Construction of Pseudorandom Binary Sequences using Additive Characters over GF(2^k) II.

JÁNOS FOLLÁTH [1]
University of Debrecen - Department of Informatics
4032 Debrecen, Egyetem tér 1.
e-mail: follath.janos@inf.unideb.hu

In a series of papers Mauduit and Sárközy introduced measures of pseudorandomness and they constructed large families of sequences with strong pseudorandom properties. In later papers the structure of families of binary sequences was also studied. In these constructions fields with prime order were used. Throughout this paper the structure of a family of binary sequences based on GF(2^k) will be studied.

Categories and Subject Descriptors: E.3 [**Data**]: Data Encryption; G.2.0 [**Mathematics of Computing**]: Discrete Mathematics—*General*; G.3 [**Mathematics of Computing**]: Probability and Statistics

General Terms: Security,Design

Additional Key Words and Phrases: Binary sequence, Character sums, Normality measure, Pseudorandom

1. INTRODUCTION

Goubin, Mauduit and Sárközy presented a new, large family of pseudorandom binary sequences based on the Legende-symbol [Goubin et al. 2004]. This construction was an extension of the sequence studied in [Mauduit and Sárközy 1997]. Although its security cannot be proved by reduction, many mathematical arguments substantiate it: it possesses the strict avalanche property [Tóth 2007], has a high family complexity [Ahlswede et al. 2003] and a subfamily having small f-correlation was also found [Gyarmati 2009], the computational complexity of the best known attacks is high and it has an extremly fast implementation [Hoffstein and Lieman 2001] and the bound on its correlation measure enables one to estimate its linear complexity profile [Brandstätter and Winterhof 2006][Andics 2005]. Accordingly this generator has mathematically substantiated security and still is faster than most provable secure pseudorandom sequence generators.

The above mentioned construction is based on a multiplicative character over prime fields. In [Folláth] a similar construction based on additive characters over fields of characteristic 2 was presented and its pseudorandom measures were studied. Throughout this paper the avalanche property, family complexity and linear complexity of this generator will be studied.

Let us define the E_{q-1} sequence as follows:

[1]The research was partially supported by the TARIPAR3 project (grant Nr. TECH_08-A2/2-2008-0086), the hungarian-slovakian project SK-8/2008 and the hungarian-croatian project HR-6/2008

Mathematics subject classification number: 68P25, 11T23.

Definition 1. Let $\mathbb{F}_q$ be a finite field of characteristic 2 and its multiplicative group of prime order. Let χ be a non principal additive character, and α a primitive element of $\mathbb{F}_q$ and let $f(x) \in \mathbb{F}_q[x]$ of odd degree $d \geq \log q$ and let the coefficients of its terms be zero if and only if the term has an even exponent. Then let (e_n) be the *trace sequence ([Folláth]) defined as*

$$E_{q-1} = \{\chi(f(\alpha^1)), \chi(f(\alpha^2)), \ldots, \chi(f(\alpha^{q-1}))\} \in \{-1, +1\}^{q-1}, \tag{1.1}$$

In [Tóth 2007] the notion of the avalanche property was adopted for pseudorandom binary sequences. In section 2 it will be proved that the generator in question posesses the strong avalanche property. In [Ahlswede et al. 2003] the authors introduced a new measure for families of pseudorandom binary sequences. The notion of f-complexity is based on a practical attack scenario, where the sequence is used as the keystream of a streamcipher. The main result of section 1 is that the f-complexity of the trace sequence is at least $\lfloor \frac{d+1}{2} \rfloor$, where d is the degree of the seed polynomial. In [Brandstätter and Winterhof 2006] the authors gave a method to estimate the linear complexity profile of a pseudorandom binary sequence when its correlation measures are low. Since the correlation measures of larger order of a trace sequence can be high, this metod cannot be applied to it. In the rest of this paper the linear complexity of the sequence is studied. It turns out, that it has a low linear complexity and consequently it is weak.

In the followings assume that $N \in \mathbb{N}$, S is a given set, to each $s \in S$ a unique binary sequence is assigned

$$E_N = E_N(s) = (e_1, \ldots, e_N) \in \{-1, 1\}^N.$$

Let $F = F(S)$ denote the family of the binary sequences:

$$F = F(S) = \{E_N(s) : s \in S\}. \tag{1.2}$$

Definition 2. If $N \in \mathbb{N}$,$E_N = (e_1, \ldots, e_N) \in \{-1, 1\}^N$ and $E'_N = (e'_1, \ldots, e'_N) \in \{-1, 1\}^N$, then the distance $d(E_N, E'_N)$ is defined by

$$d(E_N, E'_N) = |\{n : 1 \leq n \leq N, e_n \neq e'_n\}|.$$

Definition 3. A polynomial $f(x) \in \mathbb{F}_q[x]$ of the form $f(x) = \sum_{i=0}^{d} a_i x^{2i+1}$ is said to be a comb polynomial.

2. AVALANCHE PROPERTY

Definition 4. A family $F(S)$ of binary sequences possesses the strict avalanche property if

$$m(F) = \min_{\substack{s,s' \in S \\ s \neq s'}} d(E_N(s), E_N(s')) \geq \left(\frac{1}{2} + o(1)\right) N.$$

THEOREM 1. *Let S be the set of comb polynomials $f(x) \in \mathbb{F}_q$ of degree at most d. Define $E_{q-1} = E_{q-1}(f) = \{e_1, \ldots, e_{q-1}\}$ by (1.1) and $F = F(S)$ by (1.2). Then if $d = o(q^{1/2})$, the family F possesses the strong avalanche property.*

PROOF. Let $f_1(x), f_2(x) \in S$ distinct. Now $f_1(x) + f_2(x)$ is obviously of positive degree, and therefore by the theorem of Weil (Theorem 5.38 in [Lidl and Niederreiter

1997]) we have:

$$|q-1-2d(E_{q-1}(f_1),E_{q-1}(f_2))| = \left|\sum_{i=1}^{q-1}\chi(f_1(\alpha^i))\chi(f_2(\alpha^i))\right| \le (d-1)q^{1/2}.$$

Now if $q-1-2d(E_{q-1}(f_1),E_{q-1}(f_2)) \ge 0$ then:

$$\left(\frac{1}{2}-\frac{(d-1)q^{1/2}}{q-1}\right)(q-1) \le d(E_{q-1}(f_1),E_{q-1}(f_2)) \le \frac{q-1}{2} \tag{2.1}$$

holds. In the case $q-1-2d(E_{q-1}(f_1),E_{q-1}(f_2)) < 0$ the inequality

$$\frac{q-1}{2} < d(E_{q-1}(f_1),E_{q-1}(f_2)) \le \left(\frac{1}{2}-\frac{(d-1)q^{1/2}}{q-1}\right)(q-1) \tag{2.2}$$

obviously follows.

Now from (2.1) and (2.2)

$$\left(\frac{1}{2}-\frac{(d-1)q^{1/2}}{q-1}\right)(q-1) \le m(F) \le \left(\frac{1}{2}+\frac{(d-1)q^{1/2}}{q-1}\right)(q-1)$$

follows and this completes the proof. □

3. FAMILY COMPLEXITY

Definition 5. Let us define a specification of length j as an index-set $(i_1,\dots,i_j)$ together with a corresponding value-set $(\varepsilon_{i_1},\dots,\varepsilon_{i_j}) \in \{+1,-1\}^j$. We say that a binary sequence $\{e_1,\dots,e_N\}$ satisfies the specification if

$$e_{i_1}=\varepsilon_1,\dots,e_{i_j}=\varepsilon_j.$$

A specification will represent the knowledge of the adversary about the cleartext.

Definition 6. The f-complexity $\Gamma(F)$ of a family F of binary sequences $E_N \in \{-1,+1\}^N$ is defined as the greatest integer j so that for any specification of length j there is at least one $E_N \in F$ which satisfies it.

LEMMA 1. *The polynomial $f \in \mathbb{F}_q[x_1,\dots,x_n]$ is a permutation polynomial over $\mathbb{F}_q$ if and only if*

$$\sum_{(c_1,\dots,c_n)\in\mathbb{F}_q^n}\chi(f(c_1,\dots,c_n)) = 0$$

for all nontrivial additive characters χ of $\mathbb{F}_q$.

PROOF. This is Corollary 7.38 in [Lidl and Niederreiter 1997]. □

LEMMA 2. *Suppose $f \in \mathbb{F}_q[x_1,\dots,x_n]$ is of the form*

$$f(x_1,\dots,x_n) = g(x_1,\dots,x_m)+h(x_{m+1},\dots,x_n),\ 1 \le m < n.$$

If at least one of g and h is a permutation polynomial over $\mathbb{F}_q$, then f is a permutation polynomial over $\mathbb{F}_q$.

PROOF. This is a part of Theorem 7.42 in [Lidl and Niederreiter 1997]. □

LEMMA 3. *Suppose $f \in \mathbb{F}_q[x_1, \ldots, x_n]$ is of the form*

$$f(x_1, \ldots, x_n) = \sum_{i=0}^{n} \alpha_i x_1.$$

If at least one α_i is nonzero, then f is a permutation polynomial over $\mathbb{F}_q$.

PROOF. The statement follows from the fact, that $g(x) = \alpha x$ is a permutation polynomial over $\mathbb{F}_q$ and Lemma 2 by induction. □

THEOREM 2. *Let us define the binary sequence family as in Theorem 1. If A is a specification of $t \leq \lfloor \frac{d+1}{2} \rfloor$ terms and $G(A)$ denotes the subset of the family $F = F(S)$ consisting the sequences satisfying A, then*

$$|G(A)| = \frac{|F|}{2^t}$$

holds.

PROOF. Let A be a specification of t terms $(\varepsilon_1, \ldots, \varepsilon_t) \in \{+1, -1\}^t$ together with an index set $(i_1, \ldots, i_t)$ where $1 \leq i_1, < \ldots < i_t \leq q-1$. Then

$$\begin{aligned}
|G(A)| &= \sum_{E(f)\in F} \prod_{j=1}^{t} \frac{(e_{i_j} + \varepsilon_j)\varepsilon_j}{2} = \frac{\varepsilon_1 \ldots \varepsilon_t}{2^t} \sum_{E(f)\in F} \prod_{j=1}^{t} (e_{i_j} + \varepsilon_j) \\
&= \frac{\varepsilon_1 \ldots \varepsilon_t}{2^t} \sum_{E(f)\in F} \left(\sum_{r=0}^{t-1} \sum_{1\leq j_1 < \ldots < j_r \leq t} \varepsilon_{j_1} \ldots \varepsilon_{j_r} \prod_{\substack{1\leq s\leq t \\ s\notin\{j_1,\ldots,j_r\}}} e_{i_s} + \varepsilon_1 \ldots \varepsilon_t \right) \\
&= \frac{|F|}{2^t} + \frac{\varepsilon_1 \ldots \varepsilon_t}{2^t} \sum_{E(f)\in F} \sum_{r=0}^{t-1} \sum_{1\leq j_1 < \ldots < j_r \leq t} \varepsilon_{j_1} \ldots \varepsilon_{j_r} \prod_{\substack{1\leq s\leq t \\ s\notin\{j_1,\ldots,j_r\}}} e_{i_s} \\
&= \frac{|F|}{2^t} + \frac{1}{2^t} \sum_{r=1}^{t} \sum_{1\leq j_1 < \ldots < j_r \leq t} \varepsilon_{j_1} \ldots \varepsilon_{j_r} \sum_{E(f)\in F} \prod_{\substack{1\leq s\leq t \\ s\in\{j_1,\ldots,j_r\}}} e_{i_s}.
\end{aligned}$$

where $E(f) = (e_1, \ldots, e_{q-1})$. Thus

$$|G(A)| - \frac{|F|}{2^t} = \frac{1}{2^t} \sum_{u=1}^{t} \sum_{1\leq v_1 < \ldots < v_u \leq t} \sum_{E(f)\in F} \prod_{z=1}^{u} e_{i_{v_z}}. \tag{3.1}$$

Consequently it suffices to show that $\sum_{E(f)\in F} \prod_{z=1}^{u} e_{i_{v_z}} = 0$. By (1.1) it follows, that:

$$\sum_{E(f)\in F} \prod_{z=1}^{u} e_{i_{v_z}} = \sum_{E(f)\in F} \prod_{z=1}^{u} \chi(f(\alpha^{i_{v_z}})) = \sum_{E(f)\in F} \chi(f(\alpha^{i_{v_1}}) + \ldots + f(\alpha^{i_{v_z}})) \tag{3.2}$$

With the notation $\alpha_i = \sum_{z=1}^{u} \alpha^{(2i+1)i_{v_z}}$ and since $f(x) = \sum_{i=0}^{d} a_i x^{2i+1}$:

$$\sum_{E(f)\in F} \chi(f(\alpha^{i_{v_1}}) + \ldots + f(\alpha^{i_{v_z}})) = \sum_{E(f)\in F} \chi(\sum_{i=0}^{d} a_i \sum_{z=1}^{u} \alpha^{(2i+1)i_{v_z}})$$

$$= \sum_{E(f)\in F} \chi(\sum_{i=0}^{d} a_i\alpha_i) = \sum_{\substack{(a_0,\ldots,a_d)\in\mathbb{F}_q^d \\ a_i\neq 0}} \chi(\sum_{i=0}^{d} a_i\alpha_i)$$

$$= \left(\sum_{(a_0,\ldots,a_d)\in\mathbb{F}_q^d} \chi(\sum_{i=0}^{d} a_i\alpha_i) - \sum_{j=0}^{d} \sum_{\substack{a_i\in\mathbb{F}_q \\ i\neq j}} \chi(\sum_{\substack{0\le i\le d \\ i\neq j}} a_i\alpha_i) \right)$$

If $f(x_0,\ldots,x_d) = \sum_{i=0}^{d} x_i\alpha_i$ and each $f_j(x_0,\ldots,x_d) = \sum_{\substack{0\le i\le d \\ i\neq j}} x_i\alpha_i$, $0 \le j \le d$ are nonzero polynomials, then by Lemma 3 they are permutation polynomials, so from (3.1), (3.2), (3.3) and Lemma 1

$$|G(A)| = \frac{|F|}{2^t}$$

follows.

If f or any f_j is a zero polynomial, then $k, l \in \mathbb{Z}$ exist such that $\alpha_i = 0$ for $i \in I = \{k, \ldots, k+l-1\}$ where $l \ge \lfloor \frac{d+1}{2} \rfloor$. And since $\alpha_i = \sum_{z=1}^{u} \alpha^{(2i+1)i_{v_z}}$, this means that $\beta_i = \alpha^{2i+1}$ is a root of the polynomial $g(x) = \sum_{z=1}^{u} x^{i_{v_z}}$, $g(x) \in \mathbb{F}_2[x]$ for every $i \in I$. Consequently, if $m_i(x)$ denotes the minimal polynomial of β_i over $\mathbb{F}_2$, then each β_i divides $g(x)$. Since the minimal polynomials are irreducible and because of the unique factorization, $g(x)$ is a codeword in the BCH code C generated by the product $h(x) = \prod_{i=k}^{k+l} m_i(x)$. Due to the BCH bound (Theorem 8 on page 201 in [MacWilliams and Sloane 1977]) the minimal distance of C is at least $l+1$. It follows that the weight of $g(x)$ is at least $\lfloor \frac{d+1}{2} \rfloor + 1$. This contradicts the assumption that the specification A has at most $\lfloor \frac{d+1}{2} \rfloor$ terms. □

COROLLARY 1. *The f-complexity of the family F is at least $\lfloor \frac{d+1}{2} \rfloor$.*

4. LINEAR RECURRING SEQUENCES

To determine the linear complexity of our sequences we will need the following results:

LEMMA 4. *Let $s_0, s_1, \ldots$ be a kth-order homogeneous linear recurring sequence in $K = \mathbb{F}_q$ whose characteristic polynomial $f(x)$ is irreducible over K. Let α be a root of $f(x)$ in the extension field $F = \mathbb{F}_{q^k}$. Then there exists a uniquely determined $\theta \in F$ such that*

$$s_n = Tr_{F/K}(\theta\alpha^n) \text{ for } n = 0, 1, \ldots.$$

PROOF. This is Theorem 8.24. in [Lidl and Niederreiter 1997]. □

LEMMA 5. *Let $f(x)$ be a kth degree, irreducible polynomial over $K = \mathbb{F}_q$, α one of its roots in the extension field $F = \mathbb{F}_{q^k}$ and θ an element in F. Then there exists a uniquely determined $s_0, s_1, \ldots$ homogeneous linear recurring sequence over F which has $f(x)$ as its characteristic polynomial and*

$$s_n = Tr_{F/K}(\theta\alpha^n) \text{ for } n = 0, 1, \ldots.$$

PROOF. This follows from Lemma 4 and the fact that there are exactly q^k elements in F as well as homogeneous linear recurring sequences over K with the characteristic polynomial $f(x)$. □

LEMMA 6. *Let $f(x) \in \mathbb{F}_q[x]$ be monic and irreducible over $\mathbb{F}_q$, and let $s_0, s_1, \ldots$ be a homogeneous linear recurring sequence in $\mathbb{F}_q$ not all of whose terms are 0. If the sequence has $f(x)$ as characteristic polynomial, then the minimal polynomial of the sequence is equal to $f(x)$.*

PROOF. This is Theorem 8.50. in [Lidl and Niederreiter 1997]. □

Let us define the addition operation for sequences termwise.

LEMMA 7. *For each $i = 1, 2, \ldots, h$, let σ_i be a homogeneous linear recurring sequence in $\mathbb{F}_q$ with minimal polynomial $m_i \in \mathbb{F}_q$. If the polynomials $m_1(x), \ldots, m_h(x)$ are pairwise relatively prime, then the minimal polynomial of the sum $\sigma_1 + \ldots + \sigma_h$ is equal to the product $m_1(x) \ldots m_h(x)$.*

PROOF. This is Theorem 8.57. in [Lidl and Niederreiter 1997]. □

5. THE LINEAR COMPLEXITY OF THE TRACE-GENERATOR

Definition 7. Let us define the following sequence for each *Trace-sequence:*

$$e_n = \begin{cases} 1, \text{ if } \chi(f(\alpha^n)) = -1, \\ 0, \text{ otherwise.} \end{cases} \tag{5.1}$$

Remark 1. It is easy to see, that this sequence is exactly

$$e_n = Tr(f(\alpha^n)),$$

where $Tr(x)$ denotes the absolute trace function.

Definition 8. The linear complexity (LC) of a sequence is the size in bits of the shortest linear feedback shift register (LFSR) which can produce that sequence.

THEOREM 3. *The linear complexity of the binary sequence defined by (5.1) is $\frac{k(d+1)}{2}$.*

PROOF. Let $f(x) = \sum_{i=0}^{l} \theta_i x^{2i+1}$ where $l = \frac{d-1}{2}$. Then with the notation $\alpha_i = \alpha^{2i+1}$ we get

$$e_n = Tr_{F/K}(f(\alpha^n)) = \sum_{i=0}^{l} Tr_{F/K}(\theta_i \alpha^{n(2i+1)}) = \sum_{i=0}^{l} Tr_{F/K}(\theta_i \alpha_i^n).$$

For the minimal polynomial of α_i over $\mathbb{F}_2$ write $m_i(x)$. By applying Lemma 1 we obtain:

$$(e_n) = \sigma_0 + \ldots + \sigma_l,$$

where σ_i is a homogeneous linear recurring sequence in $\mathbb{F}_2$ with $m_i(x)$ as its characteristic polynomial. According to Lemma 6 the polynomial $m_i(x)$ is not only a characteristic polynomial but the minimal polynomial of σ_i. Then by Lemma 7 the sequence (e_n) itself is a homogeneous linear recurring sequence over $\mathbb{F}_2$ with minimal polynomial $M(x) = \prod_{i=0}^{l} m_i(x)$. Since each of $m_i(x)$ is of degree k, the degree of $M(x)$ and the order of $(e_n) = \sigma_0 + \ldots + \sigma_l$ is $\frac{k(d+1)}{2}$. □

6. CONCLUDING REMARKS

It turns out that despite its fair statistical attributes the Trace-generator is a pure linear feedback shift register, and as such it is vulnerable to the Berlekamp-Massey algorithm. Furthermore if it is used with fixed primitive element, even the linear recurrence is known to the adversary. Another consequence of this result is, that the other results proved for the trace sequences hold for this narrow class of linear feedback shift registers too.

In [Mauduit and Sárközy 1997] measures of pseudorandomness were introduced. There was also mentioned that even the combined measure can be low, as symmetric sequences show, whose first part has low combined measure. The Trace-generator is a nontrivial example for highly predictable sequences with low combined measure.

In [Mauduit and Sárközy 1997] the authors proposed the following two options for the combined pseudorandom measure:

$$Q(E_N) = \max_{k \le (\log N)/\log 2} Q_k(E_N)$$

$$Q^*(E_N) = \sum_{k=1}^{\infty} Q_k(E_N)/2^k$$

where

$$Q_k(E_N) = \max_{a,b,t,D} \left| \sum_{j=0}^{t} e_{a+jb+d_1} e_{a+jb+d_2} \ldots e_{a+jb+d_k} \right|.$$

As we have proved, there exist binary sequences whose $Q(E_N)$ combined measure is small, because their combined measures of small order are small, but some of their other pseudorandom properties are poor. Brandstätter and Winterhof, resp. Andics studied the linear complexity profile in terms of the correlation measures (the results of Brandstätter and Winterhof are nontrivial only if the correlation measures of large order are also small). These two arguments imply, that the $Q^*(E_N)$ combined pseudorandom measure is more convenient candidate to be the single measurement number of pseudorandomness than the other option.

REFERENCES

Ahlswede, R., Khachatrian, L., Mauduit, C., and Sárközy, A. 2003. A complexity measure for families of binary sequences. *Period. Math. Hungar. 46,* 2 (June), 107–118.

Andics, Á. 2005. On the linear complexity of binary sequences. *Ann. Univ. Sci. Budapest. Eötvös Sect. Math. 48*, 173–180.

Brandstätter, N. and Winterhof, A. 2006. Linear complexity profile of binary sequences with small correlation measure. *Period. Math. Hungar. 52,* 2 (June), 1–8.

Folláth, J. Construction of pseudorandom binary sequences I. *Period. Math. Hungar. 57,* 1, 73–81.

Goubin, L., Mauduit, C., and Sárközy, A. 2004. Construction of large families of pseudorandom binary sequences. *Journal of Number Theory 106,* 1, 56–69.

Gyarmati, K. 2009. Concatenation of pseudorandom binary sequences. *Period. Math. Hungar. 58,* 1, 99–120.

Hoffstein, J. and Lieman, D. 2001. *Cryptography and Computational Number Theory*. Progress in Computer Science and Applied Logic, vol. 20. Birkhäuser Verlag, Chapter The Distribution of the Quadratic Symbol in Function Fields and a Faster Mathematical Stream Cipher, 59–68.

Lidl, R. and Niederreiter, H. 1997. *Finite Fields*. Encyclopedia of Mathematics, vol. 20. Cambridge University Press.

MacWilliams, F. J. C. and Sloane, N. J. A. 1977. *The Theory of Error-Correcting Codes*. North-Holland Mathematical Library, vol. 16. North-Holland Publishing Company.

Mauduit, C. and Sárközy, A. 1997. On finite pseudorandom binary sequences I: Measure of pseudorandomness, the Legendre symbol. *Acta Arith. 82,* 4, 365–377.

Tóth, V. 2007. Collision and avalanche effect in families of pseudorandom binary sequences. *Period. Math. Hungar. 55,* 2 (November), 185–196.