A New Statistical Distinguisher for the Shrinking Generator

Jovan Dj. Golić and Renato Menicocci

Abstract

The shrinking generator is a well-known keystream generator composed of two linear feedback shift registers, LFSR_1 and LFSR_2, where LFSR_1 is clock-controlled according to regularly clocked LFSR_2. The keystream sequence is thus a decimated LFSR_1 sequence. Statistical distinguishers for keystream generators are algorithms whose objective is to distinguish the keystream sequence from a purely random sequence. Previously proposed statistical distinguishers for the shrinking generator are based on detecting binary linear relations in the keystream sequence that hold with a probability sufficiently different from one half. In this paper a novel approach which significantly reduces the required computation time is introduced. It is based on a probabilistic reconstruction of the bits in the regularly clocked LFSR_1 sequence that satisfy the LFSR_1 recurrence or any linear recurrence derived from low-weight multiples of the LFSR_1 characteristic polynomial. The keystream sequence length and the computation time required for a reliable statistical distinction are analyzed both theoretically and experimentally.

Key words. Stream ciphers, irregular clocking, posterior probabilities, statistical distinguishers.

1 Introduction

The shrinking generator [1] is a well-known keystream generator for stream cipher applications. It consists of only two linear feedback shift registers (LFSR's). The clock-controlled LFSR, LFSR_1, is irregularly clocked according to the clock-control LFSR, LFSR_2, which is regularly clocked. More precisely, at each time, both LFSR's are clocked once and the bit produced by LFSR_1 is taken to the output if the clock-control bit produced by LFSR_2 is equal to 1. Otherwise, the output bit is not produced. The output sequence is thus a nonuniformly decimated LFSR_1 sequence. It is recommended in [1] that the LFSR initial states and the feedback polynomials be defined by the secret key. Under certain conditions, the output sequences possess a long period, a high linear complexity, and good statistical properties.

The basic divide-and-conquer attacks on LFSR_2 and LFSR_1 are given in [9] and [3], respectively, and both require the exhaustive search through the initial states and feedback polynomials of the respective LFSR. A number of algorithms which may yield faster attacks on LFSR_1 are proposed in [4], [8], and [7].

This paper deals with statistical distinguishers to be used for distinguishing the output sequence of the shrinking generator from a purely random sequence.[1] It is shown in [2] and [5] that the output sequence may have a detectable linear statistical weakness if the feedback polynomial of LFSR_1 has low-weight polynomial multiples of moderately large degrees. More precisely, with a probability considerably different from one half, the output sequence satisfies the linear recurrences corresponding to the so-called shrunken versions of these polynomials. These linear recurrences are called the linear models for the shrinking generator. Any such linear model directly yields a linear statistical distinguisher for the shrinking generator. It is argued in [4] that this weakness may even be used for a faster reconstruction of the feedback polynomial as well as the initial state of LFSR_1.

Our objective is to introduce another type of statistical distinguishers which, instead of looking for linear relations in the output sequence, aim at estimating the linear relations in the regularly clocked LFSR_1 sequence by using the posterior probabilities of individual bits of the regularly clocked LFSR_1 sequence when conditioned on appropriate segments of the output sequence. The LFSR_1 feedback polynomial is assumed to be known. In the probabilistic model where the LFSR sequences are assumed to be independent and purely random, these probabilities can be computed by the algorithm derived in [7]. In addition, when the bits satisfying the linear relations in the LFSR_1 sequence are very close to each other so that the independence assumption behind the bitwise approach is less realistic, the algorithm from [7] is extended to deal with blocks of bits instead of individual bits. A theoretical analysis shows that the novel approach takes a significantly smaller computation time than the linear model approach from [2] and [5]. The theoretical analysis is supported by experimental results obtained by computer simulations.

Section 2 contains a concise overview of the linear model approach from [2], [4], and [5]. Algorithms for computing the posterior probabilities of individual bits and of blocks of bits in the regularly clocked LFSR_1 sequence are presented in Section 3. The new statistical distinguisher based on these probabilities is proposed and theoretically analyzed in Section 4, whereas the experimental results are described in Section 5. Conclusions are presented in Section 6.

2 Linear Models

We use the notation $A = a_1, a_2, \cdots$ for a binary sequence, A_k for its subsequence $a_k, a_{k+1}, \cdots$, A^n for its prefix $(a_i)_{i=1}^n = a_1, a_2, \cdots, a_n$, and A_k^n for its subsequence $(a_i)_{i=k}^n = a_k, a_{k+1}, \cdots, a_n$. If its length is finite, then A is called a string. Also, let $\bar{A}$ denote the bitwise binary complement

[1] A sequence of independent uniformly distributed random variables over a finite set is called purely random.

of A. Let $w(A)$ denote the number of 1's in A. For simplicity, wherever possible, we keep the same notation for random variables and their values.

Let X, C, and Y denote the output sequences of LFSR_1, LFSR_2, and the shrinking generator itself, respectively. Then Y is obtained from X by the nonuniform decimation according to C, that is, a bit x_i is deleted from X iff $c_i = 0$. It is assumed that C is a purely random sequence.

Let X satisfy the linear recurrence corresponding to a characteristic polynomial f of degree r and weight $w+1$, $f(z) = 1 + \sum_{k=1}^{w} z^{i_k}$, $1 \leq i_1 < \cdots < i_w = r$, that is, $x_t = \sum_{k=1}^{w} x_{t+i_k}$, $t \geq 1$. Here and throughout it is assumed that the summation of binary values is modulo 2. Note that f is the reciprocal of the LFSR_1 feedback polynomial, that is, the LFSR_1 characteristic polynomial or its any polynomial multiple. A shrunken polynomial of f is then defined as $\hat{f}(z) = 1 + \sum_{k=1}^{w} z^{\hat{i}_k}$, $1 \leq \hat{i}_1 < \cdots < \hat{i}_w = \hat{r}$, where $\hat{\Delta}_k = \hat{i}_k - \hat{i}_{k-1} - 1 \leq \Delta_k = i_k - i_{k-1} - 1$, $1 \leq k \leq w$, and $\hat{i}_0 = i_0 = 0$. According to the so-called linear sequential circuit approximation method applicable to arbitrary keystream generators, it is pointed out in [2] and [5] that for any $t \geq 1$, the linear equation

$$y_t \;=\; \sum_{k=1}^{w} y_{t+\hat{i}_k} \tag{1}$$

holds with probability $\frac{1+c}{2}$, where the correlation coefficient[2] c is given as

$$c \;=\; \frac{1}{2^r} \prod_{k=1}^{w} \binom{\Delta_k}{\hat{\Delta}_k}. \tag{2}$$

Note that c is the probability that the w bits on the right-hand-side of (1) originate from the corresponding w bits in X that together with the bit yielding y_t satisfy f in X. So, c is essentially the probability that the linear relation in X involving y_t is preserved in Y. The probabilistic linear recurrence defined by (1) is called a linear model as it means that Y can be generated by a nonautonomous LFSR with a nonbalanced[3] additive input and the feedback polynomial being the reciprocal of $\hat{f}$.

The correlation coefficient c reaches its maximal value, c_f, given f, if the shrunken polynomial is chosen according to $\hat{\Delta}_k^{opt} = \lfloor (\Delta_k + 1)/2 \rfloor$, $1 \leq k \leq w$. If none of $\Delta_k/2$ is very small, then c_f can be approximated by

$$c_f \;\approx\; (2\pi)^{-\frac{w}{2}} \left(\prod_{k=1}^{w} \Delta_k \right)^{-\frac{1}{2}} \;\geq\; (2\pi)^{-\frac{w}{2}} \left(\frac{r-w}{w} \right)^{-\frac{w}{2}}, \tag{3}$$

where the lower bound is reached if the feedback taps are approximately equidistant.

[2]The correlation coefficient between two binary random variables is defined as $c(a,b) = 2\Pr\{a = b\} - 1$. The correlation coefficient of a single binary variable is defined as $c(a) = c(a,0)$.

[3]A uniformly distributed random variable over a finite set is called balanced.

The weakness can be detected by the chi-square statistical test applied to the sequence $(y_t + \sum_{k=1}^{w} y_{t+\hat{i}_k})_{t=1}^{\infty}$. The required sequence length, n, and the computation time are proportional to $1/c^2$, say $10/c^2$, and the output sequence length is $n+r$. The computational step consists of w binary summations and one binary counting update. For the optimal shrunken polynomial, we have

$$\frac{1}{c_f^2} \;\leq\; (2\pi)^w \left(\frac{r-w}{w}\right)^w . \tag{4}$$

Accordingly, to minimize n, the use of the multiples of the LFSR$_1$ characteristic polynomial of relatively low weight and not too large degree is desirable.

More importantly, n can be significantly reduced by using a number, approximately $2^{-w}/c_f$, of different shrunken polynomials close to being optimal, that is, such that $|\hat{\Delta}_k - \hat{\Delta}_k^{opt}| \leq (\pi\Delta_k/2)^{\frac{1}{2}}$. The length required is thus considerably reduced and is proportional to

$$\frac{1}{c_f 2^{-w}} \;\leq\; (2\sqrt{2\pi})^w \left(\frac{r-w}{w}\right)^{\frac{w}{2}} , \tag{5}$$

whereas the computation time remains the same, that is, proportional to (4).

3 Computing the Posterior Probabilities of LFSR$_1$ Bits

We keep the same notation as in Section 2, but now assume a probabilistic model where apart from C, X is also a purely random sequence, which is independent of C. In this model the output sequence Y is also purely random.

3.1 Individual LFSR$_1$ Bits

The basic statistical distinguisher to be introduced in Section 4 is based on the posterior probabilities of individual LFSR$_1$ bits $\hat{p}_i = \Pr\{x_i = 1 \mid Y^n\} = \Pr\{x_i = 1 \mid Y^i\}$, $1 \leq i \leq n$, for appropriately defined Y^n. As proven in [7], they can be computed in $O(n^2)$ time and $O(n)$ space by

$$\hat{p}_i \;=\; \frac{1}{2}\left(\frac{1}{2} \;+\; 2^{-(i-1)} \sum_{e=0}^{i-1} \binom{i-1}{e} y_{i-e}\right) . \tag{6}$$

This is proven by simplifying a more general expression which holds in a general probabilistic model where X is a sequence of independent binary random variables. An alternative proof of (6) is given in Section 3.3 where a more general problem of computing the posterior probabilities of blocks of LFSR$_1$ bits is considered. As $\hat{p}_i$ can be numerically approximated with an arbitrarily small error by using only $O(\sqrt{i-1})$ values of e around $(i-1)/2$, the computation time can be reduced to $O(n\sqrt{n})$.

The binomial coefficients can be recursively precomputed in $O(n^2)$ time by

$$\binom{e+s}{e} = \binom{e+s-1}{e-1} + \binom{e+s-1}{e} \tag{7}$$

for $0 \le s \le n-1$, $0 \le e \le n-1-s$, and $(e,s) \ne (0,0)$, from the initial value $\binom{0}{0} = 1$.

It is also shown in [7] that $1/4 \le \hat{p}_i \le 3/4$, where the lower and upper bounds are achieved if and only if Y^i consists of all 0's and of all 1's, respectively.

3.2 Initial Blocks of LFSR$_1$ Bits

For computing the posterior probabilities of blocks of LFSR$_1$ bits to be introduced in Section 3.3, we need to compute the posterior probabilities of the initial blocks in the LFSR$_1$ sequence $\Pr\{X^n \mid Y\} = \Pr\{X^n \mid Y^n\}$. According to [3] and [7], we have

$$\Pr\{X^n \mid Y^n\} \; = \; \Pr\{Y^n \mid X^n\} \; = \; \sum_{e=0}^{n} 2^{-e} \, Q(e, n-e) \tag{8}$$

where $Q(e,s)$ is the conditional probability for prefixes of X and Y defined by

$$Q(e,s) \; = \; \Pr\{Y^s, w(C^{e+s}) = s \mid X^{e+s}\}. \tag{9}$$

It is in fact the probability that Y^s is obtained by deleting e bits from a given string X^{e+s}. This probability can be computed recursively, in $O(n^2)$ time and $O(n)$ space, by

$$Q(e,s) \; = \; \frac{1}{2} \, Q(e-1,s) \; + \; \frac{1}{2} \, \delta(x_{e+s}, y_s) \, Q(e, s-1) \tag{10}$$

for $0 \le s \le n$ and $0 \le e \le n-s$, from the initial condition $Q(0,0) = 1$, where for $e = 0$ or $s = 0$ the corresponding terms on the right-hand side of (10) are assumed to be equal to zero. Here, $\delta(i,j)$ or $\delta_{i,j}$ is the Kronecker symbol, i.e., $\delta(i,j) = 1$ if $i = j$ and $\delta(i,j) = 0$ if $i \ne j$.

Note that the statistically optimal correlation attack on LFSR$_1$ is based on $\Pr\{X^n \mid Y^n\}$ as a measure of correlation between X^n and Y^n, where X^n is produced from an assumed LFSR$_1$ initial state. The required n is proportional to the length of LFSR$_1$.

3.3 Arbitrary Blocks of LFSR$_1$ Bits

The refinement of the basic statistical distinguisher to be introduced in Section 4 is based on the posterior probabilities of blocks of m consecutive bits at variable positions in the LFSR$_1$ sequence $\Pr\{X_{i-m+1}^i \mid Y^n\} = \Pr\{X_{i-m+1}^i \mid Y^i\}$, $1 \le m \le i \le n$. For $m = 1$, they reduce to the posterior probabillities of individual LFSR$_1$ bits already considered in Section 3.1. These probabilities can also be useful for improving the fast correlation attack on LFSR$_1$ from [7]. Our objective here is to show that they can be computed efficiently by generalizing the expression (6).

The basic fact to be used is that the clock-control sequence C, in the model where X is purely random, remains to be purely random even when conditioned on the output sequence Y. This is because Y is purely random even if C is fixed, provided that X is purely random. Accordingly, we have

$$\begin{aligned}\Pr\{X_{i-m+1}^{i} \mid Y^{i}\} &= \sum_{e=0}^{i-m} \Pr\{X_{i-m+1}^{i} \mid w(C^{i-m}) = e, Y^{i}\} \cdot \Pr\{w(C^{i-m}) = e \mid Y^{i}\} \\ &= \sum_{e=0}^{i-m} \Pr\{X_{i-m+1}^{i} \mid w(C^{i-m}) = e, Y_{e+1}^{e+m}\} \cdot \Pr\{w(C^{i-m}) = e\} \\ &= \sum_{e=0}^{i-m} P(X_{i-m+1}^{i} \mid Y_{e+1}^{e+m}) \cdot \frac{1}{2^{i-m}} \binom{i-m}{e} \qquad (11)\end{aligned}$$

where we used the notation

$$P(A^{m} \mid B^{m}) \stackrel{\text{def}}{=} \Pr\{X^{m} = A^{m} \mid Y^{m} = B^{m}\}. \qquad (12)$$

More precisely, the second and the third line of (11) follow from the fact that C is purely random when conditioned on Y and from the fact that, on the condition that $w(C^{i-m}) = e$, C_{i-m+1}^{i} and X_{i-m+1}^{i} remain to be purely random and mutually independent as well as that a prefix of the string Y_{e+1}^{e+m} is obtained by decimating X_{i-m+1}^{i} according to C_{i-m+1}^{i}. The probability $P(X^{m} \mid Y^{m})$ can be computed recursively by using (8) and (10).

Accordingly, we have the following theorem, which generalizes (6).

Theorem 1 *For any given X_{i-m+1}^{i} and Y^{n} and each $m \leq i \leq n$, we have*

$$\Pr\{X_{i-m+1}^{i} \mid Y^{n}\} = \frac{1}{2^{i-m}} \sum_{e=0}^{i-m} \binom{i-m}{e} P(X_{i-m+1}^{i} \mid Y_{e+1}^{e+m}). \qquad (13)$$

If m is relatively small, then one can precompute the probabilities $P(X^{m} \mid Y^{m})$ for all 2^{2m} string pairs (X^{m}, Y^{m}). In fact, in view of the complementation property $P(X^{m} \mid Y^{m}) = P(\bar{X}^{m} \mid \bar{Y}^{m})$, one has to precompute only $2^{m-1}(2^{m}-1)$ probabilities. The posterior probabilities can then be computed in $O((2^{m}-1)n^{2})$ time and $O((2^{m}-1)n)$ space by using (13). It follows that

$$\left(\frac{1}{4}\right)^{m} \leq \Pr\{X_{i-m+1}^{i} \mid Y^{n}\} \leq \left(\frac{3}{4}\right)^{m} \qquad (14)$$

where the bounds are achieved if and only if X_{i-m+1}^{i} and Y_{i} are both constant strings, where different constants yield the lower bound and the same constants yield the upper bound.

4 Statistical Distinguisher

The posterior probabilities of individual bits and of blocks of bits in the LFSR_1 sequence are determined in Sections 3.1 and 3.3, respectively, in the probabilistic model where the LFSR_1 sequence is assumed to be purely random. Our main objective here is to show that these probabilities can also be used in the model where the LFSR_1 sequence is assumed to satisfy a linear recurrence and thus establish a basis for a new and more efficient statistical distinguisher for the shrinking generator.

4.1 Basic Statistical Test

As in Section 2, assume that X satisfies the linear recurrence $x_t = \sum_{k=1}^{w} x_{t+i_k}$, $t \geq 1$, corresponding to a characteristic polynomial $f(z) = 1 + \sum_{k=1}^{w} z^{i_k}$ of degree r and weight $w + 1$. Consider a bit y_t in the output sequence Y. It is obtained from some bit $x_{t'}$ in X, and it is not important to know t'. Therefore, the subsequence Y_{t+1} is obtained by decimating the subsequence $X_{t'+1}$. We know that the linear equation $x_{t'} = \sum_{k=1}^{w} x_{t'+i_k}$ is satisfied and we know $x_{t'}$, but we do not know the remaining w involved bits of X. However, as Y_{t+1} is obtained by decimating $X_{t'+1}$, we can determine the posterior probabilities of these w bits when conditioned on known Y_{t+1}.

Namely, we can compute $\hat{p}_{i_k} = \Pr\{x_{t'+i_k} = 1 \mid Y_{t+1}^{t+i_k}\}$ by using (6) for $i = i_k$, $1 \leq k \leq w$, in $\sum_{k=1}^{w} i_k = O(wr)$ time. Here the dependence of $\hat{p}_{i_k}$ on t is not explicitly shown for simplicity and to emphasize the fact that the position i_k rather than $t' + i_k$ is relevant. By using the numerical approximation described in Section 3.1, the computation time can be reduced to $O(w\sqrt{r})$. We can then make hard decisions on these w bits by applying the maximal posterior probability decision rule for individual bits, i.e., $\hat{x}_{t'+i_k} = 1$ if $\hat{p}_{i_k} \geq 0.5$ and $\hat{x}_{t'+i_k} = 0$ if $\hat{p}_{i_k} < 0.5$, $1 \leq k \leq w$. Finally, we can produce a hard estimate $\dot{x}_t = y_t + \sum_{k=1}^{w} \hat{x}_{t'+i_k}$ of the value of the linear equation $x_{t'} + \sum_{k=1}^{w} x_{t'+i_k}$, which is known to be equal to zero. The main point of the approach is that we can expect that $\dot{x}_t$ is biased towards zero when Y is the output sequence of the shrinking generator and is balanced when Y is a purely random sequence. This is justified by the analysis given in Section 4.2.

Accordingly, the statistical distinguisher is defined by repeating the procedure described above for every $1 \leq t \leq n$, by counting the numbers of 0's and 1's in the hard estimate sequence $\dot{X}^n = (\dot{x}_t)_{t=1}^{n}$, say n_0 and n_1, respectively, and by computing the chi-square statistic $\chi^2 = (n_0 - n_1)^2/n$ and the sign of $n_0 - n_1$. For large n, the statistic is expected to follow the chi-square distribution with one degree of freedom if Y is a purely random sequence and to grow approximately linearly with n if Y is the output sequence of the shrinking generator. So, let χ_0^2 be a threshold chosen to an assumed significance level α (e.g., $\chi_0^2 \approx 6.6349$ for $\alpha = 0.01$). Then for a sufficiently large n, the statistic exceeds the threshold with probability α if Y is a purely random sequence and with probability very close to 1 if Y is the output sequence of the shrinking generator. The minimal length and computation time required are estimated theoretically in Section 4.2 and experimentally in Section 5.

4.2 Theoretical Analysis

Given Y_{t+1}^{t+r}, the binary variable $\sum_{k=1}^{w} \hat{x}_{t'+i_k}$ has a fixed value. However, the hard estimate $\dot{x}_t = y_t + \sum_{k=1}^{w} \hat{x}_{t'+i_k}$ is a binary random variable if y_t is chosen at random. In the case when Y is a purely random sequence, y_t is balanced and independent of $\sum_{k=1}^{w} \hat{x}_{t'+i_k}$, so that $\dot{x}_t$ is balanced.

In the case when Y is the output sequence of the shrinking generator, y_t is balanced, but is dependent on $\sum_{k=1}^{w} \hat{x}_{t'+i_k}$ in such a way that the linear equation $y_t + \sum_{k=1}^{w} x_{t'+i_k} = 0$ is satisfied. Consequently, we obtain

$$\Pr\{\dot{x}_t = 0\} \;=\; \Pr\{\sum_{k=1}^{w} (\hat{x}_{t'+i_k} + x_{t'+i_k}) = 0\}. \tag{15}$$

Since $\hat{x}_{t'+i_k}$ is obtained by applying the maximal posterior probability decision rule, we have

$$\Pr\{\hat{x}_{t'+i_k} = x_{t'+i_k} \mid Y_{t+1}^{t+r}\} \;=\; \max(\hat{p}_{i_k}, 1 - \hat{p}_{i_k}) \tag{16}$$

or, in terms of correlation coefficients, we have

$$c_{i_k} \;=\; 2\Pr\{\hat{x}_{t'+i_k} = x_{t'+i_k} \mid Y_{t+1}^{t+r}\} - 1 \;=\; |1 - 2\hat{p}_{i_k}|. \tag{17}$$

Under the independence assumption, by using the well-known fact that the correlation coefficient of a binary sum of independent binary random variables equals the product of their individual correlation coefficients, we obtain

$$c \;=\; c(\dot{x}_t) \;=\; \prod_{k=1}^{w} c_{i_k} \;=\; \prod_{k=1}^{w} |1 - 2\hat{p}_{i_k}| \geq 0 \tag{18}$$

or, in terms of probabilities, $\Pr\{\dot{x}_t = 0\} = (1 + c)/2 \geq 1/2$. This shows that the random variable $\dot{x}_t$ is nonbalanced and is biased towards zero.

As the correlation coefficient c, relevant for the statistical distinction, depends on Y_{t+1}^{t+r}, it is important to estimate its expected value over purely random Y_{t+1}^{t+r}. Under the same independence assumption as above, (18) implies that

$$\bar{c} \;=\; \prod_{k=1}^{w} \bar{c}_{i_k}. \tag{19}$$

Now, by using a uniform approximation to the binomial distribution, it is shown in [7] that

$$c_i \;=\; |1 - 2\hat{p}_i| \;\approx\; \frac{1}{m(i)}\,|m_1(i) - 0.5m(i)| \tag{20}$$

where $m_1(i)$ is the number of 1's in Y^i on the segment $I(i)$ of length $m(i) \approx \sqrt{3(i-1)}$ centered around $0.5(i+1)$. Since $m_1(i)$ is binomially distributed, we have $\overline{|m_1(i) - 0.5m(i)|} \approx \sqrt{m(i)}/\sqrt{2\pi}$ and hence

$$\bar{c}_i \;\approx\; \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{m(i)}} \;\approx\; \frac{1}{2\pi\sqrt{3}}\,\frac{1}{\sqrt[4]{i}}. \tag{21}$$

Finally, in view of (21), (19) reduces to

$$\bar{c} \approx \frac{1}{(2\pi\sqrt{3})^w} \frac{1}{\sqrt[4]{\prod_{k=1}^{w} i_k}} \tag{22}$$

which is independent of t.

Another interesting measure is the expected value of c^2 over purely random Y_{t+1}^{t+r}. By using (18) and the numerical approximation (20), in view of $\overline{(m_1(i) - 0.5m(i))^2} \approx m(i)/4$, we first get

$$\overline{c_i^2} \approx \frac{1}{4m(i)} \approx \frac{1}{4\sqrt{3}} \frac{1}{\sqrt{i}} \tag{23}$$

and then

$$\overline{c^2} \approx \frac{1}{(4\sqrt{3})^w} \frac{1}{\sqrt{\prod_{k=1}^{w} i_k}}. \tag{24}$$

The length n required for the statistical distinction is proportional to $1/\bar{c}^2$ or to $1/\overline{c^2}$, where the two quantities differ only by a multiplicative constant, dependent on w, and are given by

$$\frac{1}{\bar{c}^2} \approx (2\pi\sqrt{3})^w \sqrt{\prod_{k=1}^{w} i_k}, \qquad \frac{1}{\overline{c^2}} \approx (4\sqrt{3})^w \sqrt{\prod_{k=1}^{w} i_k}. \tag{25}$$

The difference is due to the fact that the correlation coefficient c is dependent on time. If instead of runing the test forwards in time along the output sequence, we run it backwards in time, then we effectively use the reciprocal polynomial of f and hence, instead of $\prod_{k=1}^{w-1} i_k$, the product $\prod_{k=1}^{w-1}(r - i_k)$ becomes relevant. Consequently, to minimize n, we run the test forwards or backwards according to which one of the two products is smaller.

Another point that may be relevant for the statistical distinction is that the binary estimates $\sum_{k=1}^{w} \hat{x}_{t'+i_k}$, with the term y_t excluded, are strongly dependent in time. However, the terms y_t are roughly independent in time. The empirical evidence reported in Section 5 suggests that the quantity $1/\bar{c}^2$ seems to be more relevant than $1/\overline{c^2}$. It follows that

$$\frac{1}{\bar{c}^2} \approx (2\pi\sqrt{3})^w \sqrt{\prod_{k=1}^{w} i_k} \leq (2\pi\sqrt{3})^w \, r^{\frac{w}{2}}. \tag{26}$$

which is to be compared with (4) and (5). Recall that if (6) is used for computing the posterior probabilities, then the computation time is at most wrn steps, and if the binomial distribution is truncated, then it is at most $w\sqrt{3r}\,n$ steps, where each step is an integer summation. Moreover, the step reduces to a binary counting update if the binomial distribution is replaced by a uniform distribution, that is, if the numerical approximation (20) is used instead.

In any case, it follows that, in comparison with the linear model approach with multiple shrunken polynomials, the computation time is significantly reduced, whereas the required output sequence length, $n + r$, roughly remains the same. The advantage is especially significant if r is relatively large, which is expected to be the case if a polynomial multiple of the LFSR_1 characteristic polynomials is used. Namely, the minimal degree of a polynomial multiple of weight $w + 1$ of a random polynomial of degree r_1 is expected to be $O(2^{r_1/w})$ (see [6]). In this case, the computation time is thus reduced from $O(2^{r_1})$ to $O(2^{\frac{r_1}{2}(1+\frac{1}{w})})$. Note that in both the approaches the required output sequence length can be further reduced by using more than just one linear recurrence satisfied by the LFSR_1 sequence, that is, by using a number of polynomial multiples of the LFSR_1 characteristic polynomial.

4.3 Improved Statistical Test

In the basic statistical test, the decisions on the involved bits $x_{t'+i_k}$ are made on the basis of their individual posterior probabilities. The decisions are close to being optimal if these bits are close to being independent when conditioned on the output sequence (the independence assumption). This is approximately satisfied if the feedback taps are not too close to each other. If there are some taps that are very close to each other, then the basic statistical test can be improved by making joint decisions on the basis of the posterior probabilities of blocks of bits at appropriate positions in X.

Suppose that the taps at positions i_k, $k_1 \le k \le k_2$, are concentrated. We first compute the posterior probabilities of the block of bits $X_{t'+i_{k_1}}^{t'+i_{k_2}}$, that is, $\Pr\{X_{t'+i_{k_1}}^{t'+i_{k_2}} = A^{k_2-k_1+1} \mid Y_{t+1}^{t+i_{k_2}}\}$, by using Theorem 1 for $i = i_{k_2}$ and $m = k_2 - k_1 + 1$, in $(2^{k_2-k_1+1} - 1)i_{k_2}$ steps. The desired posterior probability $\Pr\{\sum_{k=k_1}^{k_2} x_{t'+i_k} = 1 \mid Y_{t+1}^{t+i_{k_2}}\}$ is then directly computed by summing up the posterior probabilities over the corresponding blocks $A^{k_2-k_1+1}$. This posterior probability is then used to make the hard decision on the binary sum $\sum_{k=k_1}^{k_2} x_{t'+i_k}$ and combined with the other posterior probabilities in the same way as in the basic statistical test.

5 Experimental Results

The theoretical estimates of the output sequence length required for the basic statistical test to be successful are based on the numerical approximation (20) to the expression (6) for individual posterior probabilities and on the independence assumption for the bits involved in the linear recurrence considered. The objective of the experiments was to obtain empirical estimates of the expected values of the correlation coefficient c and its square c^2 as well as to verify the effectiveness of the test by computing the value of the χ^2 statistic as a function of the output sequence length. This is performed for various degrees r and weights $w + 1$ of the LFSR_1 characteristic polynomial. Each polynomial $f(z) = 1 + \sum_{k=1}^{w} z^{i_k}$ is chosen so that $\prod_{k=1}^{w-1} i_k < \prod_{k=1}^{w-1}(r - i_k)$, so that it is better to run the test forwards than backwards in time, and is specified as a w-tuple $(i_1, \cdots, i_w)$.

For Fig. 1, we fixed $w = 2$ and varied r by choosing the polynomials $(17, 50), (37, 100), (97, 250), (213, 500)$, and $(319, 1000)$, whereas for Fig. 2, we fixed $r = 50$ and varied w by choosing the polynomials $(17, 50), (7, 21, 50), (3, 15, 23, 50)$, and $(5, 13, 15, 24, 50)$. For both figures, the same clock-control LFSR_2 with a primitive characteristic polynomial $(2, 27, 29, 128)$ was used. For Fig. 1, the χ^2 statistic is computed for different sequence lenghts, $n' = n + r$, with a step of 2000 bits, for both the shrinking generator sequence and a purely random sequence. Each value is averaged over 10 randomly chosen initial states of LFSR_1. For the random case, the obtained curves are independent of the polynomial and follow the chi-square probability distribution with one degree of freedom, whereas for the shrinking generator case, they increase with n roughly linearly, as $\chi^2 \approx \bar{c}^2 n$ for large n. We observed that whenever χ^2 was large, the sign of $n_0 - n_1$ was positive, as expected.

Similar experiments were run for Fig. 2, but only for the shrinking generator case. In the left part of Fig. 2 we used a step of 10^4 bits and the curves are averaged over 10 randomly chosen initial states of LFSR_1, whereas in the right part we used a step of 10^6 bits and the curves are obtained only for one initial state of LFSR_1. The sequence length is shown on the logarithmic scale in Fig. 2.

For each curve in both figures, the critical length n'_0 is computed as the minimal sequence length such that the threshold $\chi_0^2 = 6.6349$ is reached. Also, for each considered polynomial and for the same sequence lengths as in Figures 1 and 2, the stable empirical estimates of the expected value of c, $\bar{c}_{\mathrm{exp}}$, and of c^2, $\overline{c^2}_{\mathrm{exp}}$, are both obtained and the inverses $1/\bar{c}^2_{\mathrm{exp}}$ and $1/\overline{c^2}_{\mathrm{exp}}$ are computed. Here for each t, c is computed by (18) and the values are then averaged over t. In addition, the theoretical estimates are computed according to (25). All these quantities are shown in Tables 1 and 2, for the polynomials used for Figures 1 and 2, respectively.

Table 1: Empirical and theoretical estimates for $w = 2$ and various r.

r	$1/\overline{c^2}_{\mathrm{exp}}$	$1/\overline{c^2}$	$1/\bar{c}^2_{\mathrm{exp}}$	$1/\bar{c}^2$	n'_0
50	$1.42 \cdot 10^3$	$1.40 \cdot 10^3$	$3.33 \cdot 10^3$	$3.45 \cdot 10^3$	$2.80 \cdot 10^4$
100	$3.00 \cdot 10^3$	$2.92 \cdot 10^3$	$7.15 \cdot 10^3$	$7.20 \cdot 10^3$	$3.60 \cdot 10^4$
250	$7.79 \cdot 10^3$	$7.47 \cdot 10^3$	$1.88 \cdot 10^4$	$1.84 \cdot 10^4$	$1.00 \cdot 10^5$
500	$1.64 \cdot 10^4$	$1.57 \cdot 10^4$	$4.01 \cdot 10^4$	$3.87 \cdot 10^4$	$1.92 \cdot 10^5$
1000	$2.83 \cdot 10^4$	$2.71 \cdot 10^4$	$6.99 \cdot 10^4$	$6.69 \cdot 10^4$	$4.24 \cdot 10^5$

The experimental results show that there is a very good accordance between the theoretical and empirical estimates of the expected values of both c and c^2 for $w = 2, 3, 4$ as in this case the feedback taps are not very close to each other, so that the independence assumption is well justified. For $w = 5$, that is, for the polynomial $(5, 13, 15, 24, 50)$, the taps 13 and 15 are concentrated and this results in the empirical estimates $1/\bar{c}^2_{\mathrm{exp}}$ and $1/\overline{c^2}_{\mathrm{exp}}$ being about three times greater than the corresponding theoretical estimates, respectively, so that the minimal

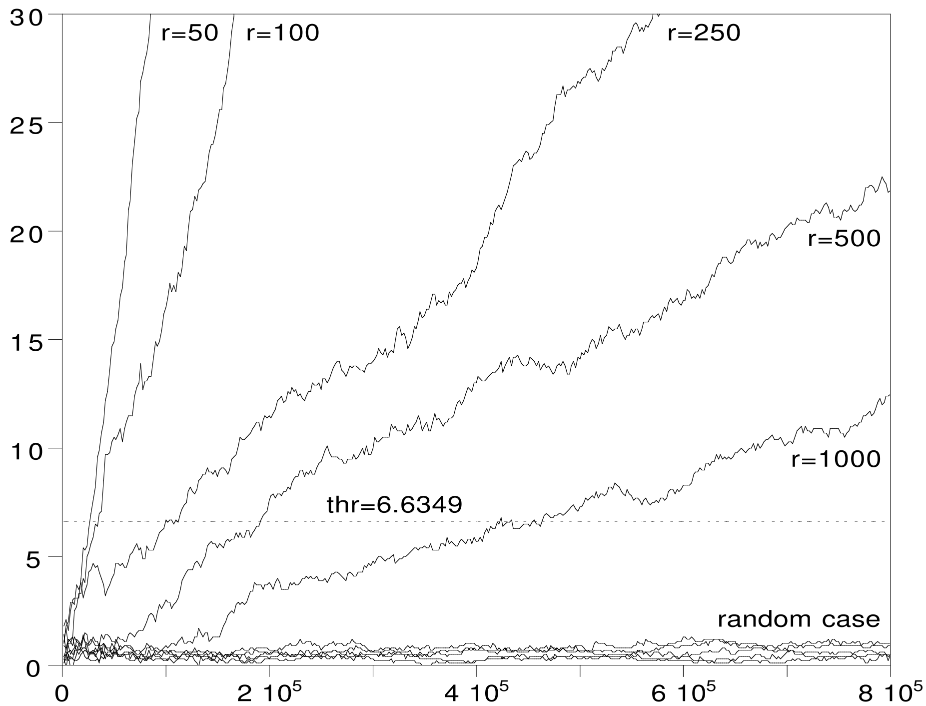

Figure 1: The χ^2 statistics for $w = 2$ and various r.

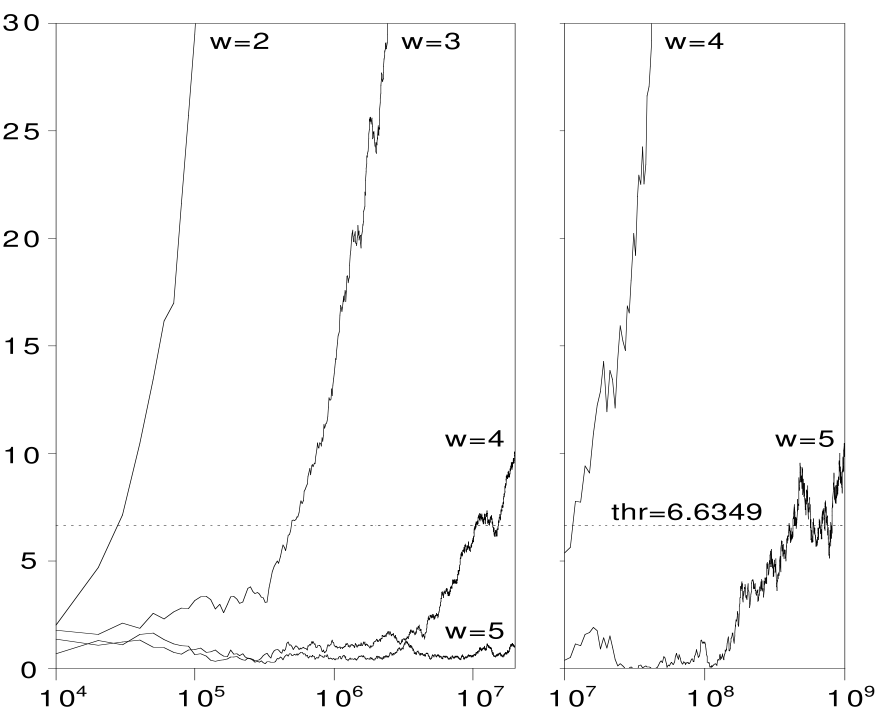

Figure 2: The χ^2 statistics for $r = 50$ and various w.

Table 2: Empirical and theoretical estimates for $r = 50$ and various w.

w	$1/\overline{c^2}_{\mathrm{exp}}$	$1/\overline{c^2}$	$1/\bar{c}^2_{\mathrm{exp}}$	$1/\bar{c}^2$	n'_0
2	$1.42 \cdot 10^3$	$1.40 \cdot 10^3$	$3.33 \cdot 10^3$	$3.45 \cdot 10^3$	$2.80 \cdot 10^4$
3	$2.82 \cdot 10^4$	$2.85 \cdot 10^4$	$9.58 \cdot 10^4$	$1.11 \cdot 10^5$	$5.00 \cdot 10^5$
4	$3.72 \cdot 10^5$	$5.24 \cdot 10^5$	$2.19 \cdot 10^6$	$3.19 \cdot 10^6$	$1.04 \cdot 10^7$
5	$5.69 \cdot 10^6$	$1.73 \cdot 10^7$	$5.58 \cdot 10^7$	$1.65 \cdot 10^8$	$4.00 \cdot 10^8$

sequence length required is in practice reduced. This length can be further reduced by using the improved statistical test treating the two bits corresponding to the taps 13 and 15 jointly.

The experimental results also indicate that the minimal output sequence length required is proportional to $1/\bar{c}^2_{\mathrm{exp}}$ where the multiplicative constant is between 5 and 10 as well as that this length is upper-bounded by $1/\bar{c}^2$ given by (26), multiplied by the same constant.

6 Conclusions

For a simplified shrinking generator in which the characteristic polynomial of the clock-controlled LFSR, LFSR_1, is assumed to be known, a new statistical distinguisher is proposed. Unlike the previously proposed statistical distinguisher which searches for the linear relations in the output sequence, which are called the linear models, the new distinguisher estimates the linear relations in the original LFSR_1 sequence by computing the posterior probabilities of individual bits or of blocks of bits in this sequence when conditioned on the known output sequence, where the output sequence is the decimated LFSR_1 sequence. Both distinguishers are based on low-weight polynomial multiples of the LFSR_1 characteristic polynomial.

The theoretical and experimental analysis show that the computation time of the new distinguisher is significantly reduced. The new approach is also effective if the LSFR_1 characteristic polynomial is randomly chosen, provided that the degree of the used low-weight polynomial multiple is close to the expected minimal degree.

References

[1] D. Coppersmith, H. Krawczyk, and Y. Mansour, "The shrinking generator," Advances in Cryptology - CRYPTO '93, *Lecture Notes in Computer Science*, vol. 773, pp. 22-39, 1993.

[2] J. Dj. Golić, "Intrinsic statistical weakness of keystream generators," Advances in Cryptology - ASIACRYPT '94, *Lecture Notes in Computer Science*, vol. 917, pp. 91-103, 1995.

[3] J. Dj. Golić and L. O'Connor, "Embedding and probabilistic correlation attacks on clock-controlled shift registers," Advances in Cryptology - EUROCRYPT '94, *Lecture Notes in Computer Science*, vol. 950, pp. 230-243, 1995.

[4] J. Dj. Golić, "Towards fast correlation attacks on irregularly clocked shift registers," Advances in Cryptology - EUROCRYPT '95, *Lecture Notes in Computer Science*, vol. 921, pp. 248-262, 1995.

[5] J. Dj. Golić, "Linear models for keystream generators," *IEEE Trans. Comput.*, vol. C-45, pp. 41-49, Jan. 1996.

[6] J. Dj. Golić, "Computation of low-weight parity-check polynomials," *Electronics Letters*, vol. 32(21), pp. 1981-1982, Oct. 1996.

[7] J. Dj. Golić, "Correlation analysis of the shrinking generator," Advances in Cryptology - CRYPTO 2001, *Lecture Notes in Computer Science*, vol. 2139, pp. 440-457, 2001.

[8] T. Johansson, "Reduced complexity correlation attacks on two clock-controlled generators," Advances in Cryptology - ASIACRYPT '98, *Lecture Notes in Computer Science*, vol. 1514, pp. 342-357, 1998.

[9] K. Zeng, C. H. Yang, and T. R. N. Rao, "On the linear consistency test (LCT) in cryptanalysis with applications," Advances in Cryptology - CRYPTO '89, *Lecture Notes in Computer Science*, vol. 435, pp. 164-174, 1990.