Abstract: Decimation-based sequence generators are a class of non-linear cryptographic generators designed to be used in hardware implementations. An inherent characteristic of such generators is that their output sequences are interleaved sequences. This profitable characteristic can be used in the cryptanalysis of those generators. In this work, emphasis is on the most representative decimation-based generator, the shrinking generator, which has been cryptanalyzed just by solving linear equation systems. Compared with previous cryptanalysis, computational complexity and intercepted sequence requirements are dramatically reduced. Although irregularly decimated generators have been conceived and designed as non-linear sequence generators, in practice they can be easily analyzed in terms of simple linear structures.

Keywords: Decimation, Shrinking generator, Interleaved sequence, Stream cipher, Secret-key cryptography

MSC: 94A55, 94A60

1 Introduction

Nowadays stream ciphers are the fastest among the encryption procedures. They are designed to generate, from a short key, a long sequence (keystream sequence) of seemingly random bits. Some well known designs in stream ciphers can be found in [1, 2]. Typically, a stream cipher consists of a keystream generator whose output sequence is bit-wise XORed with the plaintext (in emission) to obtain the ciphertext or with the ciphertext (in reception) to recover the original plaintext. References [3–5] provide a solid introduction to the study of stream ciphers.

There are many proposals of keystream generators that are based on maximal-length Linear Feedback Shift Registers (LFSRs) [6]. Such registers are linear structures characterized by their length $L$, their characteristic polynomial $p(x)$ and their initial state is (currently the key of the cryptosystem). Their output sequences, the so-called PN-sequences, are usually combined in a non-linear way in order to break their linearity and to produce new pseudorandom sequences of cryptographic application. LFSRs with dynamic feedback, clock-controlled generators, nonlinear filters or irregularly decimated generators are just some of the most popular keystream generators, see above references.
Irregularly decimated generators produce sequences with good cryptographic properties: long periods, right correlation, excellent run distribution, balancedness, simplicity of implementation, etc. The underlying idea of this kind of generators is the irregular decimation of a PN-sequence according to the bits of another one. The result of this decimation is a binary sequence that will be used as keystream sequence in the cryptographic procedure of stream cipher.

Inside the family of irregularly decimated generators, we can enumerate:

1. The shrinking generator proposed by Coppersmith, Krawczyk and Mansour [7] that involves two LFSRs.
2. The self-shrinking generator designed by Meier and Staffelbach [8] involving only one LFSR.

In addition, different linear structures based in Cellular Automata that model such generators can also be found in the literature [11–13].

This work focuses on the most representative element in the class of decimation-based sequence generators: the shrinking generator. Taking advantage of the fact that its output sequence is an interleaved sequence, a simple cryptanalytic attack has been developed. The basic ideas of this attack can be generalized to other elements in the same class of generators.

The paper is organized as follows: in Section 2 fundamentals and basic concepts are provided. In Section 3, we introduce some important properties of the shrinking generator that will be used in Section 4 to perform a recovering algorithm for the generated sequence. Section 5 compares the attack here presented with other ones found in the literature. Finally, conclusions in Section 6 end the paper.

2 Preliminaries

Notation and basic concepts are now introduced. First of all, we introduce the concept of decimation, which will be used repeatedly throughout this paper. Let \( \{u_i\} (i = 0, 1, 2, \ldots) \) be a linear recursive sequence over a finite field. The **decimation** of the sequence \( \{u_i\} \) by \( d \) is a new sequence obtained by taking every \( d \)-th term of \( \{u_i\} \) [14].

Next, the definition of interleaved sequence is provided [15].

**Definition 2.1.** Let \( g(x) \) be a polynomial of degree \( r \) over \( GF(q) \) (the Galois field of \( q \) elements) and let \( n \) be a positive integer. For any sequence \( w = \{w_k\} \) over \( GF(q) \), write \( k = in + j \) (\( i = 0, 1, 2, \ldots \ j = 0, \ldots, n - 1 \)). If all the subsequences \( w_j = \{w_{ijn+j}\} \) \( (j = 0, \ldots, n - 1) \) are generated by \( g(x) \), then \( w \) is called an interleaved sequence over \( GF(q) \) of size \( n \) associated with \( g(x) \).

We can write \( w = (w_0, w_1, \ldots, w_{n-1}) \) where each \( w_j \) \( (j = 0, \ldots, n - 1) \) is a subsequence of \( w \). In fact, each \( w_j \) is an \( n \)-decimation of the sequence \( w \) obtained from such a sequence by taking one out of \( n \) terms. In the sequel, \( GF(q) \) will be the binary field \( GF(2) \).

The **shrinking generator** (SG) was first introduced in [7]. It is made up of two maximal-length LFSRs denoted by \( R_1 \) and \( R_2 \). Let \( L_1 \) and \( L_2 \) \( (L_1 < L_2) \) be the LFSR lengths, the primitive polynomials \( p_1(x) \), \( p_2(x) \) their characteristic polynomials, and \( s_1 \) and \( s_2 \) their initial states, respectively. Moreover, let \( \{a_i\} \) and \( \{b_i\} \) be the PN-sequences generated by \( R_1 \) and \( R_2 \), respectively. In this case, the sequence \( \{a_i\} \) decimates the other sequence \( \{b_i\} \). The decimation rule is very simple: given two bits \( a_i \) and \( b_i \), the output sequence of the
generator \( \{ s_k \} \) is computed as

\[
\begin{align*}
\text{If } a_i = 1 & \text{ then } s_k = b_i \\
\text{If } a_i = 0 & \text{ then } b_i \text{ is discarded.}
\end{align*}
\]

We call the sequence \( \{ s_k \} \) as the **shrunked sequence** (SS). Assume that \( \gcd(L_1, L_2) = 1 \), then the period of SS is \( T = 2^{L_i - i}(2^{L_2} - 1) \).

The linear complexity of a sequence, denoted by \( LC \), is defined as the length of the minimum LFSR that generates such a sequence. As \( \gcd(L_1, L_2) = 1 \), then the linear complexity of the shrunked sequence is given by \( L_2 2^{L_1 - 2} < LC \leq L_2 2^{L_1 - 1} \). Moreover, its characteristic polynomial is of the form \( p(x)^m \) where \( p(x) \) is a primitive polynomial of degree \( L_2 \) and \( m \) an integer satisfying \( 2^{L_1 - 2} < m \leq 2^{L_1 - 1} \).

As usual, the key of this generator is the initial state of the both registers \( R_1 \) and \( R_2 \).

Next a simple illustrative example is introduced.

**Example 2.2.** Consider two LFSRs \( R_1 \) and \( R_2 \) with lengths \( L_1 = 2 \) and \( L_2 = 3 \), characteristic polynomials \( p_1(x) = 1 + x + x^3 \) and \( p_2(x) = 1 + x^2 + x^3 \), and initial states \( s_{i_1} = (1, 0) \) and \( s_{i_2} = (1, 0, 0) \), respectively.

The shrunked-sequence can be computed as follows:

\[
\begin{align*}
\{ a_i \} & : 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 0 \\
\{ b_i \} & : 1 \ 1 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 1 \\
\{ s_k \} & : 1 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0 \ 1 \\
\end{align*}
\]

The shrunked sequence \( \{ s_k \} \) has period 14 and it is easy to check that its characteristic polynomial is \( p(x)^2 = (1 + x + x^3)^2 \), consequently its linear complexity equals 6.

### 3 Linear properties of the shrunked sequence

In this section, we highlight some properties of the shrunked sequence, which will be used in the algorithm proposed in Section 4. As before, we consider two LFSRs \( R_1 \) and \( R_2 \) with lengths \( L_1 \) and \( L_2 \), characteristic polynomials \( p_1(x), p_2(x) \) and initial states \( s_{i_1} \) and \( s_{i_2} \). In addition, \( T_1 = 2^{L_1} - 1 \) and \( T_2 = 2^{L_2} - 1 \) are the periods of their corresponding PN-sequences \( \{ a_i \} \) and \( \{ b_i \} \), respectively.

According to Definition 2.1, the shrunked sequence \( s = \{ s_k \} \) can be written as \( s = \{ s_0, s_1, \ldots, s_{n-1} \} \) where \( n = 2^{L_1 - i} \). In fact, every subsequence \( s_j \) \( (j = 0, \ldots, n-1) \) is a PN-sequence generated by the \( L_2 \)-degree primitive polynomial \( p(x) \) defined as

\[
p(x) = \prod_{i=0}^{L_2 - 1} (x + \alpha^{e_i}),
\]

where \( e_i = 2^i \cdot T_1 \mod T_2 \) and \( \alpha \) is a root of the polynomial \( p_2(x) \). Recall that every subsequence \( s_j \) is just a decimation of \( \{ b_i \} \) by \( d = 2^{L_1} - 1 \), thus the resulting sequence is a PN-sequence too. In brief, \( e_i \) \( (i = 0, \ldots, L_2 - 1) \) are the elements of the cyclotomic coset \( 2^{L_1 - 1} \) and \( p(x) \) is the polynomial associated with such a coset [6]. The subsequences \( s_j \) \( (j = 0, \ldots, n-1) \) are called the **interleaved PN-sequences** of the shrunked sequence.

**Example 3.1.** Consider two LFSRs \( R_1 \) and \( R_2 \) with lengths \( L_1 = 3 \) and \( L_2 = 4 \), characteristic polynomials \( p_1(x) = 1 + x + x^3 \) and \( p_2(x) = 1 + x + x^4 \) and initial states \( s_{i_1} = (1, 0, 0) \) and \( s_{i_2} = (1, 0, 0, 0) \), respectively. The shrunked sequence has period \( T = 60 \) and its characteristic polynomial is \( p(x)^5 = (1 + x^3 + x^4)^5 \). Since the shrunked sequence is an interleaved sequence, it is composed of 4 PN-sequences:
All of them have the same characteristic polynomial $p(x) = 1 + x^3 + x^4$, thus there is a unique PN-sequence but shifted. This shift depends on the positions of the 1s in the PN-sequence $\{a_i\}$.

Let $\{i_0, i_1, \ldots, i_{2^{L_i-1}-1}\}$ denote the position of the $2^{L_i-1}$ ones in the PN-sequence $\{a_i\}$ and let $\delta$ be an integer such that $(2^{L_i-1}) \delta = 1 \mod (2^{L_i} - 1)$. Let also $d_j (j = 1, 2, \ldots, 2^{L_i-1} - 1)$ be the position over $s_0$ of the first element of each subsequence $s_j (j = 1, \ldots, n-1)$, respectively. If we know such positions $d_j$ over $s_0$, then we can compute the indices $i_j$ by means of the following expressions:

$$d_j = \delta \cdot i_j \mod 2^{L_i} - 1, \quad \text{for } j = 1, 2, \ldots, 2^{L_i-1} - 1. \quad (1)$$

In Example 3.1, we had four interleaved subsequences $s_0, s_1, s_2$ and $s_3$. It is easy to check that $d_1 = 9, d_2 = 5$ and $d_3 = 3$. In this case, $T_1 = 7$ and $T_2 = 15$, then $\delta = 13$. With this information, we can determine the position of the ones in $\{a_i\} (i_0 = 0, \text{ without loss of generality})$:

$$13 \cdot i_1 = 9 \mod 15 \Rightarrow i_1 = 3$$
$$13 \cdot i_2 = 5 \mod 15 \Rightarrow i_2 = 5$$
$$13 \cdot i_3 = 3 \mod 15 \Rightarrow i_3 = 6$$

Therefore, the set of indices is given by $\{0, 3, 5, 6\}$ and the PN-sequence $\{a_i\}$ is given by $\{1, 0, 0, 1, 0, 1, 1\}$.

In the algorithm proposed in Section 4, the opposite situation occurs. In that case, we know the position of the ones in the PN-sequence $\{a_i\}$ and we compute the position of the first element of each subsequence $s_j$ in $s_0$ by means of the expressions given in (1).

The presence of PN-sequences inside the shrunken sequence reveals severe dependencies among its bits. These linear relationships will be advantageously used in the proposed attack. In fact, given $N$ intercepted bits of this sequence, the goal is to determine the pair of initial states $(i_{s_1}, i_{s_2})$ of both registers.

## 4 Cryptanalytic attack

Prior to the attack’s description, the following notation is introduced:

- $is_1 = (a_{i_0}, a_{i_1}, \ldots, a_{i_{t-1}})$, $is_2 = (b_{i_0}, b_{i_1}, \ldots, b_{i_{t-1}})$
- $S = \{s_0, s_1, \ldots, s_{N-1}\}$ are the $N$ intercepted bits of the shrunken sequence. Currently, the number $N$ can be written as $N = N_1 + N_2$ where $N_1$ bits are used to compute the pair $(is_1, is_2)$ while $N_2$ bits are used to check the correctness of the previous pair.
\( \delta \) as before is an integer \( \delta \in \{1, 2, 3, \ldots, T_2 - 1\} \), such that \( T_1 \delta = 1 \mod T_2 \).

The \( N_1 \) intercepted bits are elements of any interleaved PN-sequence \( s_i \). Nevertheless, in this attack we only focus on the first interleaved PN-sequence \( s_0 \). For simplicity it will be denoted by \( \{u_i\} \) \((i = 0, 1, \ldots, 2^{L_2} - 2)\). According to the properties of the PN-sequences, any term \( u_i \) of \( \{u_i\} \) can be expressed as a function of the first \( L_2 \) bits \((u_0, u_1, \ldots, u_{L_2-1})\) by means of the modular expression

\[
q(x) = x^k \mod p(x),
\]

where \( q(x) = c_{L_2-1}x^{L_2-1} + \ldots + c_1x + c_0 \) with \( c_i \in GF(2) \). Thus,

\[
u_k = c_{L_2-1}u_{L_2-1} + \ldots + c_1u_1 + c_0u_0.
\]

This cryptanalytic attack is based on solving systems of linear equations of the form:

\[
A \mathbf{x} = \mathbf{b},
\]

where \( A \) is an \((N_1 \times L_2)\) binary coefficient matrix, \( \mathbf{x} \) is the \((L_2 \times 1)\) vector of unknowns and \( \mathbf{b} \) is the \((L_2 \times 1)\) right side vector of intercepted bits. Each initial state \( is_1 \) parametrises the coefficient matrix \( A \), then the Linear Consistency Test (LCT) [16] checks the consistency of the corresponding equation system (2). If \( is_1 \) considered is the right initial state, then the equation system certainly will be consistent. On the other hand, if \( is_1 \) is not the initial state used in the generation of the intercepted bits, then by [16, Theorem 1] the consistency probability of the system will be very small when the intercepted segment is long enough. In order to make the number of false consistency alarms as small as possible, the number of equations in (2) should exceed \( L_1 + L_2 \) significantly, see [16] and [17].

The attack is divided into two phases. In phase 1, we check the \( 2^{L_1-1} \) initial states \( is_1 \) starting by 1 (as only the 1s of \( \{a_i\} \) generate bits in the shrunken sequence) to determine a set \( Q \) of possible candidates to initial state of \( R_1 \). In phase 2, for every \( is_1 \) in \( Q \) its corresponding \( is_2 \) will be computed. The pair \( (is_1, is_2) \) able to generate all the intercepted shrunken sequence will be the key of the cryptosystem. In brief, the algorithm can be described as follows:

**INPUT** : The lengths \( L_1 \) and \( L_2 \) of both registers, the characteristic polynomials \( p_1(x), p_2(x) \) and the \( N \) intercepted bits \( S = \{s_0, s_1, \ldots, s_{N-1}\} \) of the shrunken sequence.

1. Computation of PHASE 1
2. Computation of PHASE 2

**OUTPUT** : The initial states \( is_1 \) and \( is_2 \) (key of the cryptosystem) that generate the shrunken sequence.

In the sequel, the whole attack is described in detail.

**PHASE 1:**

For each \( is_1 \) considered do:

1. Starting in \( is_1 \), generate a portion of sequence \( \{a_i\} \) until \( N_1 \) ones are obtained. Such ones will be located at positions \( i_k \) \((k = 0, 1, \ldots, N_1 - 1)\) over \( \{a_i\} \).
2. Determine \( N_1 \) positions in the sequence \( \{u_i\} \) as

\[
d_k = \delta \cdot i_k \mod T_2 \quad (k = 0, 1, \ldots, N_1 - 1).
\]

3. Assign the \( N_1 \) intercepted bits to the previous positions

\[
u_{d_k} = s_k \quad (k = 0, 1, \ldots, N_1 - 1).
\]

4. Express each \( u_{d_k} \) as a function of the first \( L_2 \) terms of \( \{u_i\} \), that is \( u_{d_k} = f_k(u_0, u_1, \ldots, u_{L_2-1}) \), by means of

\[
x^{d_k} \mod p(x) \quad (k = 0, 1, \ldots, N_1 - 1).
\]
It turns out to be a system of linear equations

\[
\begin{align*}
u_{ik} &= f_k(u_0, u_1, \ldots, u_{L-1}) = s_k \\
(k = 0, 1, \ldots, N_1 - 1) &\text{ with } N_1 \text{ equations in the } (u_0, u_1, \ldots, u_{L-1}) \text{ unknowns.}
\end{align*}
\]

2. Apply the Linear Consistency Test (LCT) to check the consistency of the previous system, 
   if the system is consistent, then include \(i_s1\) in \(Q\) 
   else \(i_s1\) is rejected.

end do

The result of this phase is the set \(Q\) of possible candidates to initial state of LFSR \(R_1\). Once the set \(Q\) has been computed, the second step of the attack is performed.

PHASE 2:
For each \(i_s1\) in \(Q\) do:

1. Express each \(b_{i_1}\) as a function of the first \(L_2\) terms of \(\{b_i\}\), \(\text{that is } b_{i_1} = g_k(b_0, b_1, \ldots, b_{L_2-1})\), by means of
   
   \[x^k \mod p_2(x) \quad (k = 0, 1, \ldots, N_1 - 1).\]

   It turns out to be a system of linear equations

   \[
   \begin{align*}
b_{i_1} &= g_k(b_0, b_1, \ldots, b_{L_2-1}) = s_k \\
(k = 0, 1, \ldots, N_1 - 1) &\text{ with } N_1 \text{ equations in the } (b_0, b_1, \ldots, b_{L_2-1}) = i_s2 \text{ unknowns.}
\end{align*}
\]

2. Apply the Linear Consistency Test (LCT) to check the consistency of the previous system, 
   if the system is not consistent, then reject \((i_s1, i_s2)\) 
   else if the pair \((i_s1, i_s2)\) can generate the shrunken sequence by using the \(N_2\) bits for checking, 
   then cryptosystem broken !!! 
   else \(i_s1\) is rejected.

end do

The result of this phase is the pair \((i_s1, i_s2)\) generating the shrunken sequence, that is the key of the cryptosystem.

A software implementation of the previous attack has been performed on a laptop device with the following specifications:

- Operative system: Arch Linux
- CPU: Dual core Intel Core i7-4510U, Cache 4096 KB, Freq. 3100 MHz
- RAM: 8 GB, Type: DDR3
- Hard Disk: Type SSD, Size 256.1 GB

Some numerical results are depicted in Table 1 where \(L_1, L_2\) are the lengths of registers \(R_1\) and \(R_2\), respectively, 
\(T\) is the period of the corresponding shrunken sequence, \(N_1\) is the number of intercepted bits for computation, 
\(c(Q)\) is the cardinality of \(Q\), that is the number of candidates to initial state of \(R_1\), and \(t\) is the running time expressed in seconds. It must be noticed that the period of the shrunken sequence is much greater than 
the number of intercepted bits needed to successfully run the algorithm within a reasonable time. For our 
computations, \(N_1 = 2 \cdot L_2\) while \(N_2\) is chosen \(N_2 = N_1\). In brief, the requirements of intercepted sequence are 
extremely low. In Table 2, the same results are shown but now the number of intercepted bits \(N_1\) equals \(L_2\). 
In this case, since \(N_1\) has been reduced, the execution time has been reduced too. Nevertheless, the number 
of candidates has grown considerably. Table 3 shows the numerical results corresponding to the verification 
of a unique initial state \(i_s1\) in the phase 1 of the algorithm. Recall that even for large values of \(L_1\) and \(L_2\) the 
execution time of such routine is very low.
The most remarkable features of the proposed attack are:

1. The low amount of intercepted bits needed for its execution. Indeed, $N_1 = n \cdot L_2$, $n$ being a small integer ($n = 2, 3, 4$), and $N_2 \leq N_1$. Thus the amount of sequence required is linear in the length of the register $R_2$.

2. The running time of the attack is dominated by phase 1 which has a time complexity of $O(2^{L_1-1} \cdot (N_1 \times L_2)^3)$, that is exponential in $L_1$ due to the number of $i$s, considered and polynomial in $L_2$. In fact, the work factor needed for each test is that of the Gauss elimination algorithm applied to the augmented matrix $(A, b)$, which is cubic in the dimension of the matrix. In any case, the cubic factor is irrelevant compared with the exponential factor.

3. Both phases 1 and 2 are fully parallelizable and some tweaks can be made to optimize the LCT step.

### Table 1. Numerical results for the algorithm

<table>
<thead>
<tr>
<th>$L_1$</th>
<th>$L_2$</th>
<th>$T$</th>
<th>$N_1$</th>
<th>$c(Q)$</th>
<th>$t$(sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5</td>
<td>248</td>
<td>10</td>
<td>1</td>
<td>0.0064</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>1008</td>
<td>12</td>
<td>1</td>
<td>0.0173</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>261888</td>
<td>20</td>
<td>1</td>
<td>0.3856</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>1048064</td>
<td>22</td>
<td>1</td>
<td>0.8552</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>4193280</td>
<td>24</td>
<td>1</td>
<td>1.8114</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>16775168</td>
<td>26</td>
<td>1</td>
<td>4.2623</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>67104768</td>
<td>28</td>
<td>1</td>
<td>9.0739</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>268427264</td>
<td>30</td>
<td>1</td>
<td>20.0681</td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>$1.0737 \times 10^9$</td>
<td>32</td>
<td>1</td>
<td>44.9963</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>$4.2949 \times 10^9$</td>
<td>34</td>
<td>2</td>
<td>98.1865</td>
</tr>
<tr>
<td>17</td>
<td>18</td>
<td>$1.7180 \times 10^{10}$</td>
<td>36</td>
<td>1</td>
<td>217.9489</td>
</tr>
<tr>
<td>18</td>
<td>19</td>
<td>$6.8719 \times 10^{10}$</td>
<td>38</td>
<td>2</td>
<td>477.1288</td>
</tr>
<tr>
<td>19</td>
<td>20</td>
<td>$2.7488 \times 10^{11}$</td>
<td>40</td>
<td>1</td>
<td>1092.7125</td>
</tr>
<tr>
<td>20</td>
<td>21</td>
<td>$1.0995 \times 10^{12}$</td>
<td>42</td>
<td>1</td>
<td>2327.2800</td>
</tr>
<tr>
<td>21</td>
<td>22</td>
<td>$4.3980 \times 10^{12}$</td>
<td>44</td>
<td>1</td>
<td>4997.0925</td>
</tr>
</tbody>
</table>

### Table 2. Numerical results for the algorithm when $N_1 = L_2$

<table>
<thead>
<tr>
<th>$L_1$</th>
<th>$L_2$</th>
<th>$T$</th>
<th>$N_1$</th>
<th>$c(Q)$</th>
<th>$t$(sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5</td>
<td>248</td>
<td>5</td>
<td>5</td>
<td>0.0046</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>1008</td>
<td>6</td>
<td>14</td>
<td>0.0099</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>4064</td>
<td>7</td>
<td>25</td>
<td>0.0216</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>16320</td>
<td>8</td>
<td>46</td>
<td>0.0513</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>65408</td>
<td>9</td>
<td>78</td>
<td>0.11969</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>261888</td>
<td>10</td>
<td>160</td>
<td>0.2478</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>1048064</td>
<td>11</td>
<td>210</td>
<td>0.7123</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>4193280</td>
<td>12</td>
<td>708</td>
<td>1.3290</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>16775168</td>
<td>13</td>
<td>1183</td>
<td>3.1078</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>67104768</td>
<td>14</td>
<td>2227</td>
<td>6.0204</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>268427264</td>
<td>15</td>
<td>4494</td>
<td>13.0011</td>
</tr>
<tr>
<td>15</td>
<td>16</td>
<td>$1.0737 \times 10^9$</td>
<td>16</td>
<td>8710</td>
<td>29.4033</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>$4.2949 \times 10^9$</td>
<td>17</td>
<td>6183</td>
<td>57.9891</td>
</tr>
<tr>
<td>17</td>
<td>18</td>
<td>$1.7180 \times 10^{10}$</td>
<td>18</td>
<td>35351</td>
<td>151.4661</td>
</tr>
</tbody>
</table>
The program makes use of SageMath, an algebraic computation systems based on Python. In order to handle polynomials over $GF(2)$, SageMath uses the libraries NLT. In order to compute with matrices over $GF(2)$, SageMath uses the libraries M4RI. In the LCT application, the system of equations is transformed into a low reduced echelon form. This step is important in the computation efficiency as the system consistency is reduced to test the existence of a row $(0, 0, \ldots, 0, 1)$ in the coefficient matrix of the system.

Table 3. Numerical results for the verification of one is 1

<table>
<thead>
<tr>
<th>$L_1$</th>
<th>$L_2$</th>
<th>$N_1$</th>
<th>$t$(sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6</td>
<td>12</td>
<td>0.00080</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>16</td>
<td>0.00112</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>22</td>
<td>0.00169</td>
</tr>
<tr>
<td>20</td>
<td>21</td>
<td>42</td>
<td>0.00911</td>
</tr>
<tr>
<td>30</td>
<td>31</td>
<td>62</td>
<td>0.01044</td>
</tr>
<tr>
<td>40</td>
<td>41</td>
<td>82</td>
<td>0.01980</td>
</tr>
<tr>
<td>50</td>
<td>51</td>
<td>102</td>
<td>0.03160</td>
</tr>
<tr>
<td>59</td>
<td>60</td>
<td>120</td>
<td>0.03547</td>
</tr>
<tr>
<td>60</td>
<td>61</td>
<td>122</td>
<td>0.03794</td>
</tr>
<tr>
<td>61</td>
<td>62</td>
<td>124</td>
<td>0.03806</td>
</tr>
<tr>
<td>62</td>
<td>63</td>
<td>126</td>
<td>0.04035</td>
</tr>
<tr>
<td>63</td>
<td>64</td>
<td>128</td>
<td>0.04108</td>
</tr>
</tbody>
</table>

5 Other attacks over the shrinking generator

Other attacks against the shrinking generator have been designed in the literature. For example, in [18], the authors proposed two fault cryptoanalysis. In that work, the attacker is supposed to have a device implementing the shrinking generator and can use it freely. They also assume that the base and control generators of the shrinking generator output bits according to the uniform distribution over $GF(2)$ and that an attacker can disturb clocking of the device, that is, he can stop the control sequence for a couple of steps, and observe the output of the generator. These attacks require injecting specific faults and restarting the device with partially the same internal state. While injecting such faults is potentially possible, it may require some design faults (so that potentially vulnerable parts of the device were placed on external layers). It shows at least that a careful examination of a chip design might be necessary. Furthermore, on the first cryptanalysis, there exists a probability of false solution and algorithm failure. As a consequence, they have to assume that the number of 0s between two 1s does not exceed a certain parameter $\maxzeros$. They proved that the probability of a false result grows rapidly with the assumed length of the gap between the 1s. That is why they assume that the control sequence does not contain a block of more than $\maxzeros$ 0s. Of course, when this assumption is false, the algorithm fails.

Several correlation attacks against the shrinking generator have been proposed too. A correlation attack was proposed in [19] and was experimentally analyzed in [20], where an exhaustive search through all initial states and all possible feedback polynomials of $R_1$ was performed. Later, in [21] the author presented a reduced complexity correlation attack based on searching for specific subsequences of the keystream sequence, whose complexity and required keystream length are both exponential in the length of $R_2$.

A few years later, in [22] Golić conducted a probabilistic correlation analysis based on a recursive computation of the posterior probabilities of individual bits of $R_2$, which revealed the possibility of implementing certain type of fast correlation attacks on the shrinking generator. A novel distinguishing attack was also proposed in [23]. In a subsequent paper [24], the author proposed an improved linear consistency attack based on an exhaustive search through all initial states of $R_1$. 

Unauthenticated
In [22], the author conjectured that the shrinking generator could be vulnerable against fast correlation attacks that would not require an exhaustive search through all possible initial states. In [25], the authors tried to answer this question with length of $R_2$ equal to 61 (as suggested in [26]). They claimed that given 140000 keystream bits, the initial state of $R_2$ with arbitrary weight characteristic polynomial of degree 61 could be recovered with success probability higher than 99% and complexity $2^{56}$, which was a good trade-off between these parameters.

In brief, the algorithm here developed presents two main advantages against other proposals. First, compared with other cryptanalytic attacks, the original key of the cryptosystem is always obtained. As pointed in [16], there is a trade-off between the number of equations to consider and the false positive ratio. Nevertheless, in our experiments we consider a minimum number of equations and in most cases only the original key was retrieved. Furthermore, with the knowledge of the LFSRs’ parameters the attacker just needs to intercept a part of the keystream sequence and perform the algorithm; our method does not need further assumptions. Second, the results given in Table 1 show that the required keystream length in our algorithm grows linearly in the length of $R_2$, in contrast with other proposals where the amount of required sequence is exponential in the length of any register.

6 Conclusions

The shrinking generator obtains an implicit non-linearity originated from the decimation process. This process is an attempt to create strong pseudorandom sequences with cryptographically good properties out of weak components. It is proved that the shrunken sequence has a long period, a desirable high linear complexity and good statistical properties. However, the linear properties presented in this work make this generator vulnerable against attacks. This paper presents a cryptanalysis over the shrinking generator based on solving linear systems. Besides, the number of intercepted bits needed to successfully perform the algorithm is substantially lower than the period of the sequence, growing linearly with the length of the register $R_2$.

Acknowledgement: The work of the first author was supported by FAPESP with process number 2015/07246-0. This research has been partially supported by Ministerio de Economía, Industria y Competitividad (MINECO), Agencia Estatal de Investigación (AEI), and Fondo Europeo de Desarrollo Regional (FEDER, UE) under project COPCIS, reference TIN2017-84844-C2-1-R, and by Comunidad de Madrid (Spain) under project reference S2013/ICE-3095-CIBERDINE-CM, also co-funded by European Union FEDER funds.

References

An improved cryptanalysis on the shrinking generator


