{Fib}_d(\frac{t-d-1}{2}), & t \not\equiv d(2). \end{cases}$

PROOF.

1. $$\begin{aligned} A^{(t)}_{0|d} &= 2 \cdot A^{(t-1)}_{1|d} && \text{by Theorem 2(2),} \\ &= 2 \cdot \textstyle\sum_{i=0}^{d-1} 2^i \cdot A^{(t-1-1-2i)}_{0|d} && \text{by Theorem 3(2,3),} \\ &= \textstyle\sum_{i=1}^{d} 2^i \cdot A^{(t-2i)}_{0|d}. \end{aligned}$$

2. By induction on t: $A^{(0)}_{0|d} = 1$ by definition, and for $1 \leq k \leq d$ one has

$$\begin{aligned} A^{(2k)}_{0|d} &= A^{(2k)}_{0|\infty} && \text{(the bound } d \text{ has no effect for } 2k \leq 2d\text{),} \\ &= 2^{2k-1} && \text{(by the counting result of Rueppel (1986, p.36)),} \\ &= 2^k \cdot \mathrm{Fib}_d(k). \end{aligned}$$

For $k \geq d$:

$$\begin{aligned} A^{(2k)}_{0|d} &= \textstyle\sum_{i=1}^{d} 2^i \cdot A^{(2k-2i)}_{0|d} && \text{by part 1,} \\ &= \textstyle\sum_{i=1}^{d} 2^i \cdot 2^{k-i} \cdot \mathrm{Fib}_d(k-i) && \text{by the induction hypothesis,} \\ &= 2^k \cdot \textstyle\sum_{i=1}^{d} \mathrm{Fib}_d(k-i) \\ &= 2^k \cdot \mathrm{Fib}_d(k). \end{aligned}$$

3. Apply part 2 to the definition. □

The combination of Theorems 3 and 4 leads to the following general formula for $A^{(t)}_{m|d}$.

THEOREM 5.

$$A^{(t)}_{m|d} = \begin{cases} 0, & |m| > d \text{ or } t \not\equiv m(2), \\ 2^{\frac{t+m}{2}} \cdot \mathrm{Fib}_d(\frac{t+m}{2}), & -d \leq m \leq 0, \quad t \equiv m(2), \\ 2^{\frac{t-m}{2}} \cdot \sum_{k=0}^{d-m} \mathrm{Fib}_d(\frac{t-m}{2} - k), & 1 \leq m \leq d, \quad t \equiv m(2). \end{cases}$$

Example. Let $d = 3$, then we get as $A^{(t)}_{m|d}$:

m	t=0	1	2	3	4	5	6	7	8	9	10	11	12	13
3				1		2		8		32		112		416
2			1		4		12		48		176		640	
1		1		4		16		56		208		768		2816
0	1		2		8		32		112		416		1536	
−1		1		2		8		32		112		416		1536
−2			1		2		8		32		112		416	
−3				1		2		8		32		112		416

DEFINITION 6. Let $A^{(t)}_{*|d} := \sum_{m=-d}^{d} A^{(t)}_{m|d}$ be the overall number of d–bound sequences of length t.

THEOREM 6.

$A^{(t)}_{*|d} = 2^{\lfloor (t-d)/2 \rfloor + 1} \cdot \mathrm{Fib}_d(\lfloor (t+d+1)/2 \rfloor)$.

PROOF. $t = 0, \ldots, d$:

$$\begin{aligned} A^{(t)}_{*|d} &= 2^t \\ &= 2^{\lfloor (t-d)/2 \rfloor + 1} \cdot 2^{\lfloor (t+d+1)/2 \rfloor - 1} \\ &= 2^{\lfloor (t-d)/2 \rfloor + 1} \cdot \mathrm{Fib}_d(\lfloor (t+d+1)/2 \rfloor). \end{aligned}$$

$t \to t+1$:

a) $t \equiv d(2)$:

Then $O^{(t+1)}_d = 2^{\frac{t-d+2}{2}} \cdot \mathrm{Fib}_d(\frac{t-d}{2})$ by Theorem 4(3), and thus

$$\begin{aligned} A^{(t+1)}_{*|d} &= 2 \cdot A^{(t)}_{*|d} - O^{(t+1)}_d \\ &= 2^{(t-d)/2+2} \cdot \mathrm{Fib}_d(\tfrac{t+d}{2}) - 2^{(t-d)/2+1} \cdot \mathrm{Fib}_d(\tfrac{t-d}{2}) \\ &= 2^{(t-d)/2+1} \cdot (\mathrm{Fib}_d(\tfrac{t+d}{2}) + \textstyle\sum_{i=1}^{d} \mathrm{Fib}_d(\tfrac{t+d}{2} - i) - \mathrm{Fib}_d(\tfrac{t-d}{2})) \\ &= 2^{(t-d)/2+1} \cdot \textstyle\sum_{i=1}^{d} \mathrm{Fib}_d(\tfrac{t+d}{2} + 1 - i) \\ &= 2^{(t-d)/2+1} \cdot \mathrm{Fib}_d(\tfrac{t+d}{2} + 1). \end{aligned}$$

b) $t \not\equiv d(2)$:

Then $O^{(t+1)}_d = 0$ by Theorem 4(3), and thus

$$\begin{aligned} A^{(t+1)}_{*|d} &= 2 \cdot A^{(t)}_{*|d} \\ &= 2^{\lfloor (t-d)/2 \rfloor + 2} \cdot \mathrm{Fib}_d(\lfloor (t+d+1)/2 \rfloor) \\ &= 2^{\lfloor (t+1-d)/2 \rfloor + 1} \cdot \mathrm{Fib}_d(\lfloor (t+d+2)/2 \rfloor). \square \end{aligned}$$

PROPOSITION 2. *Let φ_d be the positive real root of $x^d = \sum_{i=0}^{d-1} x^i$. Then*

$\mathrm{Fib}_d(t) \le {\varphi_d}^t$ *for all* $t \in \mathbf{Z}$.

PROOF.

This is shown by induction on t, with the case $t \le 0$ being trivial. □

It is clear that we always have $1 \le \varphi_d < 2$. Typical values are $\varphi_1 = 1$ and $\varphi_2 = (1+\sqrt{5})/2 = 1.618...$ (Fibonacci's golden ratio).

4 HAUSDORFF DIMENSION

We follow the introduction of the Hausdorff dimension given by Peitgen et al. (1992) for a subset $\mathcal{A}$ of the reals.

Set $h^s_\varepsilon(\mathcal{A}) = \inf\{\sum_{i=1}^{\infty} \mathrm{diam}(U_i)^s \mid \mathcal{U} = \{U_1, U_2, \ldots\},\ \mathrm{diam}(U_i) < \varepsilon\}$ for $s \ge 0$, $\varepsilon > 0$, where the infimum runs over all open covers $\mathcal{U}$ of $\mathcal{A}$, and letting $\varepsilon \to 0$: $h^s(\mathcal{A}) := \lim_{\varepsilon\to 0} h^s_\varepsilon(\mathcal{A})$.

Then $h^s(\mathcal{A}) = \begin{cases} \infty, & s < D_H(\mathcal{A}) \\ 0, & s > D_H(\mathcal{A}) \end{cases}$ for a certain real number $D_H(\mathcal{A})$.

DEFINITION 7. The Hausdorff dimension of a set $\mathcal{A}$ is defined as

$$\begin{aligned} D_H(\mathcal{A}) &= \inf\ \{s \mid h^s(\mathcal{A}) = 0\} \\ &= \sup\ \{s \mid h^s(\mathcal{A}) = \infty\}. \end{aligned}$$

($h^{D_H(\mathcal{A})}(\mathcal{A})$ may assume any value in $[0, \infty]$.)

5 THE MAIN RESULT

The space $\mathbb{F}_2^\infty$ of all infinite binary sequences can be mapped onto the interval $[0,1]$ by $\iota : (a_i)_{i=1}^{\infty} \mapsto \sum_{i=1}^{\infty} a_i 2^{-i}$. If $\mathcal{A}_d \subset \mathbb{F}_2^\infty$ is the set in Definition 2, then we study the subset $\mathcal{B}_d := \iota(\mathcal{A}_d)$ of $[0,1]$.

THEOREM 7.

$$D_H(\mathcal{B}_d) \le \frac{1 + \log_2 \varphi_d}{2}.$$

PROOF.

For fixed $t \ge 1$, consider the set of all initial strings $\underline{a}$ of length t with $|m_{\underline{a}}(n)| \le d$ for $1 \le n \le t$. The cardinality of this set is $A^{(t)}_{*|d}$. By Theorem 6 and Proposition 2 we have

$$A^{(t)}_{*|d} \le 2^{(t-d)/2+1} \cdot \varphi_d{}^{\lfloor (t+d+1)/2 \rfloor} \le C \cdot (2 \cdot \varphi_d)^{\frac{t}{2}}$$

with a constant $C > 0$ depending only on d.

Each initial string $\underline{a}$ of length t defines a cylinder set in $\mathbb{F}_2^\infty$ consisting of all infinite continuations of this string. The image of each such cylinder set under the function ι is a closed interval of length 2^{-t} in $[0,1]$. Thus, $\mathcal{B}_d$ can be covered by $A^{(t)}_{*|d}$ open intervals of length less than 2^{-t+1}. With $\varepsilon_t = 2^{-t+1}$ it follows that

$$h^s_{\varepsilon_t}(\mathcal{B}_d) \le A^{(t)}_{*|d} \cdot 2^{(-t+1)s} \le 2^s C \cdot \left(\frac{\sqrt{2\varphi_d}}{2^s}\right)^t.$$

For any $s > \frac{1}{2}(1 + \log_2 \varphi_d)$ we have $2^s > \sqrt{2\varphi_d}$. Thus, letting $t \to \infty$ (hence $\varepsilon_t \to 0$), we get

$$h^s(\mathcal{B}_d) = 0.$$

By the definition of $D_H(\mathcal{B}_d)$ it follows that $D_H(\mathcal{B}_d) < s$. Since $s > \frac{1}{2}(1 + \log_2 \varphi_d)$ is arbitrary, we obtain

$$D_H(\mathcal{B}_d) \leq \frac{1}{2}(1 + \log_2 \varphi_d).$$

□

6 REFERENCES

Dai, Z.-D. (1989) Continued fractions and the Berlekamp–Massey algorithm, E.I.S.S. Report # 89/7, Europäisches Institut für Systemsicherheit, Karlsruhe.

Lidl, R. and Niederreiter, H. (1994) *Introduction to Finite Fields and Their Applications*, revised ed., Cambridge University Press, Cambridge.

Niederreiter, H. (1988a) Sequences with almost perfect linear complexity profile, in *Advances in Cryptology – EUROCRYPT '87* (eds. D. Chaum, W.L. Price), LNCS 304, 37–51, Springer, Berlin.

Niederreiter, H. (1988b) The probabilistic theory of linear complexity, in *Advances in Cryptology – EUROCRYPT '88* (ed. C.G. Günther), LNCS 330, 191–209, Springer, Berlin.

Peitgen, H.O., Jürgens, H. and Saupe, D. (1992) *Chaos and Fractals — New Frontiers of Science*, Springer, New York, Berlin.

Rueppel, R.A. (1986) *Analysis and Design of Stream Ciphers*, Springer, Berlin.

7 BIOGRAPHY

Harald Niederreiter received his Ph.D. in mathematics at the University of Vienna in 1969. He has held research and teaching positions in the United States and visiting positions in Australia, France, and Germany. He is currently director of the Institute for Information Processing at the Austrian Academy of Sciences in Vienna. He serves on the editorial boards of 9 journals, including *Mathematics of Computation, Applicable Algebra, Acta Arithmetica*, and *Finite Fields and Their Applications.* His research interests are cryptology, number theory, applied algebra, and numerical analysis.

Michael Vielhaber received his diploma in computer science at the University of Karlsruhe (T. H.) in 1988. His research interests are cryptology and theoretical computer science.