Towards Fast Correlation Attacks on Irregularly Clocked Shift Registers

Jovan Dj. Golić *

Information Security Research Centre, Queensland University of Technology
GPO Box 2434, Brisbane Q 4001, Australia
School of Electrical Engineering, University of Belgrade
Email: golic@fit.qut.edu.au

Abstract. A theoretical framework for fast correlation attacks on irregularly clocked linear feedback shift registers (LFSRs) based on a recently established linear statistical weakness of decimated LFSR sequences is developed. When the LFSR feedback polynomial is not known, methods for the statistical weakness detection and the feedback polynomial reconstruction are proposed. When the LFSR feedback polynomial is known, an iterative procedure for fast LFSR initial state reconstruction given an observed keystream sequence is introduced. The procedure is based on appropriately defined parity-check sums and consists in iterative recomputation of the posterior probabilities for unknown elements of the decimation sequence. A convergence condition in terms of the numbers of the parity-check sums needed for successful reconstruction and the required polynomial computational complexity indicate that the proposed fast correlation attack may be realistic, especially in the constrained clocking case. The number of the feedback polynomial multiples of relatively low weight and not too large degree thus proves to be critical for the security of irregularly clocked LFSRs.

1 Introduction

Clock-controlled linear feedback shift registers (LFSRs) have become important building blocks for keystream generators in stream cipher applications, because they are known to produce sequences of long period and high linear complexity, see [12], [5], and [17]. They are also immune to fast correlation attacks [13, 19, 20, 18] on additively noised LFSR sequences. They have even been proposed as the keystream generators themselves, see [4], [14]. A clock-controlled shift register is a LFSR that is irregularly clocked according to a decimation sequence which defines the number of symbols to be deleted before the next output symbol is produced. The decimation sequence is itself a pseudorandom sequence produced by a clock-control generator, for example, by another LFSR, as is proposed in [4]. Irregular clocking is called constrained if the number of consecutive deletions is limited and unconstrained otherwise, see [9]. The secret key is assumed to

* This research was supported in part by the Science Fund of Serbia, grant #0403, through the Institute of Mathematics, Serbian Academy of Arts and Sciences.

L.C. Guillou and J.-J. Quisquater (Eds.): Advances in Cryptology - EUROCRYPT '95, LNCS 921, pp. 248-262, 1995.

control the LFSR initial state, the initial state and the structure of the clock-control generator, and, possibly the LFSR feedback polynomial as well. The objective is to reconstruct the feedback polynomial and the initial state of the clock-controlled LFSR given an observed segment of the keystream sequence.

When the LFSR feedback polynomial is known, a divide and conquer attack on the unknown decimation sequence based on the linear consistency test is proposed in [21], where the required length of the observed keystream sequence for the attack to be successful is also estimated. A similar attack for multiplexed sequences based on the collision test is independently suggested in [1], see also [2]. On the other hand, divide and conquer correlation attacks on the LFSR initial state are also possible if the observed decimated sequence is sufficiently long. Namely, the embedding correlation attack is in [23] and [10] analyzed for the constrained clocking case and in [10] for the unconstrained clocking case. The statistically optimal probabilistic correlation attack in the unconstrained clocking case is analyzed in [10] too. The attacks imply the exhaustive search over all possible decimation sequences and over all possible LFSR initial states, respectively. Although a divide and conquer effect is achieved, the computational complexity remains exponential.

The first problem to be considered in this paper is the LFSR feedback polynomial reconstruction given a known segment of the keystream sequence. We develop an approach based on a recently found linear statistical weakness of irregularly decimated LFSR sequences [11]. The weakness and the corresponding correlation coefficients are further analyzed in more detail in Section 2. In Section 3, it is then shown how to detect the statistical weakness when the feedback polynomial is not known, how to reconstruct the feedback polynomial by exhaustive search over appropriately defined shrunk polynomials, and how to reconstruct the feedback polynomial in a fast iterative way. The necessary length of the keystream sequence and the computational complexity are estimated in all the cases.

The second problem, dealt with in Section 4, is the fast reconstruction of the LFSR initial state based on an observed keystream sequence, assuming that the feedback polynomial is known. The linear statistical weakness [11] is again the starting point, and the method that we propose is based on the parity-check sums corresponding to the shrunk polynomials of specially defined polynomial multiples of the feedback polynomial. The algorithm consists in iterative recomputation of the posterior probabilities for unknown elements of the decimation sequence and is conceptually similar to the iterative probabilistic decoding algorithms used in fast correlation attacks on additively noised LFSR sequences [13, 22, 3, 15, 16]. Instead of the binary additive noise we deal with the integer decimation noise. By using the analogy with the problem considered in [16], we give a convergence condition in terms of the numbers of the parity-check sums needed for successful reconstruction. This condition along with the required polynomial computational complexity shows that the proposed fast correlation attack may be realistic, particularly in the constrained clocking case.

2 Linear Statistical Weakness

Consider a clock-controlled shift register as a keystream generator consisting of a binary linear feedback shift register (LFSR) that is irregularly clocked according to a nonnegative integer decimation sequence which defines the number of symbols to be deleted before the next output symbol is produced, see [12], [5]. It is thus assumed that the number of clocks per each output symbol is a positive integer. The decimation sequence is itself generated in a pseudorandom manner by a clock-control generator, for example, by another LFSR, as is proposed in [4]. More precisely, if $X = \{x_t\}_{t=0}^{\infty}$ denotes a regularly clocked LFSR sequence and $D = \{d_t\}_{t=0}^{\infty}$ a decimation sequence, then the output sequence $Y = \{y_t\}_{t=0}^{\infty}$ is defined as a decimated sequence

$$y_t = x\left(t + \sum_{i=0}^{t} d_i\right), \quad t \geq 0. \tag{1}$$

The secret key is assumed to control the LFSR initial state and the initial state and the structure of the clock-control generator. It may in addition be assumed that the LFSR feedback polynomial is also defined by the secret key, as in [4]. The objective is to reconstruct the initial state of the clock-controlled LFSR given an observed segment of the keystream sequence, provided that the feedback polynomial is known. If the feedback polynomial is not known, then the first objective is to determine this polynomial.

Since the decimation sequence is not known, it is reasonable to assume an appropriate probabilistic model based on partial or complete knowledge of the structure of the clock-control generator. Let thus D be a sequence of independent identically distributed nonnegative integer random variables with a probability distribution $\mathcal{P} = \{P(d)\}_{d \in \mathcal{D}}$ where $\mathcal{D}$ is the set of integers with positive probability. The deletion rate [10] is then defined as $\bar{p} = \frac{\bar{d}}{1+\bar{d}}, \bar{d} = \sum_{d \in \mathcal{D}} dP(d)$. The unknown LFSR initial state is assumed to be chosen uniformly at random. Let us now briefly review and then further analyze in more detail the linear statistical weakness of irregularly clocked LFSR sequences pointed out in [11]. Let the LFSR have length r and the feedback polynomial $f(z) = 1 + \sum_{i=1}^{r} f_i z^i = 1 + \sum_{k=1}^{w} z^{i_k}$, $1 \leq i_1 < \ldots < i_w = r$, where $W = w + 1$ is the weight of $f(z)$. Then a *shrunk* polynomial of $f(z)$ is defined as a polynomial of the form $\hat{f}(z) = 1 + \sum_{i=1}^{\hat{r}} \hat{f}_i z^i = 1 + \sum_{k=1}^{w} z^{\hat{i}_k}$, where $1 \leq \hat{i}_1 < \ldots < \hat{i}_w = \hat{r}$ and $\hat{i}_k - \hat{i}_{k-1} \leq i_k - i_{k-1}, 1 \leq k \leq w$, with $\hat{i}_0 = i_0 = 0$. The weight of $\hat{f}(z)$ is the same as the weight of $f(z)$ and its degree $\hat{r}$ is not bigger than the degree r. The degrees are equal if and only if $\hat{f}(z) = f(z)$. According to the LSCA method [8], it is shown in [11] that the linear equation

$$y_t + \sum_{k=1}^{w} y_{t-\hat{i}_k} = 0 \tag{2}$$

holds with probability $(1 + c)/2$ in the decimated sequence for any $t \geq \hat{r}$, where the corresponding correlation coefficient c depends on the probability

distribution $\mathcal{P} = \{P(d)\}_{d \in \mathcal{D}}$. For simplicity, consider the geometric distribution $P(d) = p^d(1-p), d \geq 0$, which corresponds to the case of independent deletions with probability $p > 0$. This is exactly the model to be used for the shrinking generator [4] where also $p = 1/2$. An arbitrary probability distribution $\mathcal{P}$ can be approximated by the geometric distribution by setting $p = \bar{p}$ where $\bar{p}$ is the deletion rate. The correlation coefficient [11] is in this case given by

$$c = p^{r-\hat{r}}(1-p)^{\hat{r}+1} \prod_{k=1}^{w} \binom{\Delta_k}{\hat{\Delta}_k} \tag{3}$$

where $\Delta_k = i_k - i_{k-1} - 1$ and $\hat{\Delta}_k = \hat{i}_k - \hat{i}_{k-1} - 1$, $1 \leq k \leq w$. Equation (3) has a clear combinatorial meaning in terms of the probability of decimation sequences. More precisely, the correlation coefficient is equal to the probability of the event that the bits satisfying the feedback polynomial in the shift register sequence remain undeleted in such a way that they satisfy the shrunk feedback polynomial in the decimated sequence. It is assumed that the conditional correlation coefficient is equal to one when the event occurs and to zero otherwise.

Consequently, the error sequence $\{e_t\}$ produced by applying the feedforward linear transform defined by $\hat{f}$ to the keystream sequence $\{y_t\}$, $e_t = y_t + \sum_{k=1}^{w} y_{t-\hat{i}_k}, t \geq \hat{r}$, is regarded as a sequence of nonbalanced identically distributed binary random variables with the correlation coefficient c to the constant zero variable. The variables are not independent. In order to detect the statistical weakness in the error sequence, one can apply a simple chi-square statistical test. For a sequence of length n with n_0 zeros and n_1 ones, the value of the chi-square statistic is given by

$$\chi^2 = \frac{(n_0 - n_1)^2}{n} = n\,\acute{c}^2 \tag{4}$$

where $\acute{c} = (n_0 - n_1)/n$ denotes an estimate of c. The error sequence can thus be distinguished from a purely random binary sequence with error probability less than about 10^{-3}, if n, which is approximately equal to the length of the observed keystream sequence, is equal to $10/c^2$ or larger. The same holds for the amount of computation needed.

The correlation coefficient depends on the chosen shrunk polynomial and is maximized if

$$\hat{\Delta}_k = \hat{\Delta}_k^{opt} \stackrel{\text{def}}{=} \lfloor (1-p)(\Delta_k + 1) \rfloor, \quad 1 \leq k \leq w. \tag{5}$$

The maximum value denoted as c_f is given by

$$c_f = (1-p)^{w+1} \prod_{k=1}^{w} p^{\Delta_k - \hat{\Delta}_k^{opt}} (1-p)^{\hat{\Delta}_k^{opt}} \binom{\Delta_k}{\hat{\Delta}_k^{opt}}. \tag{6}$$

The multiplicative terms in (6) are the central terms of the corresponding binomial distributions. If $2 \leq \hat{\Delta}_k^{opt} \leq \Delta_k - 2$, then the kth multiplicative term in (6) is by Stirling's formula well approximated as $(2\pi p(1-p)\Delta_k)^{-1/2}$. If $\hat{\Delta}_k^{opt}$ is equal

either to Δ_k or to $\Delta_k - 1$, that is, if $p\Delta_k < 2 - p$, then the kth multiplicative term is lower-bounded by $(1-p)^{\Delta_k}$.

Let us analyze c_f in two extreme cases, when p is relatively large and when p is relatively small. The case when p is very close to 1 can be treated analogously, but is not of practical interest. Suppose first that none of r/w, Δ_k, $p\Delta_k$, and $(1-p)\Delta_k$ is very small. Then we have

$$c_f \simeq (1-p)\left(\frac{2\pi p}{1-p}\right)^{-\frac{w}{2}}\left(\prod_{k=1}^{w}\Delta_k\right)^{-\frac{1}{2}}. \tag{7}$$

The smallest magnitude of c_f, which is the worst case for cryptanalysis, is then obtained when the feedback taps are approximately equidistant and is equal to

$$c_f \simeq (1-p)\left(\frac{2\pi p}{1-p}\frac{r}{w}\right)^{-\frac{w}{2}}. \tag{8}$$

The necessary length of the keystream sequence and the amount of computation needed to detect the weakness are then both upper-bounded by $(10/(1-p)^2)$ $(2\pi p/(1-p))^w((r-w)/w)^w$. Given w, the larger the values of r and p the smaller the correlation coefficient. Given r and p, there exists an optimal value of w that minimizes the correlation coefficient. Also, $c \to 1$ when $p \to 0$ and $c \to 0$ when $p \to 1$. For example, if $p = 1/2$ as is suggested for the shrinking generator [4], then the necessary length becomes $40\,(6.28\,(r-w)/w)^w$. Second, suppose that the deletion rate p is relatively small so that $p\Delta_k < 2 - p$ for each k. Then we obtain a lower bound

$$c_f \geq (1-p)^{r+1} \geq \left(1 - \frac{2w}{r}\right)^{r+1}. \tag{9}$$

The minimum length to detect the weakness is then at most $10/(1-p)^{2(r+1)}$ $\leq 10/(1-\frac{2w}{r})^{2(r+1)}$.

Consequently, one may conclude that the linear statistical weakness is realistic and may be relatively easy to detect unless both r and w are relatively large. If w is large, then instead of f one should use a polynomial multiple of f whose weight is relatively low and whose degree is not too large, based on the well-known fact that an LFSR sequence generated by f satisfies the linear recursions determined by all the polynomial multiples of f as well. In any case, the necessary length of the keystream sequence for the weakness detection is much smaller than 2^r. The underlying assumption is that the feedback polynomial f is known.

The required length of the observed keystream sequence can be significantly reduced by using more than just one different shrunk polynomials of a given polynomial f. The chi-square statistic is then computed on the resulting concatenation of the error sequences as a whole. There are exactly $M_f = \prod_{k=1}^{w}(\Delta_k + 1)$ different shrunk polynomials $\hat{f}$ of a given polynomial f. The sum of the correlation coefficients (3) for all the shrunk polynomials of f is readily verified to be

$$\sum_{\hat{f}} \left(p^{r-\hat{r}}(1-p)^{\hat{r}+1} \prod_{k=1}^{w} \binom{\Delta_k}{\hat{\Delta}_k} \right) = (1-p)^{w+1} \tag{10}$$

and is equal to the probability that any $w+1$ bits satisfying the linear recursion defined by the polynomial f in the regularly clocked LFSR sequence all remain in the keystream sequence as undeleted after the decimation. The average correlation coefficient over all the shrunk polynomials of f is hence $(1-p)^{w+1}/M_f$. However, the total correlation virtually remains the same, up to a multiplicative constant close to one, if the summation is carried out only over the shrunk polynomials close to the optimal one for which the correlation coefficient is equal to the maximum value c_f. A shrunk polynomial is regarded to be close to the optimal one if for each $1 \leq k \leq w$ the absolute difference $\hat{\Delta}_k - \hat{\Delta}_k^{opt}$ is at most equal to the standard deviation $(2\pi p(1-p)\Delta_k)^{1/2}$ of the binomial distribution multiplied by a positive constant close to 1. Since the corresponding average correlation coefficient $\bar{c}_f$ is then very close to c_f, there are approximately $(1-p)^{w+1}/c_f$ such shrunk polynomials. So, the required length of the keystream sequence to detect the weakness is thus reduced to $10/(c_f(1-p)^{w+1})$, whereas the amount of computation remains the same, that is, $10/c_f^2$, comparing with the single optimal shrunk polynomial case.

3 Feedback Polynomial Reconstruction

In this section, it is assumed that the LFSR feedback polynomial f is key-dependent and hence unknown, as is proposed in [4]. In order to achieve long period and good long-term statistical properties on a period, f is usually chosen to be primitive or irreducible. However, since f is time-invariant given a key, the statistical weakness analyzed in the previous section remains, but the amount of computation required to detect the weakness is in general increased. Moreover, since the correlation coefficient (3) depends on the assumed shrunk polynomial $\hat{f}$ given f and is maximized if (5) is satisfied, the estimate $\acute{c}$, see (4), of the correlation coefficient may be used as a statistic to reconstruct the feedback polynomial f. We will now consider more closely the following three cases: the weakness detection, the feedback polynomial reconstruction, and the fast reconstruction of the feedback polynomial.

In practical applications, the degree r of f is typically known with relatively small uncertainty. The first objective is to find a shrunk polynomial $\hat{f}$ of f that is close to an optimal one yielding the maximum value of the correlation coefficient (6). Its degree $\hat{r}$ is then close to $(1-p)r + pw$. For example, if $p = 1/2$, then $\hat{r} \simeq (r+w)/2$, which may not be large. For each assumed w, one should then check $\binom{\hat{r}-1}{w-1}$ possible candidates for an optimal shrunk polynomial. The best candidate is the one with the maximum value of the statistic $\acute{c}$ over all possible values of w. The weakness is detected if the chi-square value (4) corresponding to this maximum value is significant. Typically, some prior information about w

is avalaible, so that the search is not complete. For example, in hardware realizations w is not large. The required length of the keystream sequence depends on f and is increased to $10\hat{r}/c_f^2$, because the optimal shrunk polynomial is described by $\hat{r}$ bits and the necessary length for each bit of uncertainty to be resolved is $10/c_f^2$. Since f is not known, instead of c_f one can use a lower bound (8) for relatively large p, $p \geq w/r$, or a lower bound (9) for small p, $p < w/r$. The obtained maximum value of the estimate $\acute{c}$ should be consistent with the chosen lower bound. The amount of computation for each of the assumed shrunk polynomials is proportional to the keystream sequence length. Another possibility is to simultaneously check all the shrunk polynomials close to the candidate one, which reduces the required keystream sequence length, as is described in the previous section. The unknown standard deviation can be estimated by using the equidistant taps assumption.

The second objective is to reconstruct the feedback polynomial f. Once the best shrunk polynomial is determined, the most likely feedback polynomial is the one that satisfies (5). However, the decision on f based on the estimate $\acute{c}$ of the correlation coefficient (3) is not reliable. Namely, the number of the shrunk polynomial candidates that are close to the best one with respect to the chi-square statistic is generally not small. In this case, there are many close candidates for the best feedback polynomial. To distinguish between them, one should then use the appropriate polynomial multiples of these polynomials. More precisely, for every chosen candidate for f, find a number of polynomial multiples and for each of them compute an estimate $\acute{c}$ of the correlation coefficient by using different shrunk polynomials close to the optimal one. The polynomial multiples should preferably have relatively low degrees and low weights so as to obtain reliable estimates of the correlation coefficients on a keystream sequence of a given length. In addition, the polynomial multiples of different candidates for f should be sufficiently different so as to obtain only one or just a few candidates whose correlation coefficient estimates are consistent with the maximum value corresponding to (5). So, in general, the longer the observed keystream sequence, the smaller the number of the remaining candidates. However, additional precomputation is required to obtain suitable polynomial multiples. For a candidate polynomial g, this can be done by computing the residues modulo g of the polynomials of the form z^i. Note that finding low weight polynomial multiples of a feedback polynomial is crucial for the fast correlation attacks on the initial state of regularly clocked LFSRs distorted by additive binary noise, see [13, 19, 3].

The third objective is to reconstruct the feedback polynomial f in a fast way, by reducing the computational effort needed to find an optimal shrunk polynomial. The feedback polynomial reconstruction then goes along the same lines as in the second approach. First observe that the average value of the correlation coefficient (3) over all the shrunk polynomials with fixed $\hat{\Delta}_i$ is maximized if (5) holds for $k = i$. Instead of the simultaneous search through all the shrunk polynomials of a given degree and weight, one can then proceed iteratively by reconstructing the values $\hat{\Delta}_i$ one at a time, starting from $i = 1$, for example.

More precisely, the sum of the correlation coefficients (3) for all the shrunk polynomials $\hat{f}$ with fixed $\hat{\Delta}_i$ is easily seen to be

$$\sum_{\hat{f}:\hat{\Delta}_i} \left(p^{r-\hat{r}}(1-p)^{\hat{r}+1} \prod_{k=1}^{w} \binom{\Delta_k}{\hat{\Delta}_k} \right) = (1-p)^{w+1} p^{\Delta_i - \hat{\Delta}_i} (1-p)^{\hat{\Delta}_i} \binom{\Delta_i}{\hat{\Delta}_i}. \quad (11)$$

The total number of such shrunk polynomials is $M_f^{(i)} = \prod_{k \neq i}(\Delta_k + 1)$, and the number of them with the correlation coefficient close to the maximum value, given the constraint on $\hat{\Delta}_i$, is approximately $(1-p)^w / c_f^{(i)}$ where $c_f^{(i)}$ is defined as c_f without the ith multiplicative term in (6) corresponding to Δ_i, that is,

$$c_f^{(i)} = (1-p)^w \prod_{k \neq i} p^{\Delta_k - \hat{\Delta}_k^{opt}} (1-p)^{\hat{\Delta}_k^{opt}} \binom{\Delta_k}{\hat{\Delta}_k^{opt}}. \quad (12)$$

However, since f is not known, one should consider the average correlation coefficient over all $M_f^{(i)}$ shrunk polynomials, that is,

$$\bar{c}_f(\hat{\Delta}_i) = \frac{(1-p)^{w+1}}{\prod_{k \neq i}(\Delta_k + 1)} p^{\Delta_i - \hat{\Delta}_i} (1-p)^{\hat{\Delta}_i} \binom{\Delta_i}{\hat{\Delta}_i}. \quad (13)$$

The maximum value is achieved when $\hat{\Delta}_i$ satisfies (5) and is equal to $\bar{c}_f^{(i)} = c_f (1-p)^w / (M_f^{(i)} c_f^{(i)})$. This value would exactly be equal to c_f if the average was taken over the $(1-p)^w / c_f^{(i)}$ shrunk polynomials. Since f is not known, let $\underline{c}$ denote a lower bound on $c_f^{(i)} / (1-p)^w$, for any $1 \leq i \leq w$, that is given by $(2\pi p(1-p)(r-w)/(w-1))^{-(w-1)/2}$ for relatively large p, $p \geq (w-1)/(r-1)$, see (8), and by $(1-p)^{r-w}$ for small p, $p < (w-1)/(r-1)$, see (9). A lower bound $\underline{\bar{c}}$ on $\bar{c}_f^{(i)}$, for any $1 \leq i \leq w$, can be obtained similarly. Roughly speaking, one may use a lower bound $\underline{\bar{c}} = (1-p)^{w+1} \pi^{-1/2} ((r-1)/(w-1/2))^{-w+1/2}$.

The reconstruction procedure for finding an optimal shrunk polynomial is iterative and is based on estimating the correlation coefficient (13) on a given keystream sequence. For any assumed value of $\hat{\Delta}_i$, the number of different shrunk polynomials used to compute the average correlation coefficient is $1/\underline{c}$. Initially, for any possible value of $\hat{\Delta}_1$, one picks at random $1/\underline{c}$ shrunk polynomials with fixed $\hat{\Delta}_1$, estimates the average correlation coefficient on a given keystream sequence, and finds the best candidate for $\hat{\Delta}_1$ as the value for which the estimate is maximal. In the next step, one finds the best candidate for $\hat{\Delta}_2$ in a similar way except that the $1/\underline{c}$ shrunk polynomials are now chosen at random so that $\hat{\Delta}_1$ is around the best candidate value obtained in the first step. The procedure is then continued iteratively in an analogous way by using the best candidates for the values of $\hat{\Delta}_i$ from previous iterations to obtain the best estimate in the current iteration. After w steps, one then proceeds to recompute the best candidates obtained in the previous round of w steps and repeats the procedure for a number of rounds until no significant improvement is observed. It remains to determine the required length of the keystream sequence. For a reliable estimate

of the correlation coefficient (13) one needs $20 \log \hat{r}/\bar{\underline{c}}^2$ computations, so that the necessary length would then be $20\,\underline{c}\,\log \hat{r}/\bar{\underline{c}}^2$ (the total bit uncertainty of w and $\hat{\Delta}_i$ is here approximated as $2\,\log \hat{r}$). It is easy to see that in such a way no reduction in computation is achieved. However, the maint point to observe is that the reliable estimate of the correlation coefficient is not needed. What is needed for the convergence of the iterative procedure to an optimal shrunk polynomial is just a slight improvement on the unknown value of $\hat{\Delta}_i$ in each iteration step, where the first round is the most critical one. This is to a certain extent similar to iterative probabilistic decoding procedures used in fast correlation attacks [13, 19, 3, 15, 16]. Accordingly, the convergence criteria established in [3] and [16] lead us to anticipate that the required number of computations per each step is then just about $1/\bar{\underline{c}}$, so that the necessary length is then $\underline{c}/\bar{\underline{c}}$. The convergence of iterative probabilistic procedures is discussed in more detail in the next section.

4 Initial State Reconstruction

In this section, it is assumed that the LFSR feedback polynomial f of degree r is known, and the objective is a fast reconstruction of the unknown LFSR initial state and the unknown decimation sequence, given an observed segment of the keystream/decimated sequence. The problem is known to be very difficult, see [9]. A divide and conquer attack on the unknown decimation sequence based on the linear consistency test [21] proves to be successful if the observed decimated sequence is sufficiently long, also see the collision test [1, 2]. On the other hand, divide and conquer correlation attacks on the LFSR initial state are also possible if the observed decimated sequence is sufficiently long. Namely, the embedding correlation attack is analyzed in [23] and [10] for the constrained clocking case and in [10] for the unconstrained clocking case, whereas the statistically optimal probabilistic correlation attack in the unconstrained clocking case is analyzed in [10]. The attacks imply the exhaustive search over all possible decimation sequences and over all possible LFSR initial states, respectively. Despite a divide and conquer effect, the computational complexity remains exponential. Our ultimate goal is to examine whether fast correlation attacks are possible, with linear or polynomial complexity.

The starting point is the linear statistical weakness of the decimated LFSR sequences based on shrunk feedback polynomials discussed in Section 2. Our aim is to reconstruct the decimation sequence $\{d_t\}$ one term at a time based on appropriate locally applied shrunk polynomials. Note that the decimation sequence is synchronous with the decimated sequence $\{y_t\}$ and that d_t is the number of bits to be deleted before the bit y_t is produced. Alternatively, the regularly clocked LFSR sequence can be obtained from the decimated sequence by inserting d_t bits before y_t, for all observed t. The decimation sequence need not be reconstructed completely: essentially only slightly more than r consecutive terms at any point in time are required to be known to determine the LFSR initial state uniquely or almost uniquely. Of course, the known decima-

tion sequence can then be used for an attack on the secret key specifying the clock-control generator. Moreover, if the clock-control generator is an easy to reconstruct scheme, for example, another LFSR, then the decimation sequence reconstruction may incorporate the structure of the clock-control generator as well.

In the assumed probabilistic model, the decimation sequence is regarded as a sequence of independent identically distributed nonnegative integer random variables with a probability distribution $\mathcal{P} = \{P(d)\}_{d\in\mathcal{D}}$. The prior information about d_t is thus determined by the assumed probability distribution. For a statistical decision on the individual term d_t, some local statistic that would necessarily involve the two consecutive terms y_{t-1} and y_t is needed. The statistic should convey information about the value of d_t, so that the posterior probability of d_t becomes significantly different from the prior one. As in the fast correlation attacks on additively noised LFSR sequences, where the decision is being made on the individual terms of the binary noise sequence, the local information is extracted from the parity-check sums corresponding to the polynomial multiples of the feedback polynomial. They represent the codewords of the dual code of the linear code formed by the truncated LFSR sequences, see [3]. In this case, however, the parity-check sums are defined by the shrunk polynomials of the appropriate polynomial multiples of the feedback polynomial.

Let $h(z) = 1 + \sum_{k=1}^{\omega} z^{j_k}$, $1 \le j_1 < \ldots < j_\omega = m$, denote a polynomial multiple of degree m and weight $\Omega = \omega + 1$, and let $\hat{h}(z) = 1 + \sum_{k=1}^{\omega} z^{\hat{j}_k}$, $1 \le \hat{j}_1 < \ldots < \hat{j}_\omega = \hat{m}$, denote a shrunk polynomial of $h(z)$. Let $\tau_k = j_k - j_{k-1} - 1$ and $\hat{\tau}_k = \hat{j}_k - \hat{j}_{k-1} - 1$, $1 \le k \le \omega$, where $j_0 = \hat{j}_0 = 0$. The main idea is to find and use the polynomial multiples $h(z)$ such that $\tau_i = d$, for some $1 \le i \le \omega$, to check whether $d_t = d$, $d \in \mathcal{D}$. For each such $h(z)$ only the different shrunk polynomials $\hat{h}(z)$ such that $\hat{\tau}_i = 0$ are then used. The parity-check sum involving the consecutive terms y_{t-1} and y_t corresponds to $\hat{h}(z)$ in an obvious way. Each $h(z)$ may be multiply used to check the same value d, for all those i such that $\tau_i = d$. Let thus $\mathcal{H}_d$ denote a set of the polynomial multiples for checking the value d. For each polynomial $h(z)$ in $\mathcal{H}_d$ for which $\tau_i = d$, we choose a set of shrunk polynomials $\hat{h}(z)$ such that $\tau_i = 0$ and whose correlation coefficients are close to the maximum value $p^d(1-p)c_h^{(i)}$ where $c_h^{(i)}$ is obtained from the maximum value c_h by deleting the ith multiplicative term corresponding to τ_i, see (12). There are approximately $(1-p)^\omega / c_h^{(i)}$ such polynomials altogether, with the total conditional correlation coefficient $(1-p)^\omega$, see (11), and with the average conditional correlation coefficient close to $c_h^{(i)}$, where the assumed condition is that $d_t = d$. This is a very important point: although the conditional correlation coefficient associated with individual shrunk polynomials is small, the total conditional correlation coefficient may not be small because the number of the shrunk polynomials that are close to optimal is large for each feedback polynomial multiple $h(z)$.

The shrunk polynomials of different $h(z)$ in $\mathcal{H}_d$ may coincide. The conditional correlation coefficient $c(\hat{h})$ for a given polynomial $\hat{h}(z)$ is then defined as

the sum of the individual conditional correlation coefficients associated with all the polynomials from $\mathcal{H}_d$ for which $\hat{h}$ is close to an optimal shrunk polynomial. In this case, each shrunk polynomial is used only once. The conditional correlation coefficient $c(\hat{h})$ given $h(z)$ is given by the same expression as $c_h^{(i)}$, see (12), except that $\hat{\tau}_k$ is substituted for $\hat{\tau}_k^{opt}$. Since the shrunk polynomials used are close to optimal, $c_h^{(i)}$ may be a good approximation for $c(\hat{h})$ given $h(z)$ which can be further simplified in a way described in Section 2, see (7) and (9). This is especially the case if $\hat{\tau}_k$ is very close to $\hat{\tau}_k^{opt}$. If not, one can also use the normal or Poisson approximations to the binomial distribution. The underlying assumption is that the probability distribution $\mathcal{P}$ is geometric. Similar expressions can also be obtained for the case of constrained irregular clocking. Let $\hat{\mathcal{H}}_d$ denote a set of the so-obtained different shrunk polynomials of the polynomials in $\mathcal{H}_d$. A basic requirement is that the pairwise intersections between the sets $\mathcal{H}_d$, for different d, should be insignificant, so that by discarding just a few shrunk polynomials one could obtain distinct sets $\hat{\mathcal{H}}_d$. This in fact means that the correlation coefficients $c_h^{(i)}$ should not be close for polynomials $h(z)$ in different $\mathcal{H}_d$. More importantly, one should also check that for each polynomial in $\mathcal{H}_d$, the correlation coefficient $c_h^{(i)}$ is not close to the correlation coefficient $c_{h'}^{(i)}$ for any possible feedback polynomial multiple $h'(z)$ that can be used to check any other possible value of d. Preliminary experiments have shown that this is possible to achieve if the set of possible values of d is $\mathcal{D} = \{0, 1\}$ which corresponds to constrained irregular clocking. In any case, finding the appropriate shrunk polynomials can be done in precomputation time if the feedback polynomial is known. The obtained shrunk polynomials are called the parity-check polynomials.

We proceed now by describing and analyzing an iterative statistical decision procedure based on the parity-check sums computed by using the assumed parity-check polynomials. The statistically optimal decision rule for individual random variables d_t is based on the posterior probabilities. Let $\hat{P}_t(d)$ denote the posterior probability that $d_t = d$ given a set of the parity-check sums defined by the parity-check polynomials in $\hat{\mathcal{H}}_d$. Let $c(\hat{h})$ be a conditional correlation coefficient associated with a parity-check polynomial $\hat{h}$, where the assumed condition is that $d_t = d$. The basic assumption, justified by the choice of the parity-check polynomials, is that the conditional correlation coefficient is zero or very close to zero when $d_t \neq d$. Then by assuming that the observations determined by the individual parity-check sums are independent, it is not difficult to see that

$$\frac{\hat{P}(d)}{1-\hat{P}(d)} = \frac{P(d)}{1-P(d)} \prod_{\hat{h} \in \hat{\mathcal{H}}_d} (1 + c(\hat{h}))^{1-s(\hat{h})} (1 - c(\hat{h}))^{s(\hat{h})} \tag{14}$$

where $s(\hat{h})$ is the binary value of the parity-check sum defined by the parity-check polynomial $\hat{h}$, and the index t is omitted for simplicity. The posterior probability ratio (the *odds*) of the event $d_t = d$ is thus increased or decreased depending on the individual observations. The expression is similar to the one that is obtained for fast correlation attacks on additively noised LFSR sequences [16].

The iterative statistical decision procedure that we now propose is in fact conceptually similar to the iterative probabilistic decoding procedures proposed for the fast correlation attacks in [13, 22, 3, 15, 16]. The main idea for the iterative procedure is to use the posterior probabilities obtained by (14) in the previous step as the prior probabilities for the current iteration step. Of course, since the independence assumption for different d is not quite accurate, the posterior probabilities calculated by (14) should be normalized after each step. For this iterative procedure to work, we need an expression for the conditional correlation coefficient $c(\hat{h})$ in the case when the probability distribution $\mathcal{P} = \{P(d)\}_{d \in \mathcal{D}}$ is time-dependent, because the expression (3) only holds for identically distributed variables with the geometric distribution. It is assumed that the decimation random variables d_t are independent. The general expression is not included here for simplicity. The computational effort may be large, but most of it can be done in precomputation time allowing a certain degree of numerical approximation. However, the amount of computation can be drastically reduced by using various simplifications, which may even improve the effectiveness of the attack. For example, it is reasonable to approximate the updated probability distribution of the decimation sequence by a different geometric distribution for each multiplicative term in (3) with the corresponding estimate of the deletion rate. Thus we get

$$c(\hat{h}) \;=\; (1-p_0) \prod_{k \neq i} p_k^{\tau_k - \hat{\tau}_k} (1-p_k)^{\hat{\tau}_k + 1} \binom{\tau_k}{\hat{\tau}_k} \tag{15}$$

where the probabilities p_k, $0 \leq k \leq \omega$, $k \neq i$, are computed as the appropriate average values based on the updated probability distribution. Another interesting approach is to identify the terms d_t for which the posterior (updated) probability distribution is most different from the prior one, and then to use these posterior probability distributions as they are, without averaging, or even to make 'hard' decisions on these terms by using the maximum posterior probability decision rule. Other tricks may also be possible. The constrained clocking case, in which the set $\mathcal{D}$ of the possible values of d is upper-bounded or limited, can be treated analogously.

The iterative procedure is successful if it converges to a sequence close to the decimation sequence that has actually produced the given keystream sequence. The main question is whether the required number of the parity-check polynomials is realistic or not in view of the additional constraints imposed on the choice of these polynomials. As is also the case with the fast correlation attacks [13, 22, 3, 15, 16], the exact mathematical derivation of the conditions for success of the new attack does not seem to be tractable. However, we now show that a simplified analysis [16] can be adapted to deal with the new attack. The main result of the analysis in [16] is to show that the fast correlation attack on additively noised LFSR sequences is successful if *the odds improvement in the first iteration step is 'significant' in the most favourable case when all the parity-check sums are different from zero.* Here, the most favourable case is when the

parity-check sums are all equal to zero. The odds improvement is then given by

$$\prod_{\hat{h}\in\hat{\mathcal{H}}_d}(1+c(\hat{h})) \simeq 1+\sum_{\hat{h}\in\hat{\mathcal{H}}_d} c(\hat{h}) \simeq 1+\sum_{h\in\mathcal{H}_d}(1-p)^{\omega} \tag{16}$$

because the correlation coefficients are very small. If we heuristically assume that the significant initial improvement is 2, then we obtain a nice convergence condition

$$\sum_{\omega} N_{d,\omega}(1-p)^{\omega} > 1 \tag{17}$$

where $N_{d,\omega}$ denotes the number of the feedback polynomial multiples of weight $\omega+1$ in $\mathcal{H}_d$. The condition should be satisfied for all the values of d in $\mathcal{D}$ whose probability is not very close to zero. This condition applies in general, where p is the deletion rate of the decimation sequence. If it were not for the constrained choice of the feedback polynomial multiples, this condition would be less restrictive than the corresponding condition [16] for additively noised LFSR sequences. Note that for each polynomial h in $\mathcal{H}_d$, there are about $(1-p)^{\omega}/c_h^{(i)}$ parity-check polynomials in $\hat{\mathcal{H}}_d$, where $c_h^{(i)}$ is determined by (12). Therefore, the computational complexity to obtain the required parity-check sums corresponding to a polynomial h of weight $\omega+1$ and degree m is for relatively large p, $p \geq (\omega-1)/(m-1)$, upper-bounded by $(2\pi p(1-p)(m-\omega)/(\omega-1))^{(\omega-1)/2}$, see Section 3. This is feasible even if m is large provided that ω is relatively small. In the constrained clocking case where $\mathcal{D}=\{0,1\}$, the condition (17) seems to be very realistic. In general, the chances for success of the proposed fast correlation attack appear to be greater in the constrained clocking case than in the unconstrained one. From the cryptographic standpoint, it turns out that the number of the feedback polynomial multiples of relatively low weight and not too large degree should be relatively small with respect to (17). This is in accordance with a vague anticipation given in [4].

After the convergence process is completed or at any iteration step when the odds are significant, one can make 'hard' decisions on the values of d_t for some t. If the uncertainty reduction regarding the decimation sequence is large, one may then check the remaining needed values by exhaustive search, either by applying the linear consistency test [21] or by applying the embedding algorithm [23, 10]. In view of [21], one has to determine slightly more than r consecutive terms of the decimation sequence at any point in time, where r is the degree of the LFSR feedback polynomial. The LFSR state at some time t is then determined uniquely by solving the appropriate linear equations. The time t is not known if all the initial terms of the decimation sequence are not known. They can be determined by applying the embedding or the Levenshtein-like distance algorithms [23, 10].

5 Conclusion

A theory of fast correlation attacks on irregularly clocked LFSRs based on a recently determined linear statistical weakness of decimated LFSR sequences is

developed. The weakness and the corresponding correlation coefficients are analyzed in more detail. When the LFSR feedback polynomial is not known, the statistical weakness detection and the feedback polynomial reconstruction are discussed and an iterative algorithm for fast feedback polynomial reconstruction is proposed. When the LFSR feedback polynomial is known, an iterative procedure for fast LFSR initial state reconstruction given an observed keystream sequence is introduced. The procedure is based on specially defined parity-check sums corresponding to the shrunk polynomials of appropriate feedback polynomial multiples, and consists in iterative recomputation of the posterior probabilities for unknown elements of the decimation sequence. By using the analogy with the well-known fast correlation attacks on additively noised LFSR sequences, a convergence condition which determines the numbers of the parity-check sums required for successful reconstruction is derived. This condition and the corresponding polynomial computational complexity show that the proposed fast correlation attack may be realistic, especially in the constrained clocking case. Extensive computer simulations are needed to support the theory and to specify the technical details.

References

1. R. J. Anderson, "Solving a class of stream ciphers," *Cryptologia*, 14(3):285–288, 1990.
2. R. J. Anderson, "Faster attack on certain stream ciphers," *Electr. Lett.*, 29(15): 1322–1323, July 1993.
3. V. Chepyzhov and B. Smeets, "On a fast correlation attack on stream ciphers," Advances in Cryptology – EUROCRYPT '91, *Lecture Notes in Computer Science*, vol. 547, D. V. Davies ed., Springer-Verlag, pp. 176–185, 1991.
4. D. Coppersmith, H. Krawczyk, and Y. Mansour, "The shrinking generator," Advances in Cryptology – CRYPTO '93, *Lecture Notes in Computer Science*, vol. 773, D. R. Stinson ed., Springer-Verlag, pp. 22–39, 1994.
5. J. Dj. Golić and M. V. Živković, "On the linear complexity of nonuniformly decimated PN-sequences," *IEEE Trans. Inform. Theory*, 34:1077–1079, Sep. 1988.
6. J. Dj. Golić and M. J. Mihaljević, "A generalized correlation attack on a class of stream ciphers based on the Levenshtein distance," *Journal of Cryptology*, 3(3):201–212, 1991.
7. J. Dj. Golić and S. V. Petrović, "A generalized correlation attack with a probabilistic constrained edit distance," Advances in Cryptology - EUROCRYPT '92, *Lecture Notes in Computer Science*, vol. 658, R. A. Rueppel ed., Springer-Verlag, pp. 472–476, 1993.
8. J. Dj. Golić, "Correlation via linear sequential circuit approximation of combiners with memory," Advances in Cryptology – EUROCRYPT '92, *Lecture Notes in Computer Science*, vol. 658, R. A. Rueppel ed., Springer-Verlag, pp. 113–123, 1993.
9. J. Dj. Golić, "On the security of shift register based keystream generators," Fast Software Encryption – Cambridge '93, *Lecture Notes of Computer Science*, vol. 809, R. J. Anderson ed., Springer-Verlag, pp. 90–100, 1994.
10. J. Dj. Golić and L. O'Connor, "Embedding and probabilistic correlation attacks on clock-controlled shift registers," *Pre-proceedings of Eurocrypt '94*, pp. 231–243, Perugia, Italy, 1994.

11. J. Dj. Golić, "Intrinsic statistical weakness of keystream generators," *Pre-proceedings of Asiacrypt '94*, pp. 72–83, Wollongong, Australia, 1994.
12. D. Gollmann and W. G. Chambers, "Clock controlled shift registers: a review," *IEEE J. Sel. Ar. Commun.*, 7(4):525–533, 1989.
13. W. Meier and O. Staffelbach, "Fast correlation attacks on certain stream ciphers," *Journal of Cryptology*, 1(3):159–176, 1989.
14. W. Meier and O. Staffelbach, "The self-shrinking generator," *Pre-proceedings of Eurocrypt '94*, pp. 201–210, Perugia, Italy, 1994.
15. M. J. Mihaljević and J. Dj. Golić, "A comparison of cryptanalytic principles based on iterative error-correction," Advances in Cryptology – EUROCRYPT '91, *Lecture Notes in Computer Science*, vol. 547, D. V. Davies ed., Springer-Verlag, pp. 527–531, 1991.
16. M. J. Mihaljević and J. Dj. Golić, "Convergence of a Bayesian iterative error-correction procedure on a noisy shift register sequence," Advances in Cryptology – EUROCRYPT '92, *Lecture Notes in Computer Science*, vol. 658, R. A. Rueppel ed., Springer-Verlag, pp. 124–137, 1993.
17. R. A. Rueppel, "Stream ciphers," in *Contemporary Cryptology: The Science of Information Integrity*, G. Simmons ed., pp. 65–134. New York: IEEE Press, 1991.
18. T. Siegenthaler, "Decrypting a class of stream ciphers using ciphertext only," *IEEE Trans. Comput.*, 34:81–85, Jan. 1985.
19. K. C. Zeng and M. Huang, "On the linear syndrome method in cryptanalysis," Advances in Cryptology – CRYPTO '88, *Lecture Notes in Computer Science*, vol. 403, S. Goldwasser ed., Springer-Verlag, pp. 469–478, 1990.
20. K. C. Zeng, C. H. Yang, and T. R. N. Rao, "An improved linear syndrome algorithm in cryptanalysis with applications," Advances in Cryptology – CRYPTO '90, *Lecture Notes in Computer Science*, vol. 537, A. J. Menezes and S. A. Vanstone eds., Springer-Verlag, pp. 34–47, 1991.
21. K. C. Zeng, C. H. Yang, and T. R. N. Rao, "On the linear consistency test (LCT) in cryptanalysis and its applications," Advances in Cryptology – CRYPTO '89, *Lecture Notes in Computer Science*, vol. 435, G. Brassard ed., Springer-Verlag, pp. 164–174, 1990.
22. M. V. Živković, "On two probabilistic decoding algorithms for binary linear codes," *IEEE Trans. Inform. Theory*, 37:1707–1716, Nov. 1991.
23. M. V. Živković, "An algorithm for the initial state reconstruction of the clock-controlled shift register," *IEEE Trans. Inform. Theory*, 37:1488–1490, Sep. 1991.