On the Success of the Embedding Attack on the Alternating Step Generator

Jovan Dj. Golić

System on Chip, Telecom Italia Lab
Telecom Italia
Via Guglielmo Reiss Romoli 274, I-00148 Turin, Italy
jovan.golic@tilab.com

Abstract. The edit distance correlation attack on the well-known alternating step generator for stream cipher applications was proposed by Golić and Menicocci. The attack can be successful only if the probability of the zero edit distance, the so-called embedding probability, conditioned on a given segment of the output sequence, decreases with the segment length, and if the decrease is exponential, then the required segment length is linear in the total length of the two linear feedback shift registers involved. The exponential decrease for the maximal value of the embedding probability as a function of the given output segment was estimated experimentally by Golić and Menicocci. In this paper, by using the connection with the interleaving and decimation operations, the embedding probability is theoretically analyzed. Tight exponentially small upper bounds on the maximal embedding probability are thus derived. Sharp exponentially small lower and upper bounds on the minimal embedding probability are also determined.

Index Terms: Correlation attack, decimation, edit distance, embedding probability, interleaving, sequences.

1 Introduction

It is well known that stream ciphers based on irregularly clocked linear feedback shift registers (LFSR's) are suitable for hardware implementations and can achieve a reasonably high security against secret key reconstruction attacks in the known plaintext scenario. Correlation attacks are a general class of divide-and-conquer attacks that aim at reconstructing the initial states of a subset of the LFSR's and use appropriate correlation measures between the known output sequence and the output sequences of the targeted LFSR's when regularly clocked. Such attacks based on the appropriately defined edit distances and edit probabilities are introduced in [1] and [2], respectively, and further analyzed in [3]. They are applicable to LFSR's that are clocked at least once per each output bit produced. In the case of a single irregularly clocked LFSR, the edit distances measure the possibility and the edit probabilities measure the probability of obtaining the output sequence from the assumed LFSR sequence, where

M. Matsui and R. Zuccherato (Eds.): SAC 2003, LNCS 3006, pp. 262-274, 2004.

the sequence controlling the clocking is not known. Therefore, such attacks can be called the embedding and probabilistic correlation attacks, respectively, and the embedding attack is first introduced in [9].

The stop-and-go (stop/go) clocking is particularly interesting for high speed applications. At any time, a stop/go shift register is clocked once if the clock-control input bit is equal to 1 (or 0) and is not clocked at all otherwise. The well-known alternating step generator (ASG) proposed in [6] consists of two stop/go clocked binary LFSR's, LFSR_X and LFSR_Y, and a clock-control regularly clocked binary LFSR, LFSR_C. At each time, the clock-control bit defines which of the two LFSR's is clocked, and the output sequence is obtained as the bitwise sum of the two stop/go clocked LFSR sequences. Some standard cryptographic properties of the ASG, such as a long period, a high linear complexity, and approximately uniform relative frequency of short output patterns on a period are established in [6], under the assumption that the clock-control sequence is a de Bruijn sequence and that the feedback polynomials of LFSR_X and LFSR_Y are primitive. It is expected that similar results also hold if the the clock-control sequence is produced by another LFSR, LFSR_C, with a primitive feedback polynomial whose period is coprime to the periods of LFSR_X and LFSR_Y.

It is shown in [6] that the initial state of LFSR_C can be recovered by a specific divide-and-conquer attack based on the fact that if and only if the guess about the initial state of LFSR_C is correct, then the first (binary) derivative of the ASG output sequence can be deinterleaved into the first derivatives of the regularly clocked LFSR_X and LFSR_Y sequences, which are then easily tested for low linear complexity.

An edit distance correlation attack targeting LFSR_X and LFSR_Y simultaneously is proposed in [4]. The specific edit distance incorporating the stop/go clocking is defined between one output string and two input strings. The output string is a given segment of the ASG output sequence and the input strings are the segments of the output sequences of LFSR_X and LFSR_Y whose initial states are guessed. An efficient algorithm for computing the edit distance is derived in [4]. Note that the attack has a divide-and-conquer effect since the unknown initial state of LFSR_C is not guessed.

If the initial states of LFSR_X and LFSR_Y are guessed correctly, then the edit distance is equal to zero. The zero edit distance means that the given segment of the ASG output sequence can be obtained from the assumed segments of the output sequences of LFSR_X and LFSR_Y by the stop/go clocking as in the ASG. In other words, it means that the ASG-embedding is possible. If the guess is incorrect, then the probability of obtaining the zero edit distance from random input strings, i.e., the ASG-embedding probability has to decrease with the output segment length in order for the attack to be successful, and if this decrease is exponential, then the required output segment length is linear in the total length of LFSR_X and LFSR_Y. This is because the expected number of false candidates for the initial states of LFSR_X and LFSR_Y can be approximated as the product of the ASG-embedding probability and the total number of incorrect guesses.

The experimental results from [4] indicate that the decrease of the ASG-embedding probability is exponential and that the output segment length of about four total lengths of LFSR_X and LFSR_Y is sufficient for success. Theoretical derivation and analysis of the maximal, average, and minimal values of the ASG-embedding probability, as a function of the given output segment, is qualified in [4] as a nontrivial combinatorial problem. Its solution is practically important for proving that the zero-edit-distance correlation attack on the ASG is successful if the output segment length is sufficiently large.

A partial step in this direction is made in [7] where an exponentially small upper bound on the average ASG-embedding probability is established by using the connection with the interleaving operation. Of course, this does not imply an exponentially small upper bound on the maximal ASG-embedding probability, which is needed to solve the problem completely. The main objective of this paper is to provide a more effective solution to the problem by deriving exponentially small upper bounds on the maximal ASG-embedding probability. This is achieved by a mathematically more involved approach using the connection with the interleaving and decimation operations.

Although it is out of the scope of this paper, it is interesting to mention that a correlation attack on an individual LFSR, either LFSR_X or LFSR_Y, which is based on a specific edit probability is later developed in [5]. For a similar approach, more explicitly using the connection with the interleaving and decimation operations, see [8]. Note that for individual LFSR's, the edit distance attack cannot be successful, due to a result from [3] regarding the embedding attack on irregularly clocked shift registers. By experimental analysis of the underlying statistical hypothesis testing problem, it is shown in [5] that the output segment length equal to about forty lengths of the targeted LFSR is sufficient for success. Theoretical analysis of this problem is very difficult and is related to the theoretically still open capacity problem for a communication channel with deletion errors.

This paper is organized as follows. In Section 2, the mathematical definitions of the ASG-embedding probability and the related interleaving and decimation probabilities are introduced. Some basic relations among these probabilities are established in Section 3. By analyzing the decimation probability, various exponentially small upper bounds on the maximal ASG-embedding probability are derived in Section 4. By analyzing the interleaving probability, exponentially small lower and upper bounds on the minimal ASG-embedding probability are determined in Section 5. Conclusions are presented in Section 6.

2 Preliminaries

For a binary sequence $A = a_0, a_1, \cdots$ or $A = a_1, a_2, \cdots$, let $A_k = a_k, a_{k+1}, \cdots$ and let $A_k^n = a_k, a_{k+1}, \cdots, a_n$ denote a segment of length $n - k + 1$. Formally, A_k^n is empty if $k > n$ and, for simplicity, let $A^n = A_1^n$. Further, the first (binary) derivative of A is denoted by $\dot{A} = \dot{a}_0, \dot{a}_1, \cdots$, where $\dot{a}_i = a_i \oplus a_{i+1}$ and $\oplus$ stands for the binary (modulo 2) addition. The (binary) complement of A is denoted by

$\overline{A} = \overline{a}_0, \overline{a}_1, \cdots$, where $\overline{a}_i = a_i \oplus 1$. As usual, $A^m B^n$ denotes the concatenation of A^m and B^n, whereas a^n denotes a constant sequence $a, a, \cdots, a$.

Let X, Y, and C be the output sequences of LFSR_X, LFSR_Y, and LFSR_C in the ASG, respectively. Assuming the step-then-add mode of operation, the ASG output sequence $Z = z_1, z_2, \cdots$ is generated as follows. Initially, the first bits of X and Y, that is, x_0 and y_0 are produced. For each $t \geq 1$, depending on whether c_{t-1} equals 1 or 0, the next bit of X or Y is produced, respectively, and the output bit z_t is obtained as the binary sum of the last produced bits of X and Y. The ASG output sequence produced from X_0^n, Y_0^n, and C_0^{n-1} is denoted as $Z^n = Z_1^n = \mathrm{ASG}(X_0^n, Y_0^n, C_0^{n-1})$, where $n \geq 1$.

Let $W^n = \mathrm{INT}(U^n, V^n, C^n)$ denote the sequence obtained by interleaving U^n and V^n according to C^n, w_i being taken from U^n if $c_i = 1$ and from V^n if $c_i = 0$. The ASG and interleaving operations are connected by

$$\dot{Z}_1^{n-1} = \begin{cases} \mathrm{INT}(\dot{X}_1^{n-1}, \dot{Y}_0^{n-2}, C_1^{n-1}) \text{ if } c_0 = 1 \\ \mathrm{INT}(\dot{X}_0^{n-2}, \dot{Y}_1^{n-1}, C_1^{n-1}) \text{ if } c_0 = 0 \end{cases}. \tag{1}$$

Our main objective is to analyze the *ASG-embedding* probability that is according to [4] defined as the conditional probability

$$P_n(Z^n) = \Pr\{(X_0^n, Y_0^n) \,|\, (\exists C_0^{n-1})\, \mathrm{ASG}(X_0^n, Y_0^n, C_0^{n-1}) = Z^n\} \tag{2}$$

in the probabilistic model in which X and Y are assumed to be mutually independent sequences of independent uniformly distributed binary random variables. In this model,

$$P_n(Z^n) = \frac{1}{2^{2(n+1)}} \,|\{(X_0^n, Y_0^n) : (\exists C_0^{n-1})\, \mathrm{ASG}(X_0^n, Y_0^n, C_0^{n-1}) = Z^n\}|. \tag{3}$$

We similarly define the *interleaving probability* as

$$p_n(W^n) = \frac{1}{2^{2n}} \,|\{(U^n, V^n) : (\exists C^n)\, \mathrm{INT}(U^n, V^n, C^n) = W^n\}|. \tag{4}$$

Both probabilities are invariant under complementation, that is,

$$P_n(\overline{Z}^n) = P_n(Z^n) \tag{5}$$

$$p_n(\overline{W}^n) = p_n(W^n). \tag{6}$$

We are particularly interested in deriving the maximal and minimal values of the two probabilities over Z^n and W^n, respectively. They are denoted as $P_n^{\max}$, $p_n^{\max}$, $P_n^{\min}$, and $p_n^{\min}$, respectively. The empirical estimates from [4] are $P_n^{\max} \approx 0.72 \cdot 0.915^n$ and $P_n^{\min} \approx 2.7 \cdot 0.562^n$.

To this end, we introduce the decimation operation by $U^k = \mathrm{DEC}(W^n, C^n)$ where w_i is taken to the output if $c_i = 1$ and discarded if $c_i = 0$, and $k = h(C^n)$ is the Hamming weight of C^n (for $k = 0$, U^0 is empty). Let

$$D_k(W^n) = \{U^k : (\exists C^n)\, \mathrm{DEC}(W^n, C^n) = U^k\}, \quad 0 \leq k \leq n \tag{7}$$

$$\nu_n(W^n) = \frac{1}{2^n} \sum_{k=0}^{n} |D_k(W^n)|. \tag{8}$$

Clearly, $|D_k(W^n)| \geq 1$ and $|D_0(W^n)| = |D_n(W^n)| = 1$. Accordingly, $\nu_1(W^1) = 1$ and

$$(n+1)2^{-n} \leq \nu_n(W^n) \leq 1. \tag{9}$$

Since $\sum_{k=0}^{n} |D_k(W^n)|$ is the number of sequences, of various lengths, that can be obtained by decimating W^n, we can call $\nu_n(W^n)$ the *decimation probability*. Like $p_n(W^n)$, $\nu_n(W^n)$ is also invariant under complementation (see (6)). One can analogously define its maximal and minimal values, respectively.

3 Basic Relations

Lemma 1, based on the characterization (1), determines the relation between P_n and p_{n-1}. Lemma 2 specifies a basic relation between the interleaving probability p_n and the decimation probability ν_n. *Throughout the paper, in all the mathematical results stated, it is assumed that the unspecified quantities take arbitrary values* (e.g., n and Z^n in Lemma 1).

Lemma 1.

$$\frac{3}{4} p_{n-1}(\dot{Z}^{n-1}) \leq P_n(Z^n) \leq p_{n-1}(\dot{Z}^{n-1}). \tag{10}$$

The same inequalities hold for the maximal and minimal values, respectively.

Proof The set $\{(X_0^n, Y_0^n) : (\exists C_0^{n-1})\, \mathrm{ASG}(X_0^n, Y_0^n, C_0^{n-1}) = Z^n\}$ can be partitioned into three subsets according to $x_0 \oplus y_1 = z_1 \neq x_1 \oplus y_0$, $x_0 \oplus y_1 = \overline{z}_1 \neq x_1 \oplus y_0$, and $x_0 \oplus y_1 = x_1 \oplus y_0 = z_1$. In the first two subsets (X_0^n, Y_0^n) uniquely determines c_0, whereas in the third subset this is not true. In view of (1), we first get

$$\begin{aligned} &|\{(X_0^n, Y_0^n) : x_0 \oplus y_1 = z_1 \neq x_1 \oplus y_0, (\exists C_0^{n-1})\, \mathrm{ASG}(X_0^n, Y_0^n, C_0^{n-1}) = Z^n\}| \\ &\quad = |\{(X_0^n, Y_0^n) : x_0 \oplus y_1 = z_1 \neq x_1 \oplus y_0, \\ &\qquad\qquad (\exists C_1^{n-1})\, \mathrm{INT}(\dot{X}_0^{n-2}, \dot{Y}_1^{n-1}, C_1^{n-1}) = \dot{Z}^{n-1}\}| \\ &\quad = 4\, |\{(\dot{X}_0^{n-2}, \dot{Y}_1^{n-1}) : (\exists C_1^{n-1})\, \mathrm{INT}(\dot{X}_0^{n-2}, \dot{Y}_1^{n-1}, C_1^{n-1}) = \dot{Z}^{n-1}\}| \end{aligned} \tag{11}$$

since $c_0 = 0$ and both x_0 and x_n can take arbitrary values. Analogous equation holds if $x_0 \oplus y_1 = \overline{z}_1 \neq x_1 \oplus y_0$, in which case $c_0 = 1$ and y_0 and y_n can be arbitrary. Consequently, we get

$$\begin{aligned} &|\{(X_0^n, Y_0^n) : x_0 \oplus y_1 \neq x_1 \oplus y_0, (\exists C_0^{n-1})\, \mathrm{ASG}(X_0^n, Y_0^n, C_0^{n-1}) = Z^n\}| \\ &\quad = 8\, |\{(U^{n-1}, V^{n-1}) : (\exists C^{n-1})\, \mathrm{INT}(U^{n-1}, V^{n-1}, C^{n-1}) = \dot{Z}^{n-1}\}|. \end{aligned} \tag{12}$$

In the remaining case $x_0 \oplus y_1 = x_1 \oplus y_0 = z_1$, c_0 can take both values. For $c_0 = 0$, similarly as (11), we get

$$\begin{aligned}
&|\{(X_0^n, Y_0^n) : x_0 \oplus y_1 = x_1 \oplus y_0 = z_1, \\
&\qquad\qquad (\exists C_0^{n-1})\,(c_0 = 0, \mathrm{ASG}(X_0^n, Y_0^n, C_0^{n-1}) = Z^n)\}| \\
&= |\{(X_0^n, Y_0^n) : x_0 \oplus y_1 = x_1 \oplus y_0 = z_1, \\
&\qquad\qquad (\exists C_1^{n-1})\,\mathrm{INT}(\dot{X}_0^{n-2}, \dot{Y}_1^{n-1}, C_1^{n-1}) = \dot{Z}^{n-1}\}| \\
&= 4\,|\{(\dot{X}_0^{n-2}, \dot{Y}_1^{n-1}) : (\exists C_1^{n-1})\,\mathrm{INT}(\dot{X}_0^{n-2}, \dot{Y}_1^{n-1}, C_1^{n-1}) = \dot{Z}^{n-1}\}|. \quad (13)
\end{aligned}$$

Analogous equation is obtained for $c_0 = 1$, but the two corresponding subsets are not necessarily disjoint. Accordingly, we get

$$\begin{aligned}
&4\,|\{(U^{n-1}, V^{n-1}) : (\exists C^{n-1})\,\mathrm{INT}(U^{n-1}, V^{n-1}, C^{n-1}) = \dot{Z}^{n-1}\}| \\
&\quad \le |\{(X_0^n, Y_0^n) : x_0 \oplus y_1 = x_1 \oplus y_0, (\exists C_0^{n-1})\,\mathrm{ASG}(X_0^n, Y_0^n, C_0^{n-1}) = Z^n\}| \\
&\quad \le 8\,|\{(U^{n-1}, V^{n-1}) : (\exists C^{n-1})\,\mathrm{INT}(U^{n-1}, V^{n-1}, C^{n-1}) = \dot{Z}^{n-1}\}|. \quad (14)
\end{aligned}$$

Finally, (10) is a direct consequence of (3), (4), (12), and (14). □

Lemma 2.

$$2^{-n} \le p_n(W^n) \le \nu_n(W^n). \tag{15}$$

The same inequalities hold for the maximal and minimal values, respectively.

Proof The inequalities directly follow from

$$p_n(W^n) = \frac{1}{2^{2n}}\,|\cup_{k=0}^{n} \{(U^n, V^n) : (\exists C^n)\,(h(C^n) = k, \mathrm{INT}(U^n, V^n, C^n) = W^n)\}| \tag{16}$$

$$|\{(U^n, V^n) : (\exists C^n)\,(h(C^n) = k, \mathrm{INT}(U^n, V^n, C^n) = W^n)\}| = 2^n |D_k(W^n)|, \tag{17}$$

and (17) follows from the fact that when $h(C^n) = k$, then $\mathrm{INT}(U^n, V^n, C^n) = W^n$ holds iff $U^k = \mathrm{DEC}(W^n, C^n)$ and $V^{n-k} = \mathrm{DEC}(W^n, \overline{C}^n)$, whereas the remaining n bits of (U^n, V^n) are arbitrary. □

4 Maximal Probabilities

Our approach consists of obtaining exponentially small upper bounds on the maximal ASG-embedding probability $P_n^{\max}$ by deriving appropriate upper bounds on the decimation probability ν_n. An analytical method for deriving an upper bound on ν_n is presented in Section 4.1 and is further refined in Section 4.3 to determine $\nu_n^{\max}$ and thus obtain the tightest upper bound in terms of the decimation probability. A method based on a concatenation lemma and direct counting is given in Section 4.2.

4.1 Analytical Upper Bound

A run in a binary sequence W is a maximal-length segment of W that consists of equal bits. A binary sequence W^n can uniquely be represented as a sequence of r runs and let $L^r = l_1, l_2, \cdots, l_r$ denote the positive integer sequence of the run lengths. Accordingly, $\sum_{i=1}^{r} l_i = n$. Clearly, the only two sequences with the same L^r are W^n and $\overline{W}^n$.

Our objective is to derive an upper bound on $\nu_n(W^n)$ that is strictly smaller than 1 for every r and then use it to obtain an exponentially small upper bound on $P_n^{\max}$ by applying Lemmas 1 and 2. Our most involved mathematical result is the following lemma, which uses the concept of minimal decimation sequences introduced in [3].

Lemma 3.

$$\nu_n(W^n) < \frac{1}{2^n}\,\frac{\sqrt{5}+2}{\sqrt{5}}\left(\frac{\sqrt{5}+1}{2}\right)^r \prod_{i=1}^{r} l_i \le \frac{1}{2^n}\,\frac{\sqrt{5}+2}{\sqrt{5}}\left(\frac{\sqrt{5}+1}{2}\right)^r \left(\frac{n}{r}\right)^r. \quad (18)$$

Proof Every decimation sequence C^n such that $h(C^n) = k \ge 1$ can uniquely be represented as an increasing sequence of positive integers $i_1, i_2, \cdots, i_k$ where i_j is the index of the jth 1 in C^n. Now, for given W^n and U^k, there may exist different C^n such that $U^k = \mathrm{DEC}(W^n, C^n)$. However, as observed in [3], for any W^n and each U^k that can be obtained by decimating W^n, there exists a unique decimation sequence $\tilde{C}^n$ such that $U^k = \mathrm{DEC}(W^n, \tilde{C}^n)$ and that either $k = 0$ or $\tilde{C}^n$ is minimal in the sense that, for each j, i_j is minimal on the set of all C^n such that $U^k = \mathrm{DEC}(W^n, C^n)$.

Such a decimation sequence is called the minimal decimation sequence and can be recursively constructed as follows: i_1 is the index of the first bit of W^n equal to u_1, and for each $j \ge 2$, i_j is the index of the first bit w_i, $i > i_{j-1}$, equal to u_j. The constructed sequence is the unique minimal decimation sequence, because if we suppose that there exists a sequence C'^n such that $i'_j < i_j$ for at least one value of j, then we get a contradiction. Namely, from the construction it then follows that $i'_{j'} < i_{j'}$ for every $j' < j$, which is impossible for $j' = 1$.

Since, for a given W^n, there is an 1-1 correspondence between the minimal decimation sequences and all possible U^k, $\sum_{k=0}^{n} |D_k(W^n)|$ is equal to the number of the minimal decimation sequences given W^n. Each $\tilde{C}^n$ can be characterized by an r-bit sequence $D^r = d_1, \cdots, d_r$ such that $d_i = 1$ iff at least one bit from the ith run of W^n is taken to the output as well as by the numbers of bits taken to the output from each of these s runs, where $s = h(D^r)$. The constraints stemming from the definition of minimal decimation sequences are that there can exist no two consecutive 0's in D^r except at the end and that if $d_i = 1$ and $d_{i+1} = 0$, then the number of bits taken to the output from the ith run is maximal possible, that is, l_i. Let M_s denote the number of all such D^r, $s = h(D^r)$, $0 \le s \le r$.

Consequently, we obtain

$$\sum_{k=0}^{n} |D_k(W^n)| \le \sum_{s=0}^{r} M_s \prod_{i=1}^{r} l_i \le \left(\frac{n}{r}\right)^r \sum_{s=0}^{r} M_s \quad (19)$$

where, to get the right-hand inequality, we used $\sum_{i=1}^{r} l_i = n$ and the well-known inequality relating the geometric and arithmetic means, for nonnegative real numbers x_i, $1 \le i \le r$,

$$(\prod_{i=1}^{r} x_i)^{1/r} \le \frac{1}{r} \sum_{i=1}^{r} x_i. \tag{20}$$

It remains to determine $\sum_{s=0}^{r} M_s$, which is the total number of r-bit sequences that can contain consecutive 0's only at the end. It is convenient to represent this number as $\sum_{j=0}^{r} N_j$, where N_j is the number of such sequences with the additional property that j is the index of the last bit equal to 1 (where $j = 0$ means that there are no 1's at all). It follows that $N_0 = N_1 = 1$. Let $\mathcal{N}_j$ denote the set of all j-bit sequences with the jth bit equal to 1 and without consecutive 0's, $j \ge 1$. If $B^j = b_1, \cdots, b_j \in \mathcal{N}_j$, then b_{j-1} can be equal to 1 or 0. In the former case, $B^{j-1} \in \mathcal{N}_{j-1}$ and in the latter case, $B^{j-2} \in \mathcal{N}_{j-2}$. Accordingly, we get

$$N_j = N_{j-1} + N_{j-2}, \quad j \ge 2. \tag{21}$$

Therefore $N_0, N_1, N_2, \cdots$ is the Fibonacci sequence, so that $\sum_{j=0}^{r} N_j$ can be expressed as

$$\sum_{j=0}^{r} N_j = \frac{\sqrt{5}+2}{\sqrt{5}} \left(\frac{\sqrt{5}+1}{2}\right)^r + \frac{\sqrt{5}-2}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^r - 1$$
$$< \frac{\sqrt{5}+2}{\sqrt{5}} \left(\frac{\sqrt{5}+1}{2}\right)^r. \tag{22}$$

Finally, (19), (22), and (8) result in (18). □

Theorem 1.

$$P_n^{\max} \le p_{n-1}^{\max} \le \nu_{n-1}^{\max} < \frac{\sqrt{5}+2}{\sqrt{5}} \left(\frac{1}{2} e^{\frac{\sqrt{5}+1}{2e}}\right)^{n-1} < 2.08929 \cdot 0.906735^n. \tag{23}$$

Proof Lemma 1 and Lemma 2 imply that $P_n^{\max} \le p_{n-1}^{\max} \le \nu_{n-1}^{\max}$. Let

$$\alpha = \frac{\sqrt{5}+1}{2}, \quad \beta = \frac{\alpha}{e}. \tag{24}$$

Equation (18) from Lemma 3 can be put into the form

$$\nu_n(W^n) < \frac{\sqrt{5}+2}{\sqrt{5}} \left(\frac{1}{2}\, e^{f(r/n)}\right)^n \tag{25}$$

where $f(\gamma) = \gamma \ln \alpha - \gamma \ln \gamma$, $\gamma = r/n$. Accordingly, we have

$$\nu_n^{\max} < \frac{\sqrt{5}+2}{\sqrt{5}} \left(\frac{1}{2}\, e^{f_{\max}}\right)^n \tag{26}$$

where $f_{\max}$ is the maximum of $f(\gamma)$ on $(0,1]$. It is easy to prove that f'' is negative and hence f is concave and has a unique maximum reached for $\gamma = \beta$. Consequently, we get (23). □

The upper bound in (23) is smaller than 1 for $n \geq 8$.

4.2 Concatenation Upper Bounds

It is interesting that direct counting by computer simulations can be used to obtain exponentially small upper bounds on $\nu_n^{\max}$ via the following concatenation lemma for the decimation probability ν_n. This lemma and the resulting theorem have counterparts pertaining to the embedding probability for an embedding correlation attack on irregularly clocked shift registers under a constraint on the maximal number of consecutive clocks (see [9] and [3]).

Lemma 4.

$$\nu_{m+n}(A^m B^n) \leq \nu_m(A^m)\nu_n(B^n). \tag{27}$$

The same inequality holds for the maximal and minimal values, respectively.

Proof According to the definition of the decimation operation, it follows that $\mathrm{DEC}(A^m B^n, C^{m+n}) = W^k$ is true iff there exist k_1 and k_2 such that $0 \leq k_1 \leq m$, $0 \leq k_2 \leq n$, $k_1 + k_2 = k$, $W^k = U^{k_1} V^{k_2}$, $\mathrm{DEC}(A^m, C^m) = U^{k_1}$, and $\mathrm{DEC}(B^n, C_{m+1}^{m+n}) = V^{k_2}$. This then implies that

$$|D_k(A^m B^n)| \leq \sum_{k_1,k_2} |D_{k_1}(A^m)||D_{k_2}(B^n)| \tag{28}$$

where the sum is over all the permitted values of k_1 and k_2. Therefore, we obtain

$$\sum_{k=0}^{m+n} |D_k(A^m B^n)| \leq \sum_{k_1=0}^{m} |D_{k_1}(A^m)| \sum_{k_2=0}^{n} |D_{k_2}(B^n)| \tag{29}$$

which in view of (8) directly yields (27). □

Theorem 2.

$$P_n^{\max} \leq p_{n-1}^{\max} \leq \nu_{n-1}^{\max} \leq \eta_m (\nu_m^{\max})^{n/m} \tag{30}$$

$$\eta_m = \frac{1}{(\nu_m^{\max})^{1/m}} \max_{0 \leq l \leq m-1} \frac{\nu_l^{\max}}{(\nu_m^{\max})^{l/m}}. \tag{31}$$

Proof Let $n - 1 = m\lfloor (n-1)/m \rfloor + l$, $0 \leq l \leq m-1$. As a direct consequence of Lemma 4, we get

$$\nu_{n-1}^{\max} \leq \nu_l^{\max} (\nu_m^{\max})^{\lfloor (n-1)/m \rfloor} \tag{32}$$

and then (30) and (31) easily follow in light of Lemmas 1 and 2. □

By using Theorem 2, exponentially small upper bounds can be obtained from any $\nu_m^{\max} < 1$. It follows that $\nu_1^{\max} = \nu_2^{\max} = 1$, but we find that $\nu_m^{\max} < 1$ already for $m = 3$. By direct counting one can compute $\nu_m^{\max}$ for $m \geq 3$ and thus obtain various exponentially small upper bounds. For $m \geq 6$, these bounds are sharper than the bound in (23).

4.3 Tight Upper Bound

Direct counting via computer simulations for moderately small values of n shows that $\nu_n^{\max}$ is achieved by alternating sequences. An analogous observation for $P_n^{\max}$ is made in [4]. It is then reasonable to expect that this holds for any n. If this is true and if we are able to determine $\nu_n(W^n)$ for an alternating sequence W^n, we can then obtain the tightest upper bound on $P_n^{\max}$ in terms of the decimation probability. Indeed, this is achieved by Lemmas 5 and 6 and Theorem 3.

Lemma 5. *If W^n is an alternating sequence, then*

$$\nu_n(W^n) = \frac{\sqrt{5}+2}{\sqrt{5}} \left(\frac{\sqrt{5}+1}{4}\right)^n + \frac{\sqrt{5}-2}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{4}\right)^n - \frac{1}{2^n} < \frac{\sqrt{5}+2}{\sqrt{5}} \left(\frac{\sqrt{5}+1}{4}\right)^n. \quad (33)$$

Proof If W^n is an alternating sequence, then $r = n$, so that Lemma 3 directly implies that $\nu_n(W^n)$ is upper-bounded as in (33). Moreover, (19) then holds with equalities and hence (22) implies the equality in (33). □

Lemma 6.

$$\nu_{n+1}(W^n w_n) < \nu_{n+1}(W^n \overline{w}_n). \quad (34)$$

Proof According to the proof of Lemma 3, $\nu_n(W^n)$ can be expressed in terms of minimal decimation sequences as

$$\nu_n(W^n) = \frac{1}{2^n} \sum_{i=0}^{n} K_i(W^n) \quad (35)$$

where $K_i(W^n)$ is the number of the minimal decimation sequences $\tilde{C}^n$ such that i is the index of the last bit of $\tilde{C}^n$ equal to 1 (if $\tilde{C}^n = 0^n$, then $i = 0$ and $K_0(W^n) = 1$). It follows that $K_i(W^n) \geq 1$, because the decimation sequence $\tilde{C}^n = 1^i 0^{n-i}$ is always minimal. As $K_i(W^n) = K_i(W^m)$ for $m \geq n$, it suffices to compare only $K_{n+1}(W^n w_n)$ and $K_{n+1}(W^n \overline{w}_n)$.

To this end, we have to consider only the minimal decimation sequences $\tilde{C}^{n+1}$ such that $\tilde{c}_{n+1} = 1$. By virtue of the minimality of minimal decimation sequences, the first n bits of any such sequence also constitute a minimal decimation sequence.

Further, for $W^n w_n$, the last two bits are equal, so that the minimality of minimal decimation sequences implies that $\tilde{c}_n = 1$. Hence

$$K_{n+1}(W^n w_n) = K_n(W^n). \tag{36}$$

On the other hand, for $W^n \overline{w}_n$, the last bit is different from all the bits in the last run of W^n whose length is l_r. The minimality of minimal decimation sequences then implies that $\tilde{C}^{n+1}$ is minimal iff $\tilde{c}_{n+1} = 1$ and $\tilde{C}^n$ is a minimal decimation sequence such that the index, i, of the last bit of $\tilde{C}^n$ equal to 1 satisfies $n - l_r \le i \le n$. Consequently, in light of (36), we get

$$K_{n+1}(W^n \overline{w}_n) = \sum_{n-l_r}^{n} K_i(W^n) = K_{n-l_r}(W^n) + l_r K_n(W^n). \tag{37}$$

Now, as $l_r \ge 1$ and $K_{n-l_r} \ge 1$, (36) and (37) imply that $K_{n+1}(W^n w_n) < K_{n+1}(W^n \overline{w}_n)$, so that (35) then results in (34). □

Theorem 3.

$$\begin{aligned} P_n^{\max} \le p_{n-1}^{\max} \le \nu_{n-1}^{\max} \\ = \frac{\sqrt{5}+2}{\sqrt{5}} \left(\frac{\sqrt{5}+1}{4}\right)^{n-1} + \frac{\sqrt{5}-2}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{4}\right)^{n-1} - \frac{1}{2^{n-1}} \\ < \frac{\sqrt{5}+2}{\sqrt{5}} \left(\frac{\sqrt{5}+1}{4}\right)^{n-1} < 2.34165 \cdot 0.809017^n \end{aligned} \tag{38}$$

and $\nu_n(W^n) = \nu_n^{\max}$ *iff* W^n *is an alternating sequence.*

Proof In view of Lemmas 5, 1, and 2, it suffices to prove that $\nu_n(W^n) < \nu_n(A^n)$ if W^n is different from an alternating sequence A^n. This is a consequence of Lemma 6, which directly implies that $\nu_n(W^n) \le \nu_n(A^n)$ and that $\nu_n(W^n) = \nu_n(A^n)$ iff W^n is an alternating sequence. □

Theorem 3 is our strongest result as it gives the sharpest upper bound on $P_n^{\max}$. It is interesting to note that the recursions (36) and (37) from the proof of Lemma 6 enable the efficient computation of $\nu_n(W^n)$ with complexity $O(r)$ instead of $O(2^n)$.

If R is the total length of LFSR_X and LFSR_Y, then the expected number of false candidates for the initial states of LFSR_X and LFSR_Y can be upper-bounded by $2^R P_n^{\max}$. The criterion $2^R P_n^{\max} \le 1$ then yields an approximate output segment length required for a successful embedding attack as

$$n \ge 3.27\, R. \tag{39}$$

5 Minimal Probabilities

Although any upper bound on $P_n^{\max}$ is also an upper bound on the minimal ASG-embedding probability $P_n^{\min}$, our main objective here is to obtain sharper upper bounds on $P_n^{\min}$. More precisely, sharp upper and lower bounds on $P_n^{\min}$ are obtained by deriving the minimal interleaving probability $p_n^{\min}$ and by applying Lemma 1.

Lemma 7.

$$p_n^{\min} = \left(\frac{n}{2} + 1\right) \frac{1}{2^n} \tag{40}$$

and $p_n(W^n) = p_n^{\min}$ *if* W^n *is a constant sequence,* 0^n *or* 1^n.

Proof Consider a set of sequences $S_i = W^{i-1}\overline{w}_i$, $1 \le i \le n+1$, where it is assumed that $S_1 = \overline{w}_1$ and $S_{n+1} = W^n$. Now, if S_i is a prefix of U^n and W_i^n is a prefix of V^n and the remaining bits of (U^n, V^n) are arbitrary (for $i = n+1$, V^n is arbitrary), then all such (U^n, V^n) are different and clearly $\mathrm{INT}(U^n, V^n, C^n) = W^n$ for $C^n = 1^{i-1}0^{n-i+1}$. The number of such (U^n, V^n) is $\sum_{i=1}^{n} 2^{n-1} + 2^n = n2^{n-1} + 2^n$, so that $p_n^{\min} \ge n2^{-n-1} + 2^{-n}$. Further, if W^n is a constant sequence, then there are no other (U^n, V^n) that can form W^n by interleaving, because the prefixes of U^n and V^n forming W^n are then also constant sequences. Therefore, the equality is achieved and (40) thus follows. □

The following bounds are a direct consequence of Lemmas 7 and 1. Since $\nu_n^{\min} = (n+1)2^{-n}$, where $\nu_n(W^n) = \nu_n^{\min}$ iff W^n is a constant sequence, the upper bound in (41) is sharper than the upper bound, $\nu_{n-1}^{\min}$, which is a direct consequence of Lemmas 1 and 2.

Theorem 4.

$$\frac{3}{4}(n+1)\frac{1}{2^n} = \frac{3}{4}p_{n-1}^{\min} \le P_n^{\min} \le p_{n-1}^{\min} = (n+1)\frac{1}{2^n}. \tag{41}$$

6 Conclusions

The problem [4] of analyzing the ASG-embedding probability, which is conditioned on the given segment of the ASG output sequence, is theoretically solved. Exponentially small upper bounds on the maximal ASG-embedding probability, including the tightest upper bound in terms of the decimation probability, as well as sharp upper and lower bounds on the minimal ASG-embedding probability are derived. Apart from their wider theoretical merits for the analysis of interleaving and decimation operations, the results prove that the embedding attack [4] on the ASG is successful if the length of the given output segment is sufficiently large and that this length depends on the total length of the two underlying LFSR's only linearly.

References

1. J. Dj. Golić and M. Mihaljević, "A generalized correlation attack on a class of stream ciphers based on the Levenshtein distance," *Journal of Cryptology*, vol. 3(3), pp. 201-212, 1991.
2. J. Dj. Golić and S. Petrović, "A generalized correlation attack with a probabilistic constrained edit distance," in Advances in Cryptology - EUROCRYPT '92, *Lecture Notes in Computer Science*, vol. 658, pp. 472-476, 1993.
3. J. Dj. Golić and L. O'Connor, "Embedding and probabilistic correlation attacks on clock-controlled shift registers," in Advances in Cryptology - EUROCRYPT '94, *Lecture Notes in Computer Science*, vol. 950, pp. 230-243, 1995.
4. J. Dj. Golić and R. Menicocci, "Edit distance correlation attack on the alternating step generator," in Advances in Cryptology - CRYPTO '97, *Lecture Notes in Computer Science*, vol. 1294, pp. 499-512, 1997.
5. J. Dj. Golić and R. Menicocci, "Edit probability correlation attack on the alternating step generator," in 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.
6. C. G. Günther, "Alternating step generators controlled by de Bruijn sequences," in Advances in Cryptology - EUROCRYPT '87, *Lecture Notes in Computer Science*, vol. 304, pp. 5-14, 1988.
7. S. Jiang and G. Gong, "On edit distance attack to alternating step generator," unpublished manuscript.
8. T. Johansson, "Reduced complexity correlation attacks on two clock-controlled generators," in Advances in Cryptology - ASIACRYPT '98, *Lecture Notes in Computer Science*, vol. 1514, pp. 342-357, 1998.
9. M. Živković, "An algorithm for the initial state reconstruction of the clock-controlled shift register," *IEEE Trans. Information Theory*, vol. 37, pp. 1488-1490, Sept. 1991.