Linear Cryptanalysis of Stream Ciphers

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. Starting from recent results on a linear statistical weakness of keystream generators and on linear correlation properties of combiners with memory, linear cryptanalysis of stream ciphers based on the linear sequential circuit approximation of finite-state machines is introduced as a general method for assessing the strength of stream ciphers. The statistical weakness can be used to reduce the uncertainty of unknown plaintext and also to reconstruct the unknown structure of a keystream generator, regardless of the initial state. The linear correlations in arbitrary keystream generators can be used for divide and conquer correlation attacks on the initial state based on known plaintext or ciphertext only. Linear cryptanalysis of irregularly clocked shift registers as well as of arbitrary shift register based binary keystream generators proves to be feasible. In particular, the direct stream cipher mode of block ciphers, the basic summation generator, the shrinking generator, the clock-controlled cascade generator, and the modified linear congruential generators are analyzed. It generally appears that simple shift register based keystream generators are potentially vulnerable to linear cryptanalysis. A proposal of a novel, simple and secure keystream generator is also presented.

1 Introduction

Keystream generators for additive stream cipher applications can generally be realized as autonomous finite-state machines whose initial state and possibly structure as well depend on a secret key. Their practical security can be defined as immunity to various types of divide and conquer attacks on secret key based on known plaintext or ciphertext only, for a survey see [29] and [12]. Apart from that, keystream pseudorandom sequences should have large period, high linear complexity [18], [28], and should satisfy the standard key-independent statistical tests, which should prevent the reconstruction of statistically redundant plaintext from known ciphertext.

* 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.

A general structure-dependent and initial-state-independent linear statistical weakness of arbitrary binary keystream generators is pointed out and analyzed in [13]. It is based on the local properties of the keystream sequence on blocks of consecutive bits whose size is larger than the memory size, and is measured in terms of an appropriate correlation coefficient. The linear weakness can be regarded as a generalization of the linear complexity which is different from the concept of the linear complexity stability introduced in [7]. An effective method for detecting the weakness based on the *linear sequential circuit approximation* (LSCA) [11] of autonomous finite-state machines is presented in [13]. If the structure is key-independent, then the corresponding statistical test can be used for the reconstruction of a statistically redundant plaintext from known ciphertext. If the structure is key-dependent, then the same test can also be used to determine the corresponding unknown key, which presents a specific divide and conquer attack. The linear statistical weakness and the corresponding linear models and correlation coefficients are described in Section 2, including some unpublished results from [14]. Section 3 contains the basic lines of the LSCA method for arbitrary binary keystream generators. Besides its potential to determine linear statistical weaknesses in the keystream sequence, it is shown that the LSCA method can also be used to find out linear correlations between the keystream sequence and appropriate sequences depending on individual initial state variables. The linear correlations can then be used for divide and conquer attacks on the initial state of keystream generators, see [30], [21]. Linear cryptanalysis of additive stream ciphers thus essentially reduces to the LSCA method.

Section 4 is devoted to linear cryptanalysis of clock-controlled shift registers and arbitrary shift register based keystream generators. It turns out that clock-controlled shift registers possess a detectable linear statistical weakness [13], but are immune to linear correlation attacks resulting from individual linear models. Note that simultaneous use of many different linear models may open new possibilities for correlation attacks. Regularly clocked shift registers are potentially vulnerable to correlation attacks, especially if the feedback is linear and if they are autonomous. Keystream generators based on a small number of shift registers appear to be vulnerable to linear cryptanalysis.

Linear cryptanalysis is then applied to concrete additive stream ciphers including the direct stream cipher mode of block ciphers, the basic summation generator [19, 29], the shrinking generator [6, 25], the clock-controlled cascade generator [3, 17, 16, 4], and the modified linear congruential generators [4] and [2], see Section 5. In Section 6, a proposal of a novel, simple and secure keystream generator is presented, which incorporates the principles of linear congruential generators, clock-controlled generators, and combiners with memory.

2 Linear Models for Stream Ciphers

A binary autonomous finite-state machine or sequential circuit is defined by

$$S_{t+1} = \mathcal{F}(S_t), \quad t \geq 0 \qquad (1)$$

$$y_t = f(S_t), \quad t \geq 0 \tag{2}$$

where $\mathcal{F} : \mathrm{GF}(2)^M \rightarrow \mathrm{GF}(2)^M$ is a next-state vector Boolean function, $f : \mathrm{GF}(2)^M \rightarrow \mathrm{GF}(2)$ is an output Boolean function, $S_t = (s_{1t}, \ldots, s_{Mt})$ is the state vector at time t, M is the number of memory bits, y_t is the output bit at time t, and $S_0 = (s_{1,0}, \ldots, s_{M,0})$ is the initial state. A binary keystream generator can be defined as a binary autonomous finite-state machine whose initial state and the next-state and output functions are controlled by a secret key.

Given $\mathcal{F}$ and f, each output bit is a Boolean function of the initial state variables, that is, $y_t = f(\mathcal{F}^t(S_0))$ where $\mathcal{F}^t$ denotes the t-fold self-composition of $\mathcal{F}$ and $\mathcal{F}^0$ is the identity function, $t \geq 0$. If S_0 is assumed to be a uniformly distributed random variable, then the output bits become binary random variables. A basic design criterion for f and $\mathcal{F}$, related to the statistics of the output sequence, is that each bit y_t should be a balanced function of S_0. This is clearly satisfied if both $\mathcal{F}$ and f are balanced. However, the vector of $M+1$ consecutive output bits $(y_t, \ldots, y_{t-M})$ can not be a balanced function of S_0 for any $t \geq M$, since S_0 has dimension only M. Therefore, there must exist a linear function $L(y_t, \ldots, y_{t-M})$ that is a nonbalanced function of S_0 for each $t \geq M$. When the next-state function is balanced, it follows that the state vector S_t is a balanced random variable at any time $t \geq 0$, provided that S_0 is balanced. The probability distribution of the linear function $L(y_t, \ldots, y_{t-M})$, treated as a function of S_{t-M}, is then the same for each $t \geq M$ and there exists such a linear function that effectively depends on y_t. The probability distribution can be expressed in terms of the correlation coefficient to the constant zero function. This essentially means that an autonomous finite-state machine can equivalently be represented as a non-autonomous linear feedback shift register of length at most M with an additive input sequence of nonbalanced identically distributed binary random variables, that is, by a linear model

$$y_t = \sum_{i=1}^{M} a_i \, y_{t-i} + e_t, \ . \, t \geq M. \tag{3}$$

The variables are not independent. The linear function L specified by the feedback polynomial applied to the output sequence $\{y_t\}$ produces a nonbalanced sequence $\{e_t\}$. The standard chi-square frequency statistical test can then be applied to $\{e_t\}$. To distinguish this sequence from the purely random binary sequence with error probability less than about 10^{-3}, the length of the observed keystream sequence should not be larger than $10/c^2$, see [30], [20], for example. Note that $c = 1 - 2\Pr\{e_t = 1\}$. For each additional bit of uncertainty to be resolved, one needs to know an additional segment of the keystream sequence of the same length. Since the linear function L is not unique in general, the maximum effect will be achieved when the linear function with the correlation coefficient of maximum magnitude is used. If this value is smaller than $2^{-M/2}$, then the keystream generator is not vulnerable to this statistical test. However, for large M, which is often the case in practice, it appears very difficult, if not impossible, to determine the value of the maximum correlation coefficient.

Another fact is even more discouraging from the cryptographer's viewpoint. Namely, one could also consider linear functions of more than $M+1$ consecutive output bits. In particular, since every linear function can be defined as a polynomial in the generating function domain, it follows that one should consider all the polynomial multiples of the polynomials corresponding to linear functions of at most $M+1$ variables. It appears very difficult to control all the corresponding correlation coefficients. Apart from that, one may simultaneously use all the obtained linear models and thus significantly reduce the required length of the observed keystream sequence.

The following result [14] determines the total correlation between the output sequence and the all zero sequence for autonomous finite-state machines. Let the next-state function of a binary autonomous finite-state machine with M memory bits be balanced. Then for any $m \geq 1$, the sum $C(m)$ of the squares of the correlation coefficients between all nonzero linear functions of m successive output bits y_t^m and the constant zero function is the same for every $t \geq m-1$ and satisfies $\underline{C}(m) \leq C(m) \leq \bar{C}(m)$, where $\bar{C}(m) = 2^m - 1, m \geq 1$, and

$$\underline{C}(m) = \begin{cases} 0, & 1 \leq m \leq M \\ 2^{m-M} - 1, & m \geq M+1 \end{cases} . \tag{4}$$

The minimum value $\underline{C}(m)$ is achieved for all $m \geq 1$ if and only if any M consecutive output bits constitute a balanced function of the initial state variables. The maximum value $\bar{C}(m)$ is achieved for any $m \geq 1$ if and only if the output function is constant. For any m, the total correlation is distributed among $2^m - 1$ output linear functions. It then follows that for each $m > M$ the maximum absolute value of the correlation coefficients can not be smaller than approximately $2^{-M/2}$ which corresponds to the uniform distribution of correlation, provided that the minimum total correlation condition is satisfied. For the condition to hold it is necessary that the output function is balanced. Large memory size is therefore an important design criterion. It is of course clear that minimum total correlation does not guarantee the uniform distribution, as is demonstrated by a linear feedback shift register.

3 Linear Sequential Circuit Approximation

In order to find all the nonbalanced linear functions of at most $M+1$ consecutive output bits whose existence is established in the previous section, one should determine the correlation coefficients for 2^M Boolean functions of M variables. Exhaustive search method has $O(2^{2M})$ computational complexity, which is not practically possible for large M.

Taking the linear sequential circuit approximation (LSCA) approach introduced in [11] for combiners with memory, we propose a LSCA method for autonomous finite-state machines which is a feasible procedure that with high probability yields nonbalanced linear functions of at most $M+1$ consecutive output bits with comparatively large correlation coefficients. The LSCA method consists of two stages. First, find a linear approximation of the output function f

and each of the component functions of the next-state function $\mathcal{F}$. This enables one to express each of these $M+1$ functions as the sum of a linear function and a nonbalanced function, whose correlation coefficient is different from zero. In practice, both the output function and the component next-state functions effectively depend on small subsets of the state variables or can be expressed as compositions of such functions. Therefore, the computational complexity of obtaining all the linear approximations along with the corresponding nonzero correlation coefficients is considerably smaller than $O((M+1)M2^M)$, which is required by the direct application of the Walsh transform technique, see [29]. Finding good linear approximations of Boolean functions in real ciphers thus appears to be a feasible task.

Second, by virtue of the obtained linear approximations, put the basic equations (1) and (2) into the form

$$S_{t+1} = \mathbf{A}S_t + \Delta(S_t), \quad t \geq 0 \tag{5}$$

$$y_t = \mathbf{B}S_t + \varepsilon(S_t), \quad t \geq 0 \tag{6}$$

where the vectors are regarded as one-column matrices, $\mathbf{A}$ and $\mathbf{B}$ are binary matrices, and ε and all the components of $\Delta = (\delta_1, \ldots, \delta_M)$ are nonbalanced Boolean functions, called the noise functions. The main point now is to treat $\{\varepsilon(S_t)\}$ and $\{\delta_j(S_t)\}$, $1 \leq j \leq M$, as the input sequences so that (5) and (6) define a non-autonomous linear sequential circuit (LSC), see [11]. Then solve the LSC using the generating function technique and thus obtain

$$\mathbf{y} = \frac{1}{\varphi(z)}\sum_{j=1}^{M} g_j(z)\, s_{j0} + \frac{1}{\varphi(z)}\sum_{j=1}^{M} z\, g_j(z)\, \boldsymbol{\delta}_j + \boldsymbol{\varepsilon} \tag{7}$$

where $\mathbf{y}$, $\boldsymbol{\delta}_j$, and $\boldsymbol{\varepsilon}$ respectively denote the generating functions in variable z of the sequences $\{y_t\}$, $\{\delta_j(S_t)\}$, and $\{\varepsilon(S_t)\}$, and the polynomial $\varphi(z) = \sum_{i=0}^{M} \varphi_i z^i$, $\varphi_0 = 1$, is the reciprocal of the characteristic polynomial of the state-transition matrix $\mathbf{A}$. As a consequence of (7), we also get

$$\sum_{i=0}^{M} \varphi_i\, y_{t-i} = \sum_{i=0}^{M} \varphi_i\, \varepsilon(S_{t-i}) + \sum_{j=1}^{M}\sum_{i=0}^{M-1} g_{ji}\, \delta_j(S_{t-1-i}) \overset{\text{def}}{=} e_t(S_0), \quad t \geq M. \tag{8}$$

The computational complexity to obtain (7) and (8) is only $O(M^3)$. For each $t \geq M$, the noise function e_t is a sum of individual noise functions that are nonbalanced if $S_{t-i}, 0 \leq i \leq M$, are balanced. If one assumes that the next-state function $\mathcal{F}$ is balanced, then it follows that each of the individual noise functions in (8) is nonbalanced and identically distributed for every $t \geq M$, meaning that the corresponding correlation coefficients are nonzero and independent of t. In general, for random $\mathcal{F}$ and f, one should expect that the individual noise functions remain nonbalanced even if $S_{t-i}, 0 \leq i \leq M$, are not balanced functions of S_0 and that the resulting noise function e_t is also nonbalanced, for almost all $t \geq M$. This conclusion is justified by the following probabilistic result, see [13], which is also relevant for the linear cryptanalysis of block ciphers.

Lemma 1. Consider m Boolean functions of the same n variables with the correlation coefficients c_i to the constant zero function, $1 \leq i \leq n$. If the functions are chosen uniformly and independently at random, then for large 2^n the probability distribution of the correlation coefficient of their modulo two sum is asymptotically normal with the expected value $\prod_{i=1}^{m} c_i$ and the variance $O(\frac{m}{2^n})$. ◇

The described LSCA method is based on the linear approximations of the component next-state functions, which is a limitation. However, the method can be generalized to deal with the linear approximations of the linear combinations of the component next-state functions. To this end, let $\mathcal{L}$ denote an arbitrary balanced, that is, one-to-one linear function $\mathrm{GF}(2)^M \rightarrow \mathrm{GF}(2)^M$ and let $S_t' = \mathcal{L}(S_t)$ denote the transformed state vector at time t. Accordingly, one may put the equations (1) and (2) into an equivalent form $S_{t+1}' = \mathcal{L}\mathcal{F}\mathcal{L}^{-1}(S_t'), t \geq 0$, and $y_t = f\mathcal{L}^{-1}(S_t'), t \geq 0$, and then proceed with the basic LSCA method, as already defined. The linearization of the component functions of the transformed next-state function $\mathcal{L}\mathcal{F}\mathcal{L}^{-1}$ is essentially the same as the linearization of the linear combinations of the component functions of the original next-state function F.

Starting from (7), one may also develop divide and conquer correlation attacks on the individual bits of the initial state. The transfer function with respect to s_{j0} is given by $g_j'(z)/\varphi_j(z) = g_j(z)/\varphi(z)$ where $g_j'(z)$ and $\varphi_j(z)$ are relatively prime, $1 \leq j \leq M$. The denominator polynomials $\varphi_j(z)$ induce an equivalence relation among s_{j0} or just $j, 1 \leq j \leq M$. Let $\bar{\varphi}_j(z)$ denote the least common multiple of $\varphi_k(z)$ for all k not belonging to the equivalence class of j. Then for a single equation (7), the initial correlation attack on s_{j0} is possible if and only if the j-th component of the next-state function is linear ($\delta_j = 0$) and $\varphi_j(z) \nmid \bar{\varphi}_j(z)$. The degree of $\varphi_j(z)/(\varphi_j(z), \bar{\varphi}_j(z))$ determines the number of bits of uncertainty resolved by the attack on the equivalence class of s_{j0}. By subtracting the effect of the initially determined bits of the initial state from the right-hand side of (7), possibly including the noise functions as well, one then recomputes the output sequence and repeats the procedure iteratively. It follows that regularly clocked linear feedback shift registers are potentially vulnerable to correlation attacks. Note that for correlation attacks it may not be necessary to linearize the whole generator. Furthermore, if one simultaneously uses several linear sequential circuit approximations, then other possibilities for correlation attacks may exist as well.

It is desirable for the LSCA method to find a linear function/model with the maximum absolute value of the correlation coefficient. To this end, the number of noise terms in (8) should be small and their correlation coefficients should be large in magnitude. A reasonable approach is to repeat the procedure several times starting from the best linear approximations of the output and next-state functions. In fact, one should tend to find an optimum invertible set of the linear combinations of the component next-state functions that yields the maximum absolute value of the overall correlation coefficient. The power of the chi-square statistical test can be considerably improved by running the test on all the obtained linear models, rather than on a single one. In order to achieve the immunity to the LSCA attack, it follows that the memory size should be large and it appears recommendable that the output function and the linear combinations of the component next-state functions should have large distance from affine functions as well as that the component next-state functions should effectively depend on large subsets of the state variables.

4 Shift Register Based Keystream Generators

In this section, the LSCA method is applied to an arbitrary binary keystream generator consisting of regularly or irregularly clocked shift registers (SRs) combined by a function with or without memory. Clock-control sequences are produced either within the generator or by separate generators. One should first form the linear models for individual SRs: regularly clocked linear feedback SRs stay as they are, linear models for regularly clocked nonlinear feedback SRs are made by linearizing the feedback functions, and linear models for irregularly clocked SRs are formed as follows.

A clock-controlled shift register is a keystream generator consisting of a linear or nonlinear feedback shift register that is irregularly clocked according to an integer decimation sequence, which defines the number of clocks per output symbol and which is itself produced by a pseudorandom sequence generator, see [16] and [9]. More precisely, if $X = \{x_t\}_{t=0}^{\infty}$ denotes a regularly clocked shift register 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 $y_t = x(\sum_{i=0}^{t} d_i)$, $t \geq 0$. First observe that a nonlinear feedback can in principle be treated in the same way as linear, except for the additive noise function. Second, assume a realistic probabilistic model for the decimation sequence, for example, assume that D is a sequence of identically distributed 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. When $\mathcal{D}$ contains positive integers only, one can also define the deletion rate p_d as $1 - \frac{1}{\bar{d}}$, $\bar{d} = \sum_{d\in\mathcal{D}} dP(d)$.

We will distinguish between the two cases: the case with possible repetitions ($0 \in \mathcal{D}$) and the case without repetitions ($0 \notin \mathcal{D}$). In the first case, it is clear that regardless of the feedback $y_t + y_{t-1} = e_t, t \geq 1$, where the correlation coefficient of e_t is equal to $P(0)$. For the stop-and-go registers [3], for which $\mathcal{D} = \{0, 1\}$, $P(0) = 1/2$. In the second case, consider a clock-controlled linear feedback shift register of length r with the feedback polynomial $f(z) = 1 + \sum_{k=1}^{w} z^{i_k}$, $1 \leq i_1 < \ldots < i_w = r$, where $W = w + 1$ is the weight of $f(z)$. By using the LSCA method or directly, one can obtain [13] a linear model of the form

$$y_t + \sum_{k=1}^{w} y_{t-\hat{i}_k} = e_t, \quad t \geq r' \tag{9}$$

where the polynomial $\hat{f}(z) = 1 + \sum_{i=1}^{r'} \hat{f}_i z^i = 1 + \sum_{k=1}^{w} z^{\hat{i}_k}$, $1 \leq \hat{i}_1 < \ldots < \hat{i}_w = r'$, satisfies $\hat{i}_k - \hat{i}_{k-1} \leq i_k - i_{k-1}$, $1 \leq k \leq w$, where $\hat{i}_0 = i_0 = 0$. Call $\hat{f}(z)$ a ***shrunk*** polynomial of $f(z)$. A shrunk polynomial of $f(z)$ is not unique but has the same weight as $f(z)$. It is possible to obtain an expression for the correlation coefficient of e_t for an arbitrary probability distribution $\mathcal{P} = \{P(d)\}_{d\in\mathcal{D}}$. For simplicity, we give only the expression for the geometric distribution $P(d) = p^{d-1}(1-p), d \geq 1$, which corresponds to the case of independent deletions with probability p. Note that an arbitrary $\mathcal{P}$ can be approximated by this distribution by setting $p = p_d$ where p_d is the deletion rate. It follows that the correlation coefficient in this case is given by

$$c = p^{r-r'}(1-p)^{r'+1} \prod_{k=1}^{w} \binom{\Delta_k}{\hat{\Delta}_k} \tag{10}$$

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 (10) has a clear combinatorial meaning in terms of the probability of decimation sequences. Namely, 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. The coefficient is maximized if $\hat{\Delta}_k = \lfloor(1-p)(\Delta_k + 1)\rfloor$, $1 \leq k \leq w$. It follows that $c = 1$ if $p = 0$ and $c = 0$ if $p = 1$, which is natural. Suppose that r/w, Δ_k, $p\Delta_k$, and $(1-p)\Delta_k$ are all large. Then Stirling's approximation gives

$$c \sim (1-p)\left(\frac{2\pi p}{1-p}\right)^{-\frac{w}{2}} \left(\prod_{k=1}^{w} \Delta_k\right)^{-\frac{1}{2}} \tag{11}$$

which shows that the magnitude of c may be considerably larger than $2^{-r/2}$ let alone $2^{-M/2}$, M being the memory size of the whole generator. The smallest magnitude of c is obtained when the feedback taps are approximately equidistant. In this case the necessary length of the keystream sequence needed to detect the weakness is $(10/(1-p)^2)(2\pi p/(1-p))^w((r-w)/w)^w$. Given w, the larger the values of r and p the smaller the correlation coefficient, respectively. Given r and p, there exists an optimal value of w that minimizes the correlation coefficient. For a given feedback polynomial, one may use different shrunk polynomials or their polynomial multiples and thus obtain different linear models. This reduces the required length of the observed keystream sequence considerably. In addition to that, one may also use different polynomial multiples of the feedback polynomial, especially the ones with low weight, possibly much lower than for the original polynomial. So, the weakness is in general easily detected if the feedback polynomial is known.

By using the LSCA method from Section 3 one then develops a linear model for the function with memory, which is treated as a non-autonomous finite-state machine with purely random input sequences, see [11]. Consequently, one obtains a linear equation of the form (8) whose right-hand side contains the additional input sequences as well. The linear function on the left-hand side of (8) corresponds to the reciprocal $\varphi(z)$ of the characteristic polynomial of the state-transition matrix. Finally, one substitutes the outputs of the linearized SRs for the inputs to the linear sequential circuit and thus derives a linear model with the feedback polynomial being equal to the least common multiple of $\varphi(z)$ and the feedback polynomials of all the linearized SRs. For estimating the overall correlation coefficient, a reasonable assumption is that the noise sequences from the linearized SRs are mutually independent and independent from the noise sequences in the linear sequential circuit, unless the SRs are connected in a very special way. Various linear models are obtained by varying the linear models for irregularly clocked and nonlinear feedback SRs and the linear model for the function with memory. Polynomial multiples may also be used especially if they reduce the magnitude of the overall correlation coefficient. In fact, it is the polynomial multiples that make it very difficult to achieve the security against the LSCA attack.

5 Concrete Keystream Generators

We now apply the linear cryptanalysis to several types of the shift register based binary keystream generators.

Direct stream cipher mode for block ciphers

The Direct Stream Cipher (DSC) mode of operation of block ciphers can be defined as a particular case of the Output Feedback (OFB) mode in which only a single output bit is used to produce the keystream sequence and the initial state is key-controlled. The keystream bit can also be generated by a simple output function of several state bits, for example, by a modulo 2 addition. Since in this case a known plaintext does not provide pairs of block cipher inputs and outputs, the available cryptanalytic techniques for block ciphers [1] and [20] are not directly applicable. Note that some possibilities for differential and/or linear cryptanalysis of block ciphers in the CBC, CFB, or regular OFB mode have been explored in [26] and [27], assuming a partial knowledge of the ciphertext. The DSC mode of a block cipher is nothing but a stream cipher whose next-state function is defined by the block cipher, where the initial state is key-controlled. Therefore, the LSCA method for linear cryptanalysis of stream ciphers can be used. It may yield a linear statistical weakness of the keystream sequence and may be a basis for divide and conquer attacks on the secret key. The starting point of the method is to find an invertible set of the linear functions of the block cipher output with relatively large correlation coefficients to linear functions of the input. The next point is to solve the corresponding linear sequential circuit by the generating function technique. To minimize the resulting correlation coefficient, the procedure is repeated several times using different invertible linear approximations of the block cipher. The method is computationally feasible for real ciphers. As a consequence, one can also obtain novel characteristics of block ciphers such as the characteristic polynomials and the corresponding correlation coefficients. Linear cryptanalysis of concrete block ciphers in the DSC mode is not an easy task and is out of the scope of this paper. It would be interesting to investigate whether the immunity of a block cipher to the linear cryptanalysis [20] in the ECB mode implies the immunity to the linear cryptanalysis in the DSC mode.

Basic summation generator [19, 29]

The basic summation generator is a combiner with one bit of memory and two regularly clocked linear feedback SRs. Its output function is already linear, whereas its next-state function can be linearized in several ways with large correlation coefficients. A good way is to take a linear function depending on a single input with the correlation coefficient $1/2$. The feedback polynomial of the corresponding linear model is just the least common multiple of the feedback polynomials of the two input SRs, or any of its polynomial multiples. The overall correlation coefficient is $(1/2)^W$ where W is the weight of this polynomial. In addition, as was already noted in [22], the output is correlated to the sum of one input and the linear transform $1+z$ of the other with the correlation coefficient $1/2$ which is highly vulnerable to fast correlation attacks, see [21] and [5], for example. Linear cryptanalysis of a general summation generator consisting of more than two SRs, as is suggested in [19], remains an open problem.

Shrinking generator [6, 23], [25]

This is a single irregularly clocked linear feedback SR whose clock is controlled by another linear feedback SR in a manner that corresponds to independent deletions with probability $p = 1/2$. The principle has first appeared in [25], but is implicit in [10] as well. The two SRs may even be the same as was suggested in [23]. Section 4 then gives a linear model with the feedback polynomial equal to a shrunk polynomial of the feedback polynomial of the irregularly clocked SR or of any of its polynomial multiples. The correlation coefficient is given by (10) or (11) for $p = 1/2$. For a single shrunk polynomial of a polynomial of weight $W = w + 1$ and relatively large degree r, the required length of the keystream sequence to detect the weakness is thus about $40\,(6.28\,r/w)^w$, where the taps are assumed to be approximately equidistant. If for example $w = 4$, then the length is about $243\,r^4$. When the feedback polynomial is key-dependent [6], the weakness may be used to determine the corresponding key from a known keystream sequence or even from ciphertext only. The results [6] of the statistical analysis of the shrinking generator are somewhat misleading for two reasons. First, the statistical properties are not nice on blocks whose length exceeds the length of the clock-controlled SR and, second, the key-dependent feedback polynomial is assumed to be selected uniformly at random.

Clock-controlled cascade generator [3, 17, 16, 4]

This is a cascade of K linear feedback SRs with the same feedback polynomial $f(z)$ of degree r. The first SR is clocked regularly and the others are clocked either k or m times per each output bit. For $k, m > 0$, by using a model for a single irregularly clocked shift register with deletion rate $p = 1 - 2/(k+m)$, one obtains a linear model with the feedback polynomial $\hat{f}(z)f(z)$ and the overall correlation coefficient $c^{(K-1)W}$ where W is the weight of $f(z)$, $\hat{f}(z)$ is a shrunk polynomial of $f(z)$, and c is given by (10) or (11). This is an approximation: the actual c is in fact different because the irregular clocking is constrained rather than independent. For a stop-and-go cascade ($k = 0, m > 0$), instead of $\hat{f}(z)$ one should use $1 + z$ and the correlation coefficient $c = 1/2$. Instead of $f(z)$ one may also take any of its polynomial multiples. For example, if $k = 1$ and $m = 2$, then $p = 1/3$ and $c \sim 0.66\,(3.14\,r/w)^{-w/2}$, so that the required length to detect the weakness is approximately $10\,(2.25\,(3.14\,r/w)^{w})^{(w+1)(K-1)}$. If one takes into account the constrained clocking, then the required length becomes $10\;(2.25\;(4.19\;r/w)^{w})^{(w+1)(K-1)}$. This length can be reduced considerably by using many different shrunk polynomials instead of a single one.

Apart from the described statistical weakness, the linear transform $\hat{f}(z)$ of the output of the cascade is correlated to the same linear transform of the output of the first shift register, with the correlation coefficient c^{K-1}. For $k = 1$ and $m = 2$, the required keystream sequence length for the successful correlation attack on the initial state of the first shift register is then $10\,r\,(2.25\,(4.19\,r/w)^{w})^{K-1}$. This of course implies an exhaustive search through all possible initial states. Fast correlation attacks might also be feasible, see [21] and [5], for example. This is in accordance with the recent statistical analysis of a stop-and-go cascade from [24]. Both the weaknesses diminish as K increases, but the efficiency of the generator remains the same. The choice of small SR length r does not seem to be appropriate, because it might be an open gate for algebraic cryptanalysis. On the other hand, if the SR length and K are both large, then the generator is not efficient.

Modified linear congruential generators [4], [2, 6]

This type of keystream generators is based on linear recursions modulo 2^m for a positive integer m, which may be chosen to be relatively large, such as 32, as was suggested in [2]. Since linear recurring truncated integer sequences are in principle predictable, for example, see [8], various modifications have been suggested. In [4], it is suggested to use a single truncated integer sequence generated by the bitwise sum modulo 2 (which is nonlinear modulo 2^m) of two feedbacks linear modulo 2^m. In [2, 6], it is proposed to use two simple linear recursions with fixed binary coefficients without truncation and the unconstrained clock-control principle [25, 6]. Linear cryptanalysis of modified linear congruential generators should start from a linear approximation of the feedback function, which is nonlinear over the binary field. A linearized scheme is then a set of non-autonomous binary linear feedback shift registers, one for each order of significance [4], with additive inputs whose correlation coefficients can be derived by using the results from [31]. Note that the shift register for the lowest order of significance is autonomous. The shift register lengths are upper-bounded by the order of the linear recursion which is relatively small compared to the memory size. Interestingly enough, it turns out that the correlation coefficient c_i remains biased when the order of significance i increases if the number w of integers to be added is odd. For example, $c_i \to -0.3333$ when $w = 3$ and $c_i \to 0.1333$ when $w = 5$, see

[31]. This is a potential trapdoor. If w is even, then c_i tends to zero like $2^{-iw/2}$. The results are slightly different for a modified recursion [4] with nonbinary coefficients. In any case, it follows that the cryptographic strength of the output binary sequences with respect to linear cryptanalysis strongly depends on the order of significance, which is not good. Possible use of low weight polynomial multiples modulo 2^m to reduce w might also be studied. Apart from that, simple modified linear congruences are potentially vulnerable to linear cryptanalysis modulo 2^m. For example, the modulo 2 sum of two integers, as suggested in [4], is correlated to their sum modulo 2^m, and the corresponding correlation coefficient c_i then behaves like 2^{-i}.

For the generator [2], $w = 2$ and there is no truncation, so that the linear statistical weakness of low order output binary sequences is easily detectable despite the irregular clocking. Note that the output feedforward function can be linearly approximated with large correlation coefficients and hence does not make much of a difference with respect to the linear cryptanalysis over the binary field.

6 Proposal

We now propose a novel scheme which is a self-clock-controlled modified linear congruential generator with a nonlinear feedforward function with memory. The first two parts are very simple to realize in software or hardware, whereas the third part is very simple to implement in hardware. The proposal is given a name GOAL.

First pick at random a primitive binary polynomial f of degree n not smaller than 100 and of weight $W = w + 1$ not smaller than 5. The polynomial should not have 'low' degree trinomial multiples, which is easily checked, and may be controlled by a secret key. The polynomial defines a linear congruence modulo 2^{32} with w nonzero binary coefficients. The initial conditions are controlled by the secret key. The 32-bit integer feedback is circularly shifted so that the least significant bit becomes the most significant one. The modified feedback is split into two 16-bit parts which are bitwise added modulo 2 to form the 16-bit output of the modified linear congruence.

Each of the 16 binary output sequences is transformed by a combiner with 15 bits of memory and a single input and output, respectively. All the 16 combiners have the same next-state function defined as a (16×15)-bit table, while the output function is the sum modulo 2 of the input bit and one of the state bits. The table is generated at random so that the maximum squared correlation coefficient between the input and output linear functions is 'close' to 2^{-16}, which is not difficult. This criterion along with relatively large memory size is in accordance with the results from [15]. The table can be stored on a single 1Mbit chip and may be controlled by the secret key. The 16 15-bit initial state vectors may also be controlled by the key. The individual combiners may be different, but that requires more space.

A constrained (1,2)-clock-control is defined by the sum modulo 2 of all the 16 bits from a previous output of the modified linear congruence that is not used in forming the current feedback. If the control bit is 1, then the output is discarded and the congruence is computed once more and transformed by the combiners with memory to form the current 16-bit integer output. Note that the constrained clocking is cryptographically weaker than unconstrained, but is faster and does not give rise to buffer-control problems. On the average, it takes 3 modified linear congruence computations to generate each 32 output bits in our scheme.

Preliminary analysis of the proposed generator suggests the following conclusions. Algebraic properties of the modified linear congruence, as the period and the linear complexity, and the distribution over a period of blocks of output integers whose size does not exceed n, may in principle be derived. For the generator as a whole, both the period and the linear complexity are almost certainly lower-bounded by 2^{n+5} and very likely by 2^{16n} as well. Furthermore, it may well be the case that they are close to 2^{32n}, see [9]. On the other hand, if one assumes that the modified linear congruence produces a purely random integer sequence, then the self-clock-control and the function with memory, defined as above, ensure that the output sequence is also purely random.

The generator is resistant against the linear cryptanalysis modulo 2, because of the circular shift feedback operation which results in balanced correlation coefficients in a linear model with the feedback polynomial $f(x^{32})$ of large degree and because of the clock-control and the function with memory applied before the clock-control. It is also immune to the linear cryptanalysis modulo 2^{32}, because of the circular shift operation which is nonlinear modulo 2^{32} and the feedforward bitwise addition modulo 2 which reduces 32-bit integers to 16-bit integers and because of the clock-control and the function with memory. Other divide and conquer attacks are very unlikely since the internal state variables of the proposed keystream generator as a binary autonomous finite-state machine are well mixed both in the modified linear congruence and in the clock-control. Finally, by changing the parameters it is easy to increase or decrease the security of the proposed generator. For example, instead of selecting a (16×15)-bit table at random, one may choose a table easy to realize in software.

7 Conclusion

By combining the recent results on a linear statistical weakness of arbitrary keystream generators [13] and on linear correlation properties of combiners with memory [11], a novel general method for assessing the strength of stream ciphers is proposed. The method is based on the linear sequential circuit approximation of finite-state machines and is called the linear cryptanalysis of stream ciphers. It results in a linear statistical weakness of the keystream sequence on blocks of consecutive output bits whose size is larger than the memory size as well as in correlations between feedforward linear transforms of the keystream sequence and linear transforms of the individual initial state variables. The statistical weakness can be used to reduce the uncertainty of unknown plaintext and also to reconstruct the unknown structure of a keystream generator, regardless of the initial state. Linear correlations can be used for divide and conquer attacks on the initial state of keystream generators based on known plaintext or ciphertext only, see [30], [21]. The effectiveness of linear cryptanalysis can be measured in terms of the corresponding correlation coefficients. Linear cryptanalysis of block ciphers [20] proves to be a special case of linear cryptanalysis of stream ciphers.

Linear cryptanalysis of irregularly clocked shift registers as well as of arbitrary binary keystream generators based on regularly or irregularly clocked shift registers, with linear or nonlinear feedback, combined by a function with or without memory is shown to be feasible. It turns out that clock-controlled shift registers possess a detectable linear statistical weakness, but are immune to linear correlation attacks resulting from individual linear models. However, simultaneous use of many different linear models may open new possibilities for correlation attacks. Regularly clocked shift registers are potentially vulnerable to correlation attacks, especially if the feedback is linear and if they are autonomous. In particular, the direct stream cipher mode of block ciphers, the basic summation generator, the shrinking generator, the clock-controlled cascade generator, and the modified linear congruential generators are analyzed. One may generally conclude that simple shift register based keystream generators are potentially vulnerable to linear cryptanalysis, especially if the number of shift registers is relatively small. A proposal of a novel, simple and secure keystream generator based on a modified linear congruential scheme, a self-clock-control principle, and combiners with memory is also presented.

References

1. E. Biham and A. Shamir, "Differential cryptanalysis of DES-like cryptosystems," *Journal of Cryptology*, 4(1):3–72, 1991.
2. U. Blöcher and M. Dichtl, "Fish: a fast software stream cipher," Fast Software Encryption – Cambridge '93, *Lecture Notes of Computer Science*, vol. 809, R. Anderson ed., Springer-Verlag, pp. 41–44, 1994.
3. W. G. Chambers and D. Gollmann, "Lock-in effect in cascades of clock-controlled shift registers," Advances in Cryptology – EUROCRYPT '88, *Lecture Notes in Computer Science*, vol. 330, C. G. Günther ed., Springer-Verlag, pp. 331–342, 1988.
4. W. G. Chambers, "Two stream ciphers," Fast Software Encryption – Cambridge '93, *Lecture Notes of Computer Science*, vol. 809, R. Anderson ed., Springer-Verlag, pp. 51–55, 1994.
5. 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.
6. 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.
7. C. Ding, G. Xiao, and W. Shan, *The Stability Theory of Stream Ciphers. Lecture Notes in Computer Science*, vol. 561, Springer-Verlag, 1991.

8. A. M. Frieze, J. Hastad, R. Kannan, J. C. Lagarias, and A. Shamir, "Reconstructing truncated integer variables satisfying linear congruences," *SIAM J. Comput.*, 17:262–280, 1988.
9. 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.
10. 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.
11. 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.
12. 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.
13. J. Dj. Golić, "Intrinsic statistical weakness of keystream generators," *Pre-proceedings of Asiacrypt '94*, pp. 72–83, Wollongong, Australia, 1994.
14. J. Dj. Golić, "Linear models for keystream generators," to appear in *IEEE Trans. Computers.*
15. J. Dj. Golić, "Correlation properties of a general binary combiner with memory," to appear in *Journal of Cryptology.*
16. D. Gollmann and W. G. Chambers, "Clock controlled shift registers: a review," *IEEE J. Sel. Ar. Commun.*, 7(4):525–533, 1989.
17. D. Gollmann and W. G. Chambers, "A cryptanalysis of step$_{k,m}$-cascades," Advances in Cryptology – EUROCRYPT '89, *Lecture Notes in Computer Science*, vol. 434, J.-J. Quisquater and J. Vandewalle eds., Springer-Verlag, pp. 680–687, 1990.
18. J. L. Massey, "Shift-register synthesis and BCH decoding," *IEEE Trans. Inform. Theory*, 15:122–127, Jan. 1969.
19. J. L. Massey and R. A. Rueppel, "Method of, and apparatus for, transforming a digital sequence into an encoded form" U. S. Patent No. 4,797,922, 1989.
20. M. Matsui, "Linear cryptanalysis method for DES cipher," Advances in Cryptology – EUROCRYPT '93, *Lecture Notes in Computer Science*, vol. 765, T. Helleseth ed., Springer-Verlag, pp. 386–387, 1994.
21. W. Meier and O. Staffelbach, "Fast correlation attacks on certain stream ciphers," *Journal of Cryptology*, 1(3):159–176, 1989.
22. W. Meier and O. Staffelbach, "Correlation properties of combiners with memory in stream ciphers," *Journal of Cryptology*, 5(1):67–86, 1992.
23. W. Meier and O. Staffelbach, "The self-shrinking generator," *Pre-proceedings of Eurocrypt '94*, Perugia, Italy, pp. 201–210, 1994.
24. R. Menicocci, "Short Gollmann cascade generators may be insecure," CODES AND CYPHERS, Cryptography and Coding IV, P. G. Farrell ed., The Institute of Mathematics and its Applications, pp. 281–297, 1995.
25. M. J. Mihaljević, "An approach to the initial state reconstruction of a clock-controlled shift register based on a novel distance measure," Advances in Cryptology – AUSCRYPT '92, *Lecture Notes in Computer Science*, vol. 718, J. Seberry and Y. Zheng eds., Springer-Verlag, pp. 349–356, 1993.
26. K. Ohta and M. Matsui, "Differential attack on message authentication codes," Advances in Cryptology – CRYPTO '93, *Lecture Notes in Computer Science*, vol. 773, D. R. Stinson ed., Springer-Verlag, pp. 200–211, 1994.

27. B. Preneel, M. Nuttin, V. Rijmen, and J. Buelens, "Cryptanalysis of the CFB mode of the DES with a reduced number of rounds," Advances in Cryptology – CRYPTO '93, *Lecture Notes in Computer Science*, vol. 773, D. R. Stinson ed., Springer-Verlag, pp. 212–223, 1994.
28. R. A. Rueppel, *Analysis and Design of Stream Ciphers*. Berlin: Springer-Verlag, 1986.
29. R. A. Rueppel, "Stream ciphers," in *Contemporary Cryptology: The Science of Information Integrity*, G. Simmons ed., pp. 65–134. New York: IEEE Press, 1991.
30. T. Siegenthaler, "Decrypting a class of stream ciphers using ciphertext only," *IEEE Trans. Comput.*, 34:81–85, Jan. 1985.
31. O. Staffelbach and W. Meier, "Cryptographic significance of the carry for ciphers based on integer addition," Advances in Cryptology – CRYPTO '90, *Lecture Notes in Computer Science*, vol. 537, A. J. Menezes and S. A. Vanstone eds., Springer-Verlag, pp. 601–614, 1991.