On the Statistically Optimal Divide and Conquer Correlation Attack on the Shrinking Generator

Shahram Khazaei*, Mahmood Salmasizadeh**, Javad Mohajeri**
*Department of Electrical Engineering
Sharif University of Technology, Iran
shahram_kh@mehr.sharif.edu
**Electronic Research Center
Sharif University of Technology, Iran
(salmasi, mohajer)@sharif.edu

Abstract

The shrinking generator is a well-known key stream generator composed of two LFSR's, LFSR$_x$ and LFSR$_c$, where LFSR$_x$ is clock-controlled according to the regularly clocked LFSR$_c$. In this paper we investigate the minimum required length of the output sequence for successful reconstruction of the LFSR$_x$ initial state in an optimal probabilistic divide and conquer correlation attack. We extract an exact expression for the joint probability of the prefix of length m of the output sequence of LFSR$_x$ and prefix of length n of the output sequence of the generator. Then we use computer simulation to compare our probability measure and two other probability measures proposed in [5] and [3] in the sense of minimum required output length. Our simulation results show that our measure reduces the required output length.

Keywords. stream ciphers, clock-controlled generators, shrinking generator, divide and conquer attack, optimal correlation attacks, deletion channel, joint probability.

1 Introduction

Stream ciphers are commonly used as building blocks in secure communications. A binary stream cipher produces a key stream using a secret key, which controls the initial state of the stream cipher and possibly its structure. The output of the key stream generator is bit-wise added to the plaintext sequence to produce the ciphertext sequence. From a cryptanalysis point of view, a stream cipher must be resistant against a known-plaintext attack. In a known-plaintext attack, the cryptanalyst is given a segment of the key stream, and the goal is to get some information about the key, the initial state, the structure or even the unseen key stream bits, faster than exhaustive search over all possible keys.

The shrinking generator [1] is a well-known stream cipher composed of two linear feedback shift registers (LFSR's), LFSR$_x$ and LFSR$_c$. We denote the length of these LFSR's by r_x and r_c, and denote their regular output sequences by $X = \{x_t\}_{t=1}^{\infty}$ and $C = \{c_t\}_{t=1}^{\infty}$, respectively. The output sequence of the generator, $Y = \{y_t\}_{t=1}^{\infty}$, is the sequence obtained from X, by removing all x_t's for which $c_t = 0$.

A basic divide and conquer correlation attack on the unknown LFSR$_c$ initial state based on the linear consistency test [8] is successful if the observed output sequence is sufficiently long. The time complexity of this attack is $O(r_x^3 2^{r_c})$. On the other hand, in a probabilistic model where X and C are considered as purely random and independent binary sequences, the probabilistic correlation attack is applied by targeting LFSR$_x$. The statistically optimal correlation attack [5], which minimizes the required output length, is based on MAP decoding in a communication channel with independent deletion synchronization errors. The MAP decoding is applied by efficient computing the joint probability of X^m and Y^n, where X^m and Y^n are prefixes of length m and n of X and Y, respectively.

The channel capacity argument in [5] shows that for successful reconstruction of the LFSR$_x$ initial state, the required output length of the output sequence, n, is linear in r_x; and it has been conjectured that $n \approx 4r_x$ is sufficient for successful reconstruction of LFSR$_x$ initial state. Also, an efficient method for computing the joint probability of X^m and Y^n has been introduced. However, what has been introduced in [5] is not exactly the joint probability of X^m and Y^n, simulation results in [7] show that it is a good measure for successful reconstruction of the LFSR$_x$ initial state.

The exact value of the joint probability of X^n and Y^n has been computed in [3] and it has been suggested to compare this measure with the measure proposed in [5] with respect to the minimum required output sequence length. In this paper we extract an exact expression for the joint probability of X^m and Y^n. Then we use computer simulation to compare our measure and the two measures proposed in [5] and [3] with respect to minimum required output length for successful reconstruction of the LFSR$_x$ initial state.

Correlation attack using all three measures requires exhaustive search over all LFSR$_x$ possible initial states. Reduced complexity methods which instead require more length of the output sequence have been discussed in [3] and [6].

The paper is organized as follows. In section 2 we restate two measures proposed in [5] and [3], and our measure. Section 3 contains the simulation results. Conclusions are given in section 4. Proof of a theorem, which exactly and efficiently computes the joint probability of X^m and Y^n, is given in appendix.

2 Definition of the Three Measures and Their Usage in a Divide and Conquer Correlation Attack

Let first introduce some notations. The notation $A = \{a_t\}_{t=1}^{\infty}$ is used for a general binary sequence, $A_k = \{a_t\}_{t=k}^{\infty}$ for its subsequence, $A^n = \{a_t\}_{t=1}^{n}$ for its prefix of length n and $A_k^n = \{a_t\}_{t=k}^{n}$ for a segment of it. The number of ones in A^n is denoted by $w(A^n)$.

Let X, C, Y denote the output sequence of LFSR$_x$, LFSR$_c$, and the shrinking generator itself, respectively. We assume a probabilistic model where X and C are independent and purely random binary sequences. It then follows that the output sequence Y is also purely random. In this model the act of shrinking generator is equivalent to obtaining Y by sending X in a deletion channel which deletes the bits of X independently with probability equal to one half. Given an observed output sequence Y^n the MAP decoding, which is optimal decoding, finds an input sequence X that maximizes the joint probability $\Pr\{X, Y^n\}$ or equivalently the conditional probability $\Pr\{Y^n|X\}$. The prefix Y^n is obtained from prefix X^m with a probability that can be arbitrarily close to 1 by increasing m. As this probability is more than 0.99 for $m = \lfloor 2n + 3\sqrt{n} \rfloor$, for this choice of m, $\Pr\{Y^n|X^m\}$ is a very good approximation for $\Pr\{Y^n|X\}$. Thus for practical respects, we use the joint $\Pr\{X^m, Y^n\}$ probability for MAP decoding where $m = \lfloor 2n + 3\sqrt{n} \rfloor$.

2.1 Three Measures

In the following, we restate the measure which proposed in [5] as the joint probability of X^m and Y^n, the exact expression for the joint probability of X^n and Y^n which proposed in [3], and our exact expression for the joint probability of X^m and Y^n. To this end we use the following partial probability which has been defined in [3] for prefixes of X and Y

$$Q(e, s) = \Pr\{Y^s, w(C^{e+s}) = s | X^{e+s}\}. \tag{1}$$

The above partial probability can be recursively computed using the following lemma which has been proved in [3].

Lemma 1 *The partial probabilities $Q(e, s)$ is recursively determined by* [1]

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

for $0 \leqslant s \leqslant n$, $0 \leqslant e \leqslant m - s$, and $(e, s) \neq (0, 0)$, from the initial value $Q(0, 0) = 1$ [3]. (The terms corresponding to impermissible value of e *or* s *are assumed to be equal to zero.)*

[1] $\delta(i, j)$ is the Kronecker function, i.e., $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$.

The measure proposed in [5] in terms of Q is $2^{-n}Q(m-n,n)$ which is equal to $2^{m-n}\Pr\{X^m, Y^n, w(C^m)=n\}$ and obviously is not the joint probability of X^m and Y^n. The exact value of the joint probability of X^n and Y^n, proposed in [3], is computed using the following theorem.

Theorem 1 *For any given Y^n and X^n, we have*

$$\Pr\{X^n, Y^n\} = 2^{-n}\sum_{e=0}^{n} 2^{-e}Q(e, n-e) \tag{3}$$

where the partial probability $Q(e,s)$ is determined by Lemma 1 [3].

Our exact and efficient expression for computing the joint probability of X^m and Y^n is given by the following theorem which is proved in appendix.

Theorem 2 *For any given Y^n and X^m, where $m > n$, we have*

$$\Pr\{X^m, Y^n\} = 2^{-m-1}\sum_{e=0}^{m-n-1} Q(e,n) + 2^{-n}\sum_{e=m-n}^{m} 2^{-e}Q(e, m-e) \tag{4}$$

where the partial probability $Q(e,s)$ is determined by Lemma 1.

2.2 Hypotheses Testing

To find the correct initial state of the LFSR$_x$, we consider two hypothesis; H$_1$ correspond to an incorrect guess for the initial state, and H$_0$ correspond to the correct guess for the initial state. In a probabilistic model we suppose that under the hypothesis H$_1$, X and Y^n are purely random and independent, and under the hypothesis H$_0$ we assume that X is purely random and Y^n is the prefix of length n of the sequence obtained from X according to the deletion channel. We need a decision measure, $Z(n)$, which its distribution under two hypotheses are distinguishable. The variable n, is entered as an argument to show the available length of the output sequence. In hypothesis testing we deal with two error probabilities, p_{mis} and p_{fa}, which p_{mis} is the error probability of deciding H$_1$ is true while H$_0$ is indeed true, and p_{fa} is is the error probability of deciding H$_0$ is true while H$_1$ is indeed true. Ideally, we like to minimize both the p_{mis} and p_{fa} but normally there is a trade-off. Using *Neyman-Pearson lemma* [2], p_{fa} and p_{mis} are jointly minimized by considering a threshold T and choosing H$_0$ if and only if $\frac{\Pr\{Z(n)|H_0\}}{\Pr\{Z(n)|H_1\}} > T$. In most situation, the importance of two error probabilities are not equal. In our case, that is finding the correct initial state of LFSR$_x$, fixing p_{mis} to a value which need not to be very small (e.g. 0.1), p_{fa} must be around 2^{-r_x} for approximately unique reconstruction of LFSR$_x$ initial state. Given the decision measure $Z(n)$, corresponding p_{fa} determines the required output length for successful reconstruction of the LFSR$_x$ initial state. If p_{fa} decreases exponentially with n, the required n is linear in r_x.

A theoretical analysis of the separation between two probability distributions using each one of the three probability measures introduced in section 2.1 as a decision measure, seems to be very difficult. If there was such a theoretical analysis, it would be more reliable than the channel capacity argument discussed in [5]. Instead, we employ experimental computer simulation similar to [7]. The results are given in the following section.

3 Simulation Results

In this section we first use the computer simulation for estimating p_{fa} by fixing $p_{mis} = 0.1$ for all three measures introduced in section 2.1; that is

$$\begin{aligned} Z_1(n) &= 2^{m-n}\Pr\{X^m, Y^n, w(C^m) = n\} = 2^{-n}Q(m-n, n) && (5)\\ Z_2(n) &= \Pr\{X^n, Y^n\} && (6)\\ Z_3(n) &= \Pr\{X^m, Y^n\} && (7) \end{aligned}$$

where $m = \lfloor 2n + 3\sqrt{n} \rfloor$.

To this end, for each probability measure $Z_i (1 \leq i \leq 3)$, we choose 10000 purely random and independent pairs X^m and $C^m (X^n$ and C^n for $Z_2)$ and compute the corresponding output sequence in deletion channel model. If the output length is more than n, we consider the first n bits as Y^n, and if the output length is less than n, we pad it with some random bits to increase its length to n. This simulates the hypothesis H_0. Then for each pair X^m and $Y^n (X^n$ and Y^n for $Z_2)$, we compute the decision measure Z_i. After that in accordance with $p_{mis} = 0.1$, we compute a threshold th_i such that 9000 out of 10000 pairs have decision measure greater than th_i. To estimate p_{fa_i}, we choose 10000 purely random and independent pairs X^m and $Y^n (X^n$ and Y^n for $Z_2)$ and compute the probability measure Z_i for each pair. The estimated p_{fa_i} is percent of pairs which have probability measure greater than th_i. Table 1 contains our simulation results.

Table 1: Estimated p_{fa_i} for three measures for $n = 25 : 25 : 350$

$\boldsymbol{n}$	**25**	**50**	**75**	**100**	**125**	**150**	**175**
$\boldsymbol{p_{fa_1}}$	0.4294	0.3286	0.1932	0.1421	0.0793	0.0373	0.0289
$\boldsymbol{p_{fa_2}}$	0.3525	0.2025	0.1534	0.0908	0.0612	0.0312	0.0212
$\boldsymbol{p_{fa_3}}$	0.2057	0.0705	0.0329	0.0113	0.0049	0.0027	0.0013
$\boldsymbol{n}$	**200**	**225**	**250**	**275**	**300**	**325**	**350**
$\boldsymbol{p_{fa_1}}$	0.0179	0.0116	0.0055	0.0029	0.0014	0.0005	0.0000
$\boldsymbol{p_{fa_2}}$	0.0140	0.0081	0.0066	0.0036	0.0016	0.0007	0.0000
$\boldsymbol{p_{fa_3}}$	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Now, we approximate each p_{fa_i} by a curve $a_i 2^{-b_i n}$, where a_i and b_i are positive values. For being p_{fa_i} around 2^{-r_x}, the required output length must be approximately $n_i \simeq r_x/b_i$. From table 1 it follows that[2]

$$\begin{aligned} p_{fa_1} &\approx 1.10 \times 2^{-0.0314n} \rightarrow n_1 \approx 32r_x \\ p_{fa_1} &\approx 0.68 \times 2^{-0.0283n} \rightarrow n_2 \approx 35r_x \\ p_{fa_1} &\approx 0.34 \times 2^{-0.0475n} \rightarrow n_3 \approx 21r_x \end{aligned}$$

As we see our measure leads to less required output length to successful reconstruction of LFSR$_x$ initial state. Decreasing p_{mis}, increases the coefficients a_i's and does not lead to significant changes on coefficient b_i's. This is the reason that we fix p_{mis} to 0.1 and claim it is small enough, see [7] and [4] for similar cases. In [7] Success of correlation attack on a shrinking generator with special parameters using probability measure Z_1 has been confirmed. We do not follow this method using the other measures. Instead, we try to guess the distribution of Z_i under hypotheses H$_0$ and H$_1$. The values of Z_i are spread on a wide range. For a similar situation in [4], it has been noted that their measure, that is measure in [4] for another generator according to insertion channel, does not have normal distribution. However, in our case Z_i does not have normal distribution, the distribution of log(Z_i) can be approximated by normal distribution very well. Figure 1 shows estimated distribution of Z_3 for $n = 200$. As just the shape of distributions are important, we have not shown the axes number. Also, we have plotted the normal distribution with the same mean and variance for comparison. As we see, the approximation is very good.

4 Conclusion

We derived the exact expression for the joint probability of the prefix of length m of the output sequence of LFSR$_x$ and prefix of length n of the output sequence of the shrinking generator. Computer simulation shows that the required output length for successful reconstruction of the LFSR$_x$ initial state using this expression as a measure of correlation is less than two probability measures proposed in [5] and [3]. In the cryptographic point of view this improvement is not very important. However in coding point of view, justification of the gap between $4r_x$ and $21r_x$, which respectively come from channel capacity argument [5] and our simulation results, seems to be more important. The reason is obvious, the communication channel with independent deletion synchronization errors is such a bad channel that a *random coding* can not provide near capacity performance on it.

[2] Simulations in [7] shows that $n_1 \approx 20r_x$ which differs from our results. We ran and checked our simulation several times, but there was not significant change in the results. Any way, it is obvious that our measure needs less output length.

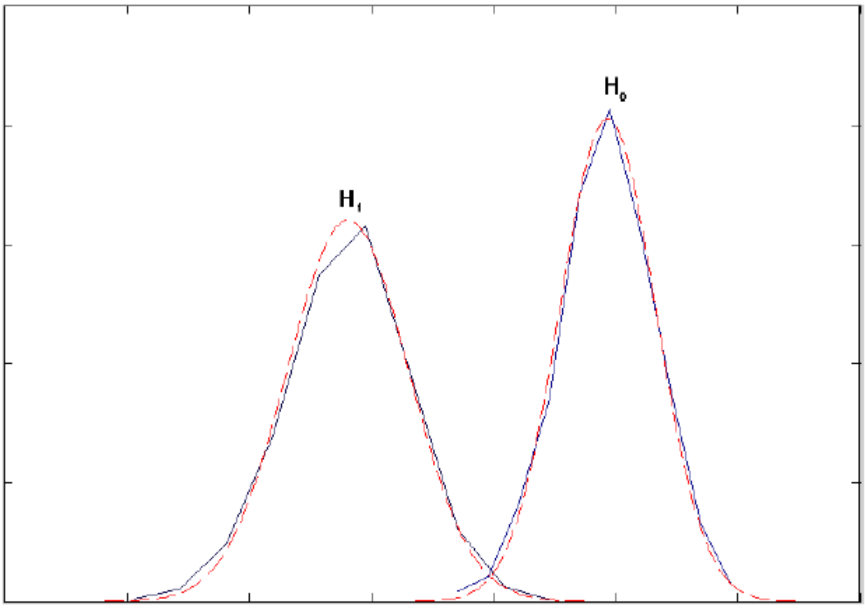

Figure 1: Figure 1: Estimated distributions of $\log(Z_3)$ for n=200

Appendix

Proof of Theorem 2.

As X is purely random we have $\Pr\{X^m, Y^n\} = 2^{-m}\Pr\{Y^n|X^m\}$, so we first compute $\Pr\{Y^n|X^m\}$. Using Total probability Theorem, we have

$$\Pr\{Y^n|X^m\} = \sum_{e=0}^{m} \Pr\{Y^n, w(C^m) = e|X^m\}. \tag{8}$$

To simplify $\Pr\{Y^n, w(C^m) = e|X^m\}$ where $0 \leq e \leq m$, we consider two situations, $0 \leq e \leq n$ and $n+1 \leq e \leq m$.

a) $0 \leq e \leq n$

In this case by separating Y^n, we have

$$\begin{aligned} \Pr\{Y^n, w(C^m) = e|X^m\} &= \Pr\{Y^e, Y^n_{e+1}, w(C^m) = e|X^m\} & (9) \\ &= \Pr\{Y^e, w(C^m) = e|X^m\}\Pr\{Y^n_{e+1}|Y^e, w(C^m) = e, X^m\} & (10) \\ &= Q(m-e, e)2^{-(n-e)} & (11) \end{aligned}$$

(10) follows from (9) according to the Chain Rule. Conditioned on $w(C^m) = e$, the string Y^n_{e+1} is obtained from X_{m+1} according to C_{m+1}, where X_{m+1} and

C_{m+1} are still mutually independent and purely random, even when conditioned on X^m and Y^e. Therefore, Y_{e+1}^n is purely random on the conditions $Y^e, w(C^m) = e$ and X^m. So (10) simplifies as (11).

b) $n+1 \leq e \leq m$

In this case we can not partition Y^n as (9). We partition the set of all C^m where $w(C^m) = e$ to some subsets which satisfy the conditions $w(C^k) = n$, $c_{k+1} = 1$ and $w(C_{k+2}^m) = e - n - 1$ over all allowed values for k, that is the overlap of three sets $n \leq k$ and $m-k-1 \leq e-n-1$ and $0 \leq k \leq m-1$. Using Total Probability Theorem, we have have

$$\Pr\{Y^n, w(C^m) = e|X^m\} = \sum_{k=n}^{m+n-e} \Pr\{Y^n, w(C^k) = n, c_{k+1} = 1, w(C_{k+2}^m) = e{-}n{-}1|X^m\}. \tag{12}$$

Using Chain Rule, we have

$$\begin{aligned}
&\Pr\{Y^n, w(C^k) = n, c_{k+1} = 1, w(C_{k+2}^m) = e-n-1|X^m\} = \\
&\Pr\{c_{k+1} = 1, w(C_{k+2}^m) = e-n-1|X^m\} \times \\
&\Pr\{Y^n, w(C^k) = n|X^m, c_{k+1} = 1, w(C_{k+2}^m) = e-n-1\}
\end{aligned} \tag{13}$$

As X and C are mutually independent and purely random, the second term of (13) is equal to

$$\Pr\{c_{k+1} = 1, w(C_{k+2}^m) = e-n-1\} = \frac{1}{2}\binom{m-k-1}{e-n-1}. \tag{14}$$

As $w(C^k)$ is independent of X and c_{k+1}, and conditioned on $w(C^k) = n$, Y^n is independent of X_{k+1} and C_{k+1}, the last term of (13) is simplified as

$$\Pr\{Y^n, w(C^k) = n|X^k\} = Q(k-n, n) \tag{15}$$

Using (14) and (15), for case (b) we have

$$\Pr\{Y^n, w(C^m) = e|X^m\} = \sum_{k=n}^{m+n-e} \frac{1}{2} Q(k-n, n)\binom{m-k-1}{e-n-1} \tag{16}$$

Combining the results of (a) and (b), and using some arithmetic relations, we have

$$\begin{aligned}
\Pr\{Y^n|X^m\} &= \sum_{e=0}^{n} Q(m-e, e)2^{-(n-e)} + \sum_{e=n+1}^{m} \sum_{k=n}^{m+n-e} \frac{1}{2} Q(k-n, n)\binom{m-k-1}{e-n-1} \\
&= \sum_{e=m-n}^{m} Q(e, m-e)2^{m-n-e} + \frac{1}{2}\sum_{e=0}^{m-n-1} Q(e, n)
\end{aligned} \tag{17}$$

(4) immediately follows from (17).

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. T. Cover and J.A. Thomas, "Elements of information Theory," Wiley series in Telecomunication, Wiley, 1991

3. J. Dj. Golic, "Correlation analysis of the Shrinking Generator," Advances in Cryptology– CRYPTO 2001, LNCS vol 2139, Springer-Verlag, 2001, pp. 440-457.

4. J. Dj. Golic and R. Menicocci, "Edit probability correlation attack on the alternating step generator," Sequences and their Applications - SETA '98, Discrete Mathematics and Theoretical Computer Science, C. Ding, T. Helleseth, and H.Niederreiter eds., Springer-Verlag, pp. 213-227, 1999.

5. J. Dj. Golic 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.

6. 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.

7. L. Simpson, J. Dj. Golic, and E. Dawson, "A probabilistic correlation attack on the shrinking generator," Information Security and Privacy - Brisbane '98, Lecture Notes in Computer Science, vol. 1438, pp. 147-158, 1998.

8. 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.