Linear Statistical Weakness of Alleged RC4 Keystream Generator

Jovan Dj. Golić *

School of Electrical Engineering, University of Belgrade
Bulevar Revolucije 73, 11001 Beograd, Yugoslavia

Abstract. A keystream generator known as RC4 is analyzed by the linear model approach. It is shown that the second binary derivative of the least significant bit output sequence is correlated to 1 with the correlation coefficient close to $15 \cdot 2^{-3n}$ where n is the variable word size of RC4. The output sequence length required for the linear statistical weakness detection may be realistic in high speed applications if $n \leq 8$. The result can be used to distinguish RC4 from other keystream generators and to determine the unknown parameter n, as well as for the plaintext uncertainty reduction if n is small.

1 Introduction

Any keystream generator for practical stream cipher applications can generally be represented as an autonomous finite-state machine whose initial state and possibly the next-state and output functions as well are secret key dependent. A common type of keystream generators consists of a number of possibly irregularly clocked linear feedback shift registers (LFSRs) that are combined by a function with or without memory. Standard cryptographic criteria such as a large period, a high linear complexity, and good statistical properties are thus relatively easily satisfied, see [16], [17], but such a generator may in principle be vulnerable to various divide-and-conquer attacks in the known plaintext (or ciphertext-only) scenario, where the objective is to reconstruct the secret key controlled LFSR initial states from the known keystream sequence, for a survey see [17] and [6]. Most the attacks require an exhaustive search over the initial states of a subset of the LFSRs, with the exception of a small number of faster cryptanalytic attacks which may work for long LFSRs as well, such as fast correlation attacks [13] based on iterative probabilistic decoding, the conditional correlation attack [14] based on information set decoding, and the inversion attack [10], all on regularly clocked LFSRs, and a specific fast correlation attack on irregularly clocked LFSRs whose theoretical framework is developed in [8]. In practice, the initial state is for resynchronization purposes also made dependent on a

* This work was done while the author was with the Information Security Research Centre, Queensland University of Technology, Brisbane, Australia. This research was supported in part by the Science Fund of Serbia, grant #04M02, through the Mathematical Institute, Serbian Academy of Science and Arts.

W. Fumy (Ed.): Advances in Cryptology - EUROCRYPT '97, LNCS 1233, pp. 226-238, 1997.

randomizing key, which is typically sent in the clear before every new message to be encrypted. This may open new possibilities for cryptanalytic attacks, see [2].

In the open literature, there is a very small number of proposed keystream generators that are not based on shift registers. For example, an interesting design approach, which may have originated from the table-shuffling principle [12], is to use a relatively big table that slowly varies in time under the control of itself. A keystream generator [15] publicized in [18] and known as RC4 (although a public confirmation is still missing) is such an example, which is according to [18] widely used in many commercial products, including Lotus Notes, Apple Computer's AOCE, Oracle Secure SQL, and the Cellular Digital Packet Data specification [1]. Another, somewhat similar example is a keystream generator called ISAAC [11]. Of course, one may also use a set of tables controlling each other, but this may lead to some divide-and-conquer attacks. The resulting schemes are hardly analyzable, and about the only known theoretical argument [4] concerns the period of the internal state sequence, but has probabilistic rather than deterministic nature. Namely, if the internal memory size is M and if the next-state function is randomly chosen according to the uniform distribution, then the average cycle and tail lengths are both around $2^{M/2}$, whereas if the next-state function is in addition required to be invertible, then the internal state period (cycle length) is uniformly distributed between 1 and 2^M, with the average value 2^{M-1}.

The statistical properties of the keystream sequence are typically measured by standard statistical tests, and for some sequences, including the LFSR ones, theoretical results can be derived as well. For keystream generators like RC4 such theoretical results are difficult to establish. The results typically deal with the relative frequency of occurrence of blocks of successive symbols within a period, where the block size is assumed to be smaller than the internal memory size. However, it is shown in [7], [9] that for block sizes bigger than M, a linear statistical weakness or a so-called linear model always exists and can be efficiently determined by the linear sequential circuit approximation (LSCA) method [5]. The linear statistical weakness is a linear relation among the keystream bits that holds with probability different from one half. It turns out [9] that for many practical schemes, including the clock-controlled LFSRs, the keystream sequence length needed to detect the weakness is considerably shorter than the period. Although the weakness may not lead to a significant plaintext uncertainty reduction, it is structure dependent and can be used as such to distinguish between different types of keystream generators and for secret key reconstruction as well. As well, linear models of individual components of a keystream generator can be utilized in correlation attacks, whereas multiple linear models can also be used to mount fast correlation attacks [8] on clock-controlled LFSRs.

The main objective of this paper is to derive linear models for RC4 by using the LSCA method [5], [9]. The LSCA method consists in determining and solving a linear sequential circuit that approximates a given keystream generator and yields linear models with comparatively large correlation coefficient c, where the probability of the corresponding linear relation among the keystream bits is $(1+c)/2$. It also gives an estimate of c, but sometimes, as in the case of RC4, special techniques have to be developed to obtain more accurate estimates of c.

Given a parameter n, the internal state of RC4 consists of a balanced table (permutation) of 2^n binary words of dimension n and two pointer binary words of the same dimension, n, which, at each time, define the positions of two words in the table to be swapped to produce the table at the next time. The internal memory size[1] is thus practically given as $M = n2^n + 2n$. One of the pointers is updated by using the table content at the position defined by the other, which is in turn updated in a known way by a counter. Initially, the two pointer words are set to zero and the table content is defined by the secret key in a specified way. At each time, the output of RC4 is a binary word of dimension n which is taken from an appropriate position in the table. The output word is then bitwise added to the plaintext word to give the ciphertext word.

Let $z = (z_t)_{t=1}^{\infty}$ denote the least significant bit output sequence of RC4 and let $\dot{z} = (\dot{z}_t = z_t + z_{t+1})_{t=1}^{\infty}$ and $\ddot{z} = (\ddot{z}_t = z_t + z_{t+2})_{t=1}^{\infty}$ denote its first and second binary derivatives, respectively. Our main results are to show that $\dot{z}$ is correlated neither to 1 nor to 0 and that $\ddot{z}$ is correlated to 1 with the correlation coefficient close to $15 \cdot 2^{-3n}$ for large 2^n. Since the output sequence length needed to detect a statistical weakness with the correlation coefficient c is $O(c^{-2})$, the required length is around $64^n/225$. For example, if $n = 8$, as recommended in most applications, the required length is close to $2^{40} \approx 10^{12}$. Experimental results agree well with the above theoretical predictions. As the resulting correlation coefficient is significantly bigger than $2^{M/2}$, $M = n2^n + 2n$, the determined linear model should be regarded as a statistical weakness, at least on a theoretical level. Moreover, the output sequence length required for the detection may even be realistic in high speed applications if $n \leq 8$. Also note that the second binary derivative weakness involves only three successive least significant output bits which is much smaller than the memory size. The weakness is a consequence of a very simple next-state function of RC4. It is also shown that similar linear relations hold for other output bits as well, but the correlation coefficients are smaller.

In Section 2, a more detailed description of the RC4 keystream generator is presented. In Section 3, some relevant correlation properties of random boolean functions are derived, while the linear models of RC4 and the corresponding correlation coefficients are determined in Section 4. A summary and conclusions are given in Section 5. Central moments of an underlying discrete probability distribution needed for estimating the correlation coefficients are evaluated in the Appendix.

[1] The effective internal memory size is slightly smaller and is according to Stirling's approximation given as $\log 2^n! + 2n \approx 2^n(n - \log e) + 5n/2 + \log\sqrt{2\pi}$. All the logarithms are to the base 2 throughout.

2 Description of RC4

We will follow the description given in [18]. RC4 is in fact a family of algorithms indexed by a parameter n, which is a positive integer typically recommended to be equal to 8. The internal state of RC4 at time t consists of a table $S_t = (S_t(l))_{l=0}^{2^n-1}$ of 2^n n-bit words and of two pointer n-bit words i_t and j_t. So, the internal memory size[1] is $M = n2^n + 2n$. Let the output n-bit word of RC4 at time t be denoted by Z_t. As usual, we keep the same notation for the binary and integer representations of n-bit words, where, for example, the least significant bit is the leftmost one. Let initially $i_0 = j_0 = 0$. Then the next-state and output functions of RC4 are for every $t \geq 1$ defined by

$$i_t = i_{t-1} + 1 \tag{1}$$

$$j_t = j_{t-1} + S_{t-1}(i_t) \tag{2}$$

$$S_t(i_t) = S_{t-1}(j_t), \quad S_t(j_t) = S_{t-1}(i_t) \tag{3}$$

$$Z_t = S_t(S_t(i_t) + S_t(j_t)) \tag{4}$$

where all the additions are modulo 2^n. It is assumed that all the words except for the swapped ones remain the same (swapping itself is effective only if $i_t \neq j_t$). The output n-bit word sequence is $Z = (Z_t)_{t=1}^{\infty}$.

The initial table S_0 is defined in terms of the key string $K = (K_l)_{l=0}^{2^n-1}$ by using the same next-state function starting from the table (identity permutation) $(l)_{l=0}^{2^n-1}$. More precisely, set $j_0 = 0$ and for every $1 \leq t \leq 2^n$, compute $j_t = (j_{t-1} + S_{t-1}(t-1) + K_{t-1}) \bmod 2^n$ and then swap $S_{t-1}(t-1)$ with $S_{t-1}(j_t)$. The last produced table represents S_0. The key string K is composed of the secret key, possibly repeated, and of the randomizing key which is sent in the clear for resynchronization purposes.

There are no published results regarding RC4. The known pointer sequence $\{i_t\}_{t=0}^{\infty}$ ensures that every element in the table is affected by swapping at least once in any 2^n successive times and, also, that the next-state function is invertible (one-to-one). Accordingly, the state diagram consists of cycles only, which, according to [4], can be expected to have average length close to 2^{M-1} and are very unlikely to be short if $n \geq 5$. Of course, since the next-state function of RC4 is not randomly chosen, this remains to be proved, if possible at all.

3 Correlation Properties of Random Boolean Functions

The correlation coefficients of the linear models of RC4 to be determined in the next section are related to certain correlation properties of random boolean functions. These properties provide insight into the linear statistical weaknesses of RC4 and are as such pointed out in this section. Note that the correlation properties of boolean functions for cryptographic applications are first introduced in [19]. Let f denote an arbitrary boolean function of n variables and let $f(X)$ denote the value of f at a point $X = (x_0, \ldots, x_{n-1}) \in \{0,1\}^n$. We will use the same notation, X, for the integer representation of X too, that is, for $\sum_{i=0}^{n-1} x_i 2^i$. A boolean function f is called balanced if it has the same number of zeros and ones in its truth table. In the probabilistic analysis to follow, we will, for simplicity, keep the same notation for random variables and their values. As usual, the correlation coefficient between any two binary random variables x and y is defined as $c = \Pr\{x = y\} - \Pr\{x \neq y\}$. The correlation coefficient of a single binary random variable x is defined as the correlation coefficient between x and the constant zero variable. Accordingly, let for any two boolean functions f and g, $c(f,g)$ denote the correlation coefficient between $f(X)$ and $g(X)$, and let $c(f)$ stand for $c(f,0)$, where X is uniformly distributed. A basic result to be used is that the correlation coefficient of a sum of independent binary random variables is equal to the product of their individual correlation coefficients, see [9] (addition of binary variables is modulo 2 throughout).

Proposition 1. *Let X and Y be two independent uniformly distributed n-dimensional binary random variables and let f be a uniformly random boolean function of n variables. Let l be an arbitrary linear boolean function of n variables (including the constant zero function). Then the correlation coefficient c between $f(X) + f(Y)$ and $l(X) + l(Y)$ is equal to $1/2^n$. (Instead of being linear, l may be any boolean function of n variables.)*

Proof. Let c_f denote the correlation coefficient between $f(X)+l(X)$ and $f(Y)+l(Y)$ for any fixed f. The correlation coefficient c is then equal to the expected value of c_f over uniformly random f. The correlation coefficient c_f is clearly equal to the correlation coefficient of $f(X) + l(X) + f(Y) + l(Y)$ which is in turn equal to the product of the correlation coefficients of $f(X) + l(X)$ and $f(Y) + l(Y)$, as X and Y are independent. Since the two are equal, we get that $c_f = c(f,l)^2$. Since l is fixed, c is then equal to the expected value $E(c(f)^2)$, where $c(f)$ is itself given as $2^{1-n}(k - 2^{n-1})$ with k being the number of zeros in the truth table of f. As k has the binomial distribution $\left\{\binom{2^n}{k}2^{-2^n}\right\}_{k=0}^{2^n}$, it follows that $E(c(f)^2) = 2^{2(1-n)}\mathrm{Var}(k) = 2^{-n}$, because the variance $\mathrm{Var}(k)$ is equal to 2^{n-2}. □

Proposition 2. *Let X and Y be two independent uniformly distributed n-dimensional binary random variables and let f be a uniformly random balanced boolean function of n variables. Let l be an arbitrary nonzero linear boolean function of n variables. Then the correlation coefficient of $f(X) + f(Y)$ is equal to zero and the correlation coefficient c between $f(X) + f(Y)$ and $l(X) + l(Y)$ is equal to $1/(2^n - 1)$. (Instead of being linear, l may be any balanced boolean function of n variables.)*

Proof. First note that for any balanced f, the correlation coefficient of $f(X)$ is equal to zero. Then the correlation coefficient of $f(X) + f(Y)$ is equal to zero since, for any fixed f, it is the product of two zero correlation coefficients. Second, proceeding along similar lines as in the proof of Proposition 1, we get that $c = E(c(f,l)^2)$. Since l is balanced and fixed, $c(f,l)$ is given as $2^{2-n}(k - 2^{n-2})$ where k is the number of zeros in the half of the truth table of f where $l(X) = 0$. The probability distribution of k is $\left\{ \binom{2^{n-1}}{k}^2 / \binom{2^n}{2^{n-1}} \right\}_{k=0}^{2^{n-1}}$ with the variance $\mathrm{Var}(k) = 2^{2(n-2)}/(2^n - 1)$. Hence $E(c(f,l)^2) = 2^{2(2-n)}\mathrm{Var}(k) = 1/(2^n - 1)$. □

Proposition 3. *Let l be an arbitrary nonzero linear boolean function of n variables and let f be a uniformly random balanced boolean function of n variables such that $c(f,l) = c$ where c is a given constant. Then the correlation coefficient of $f(X) + l(X)$ is equal to c for any fixed X. (Instead of being linear, l may be any balanced boolean function of n variables.)*

Proposition 4. *Let X be a uniformly distributed n-dimensional binary random variable and let f be a uniformly random balanced boolean function of n variables. Let $X+1$ denote the integer addition modulo 2^n of X and 1. Then the correlation coefficient of $f(X) + f(X+1) + 1$ is equal to $1/(2^n - 1)$. Furthermore, let l be a linear function defined as $l(X) = x_0$ and let f be in addition such that $c(f,l) = c$ where c is a given constant. Then the correlation coefficient of $f(X)+f(X+1)+1$ is equal to c^2 for any fixed X. (Instead of $X + 1$, one may take any permutation $P(X)$ such that $P(X) \neq X$, $X \in \{0,1\}^n$, but then a balanced function l has to be defined appropriately.)*

4 Linear Models

The essence of the linear sequential circuit approximation (LSCA) method [5], [9] applied to binary keystream generators is in finding good linear approximations to the output and the component next-state functions and in solving the resulting linear sequential circuit. Its objective is to obtain feedforward linear transforms (i.e., linear sequential transforms with finite input memory) of the output sequence that are correlated to linear transforms of the initial state variables (to be used in correlation attacks) and, in particular, to the constant zero sequence, in which case the output linear transform defines a linear relation among the output bits that holds with probability different from one half. The resulting probabilistic linear recursion is called a linear model [9]. Estimating the correlation coefficients can be a problem on its own. In the underlying probabilistic model, the initial state is assumed to be random and uniformly distributed, and if the next-state function is one-to-one, then the internal state at any time is also uniformly distributed, so that the resulting correlation coefficients are time independent, see [9].

In the case of RC4, the next-state function is one-to-one and the balanced initial table S_0 (each n-bit word appears exactly once) can be assumed to be uniformly random, but the initial pointer words i_0 and j_0 are both fixed to zero. It follows that for every $t \geq 0$, the table S_t is uniformly random and balanced, whereas i_t is deterministic and known and j_t is uniformly distributed for $t \geq 1$, but dependent on S_t. As a consequence, while the dependence between j_t and S_t is insignificant, the deterministic nature of i_t may in principle lead to linear models with time dependent correlation coefficients. A related approach is to fix the initial state and to consider the same linear relation at random times, in which case the average value of the correlation coefficient over time is relevant. If the tail and cycle lengths combined are big (as one should expect for RC4), then the obtained correlation coefficient should be close to the value corresponding to a fixed time and a random initial state.

Since RC4 has n binary outputs, one should first decide on a linear combination of these outputs to be linearly approximated. To maximize the correlation coefficients, we will consider the individual binary outputs. Let $Z_t^{(k)}$, $i_t^{(k)}$, $j_t^{(k)}$, and $S_t^{(k)}$ denote the kth components of Z_t, i_t, j_t, and S_t, respectively, $0 \leq k \leq n-1$, where $k = 0$ corresponds to the least significant bit of the corresponding n-bit words. Note that S_t defines a uniformly random balanced vectorial boolean function $\{0,1\}^n \to \{0,1\}^n$, so that $S_t^{(k)}$ is a uniformly random balanced boolean function of n variables. As the linearization of Z_t and j_t necessarily involves finding linear approximations to S_t, the problem is to find such approximations leading to the correlation coefficients that do not vanish for a random S_t. The main point of the LSCA method applied to RC4 is that S_t can be approximated by S_{t-1}, because of the slow change of the table due to swapping. Another point is that $S_{t-1}^{(k)}$ can be approximated by any linear function of its inputs, but to maximize the overall correlation coefficient, $S_{t-1}^{(k)}$ is approximated by its kth binary input. As before, all the additions of l-bit words are integer additions modulo 2^l (usually, $l = 1$ or $l = n$).

As a result, we get $Z_t^{(k)} \approx S_{t-1}^{(k)}(i_{t-1}+1) + S_{t-1}^{(k)}(j_{t-1} + S_{t-1}(i_{t-1}+1)) \approx j_{t-1}^{(k)}$, where $S_{t-1}^{(k)}$ is linearized exactly twice. It then follows that $Z_t^{(k)} + Z_{t+1}^{(k)} \approx j_{t-1}^{(k)} + j_t^{(k)} \approx i_t^{(k)}$, where $i_t^{(k)}$ is known for every $t \geq 1$. The total number of linear approximations needed is five. In order for the overall correlation coefficient not to vanish, the total number of linear approximations to $S_{t-1}^{(k)}$ should be even, because positive and negative correlation coefficients would otherwise cancel out. More precisely, Proposition 2 can be extended to deal with an arbitrary number of linear approximations, in which case the resulting correlation coefficient is related to the central moments of the probability distribution considered in the Appendix, and the odd central moments are necessarily equal to zero. So, the first binary derivative of any binary component of the n-dimensional output sequence does not represent a linear model with a nonzero correlation coefficient.

Further, by adding two successive bits of the first binary derivative sequence we get that $Z_t^{(k)} + Z_{t+2}^{(k)} \approx i_t^{(k)} + i_{t+1}^{(k)}$, which is further equal to 1 if $k = 0$ and can be approximated as 0 if $1 \leq k \leq n-1$. The total number of linear approximations needed for this is at most ten and will be shown be equal to six. Accordingly, the second binary derivative of any binary component of the output sequence defines a linear model with a nonzero correlation correlation coefficient, to be determined in the sequel. The most significant correlation coefficient is obtained for the least significant bit, that is, for $k = 0$. Other linear models for RC4 should have smaller or much smaller correlation coefficients.

Our objective now is to estimate the correlation coefficient between the second binary derivative $\ddot{Z}_t^{(0)} = Z_t^{(0)} + Z_{t+2}^{(0)}$ and 1, for any $t \geq 1$. Letting $F = S_t$, $F' = S_{t+1}$, $F'' = S_{t+2}$, $X = i_t$, and $Y = j_t$, we have

$$\begin{aligned}\ddot{Z}_t^{(0)} &= F^{(0)}(F(X) + F(Y)) + \\ &\quad F''^{(0)}(F''(X+2) + F''(Y + F(X+1) + F'(X+2))) \end{aligned} \tag{5}$$

where Y is uniformly distributed, F is a uniformly random balanced vectorial boolean function, and F' and F'' are obtained from F by one and two random swappings of two n-bit words, respectively, whereas X is fixed for any particular t and is uniformly distributed for a random t.

The direct computation of the correlation coefficient by using (5) is not possible since the functions F, F', and F'' are random. The starting point of our approach is forming the following series of linear approximations:

$$\begin{aligned}\ddot{Z}_t^{(0)} &\approx F^{(0)}(X) + F^{(0)}(Y) + \\ &\quad F''^{(0)}(F''(X+2) + F''(Y + F(X+1) + F'(X+2))) && (6)\\ &\approx F^{(0)}(X) + F^{(0)}(Y) + \\ &\quad F''^{(0)}(X+2) + F''^{(0)}(Y + F(X+1) + F'(X+2)) && (7)\\ &\approx F^{(0)}(X) + F^{(0)}(Y) + F''^{(0)}(X+2) + \\ &\quad Y^{(0)} + F^{(0)}(X+1) + F'^{(0)}(X+2) && (8)\\ &\approx F^{(0)}(X) + Y^{(0)} + F''^{(0)}(X+2) + \\ &\quad Y^{(0)} + F^{(0)}(X+1) + F'^{(0)}(X+2) && (9)\\ &\approx F^{(0)}(X) + F^{(0)}(X+1) && (10)\\ &\approx 1. && (11)\end{aligned}$$

The next point is to observe that the correlation coefficients of the individual linear approximations can be computed if conditioned on the random functions in an appropriate way. Let $c_f = c(F^{(0)}, X^{(0)})$, $c'_f = c(F'^{(0)}, X^{(0)})$, and $c''_f = c(F''^{(0)}, X^{(0)})$ be the correlation coefficients between $F^{(0)}$ and $X^{(0)}$, $F'^{(0)}$ and $X^{(0)}$, and $F''^{(0)}$ and $X^{(0)}$, respectively, where the subscript f indicates the dependence upon a particular balanced boolean function f (here $f = F^{(0)}$). Then the linear approximations (6), (7), (8), and (9) hold with the correlation coefficients c_f, c''_f, c''_f, and c_f, respectively, where $F^{(0)}$, $F'^{(0)}$, and $F''^{(0)}$ are fixed and X is either uniformly distributed or fixed. The linear approximation (10) holds for any fixed X with the correlation coefficient $\varepsilon'_{m'} = 1 - m'2^{1-n}$ (conditioned on m') if $F'^{(0)}$ is a uniformly random balanced boolean function and if $F''^{(0)}$ is produced from $F'^{(0)}$ by a random effective change, due to swapping, of m' bits, where, as before, m' takes values 0 and 2, each with probability 1/2. The linear approximation (11) holds for any fixed X with correlation coefficient c_f^2 if $F^{(0)}$ is a uniformly random balanced boolean function with a fixed correlation coefficient c_f to $X^{(0)}$, see Proposition 4.

Now, let m denote the number of bits where $F^{(0)}$ and $F''^{(0)}$ are effectively different. Under the *independence assumption* that the individual linear approximations are independent when conditioned on c_f, m', and m, the correlation coefficient between $\ddot{Z}_t^{(0)}$ and 1 is given as $c_f^4 c_f''^2 \varepsilon'_{m'}$, where $c_f'' = c_f \varepsilon_m$, $\varepsilon_m = 1 - m2^{1-n}$, if $F^{(0)}$ is a uniformly random balanced boolean function with a fixed correlation coefficient c_f to $X^{(0)}$, where X is either uniformly distributed or fixed. The resulting correlation coefficient conditioned on c_f, m', and m is thus equal to $c_f^6 \varepsilon_m^2 \varepsilon'_{m'}$. Note that the independence assumption seems to be the only tractable way of combining the individual linear approximations.

Consequently, the overall correlation coefficient is then given as

$$c = E(c_f^6) \cdot E(\varepsilon_m^2) \cdot E(\varepsilon'_{m'}) \tag{12}$$

where the expectations are over random c_f, m, and m', respectively (for simplicity, it is assumed that the random variables m' and m are independent). From the proof of Proposition 2, recall that c_f can be expressed as $2^{2-n}(k - 2^{n-2})$ where k (standing for the number of zeros in the half of the truth table of $f = F^{(0)}$ where $X^{(0)} = 0$) has the probability distribution

$$\Pr\{k\} = \frac{\binom{2^{n-1}}{k}^2}{\binom{2^n}{2^{n-1}}}, \quad 0 \leq k \leq 2^{n-1}. \tag{13}$$

The random variable m' takes values 0 and 2 each with probability 1/2, so that

$$E(\varepsilon'_{m'}) = \varepsilon'_{E(m')} = 1 - 2^{1-n} \tag{14}$$

which tends to 1 as 2^n increases.

The probability distribution of m is not straightforward to derive. By careful combinatorial analysis, one can prove the following result.

Lemma 5. *Let f be a uniformly random balanced boolean function of n variables and let f'' be a boolean function obtained from f first by swapping the bits defined by input variables X and Y and, then, by additional swapping the bits defined by $X + 1$ and Y', where X is fixed or random and Y and Y' are independent uniformly distributed n-dimensional binary random variables. Let m be the number of bits where f and f'' are different and let $N = 2^n$. Then m is a random variable with the following probability distribution*

$$\Pr\{m = 0\} = \frac{N^2 - N + 2}{4N(N-1)} \tag{15}$$

$$\Pr\{m = 2\} = \frac{2N^2 + N - 6}{4N(N-1)} \tag{16}$$

$$\Pr\{m = 4\} = \frac{(N-2)^2}{4N(N-1)}. \tag{17}$$

The expected value of m is given by

$$E(m) \;=\; \frac{4N^2 - 7N + 2}{2N(N-1)}. \tag{18}$$

Note that $E(m) < 2$ since effective changes in two successive swappings can cancel out, but as N increases, we have that $\Pr\{m = 0\} \sim 1/4$, $\Pr\{m = 2\} \sim 1/2$, $\Pr\{m = 4\} \sim 1/4$, and $E(m) \sim 2$, as should be expected. Accordingly, we get

$$E(\varepsilon_m^2) \;=\; \frac{N^4 - 9N^3 + 38N^2 - 64N + 40}{N^3(N-1)} \tag{19}$$

which, of course, tends to 1 as $N = 2^n$ increases.

Finally, it remains to compute the main product factor in (12), that is, $E(c_f^6)$. According to (13), we then have

$$E(c_f^6) \;=\; 2^{-6(n-2)}\mu_6 \tag{20}$$

where μ_6 is the 6th central moment of the probability distribution (13), that is,

$$\mu_6 \;=\; \sum_{k=0}^{2^{n-1}} (k - 2^{n-2})^6 \frac{\binom{2^{n-1}}{k}^2}{\binom{2^n}{2^{n-1}}} \sim 15 \cdot 2^{3(n-4)}, \tag{21}$$

see the Appendix. It is crucial to observe that the exponent, 6, is even, so that μ_6 is necessarily different from zero.

The equation (12) together with (20), (21), (19), and (14) then determines the overall correlation coefficient c which can be easily computed for any n of interest, and, as 2^n increases we have

$$c \;\sim\; 15 \cdot 2^{-3n}. \tag{22}$$

The necessary sequence length to detect with high probability the second binary derivative statistical weakness is $O(c^{-2})$ [9], that is, neglecting a small constant less than 10,

$$L \;\approx\; 2^{6n}/225 \;\approx\; 2^{6n-7.814} \;\approx\; 10^{1.8n-2.35}. \tag{23}$$

As the memory size of RC4 is $M = n2^n + 2n$, we get $L \approx (M/(2.466 \log M))^6$.

For example, for $n = 4, 6, 8$, we computed the following values for μ_6 and c: $\mu_6 \approx 16.1716$ and $c \approx 2.2 \cdot 10^{-3}$, $\mu_6 \approx 975.762$ and $c \approx 4.97 \cdot 10^{-5}$, $\mu_6 \approx 61682.916$ and $c \approx 8.67 \cdot 10^{-7}$, respectively. In fact, for $n \geq 4$, the approximation to μ_6 included in (21) is also very good. The estimates of c obtained by computer simulations for $n = 4$ and $n = 6$ are $\hat{c} = 1.34 \cdot 10^{-3}$ and $\hat{c} = 1.95 \cdot 10^{-5}$, respectively. The first estimate is an average value for 5 output sequences each of length 10^{11} and the second one is an average value for 10 output sequences each of length 10^{11}, where each sequence is produced from a randomly chosen initial state. One may observe that the estimates are roughly by 50% smaller that the values predicted by theory. This shows that the influence of the utilized linear approximations being dependent is relatively small. The difference may also be due to the fact that the correlation coefficient estimates are essentially obtained by averaging over time rather than over random initial states.

5 Conclusions

The linear model approach aiming at finding linear relations among the keystream bits that hold with probability different from one half is applied to the RC4 keystream generator. It is first shown by the linear sequential circuit approximation method that the first and the second binary derivative of the least significant bit output sequence may yield such linear relations. A specific technique involving correlation properties of random balanced boolean functions is then developed to study the corresponding correlation coefficients. It is thus proven that the correlation coefficient for the first binary derivative is equal to zero and, more importantly, that the correlation coefficient between the second binary derivative and 1 is around $15 \cdot 2^{-3n}$ where n is the word size of RC4. The theoretical result derived agrees well with the experimental results obtained by computer simulations.

The output sequence length needed to detect the corresponding linear statistical weakness is then around $64^n/225$, which is significantly smaller than 2^M, where $M = n2^n + 2n$ is the memory size, and may even be realistic in high speed applications. Although the resulting plaintext uncertainty reduction may not be practically important unless n is small, the determined linear model can be used to distinguish RC4 from other keystream generators and, also, to recover the unknown parameter n. Whether the linear model indicates that the initial state reconstruction from the known output sequence is also possible remains to be further investigated.

Appendix

Consider a discrete probability distribution $\left\{\binom{2\nu}{k}^2/\binom{4\nu}{2\nu}\right\}_{k=0}^{2\nu}$ where ν is a positive integer. For any positive integer r, the central moment μ_r of this probability distribution is defined as

$$\mu_r = \sum_{k=0}^{2\nu}(k-\nu)^r \frac{\binom{2\nu}{k}^2}{\binom{4\nu}{2\nu}}. \tag{24}$$

Our objective here is to study the asymptotics of μ_r as ν increases. First note that $\mu_r = 0$ if r is odd. Assume then that r is even. By using the well-known normal approximation to the binomial coefficients, obtained by Stirling's formula $n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}}e^{-n}$, along with a uniform convergence argument regarding this approximation (e.g., see [3, pp. 179-186]), it is easy to see that

$$\mu_r \sim \frac{\nu^{r/2}}{2^r\sqrt{2\pi}}\int_{-\infty}^{\infty} x^r e^{-x^2/2}dx \tag{25}$$

as $\nu \to \infty$. For r even, this reduces to

$$\mu_r \sim \frac{\nu^{r/2}}{2^r}\sqrt{\frac{2^r}{\pi}}\,\Gamma\left(\frac{r+1}{2}\right) \tag{26}$$

where $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$ is the well-known gamma function. Finally, we obtain

$$\mu_r \sim \frac{\nu^{r/2}}{2^r}(r-1)!! \tag{27}$$

where $(r-1)!! = 1 \cdot 3 \cdots (r-1)$.

Acknowledgments

The author is grateful to Lars Knudsen and Andrew Klapper for providing the correlation coefficient estimates by computer simulations. Part of this work was carried out while the author was on leave at the Isaac Newton Institute for Mathematical Sciences, Cambridge, United Kingdom.

References

1. Ameritech Mobile Communications et al., "Cellular digital packet data system specifications, part 406: airlink security," CDPD Industry Input Coordinator, Costa Mesa, Calif., July 1993.
2. J. Daemen, R. Govaerts, and J. Vandewalle, "Resynchronization weakness in synchronous stream ciphers," Advances in Cryptology - EUROCRYPT '92, *Lecture Notes in Computer Science*, vol. 765, T. Helleseth ed., Springer-Verlag, pp. 159-167, 1994.
3. W. Feller, *An Introduction to Probability Theory and its Applications*. New York: Wiley, 3. edition, vol. 1, 1968.
4. P. Flajolet and A. M. Odlyzko, "Random mapping statistics," Advances in Cryptology - EUROCRYPT '89, *Lecture Notes in Computer Science*, vol. 434, J.-J. Quisquater and J. Vandewalle eds., Springer-Verlag, pp. 329-354, 1990.
5. 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.
6. J. Dj. Golić, "On the security of shift register based keystream generators," Fast Software Encryption - Cambridge '93, *Lecture Notes in Computer Science*, vol. 809, R. J. Anderson ed., Springer-Verlag, pp. 90-100, 1994.
7. J. Dj. Golić, "Intrinsic statistical weakness of keystream generators," Advances in Cryptology - ASIACRYPT '94, *Lecture Notes in Computer Science*, vol. 917, J. Pieprzyk and R. Safavi-Naini eds., Springer-Verlag, pp. 91-103, 1995.
8. J. Dj. Golić, "Towards fast correlation attacks on irregularly clocked shift registers," Advances in Cryptology - EUROCRYPT '95, *Lecture Notes in Computer Science*, vol. 921, L. C. Guillou and J.-J. Quisquater eds., Springer-Verlag, pp. 248-262, 1995.

9. J. Dj. Golić, "Linear models for keystream generators," *IEEE Trans. Computers*, vol. C-45, pp. 41-49, Jan. 1996.
10. J. Dj. Golić, "On the security of nonlinear filter generators," Fast Software Encryption - Cambridge '96, *Lecture Notes in Computer Science*, vol. 1039, D. Gollmann ed., Springer-Verlag, pp. 173-188, 1996.
11. R. J. Jenkins Jr., "ISAAC," Fast Software Encryption - Cambridge '96, *Lecture Notes in Computer Science*, vol. 1039, D. Gollmann ed., Springer-Verlag, pp. 41-49, 1996.
12. M. D. MacLaren and G. Marsaglia, "Uniform random number generation," *J. ACM*, vol. 15, pp. 83-89, 1965.
13. W. Meier and O. Staffelbach, "Fast correlation attacks on certain stream ciphers," *Journal of Cryptology*, vol. 1(3), pp. 159-176, 1989.
14. W. Meier and O. Staffelbach, "Correlation properties of combiners with memory in stream ciphers," *Journal of Cryptology*, vol. 5(1), pp. 67-86, 1992.
15. R. L. Rivest, "The RC4 encryption algorithm," RSA Data Security, Inc., Mar. 1992.
16. R. A. Rueppel, *Analysis and Design of Stream Ciphers*. Berlin: Springer-Verlag, 1986.
17. R. A. Rueppel, "Stream ciphers," *Contemporary Cryptology: The Science of Information Integrity*, G. Simmons ed., pp. 65-134. New York: IEEE Press, 1991.
18. B. Schneier, *Applied Cryptography*. New-York: Wiley, 1996.
19. T. Siegenthaler, "Correlation immunity of nonlinear combining functions for cryptographic applications," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 776-780, Sept. 1984.