On the Stability of m-Sequences

Alex J. Burrage[1], Ana Sălăgean[1], and Raphael C.-W. Phan[2]

[1] Computer Science, Loughborough University, Leicestershire, LE11 3TU, UK
a.burrage@lboro.ac.uk, a.m.salagean@lboro.ac.uk
[2] Electronic, Electrical & Systems, Loughborough University, Leicestershire LE11 3TU, UK r.phan@lboro.ac.uk

Abstract. We study the stability of m-sequences in the sense of determining the number of errors needed for decreasing the period of the sequences, as well as giving lower bounds on the k-error linear complexity of the sequences. For prime periods the results are straightforward so we concentrate on composite periods. We give exact results for the case when the period is reduced by a factor which is a Mersenne number and for the case when it is reduced by a prime p such that the order of 2 modulo p equals $p-1$. The general case is believed to be difficult due to its similarity to a well studied problem in coding theory. We also provide results about the relative frequencies of the different cases. We formulate a conjecture regarding the minimum number of errors needed for reducing the period at all. Finally we apply our results to the LFSR components of several well known stream ciphers.

1 Introduction

Linear feedback shift registers (LFSRs) are frequently used in stream ciphers (see, for example [7] [1] [2] [5] [16]) due to their well understood properties and simplicity of construction in hardware. m-Sequences were first discussed by Golomb [6], and have many interesting and well studied properties.

However, to our knowledge, one property that has not been studied is the k-error linear complexity of such sequences. The k-error linear complexity of a periodic sequence s is the minimum linear complexity of the sequences that can be obtained from s by changing up to k terms in each period (see [3], [14]). The stability of the linear complexity (i.e. high k-error linear complexity for small values of k) is an important criterion [3] in the design of stream ciphers because if a sequence has a low k-error linear complexity, an attacker could potentially recover easily all but k terms of the sequence.

The k-error linear complexity and the error linear complexity spectrum are very difficult to determine for a general sequence but for some classes of sequences polynomial time algorithms have been found [14] [12] [10] [9]. However, these classes of sequences are usually chosen so that the k-error linear complexity is easy to determine, rather than being chosen because they are used in cryptographic primitives. We begin to rectify that in this paper by analyzing the k-error complexity of m-sequences, specifically by finding lower bounds on the

minimum number of errors that are required to reduce the complexity of the sequence.

By analogy to the k-error linear complexity, we can define the k-error period of a sequence, i.e. the minimal period that we can obtain for a sequence by changing up to k terms in each period. As the period of m-sequences is maximal among all sequences of a given linear complexity, m-sequences are often used as components of stream cipher in order to ensure a large period, see [8]. Therefore it is perhaps even more important in such situations to guarantee the stability of the period rather than of the linear complexity. Moreover for m-sequences we cannot reduce the linear complexity without reducing the period and therefore the minimum number of errors needed for reducing the period is a lower bound for the minimum number of errors needed for reducing the linear complexity (Proposition 1).

The case of m-sequences with prime period (Section 3.1) is relatively easy and we obtain a closed form expression for the k-error linear complexity. When the period is composite, the problem is related to the problem of determining the weight enumerator of minimal cyclic codes (Section 3.2). This is a well studied and as yet not fully solved problem (see for example [4]). The weight enumerator is known for certain particular cases, but for the general case it seems no closed form or algorithm better than brute force is known [4]. For two particular cases we give exact formulae for the minimum number of errors needed to reduce the period of the m-sequence by a factor q: the case when q is a Mersenne number (Sections 3.3), and the case when q is a prime such that the order of 2 modulo q equals $q-1$ (Section 3.4). Other particular cases could be treated by looking at those minimal cyclic codes for which the weight enumerator is known. In Section 3.5 we formulate a conjecture regarding the minimum number of errors needed to reduce the period of an m-sequence at all. We show that if the conjecture is true we can determine this number for at least 76% of m-sequences. Finally, in Section 4 we study how these results relate to several cipher systems (the eStream candidates Grain and DECIM^{v2}, LILI-128 and SSC2) and show that their LFSR component is secure from the point of view of the stability of its period and linear complexity.

2 Preliminaries

We start by recalling a number of definitions and results. While this paper is only concerned with binary sequences, all the results in this section hold for any field K unless otherwise specified.

Definition 1. *Given a degree n monic polynomial $f = x^n + c_{n-1}x^{n-1} + \ldots + c_0 \in K[x]$ and n initial values, $s_0, s_1, \ldots, s_{n-1} \in K$, we can recursively generate an infinite sequence $s = (s_0, s_1, \ldots)$, by using the following linear recurrence relation:*

$$s_i = -s_{i-n}c_0 - s_{i-n+1}c_1 - \ldots - s_{i-1}c_{n-1}$$

for $i \geq n$. Note that it is possible to generate identical sequences using different polynomials. We refer to f as a characteristic polynomial for s, and any sequence which can be generated in this way as a linear recurrent sequence.

Definition 2. *For any linear recurrent sequence s, the characteristic polynomial of the lowest degree is referred to as the* minimum polynomial*, and its degree the* linear complexity *of the sequence. We will denote the linear complexity of s as* $\mathrm{LC}(s)$*. Note that the all zero sequence has complexity zero.*

Definition 3. *A sequence s is called periodic if there exists an integer t such that $s_i = s_{i+t}$ for all $i = 0, 1, \ldots$. We call t a period of s. We call the smallest such t the minimal period of s and denote it by* $\mathrm{P}(s)$*.*

We will identify a sequence s of period t with the finite sequence $(s_0, s_1, \ldots, s_{t-1})$. We can then talk about the Hamming weight of the sequence s, denoted $\mathrm{wt}(s)$. Note that we do not restrict the period to being the minimal period, so the period needs to be specified in order to determine the weight. Similarly we can talk about the Hamming distance between two sequences s and s' of the same period, denoted $\mathrm{d}(s, s')$.

Definition 4. *Let f be a primitive polynomial of degree n. Then any non-zero sequence generated by f is called an m-sequence.*

Note that for any binary sequence generated by a polynomial of degree n, the period can range between n and $2^n - 1$. The maximum is achieved exactly when the sequence is an m-sequence.

The k-error linear complexity of a sequence is a parameter that generalizes the linear complexity:

Definition 5. *[14] [3] Let s be an infinite periodic sequence with period t and $0 \leq k \leq t$. The k-error linear complexity of s is defined as:*

$$\mathrm{LC}_k(s) = \min\{\mathrm{LC}(s') : s' \textit{sequence of period } t, d(s, s') \leq k\}.$$

Definition 6. *[10] The error linear complexity spectrum of a sequence s of period t is a list of pairs, $(k, \mathrm{LC}_k(s))$, where k takes all values in the range $0 \leq k \leq t$. A critical point in the spectrum is one where* $\mathrm{LC}_k(s) < \mathrm{LC}_{k-1}(s)$*.*

Note that knowing the critical points of the error linear complexity spectrum is enough to generate the whole spectrum. The extreme (and trivial) cases are when we change all non-zero elements into zeros, obtaining the all-zero sequence; also, for binary sequences we can change all zeros into ones obtaining a sequence consisting only of ones.

Lemma 1. *Let s be an infinite periodic sequence. Then* $\mathrm{LC}_{\mathrm{wt}(s)}(s) = 0$*, so the last critical point in the complexity spectrum is* $(\mathrm{wt}(s), 0)$*. If s is a binary sequence,* $\mathrm{LC}_{\mathrm{P}(s)-\mathrm{wt}(s)}(s) \leq 1$ *so if* $\mathrm{wt}(s) > \mathrm{P}(s)/2$ *the penultimate critical point in the spectrum is* $(\mathrm{P}(s) - \mathrm{wt}(s), 1)$*.*

We will need a few other parameters related to k-error linear complexity:

Definition 7. *The smallest k such that $\mathrm{LC}_k(s) < \mathrm{LC}(s)$ will be called the* complexity reduction value*, denoted $\mathrm{RLC}(s)$. The minimum number of errors required in each period to reduce the linear complexity of s to a value that is at most c will be denoted by $\mathrm{ELC}_c(s)$.*

Note that for any fixed sequence s, if we consider $\mathrm{ELC}_c(s)$ as a function of c, and $\mathrm{LC}_k(s)$ as a function of k, then $\mathrm{ELC}_c(s)$ is the minimum of the preimage of c under $\mathrm{LC}_k(s)$.

Example 1. Consider the binary sequence s whose minimal period is $(0,0,0,1,0,$-$0,1,1,0,1,0,1,1,1,1,0)$. Then s has the following complexity spectrum: (0,15), (1,15), (2,10), (3,10), (4,5), (5,5), (6,2), (7,2), (8,0), (9,0), (10,0), (11,0), (12,0), (13,0), (14,0), (15,0), (16,0). The critical points on this spectrum are (0,15), (2,10), (4,5), (6,2), (8,0) and the complexity reduction value is 2. We also have, for example, $\mathrm{ELC}_5(s) = 4$.

We define for the period of a sequence analogues of the k-error linear complexity parameters:

Definition 8. *For an infinite periodic sequence s we define the k-error period to be:*

$$\mathrm{P}_k(s) = \min\{\mathrm{P}(s') : s' \text{ sequence of period } \mathrm{P}(s), d(s,s') \leq k\}$$

Note s' above must have (possibly not minimal) period equal to $\mathrm{P}(s)$. The smallest k such that $\mathrm{P}_k(s) < \mathrm{P}(s)$ will be called the period reduction value*, denoted $\mathrm{RP}(s)$. We will denote by $\mathrm{EP}_c(s)$ the number of errors required to reduce the period of s to at most c.*

Note that $\mathrm{EP}_c(s)$ and $\mathrm{P}_k(s)$ have the same relation as $\mathrm{ELC}_c(s)$ and $\mathrm{LC}_k(s)$.

We recall some terminology from number theory.

Definition 9. *A Mersenne number is any number of the form $2^n - 1$ for some positive integer n. A Mersenne prime is a Mersenne number that is prime.*

Note that in the definition of a Mersenne number we do not require that either n or $2^n - 1$ be prime, by following the terminology used in, for example, the Online Encyclopedia of Integer Sequences.

Definition 10. *Let p be a prime. The multiplicative order of 2 modulo p is the smallest integer u such that $2^u \equiv 1 \bmod p$. We will refer to this as simply the order of $2 \bmod p$ and denote it as $\mathrm{ord}_p(2)$.*

Recall that for any integer v, $2^v \equiv 1 \bmod p$ iff $\mathrm{ord}_p(2)|v$. Also, we recall Fermat's Little Theorem, which states that $\mathrm{ord}_p(2)$ must divide $p - 1$.

3 k-Error Complexity and Period for Various Classes of m-Sequence

In this paper we study binary m-sequences. We aim to determine the number of errors needed for reducing the complexity of such sequences and the number of errors needed for reducing the period. These two problems are closely related:

Proposition 1. *For any given m-sequence s, the period reduction value is a lower bound on the complexity reduction value. Moreover if s' is any linearly recurrent sequence with* $\mathrm{LC}(s') < \mathrm{LC}(s)$ *then* $\mathrm{P}(s') < \mathrm{P}(s)$.

Proof. For any sequence of linear complexity n, its period length must be less than or equal to $2^n - 1$, and this period length is exactly achieved by any m-sequence of linear complexity n. Therefore, if we wish to consider sequences with smaller linear complexity than a given m-sequence, they must have smaller period as well. □

3.1 Prime period

We will first deal with the relatively easy case when the period of the m-sequence is prime. In this case, we are able to determine not just the period reduction value, but the full error linear complexity spectrum, which trivially gives the complexity reduction value.

Theorem 1. *Consider an m-sequence, s, with* $\mathrm{P}(s) = m = 2^n - 1$. *If m is a Mersenne prime, then the critical points of the k-error complexity spectrum of s are:* $(0, n)$, $((m-1)/2, 1)$, $((m+1)/2, 0)$.

Proof. From Lemma 1 we know that the spectrum will contain at least the three critical points listed in the statement. By Proposition 1, the only way to reduce the complexity of s is by reducing its period to a factor of m, i.e. to 1, as m is prime. That means s can only become a sequence of all ones (requiring $(m-1)/2$ changes) or a sequence of all zeros (requiring $(m+1)/2$ changes). □

This result shows that both the complexity reduction value and the period reduction value are almost half the period length, which implies that such sequences are very secure from this point of view, in fact, by Lemma 1 they are as secure as possible. Since it is possible to construct an m-sequence of length equal to any Mersenne number, the frequency of such sequences is dependent on the frequency of Mersenne primes among Mersenne numbers. There are no known results about this, but the widely believed Lenstra-Pomerance-Wagstaff Conjecture [13] implies that the proportion of Mersenne primes less than x as a proportion of all Mersenne numbers is $\log\log x / \log x$. This implies that the frequency of these sequences is low, and decreases as we consider longer sequences. Out of the smallest 200 lengths for m-sequences, 13 of them are prime, a proportion of 0.07.

3.2 Reducing the period of an m-sequence by an arbitrary factor

For treating the case when the period is composite, reducing the period to a factor r of the original period can be visualized by writing the sequence row-wise in a table of r columns and aiming to make each column of the table contain one single value. We formalize this as follows:

Definition 11. *For a periodic sequence s, with $\mathrm{P}(s) = m$ and $m = qr$ for some integers q, r, we define the* r-decimation matrix of s *to be the q by r matrix T with entries: $T_{i,j} = s_{ir+j}$ for $i = 0, \ldots, q-1$ and $j = 0, \ldots, r-1$. That is, we construct T by sequentially filling its rows with the values of s. It will often be useful for us to refer to the columns of T as sequences themselves.*

Note that using the notation above, the columns of T are r-regular, improper decimations of s.

Lemma 2. *The minimum number of errors needed for reducing the period $m = qr$ of a binary sequence s from m to r equals: $\mathrm{EP}_r(s) = \sum_{i=0}^{r-1} \min\{\mathrm{wt}(T_i), q - \mathrm{wt}(T_i)\}$ where T_i are the columns of the r-decimation matrix of s.*

Proof. The number of errors needed to make the column T_i contain only zeros is $\mathrm{wt}(T_i)$ and to contain only ones is $q - \mathrm{wt}(T_i)$. If each column contains only one value, then the period of the sequence has been reduced to r. □

Note that an algorithm for computing by brute force the weight of the columns of the decimation matrix is linear in the period length of the sequence. However, for m-sequences the period length is exponentially higher than the degree of the generator polynomial, so a more efficient algorithm, or a closed formula, would be preferable.

Theorem 2. *Let s be a sequence with $\mathrm{P}(s) = m = p_1^{e_1} p_2^{e_2} \ldots p_r^{e_r}$ (with p_i prime for all i). Then $\mathrm{EP}_{m/p_i p_j}(s) \geq \mathrm{EP}_{m/p_i}(s)$.*

Proof. Note that a sequence of period $m/p_i p_j$ also has period m/p_i. Therefore $\mathrm{EP}_{m/p_i p_j}(s)$ errors can change s into a sequence of period m/p_i, and so the result follows. □

Corollary 1. *Let s be a sequence with $\mathrm{P}(s) = m = p_1^{e_1} p_2^{e_2} \ldots p_r^{e_r}$ (with p_i prime for all i). Then $\mathrm{RP}(s) = \min_i\{\mathrm{EP}_{m/p_i}\}$.*

Corollary 1 implies that to determine the period reduction value for an m-sequence, we will only need to consider reducing the period by a prime factor.

We recall the following results which shed light on the structure of the decimation matrix:

Lemma 3. *[15] Assume that T is the r-decimation matrix for an m-sequence s with $\mathrm{P}(s) = qr$. Then the columns of T are all generated by a single, irreducible polynomial.*

Lemma 4. *[15] Let s be an m-sequence of period m, and assume $m = qr$. If $r = 2^n - 1$ for some n (that is, r is a Mersenne number) then each column of the q-decimation matrix of s will either be an m-sequence or the all zero sequence. Further, each of the m-sequences will be identical, up to a cyclic shift.*

The following is a closely related result from coding theory:

Theorem 3. *([11, Theorem 11, Ch. 8, §4]) Let q, n be integers such that n is minimal such that $q|2^n - 1$. Let s be an m-sequence with period length $2^n - 1$. The columns of the $(2^n - 1)/q$-decimation matrix of s are a set of representatives (with respect to the equivalence relation of cyclic shifting) for a minimal $[q, n]$ cyclic code.*

Please note that the original version of the above theorem uses a matrix which is the transpose of our decimation matrix.

Recall that a cyclic code of length q with generator polynomial $g|x^q - 1$ can be equivalently viewed as the set of sequences of period q with characteristic polynomial equal to the reciprocal of the parity check polynomial $(x^q - 1)/g$. We can define an equivalence relation on the set of sequences generated by a fixed polynomial: two sequences of period t are equivalent if when represented as a finite sequence of length t one can be obtained from the other by a cyclic shift. We can therefore formulate the theorem above equivalently as follows:

Corollary 2. *Let q, n be integers such that n is minimal such that $q|2^n - 1$. Let s be an m-sequence with period length $2^n - 1$. The columns of the $(2^n - 1)/q$-decimation matrix for s are a set of representatives for the set of sequences generated by a fixed irreducible polynomial of degree n and order q.*

We extend the results above to the case when n is not minimal with the property that $q|p^n - 1$:

Theorem 4. *Let q, n be integers such that $q|2^n - 1$. Let n' be minimal such that $q|2^{n'} - 1$. Let s be an m-sequence with period length $2^n - 1$. In the $(2^n - 1)/q$-decimation matrix of s there are $\frac{2^{n-n'}-1}{q}$ all-zero columns and $2^{n-n'}$ columns from each of the $\frac{2^{n'}-1}{q}$ equivalence classes of the sequences generated by a fixed irreducible polynomial of degree n' and order q.*

Proof. We decimate the sequence s in two stages. Let B be the $(2^n - 1)/(2^{n'} - 1)$ decimation matrix of s. Each column of B has length $2^{n'} - 1$ and so by Lemma 4 it is either the all-zero column or one fixed m-sequence (possibly shifted). As in [15, Section IV B] we can count how many of each we have and show that there are $\frac{2^{n-n'}-1}{2^{n'}-1}$ all-zero columns and $2^{n-n'}$ m-sequences. To obtain the $(2^n - 1)/q$-decimation matrix T of s we can think of concatenating the first $(2^{n'} - 1)/q$ rows of B to obtain the first row of T, then concatenating the next $(2^{n'} - 1)/q$ rows of B to obtain the second row of T and so on. Looking at a particular column of B, say column j, we see that its elements end up as columns $j, j + (2^n - 1)/(2^{n'} -$

$1), j + 2(2^n - 1)/(2^{n'} - 1), \ldots$ of T. Moreover, these columns of T are exactly a $(2^{n'} - 1)/q$-decimation matrix for the sequence in column j of B. If this sequence is an m-sequence, then by Theorem 3 the resulting columns of T are exactly a set of representatives for the equivalence classes of the sequences generated by a fixed irreducible polynomial. If column j of B is all-zero, then obviously the corresponding columns in T are also all-zero. □

We have therefore:

Corollary 3. *Let s be an m-sequence of length $2^n - 1$ and let q be a factor of $2^n - 1$. Let n' be minimal such that q is a factor of $2^{n'} - 1$. Then:*

$$\mathrm{EP}_{(2^n-1)/q}(s) = 2^{n-n'} \mathrm{EP}_{(2^{n'}-1)/q}(s).$$

Recall that the weight enumerator (or weight distribution) of a code C of length m can be defined as the list of integers $A_0, A_1, \ldots, A_m$ with A_i equal to the number of codewords in C that have Hamming weight equal to i. In a minimal $[q, n]$ cyclic code each of the q cyclic shifts of a non-zero codeword are distinct i.e. there are exactly q codewords in each equivalence class, all of the same weight. Therefore, as a consequence of Theorem 3 and Corollary 3 we have:

Corollary 4. *Let s be an m-sequence of length $2^n - 1$ and let q be a factor of $2^n - 1$. Let n' be minimal such that q is a factor of $2^{n'} - 1$. If $A_0, A_1, \ldots, A_m$ is the weight enumerator of a minimal $[q, n']$ cyclic code, then*

$$\mathrm{EP}_{(2^n-1)/q}(s) = 2^{n-n'} \sum_{i=1}^{q} \min\left\{ \frac{A_i}{q}, q - \frac{A_i}{q} \right\}.$$

The problem of finding the weight enumerator for a general cyclic code is a well studied, and yet unsolved, problem in coding theory (see, for example, [4]). There are a number of particular cases for which the problem has been solved, and we examine some of them in the next sections; others can be similarly transferred. However the general case seems difficult as no better solution than brute force (i.e. going through the decimation matrix and determining the weight of each column) is known. In view of the corollary above, we suspect that determining $\mathrm{EP}_c(s)$ for s an m-sequence is an equally difficult problem.

For m-sequences of large length, Corollaries 3 and 4 above will allow us to reduce the problem to one for an m-sequence of smaller length, for which we can either find the weight enumerator if it falls in one of the particular cases for which the weight enumerator is known, or if a brute force approach is the only option, the chances of success are higher due to the shorter length of the sequence. Examples will be given in Section 4.

3.3 Reducing the period by a Mersenne number

When the period length of an m-sequence is a Mersenne number $2^n - 1$ with n not prime, a large number of factors of the period are Mersenne numbers

themselves and can easily be obtained by factorizing the exponent n. Namely, if n' is a factor of n, then $2^{n'}-1$ is a factor of 2^n-1. In this section we will compute the number of errors needed to reduce the period of an m-sequence by a Mersenne number.

Theorem 5. *Consider an m-sequence s with $\mathrm{P}(s)=m$ and assume m has a factor of the form $2^{n'}-1$. Then*

$$\mathrm{EP}_{m/(2^{n'}-1)}(s)=(m+1)\frac{2^{n'-1}-1}{2^{n'}}=\mathrm{wt}(s)(1-\frac{1}{2^{n'-1}})$$

.

Proof. Let T_i be the i-th column of the q-decimation matrix T of s. By Lemma 4 each T_i is either an m-sequence (and therefore $\mathrm{wt}(T_i)=2^{n'-1}$) or the all zero sequence (and so $\mathrm{wt}(T_i)=0$). As in the proof of Theorem 5 we can count how many of each we have: $(m+1)/2^{n'}$ columns of T are m-sequences and the rest are all-zero sequences. Applying Lemma 2, $\mathrm{EP}_{m/(2^{n'}-1)}(s)=((m+1)/2^{n'})(2^{n'-1}-1)$. □

Note that Theorem 5 can also be viewed as a particular case of Corollary 3 for $q=2^{n'}-1$, as in that case $\mathrm{EP}_{(2^{n'}-1)/q}(s)=\mathrm{EP}_1(s)=2^{n'-1}-1$ (to reduce the period of an m-sequence to 1 we need to change all zeros into ones).

Example 2. The most important cases will be when the period length m is reduced by a small factor. Theorem 5 shows that $\mathrm{EP}_{m/3}(s)=(1/4)(m+1)$, $\mathrm{EP}_{m/7}(s)=(3/8)(m+1)$, $\mathrm{EP}_{m/15}(s)=(7/16)(m+1)$ and $\mathrm{EP}_{m/31}(s)=(15/32)(m+1)$.

Corollary 5. *If the period of an m-sequence s is being reduced by a factor that is a Mersenne prime q, the smallest number of errors required will be when $q=3$, in which case $(\mathrm{P}(s)+1)/4$ errors are required, i.e. half of the weight of s. As q increases, the number of errors approaches the weight of the sequence.*

Example 3. Let s be an m-sequence of period to 4095. Then $\mathrm{EP}_{4095/3}(s)=\mathrm{EP}_{1365}(s)=1024=\mathrm{wt}(s)/2$ and $\mathrm{EP}_{4095/7}(s)=\mathrm{EP}_{585}(s)=1536=(3/4)\,\mathrm{wt}(s)$.

3.4 Reducing the period by a prime p with $\mathrm{ord}_p(2)=p-1$

When p is a prime such that $\mathrm{ord}_p(2)=p-1$, the factorization of x^p-1 into irreducible factors is the trivial factorization $x^p-1=(x-1)(x^{p-1}+x^{p-2}+\ldots+1)$ i.e. $x^{p-1}+x^{p-2}+\ldots+1$ is an irreducible polynomial of degree $p-1$. Therefore, if s is an m-sequence of length $2^{p-1}-1$, by Corollary 2 the columns of its $(2^{p-1}-1)/p$-decimation matrix are sequences generated by this irreducible polynomial, i.e. sequences obtained by a parity-check bit type equation i.e. the sum of the bits equals 0. For each even weight $2i$ there are $\binom{p}{2i}$ such sequences, i.e. $\frac{1}{p}\binom{p}{2i}$

inequivalent sequences (equivalence under cyclic shifts). Hence by Lemma 2 we have

$$\mathrm{EP}_{(2^{p-1}-1)/p}(s) = \sum_{i=1}^{\frac{p-1}{2}} \frac{1}{p}\binom{p}{2i} \min(2i, p-2i).$$

With some combinatorial manipulation we will obtain:

Theorem 6. *Let p be a prime such that $\mathrm{ord}_p(2) = p-1$. Let s be an m-sequence of length $2^{p-1}-1$. Then:*

$$\mathrm{EP}_{(2^{p-1}-1)/p}(s) = 2^{p-2} - \frac{1}{2}\binom{p-1}{\frac{p-1}{2}}.$$

Proof.

$$\begin{aligned}
\mathrm{EP}_{(2^{p-1}-1)/p} &= \sum_{i=1}^{\frac{p-1}{2}} \frac{1}{p}\binom{p}{2i} \min(2i, p-2i) \\
&= \frac{1}{p}\left(\sum_{i=1}^{\lfloor \frac{p-1}{4} \rfloor} 2i\binom{p}{2i} + \sum_{i=\lfloor \frac{p-1}{4} \rfloor + 1}^{\frac{p-1}{2}} (p-2i)\binom{p}{2i} \right) \\
&= \frac{1}{p}\left(\sum_{i=1}^{\lfloor \frac{p-1}{4} \rfloor} 2i\binom{p}{2i} + \sum_{i=\lfloor \frac{p-1}{4} \rfloor + 1}^{\frac{p-1}{2}} (p-2i)\binom{p}{p-2i} \right) \\
&= \frac{1}{p} \sum_{j=1}^{\frac{p-1}{2}} j\binom{p}{j} \\
&= \sum_{j=1}^{\frac{p-1}{2}} \binom{p-1}{j-1}.
\end{aligned}$$

The last expression is the sum of the combinatorial coefficients in Pascal's triangle on the row $p-1$ up to but excluding the middle element. If we would also add half of the middle element we would obtain exactly half of the total sum of the row, i.e. 2^{p-2}. Hence the result in the theorem follows. □

Combining this Theorem with Corollary 3 we obtain:

Corollary 6. *Let s be an m-sequence of length 2^n-1 and let p be a prime factor of 2^n-1. If $\mathrm{ord}_p(2) = p-1$ then*

$$\mathrm{EP}_{(2^n-1)/p}(s) = 2^{n-1} - 2^{n-p}\binom{p-1}{\frac{p-1}{2}} = \mathrm{wt}(s)(1 - \frac{1}{2^{p-1}}\binom{p-1}{\frac{p-1}{2}}).$$

3.5 The minimum number of errors needed for reducing the period of an m-sequence

In the previous sections we examined the number of errors needed for reducing the period of an m-sequence to specific factors of the original period. In this section we examine the period reduction value, i.e. the minimum number of errors needed for reducing the period of an m-sequence at all. By Corollary 1 this is the minimum among $\mathrm{EP}_{m/p}(s)$ for the different prime factors p of the period m. We can determine the minimum among different $\mathrm{EP}_{m/p}(s)$ for those p which fall in the cases of Theorem 5 and Corollary 6:

Corollary 7. *Let s be an m-sequence with $\mathrm{P}(s) = m$, and let $p_1 < p_2$ be factors of m. Moreover assume that p_1 and p_2 are such that both satisfy condition (i) or both satisfy condition (ii) or p_1 satisfies condition (ii) and p_1 satisfies condition (i) below:*

(i) being a Mersenne number

(ii) being a prime p such that $\mathrm{ord}_p(2) = p - 1$.

Then $\mathrm{EP}_{m/p_1}(s) < \mathrm{EP}_{m/p_2}(s)$.

Proof. The first situation is immediate, for the second we use the combinatorial inequality $4\binom{2t}{t} > \binom{2(t+1)}{t+1}$. For the last situation using a Stirling inequality $\binom{2t}{t} \geq \frac{2^{2t-1}}{\sqrt{t}}$ it suffices to prove that $\frac{1}{\sqrt{2(p_1-1)}} > \frac{2}{p_2+1}$, which can be easily verified.

It would be tempting to conjecture that $\mathrm{EP}_{m/p_1}(s) < \mathrm{EP}_{m/p_2}(s)$ for any prime factors $p_1 < p_2$ of m. (Indeed we conjectured that in the preliminary version of this paper.) However, this is not true, for example for $m = 2^{180} - 1$, $p_1 = 31$ (which falls into case (i) in Corollary 7) and $p_2 = 37$ (which falls into case (ii) in Corollary 7) we compute $\mathrm{EP}_{m/31}(s) = 2^{180}(15/32) \approx 7.2 * 10^{53}$ using Theorem 5 and $\mathrm{EP}_{m/37}(s) = 2^{179} - 2^{143} * \binom{36}{18} \approx 5.7 * 10^{53}$ using Corollary 6. Hence the reverse inequality $\mathrm{EP}_{m/p_1}(s) > \mathrm{EP}_{m/p_2}(s)$ holds in this example. However, for our purposes we are only interested in finding the minimum $\mathrm{EP}_{m/p}(s)$ and in this example neither of these primes achieves it, as $\mathrm{EP}_{m/3}(s) = 2^{178}$ is lower than both $\mathrm{EP}_{m/p_1}(s)$ and $\mathrm{EP}_{m/p_2}(s)$.

Proposition 2. *Let s be an m-sequence with $\mathrm{P}(s) = 2^n - 1 = m$, and let p be a prime such that p divides m and $\mathrm{ord}_p(2) = p - 1$. Then the smallest prime factor of m is 3 and $\mathrm{EP}_{m/3}(s) < \mathrm{EP}_{m/p}(s)$.*

Proof. The prime factors of m are exactly those primes q with $\mathrm{ord}_q(2) | n$. For the particular p in the statement we have therefore $(p-1)|n$. Since p is odd (obviously 2 is never a factor of $2^n - 1$) that means n is even. On the other hand, $\mathrm{ord}_3(2) = 2$, so 3 must be a factor (the smallest one) of $2^n - 1$ whenever n is even.

Using Corollary 6 the inequality becomes $\binom{p-1}{\frac{p-1}{2}} < 2^{p-2}$ which can be easily proved. □

The results of this section together with an exhaustive search computation for all sequences up to length $16383 = 2^{14} - 1$ led us to the following conjecture:

Conjecture 1. Let s be an m-sequence with $\mathrm{P}(s) = m$, and p_1 be the smallest prime factor of m. Then if p_2 is any prime prime factor of m, $\mathrm{EP}_{m/p_1}(s) \leq \mathrm{EP}_{m/p_2}(s)$.

Corollary 1 becomes:

Corollary 8. *Let s be an m-sequence of period m and let p be the smallest prime factor of m. If Conjecture 1 holds then* $\mathrm{RP}(s) = \mathrm{EP}_{m/p}(s)$.

Using Theorem 5 we obtain therefore:

Corollary 9. *Let s be an m-sequence of period m. If the smallest factor of m is a Mersenne prime $2^{n'} - 1$ and Conjecture 1 holds then the minimum number of errors needed to reduce the period of s is* $\mathrm{RP}(s) = (m+1)\frac{2^{n'-1}-1}{2^{n'}}$.

We will estimate now what proportion of m-sequences are covered by Corollary 9, i.e. the proportion of Mersenne numbers that admit a Mersenne prime as their smallest factor. As previously stated, for a given prime p, the Mersenne numbers that are multiples of p are exactly those of the form $2^v - 1$ with $\mathrm{ord}_p(2)|v$. Table 1 contains $\mathrm{ord}_p(2)$ for small values of p. Note that $\mathrm{ord}_{2^n-1}(2) = n$ when $2^n - 1$ is prime.

Prime p	$\mathrm{ord}_p(2)$
3	2
5	4
7	3
11	10
13	12
17	8
19	18
23	11
29	28
31	5
37	36
41	20
43	14

Table 1. $\mathrm{ord}_p(2)$ for small prime p

We can compute how many Mersenne numbers have a particular prime as their smallest factor. Consider all sequences whose period length is divisible by 3. Since the $\mathrm{ord}_3(2) = 2$, this is half of all sequences, and whenever 3 divides the period length, it must be the smallest factor (since the period length is odd). Now consider all sequences whose period length is divisible by 5. The order of 2 mod 5 is 4, which implies that if 5 divides the period length, then so does 3,

and so 5 cannot be the smallest factor, and so we do not have to consider it. Now consider all sequences whose period length is divisible by 7. Since $\mathrm{ord}_7(2) = 3$, we have that 1/3 of all period lengths are divisible by 7. However, half of those are divisible by 3 as well, in which case 7 will not be the smallest factor. Therefore, 1/6-th of all period lengths will have 7 as their smallest factor. We can continue on in this way, using the inclusion-exclusion principle to determine how many period lengths have p as their smallest prime factor, and so determine how many sequences have the smallest factor of their period length as a Mersenne Prime. The results of these calculations for the first few Mersenne primes are contained in Table 2

Prime p	Proportion of Mersenne numbers having p as smallest factor
3	1/2
7	1/6
31	2/33
127	16/483

Table 2. Proportion of Mersenne numbers with certain factors

Adding up these results allows us to say that the proportion of m-sequences whose period length has a Mersenne prime as its smallest factor is at least 0.76, and that for each of these sequences, the period reduction value is at least $(\mathrm{P}(s) + 1)/4$ (achieved for smallest factor 3). Note that this is a large proportion of errors, as usually the largest number of errors considered is $\mathrm{P}(s)/20$ or possibly $\mathrm{P}(s)/10$. Out of the smallest 200 lengths of m-sequences, 146 of them have a Mersenne prime as their smallest factor, a proportion of 0.73. From a cryptographic standpoint this implies that using m-sequences as primitives in a cipher scheme to ensure a minimum period of the output or to provide a lower bound on linear complexity is a very secure method by this measure, since an unreasonably large number of the bits need to be changed to reduce the period or the linear complexity at all.

We can also determine the proportion of sequences which have a period length that is composite, but that does not have a Mersenne Prime as its smallest factor. Since the proportion of sequences that have prime period length will become arbitrarily small as the lengths considered increases, and we have seen that at least 0.76 of all sequences have period length that is divisible by a Mersenne Prime, the proportion that do not cannot be more than 0.24. Out of the smallest 200 lengths for m-sequences, 41 were composite with their smallest factor not a Mersenne prime, a proportion of 0.21. The smallest of these is the m-sequence of length $2047 = 2^{11} - 1$, and we have calculated by brute force that the period reduction value is 869, which is a large proportion of the weight of the sequence, which is 1024. The next smallest example will occur for the m-sequence of length $8388607 = 2^{23} - 1$.

4 Application to Grain and other Stream Ciphers

We will now apply our results to one of the eStream Candidates, namely Grain [7]. Grain is composed of a linear feedback shift register and a non-linear feedback shift register, whose outputs are combined using a non-linear function. We will be looking at the LFSR, which has 80 registers, and a primitive feedback polynomial:

$$f(x) = 1 + x^{18} + x^{29} + x^{42} + x^{57} + x^{67} + x^{80}$$

Therefore it generates an m-sequence of period length $2^{80} - 1 (\approx 1.2 * 10^{24})$ which we will refer to as s. It would clearly require a very large amount of processing power to compute by brute force either the k-error complexity or the k-error period for s, even for small values of k, but we can use the results of this paper to study the security of this sequence. To reduce the period of the sequence, we need to know the factors of its period length:

$$\begin{aligned} 2^{80} - 1 &= 3 * 5^2 * 11 * 17 * 31 * 41 * 257 * 61681 * 4278255361 \\ &= p_1 * p_2^2 * p_3 * p_4 * p_5 * p_6 * p_7 * p_8 * p_9 \end{aligned}$$

We can see that the smallest factor (3) is a Mersenne Prime. Therefore if Conjecture 1 holds the period reduction value for s is $2^{80}/4 = 2^{78} \approx 3.0 * 10^{23}$ and by Proposition 1 this is also a lower bound on the complexity reduction value. This is half the weight of s, and implies that the sequence is very secure from this point of view. We also note that the LFSR primitive is included as part of Grain not to provide complexity for the output, but to ensure that the output of the cipher has a very high minimum period. Therefore, while we have not calculated the complexity reduction value of s in this case it is more important to calculate the period reduction value which we have done.

We can go further than this, and determine the number of errors required to reduce the period to several different values. To reduce the period by 3, 15 or 31, we apply Theorem 5, to reduce the period by 5 or 11, we apply Corollary 6 and to reduce the period by 17 we apply Corollary 3 and we calculate $\mathrm{EP}_{(2^8-1)/17}$ from the decimation tables by brute force.

$$\mathrm{EP}_{(2^{80}-1)/3}(s) = (1/4)2^{80} = 2^{78} \approx 3.0 * 10^{23}$$

$$\mathrm{EP}_{(2^{80}-1)/5}(s) = 2^{80-1} - 2^{80-5}\binom{4}{2} = 10 * 2^{75} \approx 3.8 * 10^{23}$$

$$\mathrm{EP}_{(2^{80}-1)/11}(s) = 2^{80-1} - 2^{80-11}\binom{10}{5} = 386 * 2^{70} \approx 4.6 * 10^{23}$$

$$\mathrm{EP}_{(2^{80}-1)/15}(s) = (7/16)(2^{80}) = 7 * 2^{76} \approx 5.3 * 10^{23}$$

$$\mathrm{EP}_{(2^{80}-1)/17}(s) = 2^{80-8}\mathrm{EP}_{(2^8-1)/17} = 102 * 2^{72} \approx 4.8 * 10^{23}$$

$$\mathrm{EP}_{(2^{80}-1)/31}(s) = (15/32)(2^{80}) = 15 * 2^{75} \approx 5.7 * 10^{23}$$

We will also briefly provide some results that can be obtained by applying the results in this paper to other ciphers. Firstly, SSC2 [16] uses an LFSR to generate an m-sequence of length $2^{127}-1$, which is prime. Therefore, by Theorem 1 the complete error linear complexity spectrum for this sequence is: $(0, 127)$, $(2^{126}-1, 1)$, $(2^{126}, 0)$ and provided Conjecture 1 holds the period reduction value is $2^{126}-1$.

DECIM^{v2} [1] uses an LFSR to generate an m-sequence of length

$$\begin{aligned} 2^{192}-1 &= 3^2 * 5 * 7 * 13 * 17 * 97 * 193 * 241 * 257 * 641 * 673 * 65537 * \\ &\quad 6700417 * 22253377 * 18446744069414584321 \\ &\approx 6.3 * 10^{57} \end{aligned}$$

Using Theorem 5 to reduce the period by 3 or 7 and Corollary 6 to reduce the period by 5 or 13 we can calculate that:

$$\mathrm{EP}_{(2^{192}-1)/3}(s) = 2^{190} \approx 1.6 * 10^{57}$$

$$\mathrm{EP}_{(2^{192}-1)/5}(s) = 5 * 2^{188} \approx 2.0 * 10^{57}$$

$$\mathrm{EP}_{(2^{192}-1)/7}(s) = (3/8) * 2^{192} \approx 2.4 * 10^{57}$$

$$\mathrm{EP}_{(2^{192}-1)/13}(s) = 793 * 2^{181} \approx 2.4 * 10^{57}$$

If Conjecture 1 holds, the period reduction value for s is 2^{190}.

LILI-128 [2] uses two LFSRs both of which generate m-sequences. The first is of length

$$2^{39}-1 = 7 * 79 * 8191 * 121369 \approx 5.5 * 10^{11}$$

and so by Theorem 5 we can say that

$$\mathrm{EP}_{(2^{39}-1)/7}(s) = 2^{39} * (3/8) \approx 2.1 * 10^{11}$$

The second is of length $2^{89}-1$ which is prime, and so by Theorem 1 its error complexity spectrum is $(0, 89)$, $(2^{88}-1, 1)$, $(2^{88}, 0)$ and provided Conjecture 1 holds the period reduction value is $2^{88}-1$.

5 Conclusions

In this paper we have studied the k-error linear complexity and k-error period of m-sequences. We have shown that although the general problem of determining these values is likely to be difficult, there are certain cases where we can find results. We have fully solved the case where the period length of the m-sequence is prime. We have shown how in general the problem of determining the period reduction value can be reduced to an equivalent problem for smaller sequences and we have provided a closed form expression for the number of errors needed to reduce the period by a Mersenne number or by a prime p where $\mathrm{ord}_p(2) = p-1$. Subject to a conjecture we have provided results for the number of errors needed to reduce the period for a large proportion (76%) of m-sequences. Finally we have applied these results to several stream cipher primitives, namely Grain, SSC2, DECIM^{v2} and LILI-128.

References

1. C. Berbain, O. Billet, A. Canteaut, N. Courtois, B. Debraize, H. Gilbert, L. Goubin, A. Gouget, L. Granboulan, C. Lauradoux, M. Minier, T. Pornin, and H. Silbert. DECIM^{v2}. *eStream Candidate, ECRYPT Stream Cipher Workshop SKEW 2005*.
2. E. Dawson, A. Clark, J. Golić, W. Millan, L. Penna, and L. Simpson. The LILI-128 Keystream Generator. *Proc. 1st NESSIE Workshop*, 2000.
3. C. Ding, G. Xiao, and W. Shan. *The Stability Theory of Stream Ciphers*, volume 561 of *LNCS*. Springer Verlag, 1991.
4. C. Ding. The Weight Distribution of Some Irreducible Cyclic Codes. *IEEE Transactions on Information Theory*, 55(3):955–960, March 2009.
5. P. Ekdahl and T. Johansson. Another Attack on A5/1. *IEEE Transactions on Information Theory*, 49(1), 2003.
6. S. W. Golomb. *Shift Register Sequences*. Aegean Park Press, 1982.
7. M. Hell, T. Johansson, and W. Meier. Grain - A Stream Cipher for Constrained Environments, 2005. *International Journal of Wireless and Mobile Computing, Special Issue on Security of Computer Network and Mobile Systems*, 2006
8. H. Hu. Periods on Two Kinds of Nonlinear Feedback Shift Registers with Time Varying Feedback Functions *Technical Reports, Center for Applied Cryptographic Research*, 2011
9. T. Kaida, S. Uehara, and K. Imamura. An algorithm for the k-error linear complexity of sequences over $GF(p^m)$ with period p^n, p a prime. *Inform. Comput.*, 151:134–147, 1999.
10. A. G. B. Lauder and K. G. Paterson. Computing the error linear complexity spectrum of a binary sequence of period 2^n. *IEEE Transactions on Information Theory*, 49:273–280, 2003.
11. F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, 1978.
12. W. Meidl, H. Aly, and A. Winterhof. On the k-error linear complexity of cyclotomic sequences. *Journal Mathematical Cryptography*, 2007.
13. C. Pomerance. Recent developments in primality testing. *The Mathematical Intelligencer*, 3:97–105, 1981.
14. M. Stamp and C. F Martin. An algorithm for the k-error linear complexity of binary sequences of period 2^n. *IEEE Transactions on Information Theory*, 39:1398–1401, 1993.
15. F. Surböck and H. Weinrichter. Interlacing Properties of Shift-Register Sequences with Generator Polynomials Irreducible over GF(p). *IEEE Transactions on Information Theory*, 24(3):386–389, 1978.
16. M. Zhang, C. Carroll, and A. H. Chan. The Software-Oriented Stream Cipher SSC2. *Fast Software Encyrption Workshop*, 2000.