Application of linear consistency test in a ciphertext-only attack on irregularly clocked linear feedback shift registers

Slobodan Petrović
Department of Computer Science and Media Technology
Gjøvik University College, p.o. box 191, N-2802 Gjøvik, Norway
Email: slobodan.petrovic@hig.no

***Abstract*—Linear Consistency Test (LCT) is a well-known algebraic method of cryptanalysis of stream ciphers. In this paper, we use LCT in an attack on a noised irregularly clocked linear feedback shift register (LFSR). We show that it is possible to reconstruct the initial states of both the clocked and the clocking LFSR in this scheme by using an essentially algebraic attack method, such as LCT, as a building block in an attack scenario with noise, which is a correlation attack by its nature. An advantage of the attack compared to other attacks against the same scheme is in the fact that it is not necessary to use search in the initial state reconstruction of the clocked LFSR, which significantly improves the efficiency of the attack. It is shown experimentally that the attack is successful for moderate levels of noise in the system.**

I. Introduction

Irregularly clocked linear feedback shift registers (LFSRs) have become usual primitives in many pseudorandom sequence generators due to their good cryptographic properties: long periods, high linear complexities, good statistical properties etc. [2]. In a scheme of this kind, a clocking LFSR, LFSR_s, produces the decimation sequence that determines which of the bits generated by the clocked LFSR, LFSR_u, will be skipped/sent to the output, see Fig. 1. It is well known (see for example [5]) that such generators are vulnerable to generalized correlation attacks, i.e. attacks exploiting existence of generalized correlation between the noised output sequence and some internal sequences. Generalized correlation in this case means that there exists an initial state of the clocked LFSR that produces the output sequence without irregular clocking, whose so-called edit distance (see for example [10], [12]) to the output sequence is less than a threshold given in advance.

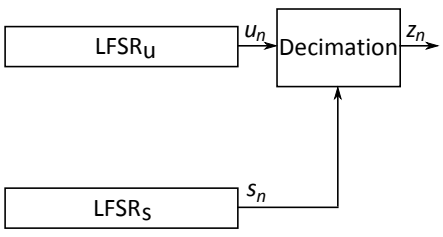

Fig. 1. An irregularly clocked LFSR

By means of the generalized correlation attack it is possible to reconstruct the initial state of the clocked register LFSR_u. The next step is to reconstruct the clocking sequence and consequently the initial state of the clocking register LFSR_s. There are several ways to achieve this (see for example [3], [7], [11], [14]). In any case, the total time complexity of the attack reconstructing the initial states of both the clocked and the clocking LFSR of the scheme depends on the lengths of both registers.

Algebraic attacks against pseudorandom generator schemes employing irregularly clocked LFSRs have been described as well. To launch an algebraic attack, the output sequence of the pseudorandom sequence generator must be known, which means that a known-plaintext attack scenario is considered. This scenario is not very realistic in the stream cipher environment, since the sequence that is usually intercepted by the cryptanalyst is the ciphertext sequence, which is a noised version of the output sequence of the generator.

A typical algebraic attack used in the known-plaintext attack scenario is the attack employing the Linear Consistency Test [13]. LCT is a key recovery attack, which uses some guessed bits from the internal state of a pseudorandom sequence generator to determine the unknown bits of the key and to accept or recject the guessed initial state. The LCT procedure is as follows: First, a candidate subkey is guessed. Then a system of equations parameterized by this subkey is set up. If the candidate subkey coincides with the very subkey used in generating the intercepted sequence, then this set of equations will be consistent. But if the candidate subkey is not the subkey used then, by a theorem proved in [13], the consistency probability of the system will be very small if the intercepted sequence is long enough. The consistency of the system of equations is tested for all the possible choices of the candidate subkey, and the right subkey is detected whenever the corresponding system is found to be consistent.

The system of equations can be solved for example by means of the Gaussian algorithm. In [8] and [9], a version of LCT with improved efficiency is described. The improvement is achieved by using low-weight cyclic equations instead of the Gaussian algorithm to check for consistency of the obtained system of equations.

In this paper, we apply the Linear Consistency Test in a

ciphertext-only attack scenario, where the output sequence of the pseudorandom sequence generator employing irregular clocking is noised (in practice it usually means that it is bitwise summed modulo 2 with the plaintext sequence) and the noised sequence is intercepted by the cryptanalyst. We show that the initial states of both LFSR$_s$ and LFSR$_u$ can be reconstructed in the presence of noise and that the time complexity of the attack depends only on the length of LFSR$_s$, unlike all the other known attacks on such schemes.

The structure of the paper is as follows: In Section II, we describe the particular pseudorandom sequence generator employing irregular clocking analyzed in this paper. Then, in Section III we give the details of the ciphertext-only attack. In Section IV, the experimental results obtained on the analyzed generator with several different parameters are given. Section V concludes the paper.

II. The analyzed generator

Irregular clocking is realized in practice in several ways. Examples are the Binary Rate Multiplier [2], the Shrinking Generator [4], the Alternating Step Generator [6] etc. In this paper, we apply the ciphertext-only attack employing LCT on the Binary Rate Multiplier (BRM). The attack is applicable on the other generators employing irregularly clocked LFSRs as well.

The Binary Rate Multiplier consists of 2 LFSRs, the clocking LFSR, LFSR$_s$ and the clocked LFSR, LFSR$_u$, where the clocking of the LFSR$_u$ is determined by the integer decimation sequence produced by k positions of the LFSR$_s$, see Fig. 2.

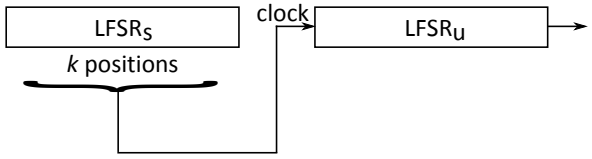

Fig. 2. The Binary Rate Multiplier (BRM)

The binary rate multiplier can be viewed at as a black box, see Fig. 3. The binary sequence u_n is the output sequence of LFSR$_u$ without irregular clocking. The integer sequence s_n is determined by the k positions of LFSR$_s$. The maximum value of a symbol from this sequence determines the maximum number of bits from the sequence u_n that can be skipped in the decimation process. The binary sequence z_n is the decimated binary sequence u_n, which is the output sequence of the whole BRM.

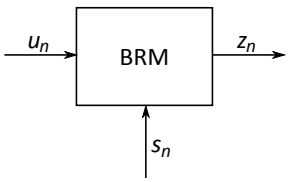

Fig. 3. Operation of the Binary Rate Multiplier, see text

The output sequence z_n is obtained in the following way:

$$z_n = u_{g(n)} \tag{1}$$

where

$$g(n) = n + \sum_{i=0}^{n} s_i \tag{2}$$

In [2] it is shown that the maximum linear complexity achievable with BRM is $l_u P_s$, where l_u is the length of LFSR$_u$ and P_s is the period of LFSR$_s$. This linear complexity is achieved if the lengths of LFSR$_s$ and LFSR$_u$ are equal and both LFSRs have primitive feedback polynomials. Since it is relatively easy to obtain sequences of high linear complexity by means of BRM, this scheme has found many applications in the design of stream ciphers.

III. Attack details

In [13], the following theorem was proved:

Theorem 1

Let $\mathbf{A} = [a_{ij}]$ be an $m \times n$ random binary matrix with entries satisfying, independently from each other, the distribution $Prob(a_{ij} = 0) = 0.5$. Let $\mathbf{b}$ be any given non-zero binary vector of dimension m, $m > n$. Then the probability for the linear system $\mathbf{Ax} = \mathbf{b}$ to be consistent is

$$Prob(\mathbf{Ax} = \mathbf{b} \text{ is consist.}) < \frac{1}{2^{m-n}} \left(1 + \frac{1}{2^{m+1}}\right)^n. \tag{3}$$

□

Obviously, if m is large enough, the consistency probability of a random system as above is very small. This result is applied in the Linear Consistency Test (LCT) in cryptanalysis. Let us fix a subkey K_1 of the key K of a pseudorandom sequence generator, $| K_1 | < | K |$. If the design of the analyzed generator is such that a linear system parameterized by the subkey K_1 can be assigned to it, then it is possible to reconstruct the bits of K_1 by means of the following procedure (the LCT):

1. Guess a value for K_1.
2. Set up a linear system $\mathbf{A}(K_1)\mathbf{x} = \mathbf{b}$ such that $\mathbf{A}(K_1)$ is determined by the analyzed generator and $\mathbf{b}$ is determined by the intercepted output sequence of the generator.
3. Check the consistency of the obtained system. This can be done in many ways, for example by means of the Kronecker-Capelli theorem and Gaussian elimination for determining the rank of the original and the extended matrix of the system. If the system is consistent then the guess of K_1 is certainly right. If the system is inconsistent then by Theorem 1 the probability that the guess of K_1 is still right is very small.
4. We guess other values for K_1 until we get a consistent system. Then we can reduce the dimension of the system by $| K_1 |$, define another subkey to guess, repeat the whole procedure until we reconstruct all the bits of the key K.

It is shown in [13] that the number of equations in the parameterized system should exceed $|\mathbf{x}| + |K_1|$ significantly in order to reduce the number of false consistency alarms to a small value. Then the solution of the system will be unique with probability very close to 1.

The key step in the attack is to determine the system of equations that is linear. Interestingly enough, a linear system suitable for application of LCT is easily assigned to a pseudorandom generator employing irregularly clocked LFSRs, as shown in [8], [9]. There it is also shown that the LCT can be significantly speeded-up by using low-weight cyclic equations determined by the feedback polynomial of the LFSR$_u$, which eliminates the need for Gaussian algorithm.

The linear system assigned to a BRM is obtained by guessing the initial state of the LFSR$_s$. Let l_s and l_u be the lengths of LFSR$_s$ and LFSR$_u$, respectively. If l_s initial bits of LFSR$_s$ are guessed and the corresponding output sequence s_n of LFSR$_s$ is generated, that sequence determines the positions of the skipped bits from the output sequence u_n of LFSR$_u$ without irregular clocking. Since the output sequence z_n of the BRM is known, i.e. intercepted, and the feedback polynomials of both LFSR$_s$ and LFSR$_u$ are known, by guessing the initial state of LFSR$_s$ we obtain a linear system of equations in l_u unknowns, whose consistency is to be checked by means of the LCT.

The attack described above is a pure algebraic, i.e. known plaintext attack. If the plaintext is not known to the cryptanalyst then the only information available to him is the ciphertext sequence and the feedback polynomials of LFSR$_s$ and LFSR$_u$. In that case, the intercepted bits represent the output sequence from the BRM degraded by a noise sequence (plaintext here is considered noise). For lower probabilities of "1" in the noise sequence, it is then possible to iterate the LCT with the same guess for the initial state of LFSR$_s$, starting building the system of equations from another position in the intercepted ciphertext sequence each time. In such a way, the output bits of BRM degraded by the noise will not be present on the right-hand side of the system of equations if the number of LCT repetitions is high enough for each guessed initial state of LFSR$_s$.

The discussion above gives rise to the following ciphertext-only attack against pseudorandom generator schemes employing irregular clocking in general and BRM in particular:

1. Guess the initial state of the clocking sub-generator.
2. Generate the corresponding output sequence of the clocking sub-generator.
3. Repeat N times, N odd, each time starting from the next bit of the intercepted ciphertext sequence (the first time we start from the 1st intercepted bit) in the proces of building the linear system:
   - 3.1. Build the system of linear equations.
   - 3.2. Test the consistency of the obtained system.
   - 3.3. Update the number of cases where a consistent system was obtained.
4. If in the majority of the linear consistency tests a consistent system was obtained, the guessed initial state of the clocking sub-generator is accepted.

Example

Suppose we use a BRM with 0/1 clocking ($k = 1$, see Fig. 2), where the feedback polynomial of LFSR$_s$ is $f_s(x) = 1 + x + x^4$ and the feedback polynomial of LFSR$_u$ is $f_u(x) = 1 + x^3 + x^4$. The clocking signal is taken from the 1st position of LFSR$_s$. If the initial state of LFSR$_s$ is 0101 and the initial state of LFSR$_u$ is 1100, we get the following sequences in the BRM:

$$\begin{aligned} s &= 110010001111010\ldots \\ u &= 010111100010011\ldots \\ z &= 1111000101\ldots \end{aligned}$$

If the cryptanalyst guesses the initial state of LFSR$_s$ right (i.e. he guesses the state 0101 for LFSR$_s$), he gets the following sequence $\hat{u}$ containing variables representing the skipped bits from u in the process of irregular clocking in the BRM:

$$\hat{u} = x_1 1 x_2 111 x_3 0001 x_4 0 x_5 1\ldots$$

From the sequence $\hat{u}$ and the feedback polynomial f_u, we get the following system of (parity check) equations:

$$\begin{array}{lll} x_1 + 1 = 1 & 1 + x_2 = 1 & x_2 + 1 = x_3 \\ 1 + 1 = 0 & 1 + 1 = 0 & 1 + x_3 = 0 \\ x_3 + 0 = 1 & 0 + 0 = x_4 & 0 + 0 = 0 \\ 0 + 1 = x_5 & 1 + x_4 = 1 & \ldots \end{array}$$

The obtained system is consistent and we conclude that the cryptanalyst's guess of the initial state of LFSR$_s$ was right.

Suppose now that the cryptanalyst only has access to the ciphertext and that in addition he knows the feedback polynomials LFSR$_s$ and LFSR$_u$. Then instead of the sequence z_n the cryptanalyst intercepts the sequence z'_n, which is z_n degraded by noise. The noise sequence is a random binary sequence, in which the probability of "1" is less than 0.5. In this particular example, suppose that the 2. and the 8. bit of the sequence z were changed by the noise. Then the intercepted sequence z' is

$$z' = 1011000001\ldots$$

and the sequence $\hat{u}'$ containing variables representing the skipped bits from u in the process of irregular clocking in the BRM:

$$\hat{u}' = x_1 1 x_2 011 x_3 0000 x_4 0 x_5 1\ldots$$

From the sequence $\hat{u}'$ and the feedback polynomial f_u, we get the following system of parity check equations:

$$\begin{array}{lll} x_1 + 1 = 1 & 1 + x_2 = 1 & x_2 + 0 = x_3 \\ 0 + 1 = 0 & 1 + 1 = 0 & 1 + x_3 = 0 \\ x_3 + 0 = 0 & 0 + 0 = x_4 & 0 + 0 = 0 \\ 0 + 0 = x_5 & 0 + x_4 = 1 & \ldots \end{array}$$

The obtained system is obviously inconsistent, even though the cryptanalyst's guess of the initial state of LFSR$_s$ was right. To overcome this, we try starting building the system from the

2. intercepted bit, the 3. and so on N times and each time we check the consistency of the obtained system of parity check equations. If in the majority of attempts to build a system we get a consistent system, we accept the guessed initial state of LFSR$_s$ as the right one.

□

IV. Experimental work

In [1] it was observed that if the guess of the initial state of LFSR$_s$ is wrong, the number of consistent systems obtained with N LCT iterations, as shown in Section III, is zero with high probability. On the other hand, if the guess of the initial state of LFSR$_s$ is right, the probability of getting a consistent system in the majority of N LCT iterations is high, but it still may happen that we get a low number of consistent systems or even 0 consistent systems even though the guess of the initial state of LFSR$_s$ is right. In that case, it is worth increasing the value of N and consequently using more intercepted ciphertext bits to build the system, as it was also shown in [1].

The goal of the experiments was to show that by increasing the number of LCT iterations N, the probability of getting 0 consistent systems when the guess of the initial state of LFSR$_s$ is wrong remains at a high level and the probability of getting 0 consistent systems when the guess of the initial state of LFSR$_s$ is right decreases significantly. To this end, for the lengths of 4 and 7 (equal lengths of LFSR$_s$ and LFSR$_u$) the following experiments were performed:

Experiment 1

In this experiment, we use the right guess for the initial state of LFSR$_s$ and we determine the minimum value of N for which we get the number of consistent systems equal to 0, for each of the lengths of LFSRs given above and for the probabilities of "1" in the noise sequence of 0.1 and 0.15. The results were obtained on the fixed (correct) initial state of LFSR$_s$ and 1000 random combinations of altered output bits from the BRM by the noise for LFSRs of length 4, and 100 for LFSRs of length 7. The results are presented in Fig. 4 – 7.

Experiment 2

In this experiment, we use a wrong guess for the initial state of LFSR$_s$ and we observe the percentage of the cases in which we get the number of consistent systems equal to 0, for each of the lengths of LFSRs given above and for the probabilities of "1" in the noise sequence of 0.1 and 0.15. The results were obtained on the fixed (incorrect) initial state of LFSR$_s$ and 1000 random combinations of altered output bits from the BRM by the noise for LFSRs of length 4, and 100 for LFSRs of length 7. The results are presented in Fig. 8 – 11.

From the figures presented above we can observe that with the increase of N the probability that the right guess of the initial state of LFSR$_s$ will be detected increases and after certain threshold value of N this probability becomes very close to 1. The threshold grows with the lengths of the LFSRs in the BRM as well as with the noise level. We also observe that the increase of N does not practically affect the

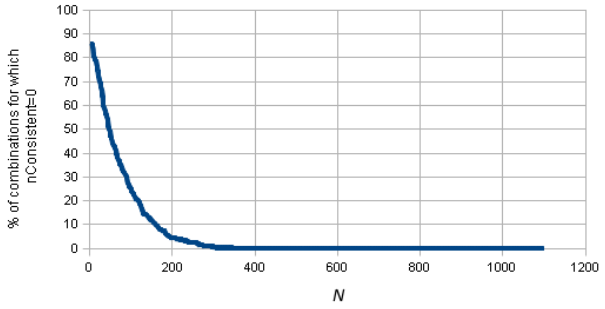

Fig. 4. Right guess detection accuracy (see text); $p = 0.1$, $l_s = 4$

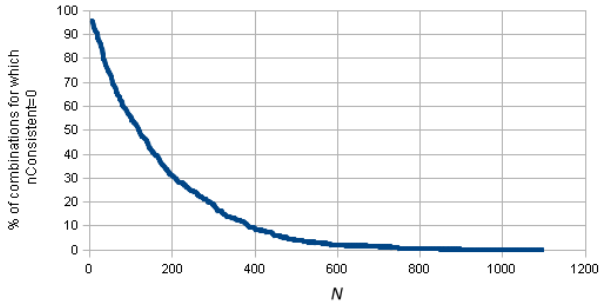

Fig. 5. Right guess detection accuracy (see text); $p = 0.15$, $l_s = 4$

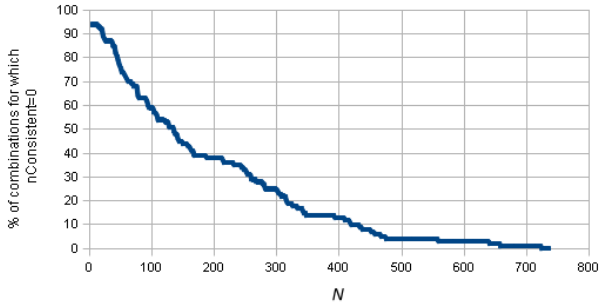

Fig. 6. Right guess detection accuracy (see text); $p = 0.1$, $l_s = 7$

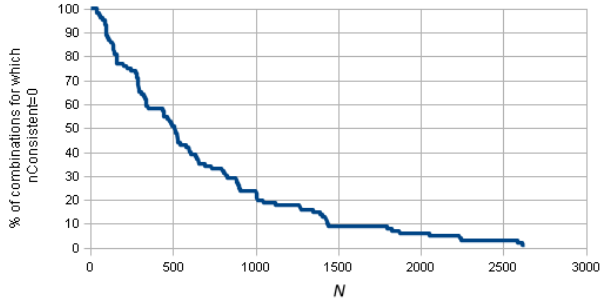

Fig. 7. Right guess detection accuracy (see text); $p = 0.15$, $l_s = 7$

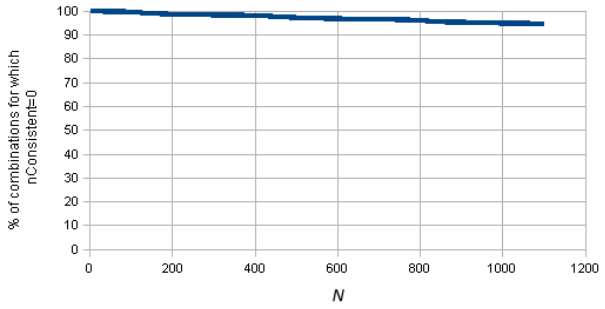

Fig. 8. Wrong guess detection accuracy (see text); $p = 0.1$, $l_s = 4$

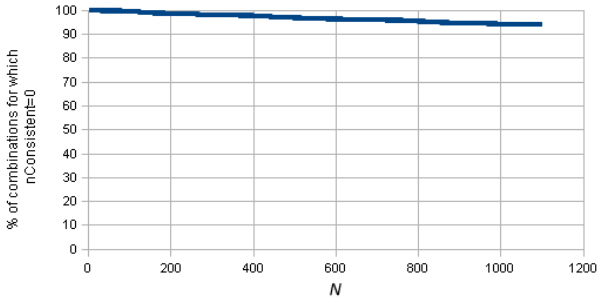

Fig. 9. Wrong guess detection accuracy (see text); $p = 0.15$, $l_s = 4$

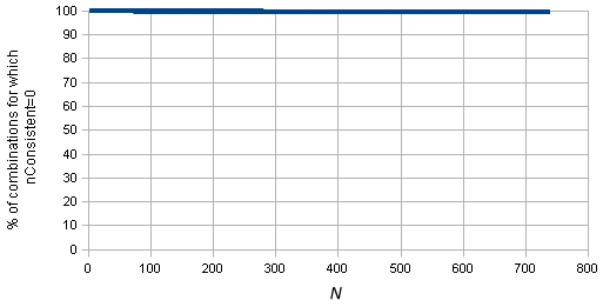

Fig. 10. Wrong guess detection accuracy (see text); $p = 0.1$, $l_s = 7$

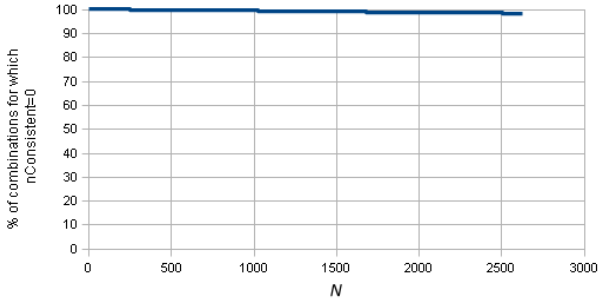

Fig. 11. Wrong guess detection accuracy (see text); $p = 0.15$, $l_s = 7$

probability of detection of a wrong guess of the initial state of LFSR$_s$. This probability is always very close to 1.

The experimental results presented above indicate that the iterative use of LCT in the ciphertext-only attack presented in Section III gives practically useful outcome of such an attack for small to moderate levels of noise.

V. Conclusion

In this paper, a new ciphertext-only attack on pseudorandom sequence generator schemes employing irregularly clocked linear feedback shift registers is presented. The attack makes use of Linear Consistency Testing (LCT), a well known algebraic attack method, in an attack that is essentially a correlation attack. It is shown experimentally that the attack is successful if the level of noise in the statistical model of the scheme is moderate. An advantage of this attack compared to other attacks against this class of pseudorandom sequence generator schemes is the fact that it is not necessary to reconstruct the initial state of the clocked LFSR by search, which significantly improves the efficiency of the attack.

References

[1] G. Bu, "Linear consistency test (LCT) in cryptanalysis of irregularly clocked LFSRs in the presence of noise", Master thesis, Gjøvik University College, Gjøvik, Norway, 2011.

[2] W. Chambers and S. Jennings, "Linear equivalence of certain BRM shift-register sequences", Electronics Letters, vol. 20, no. 24, pp. 1018–1019, 1984.

[3] W. Chambers and J. Golić, "Fast reconstruction of clock-control sequence", Electronics Letters, vol. 38, no. 20, pp. 1174–1175, 2002.

[4] D. Coppersmith, H. Krawczyk, and Y. Mansour, "The shrinking generator", in Proceedings of CRYPTO '93, Lecture Notes in Computer Science LNCS 773, pp. 22–39, Springer–Verlag, 1994.

[5] J. Golić and M. Mihaljević, "A generalized correlation attack on a class of stream ciphers based on the Levenshtein distance", Journal of Cryptology, vol. 3, no. 3, pp. 201–212, 1991.

[6] C. Günther, "Alternating step generators controlled by de Bruijn sequences", in Proceedings of EUROCRYPT '87, Lecture Notes in Computer Science LNCS 304, pp. 5-14, Springer–Verlag, 1988.

[7] T. Johansson, "Reduced complexity correlation attacks on two clock-controlled generators", in: Ohta K. (Ed.), Advances in Cryptology: Proceedings of ASIACRYPT '98, Lecture Notes in Computer Science LNCS 1514, pp. 342–356, Springer–Verlag, 1998.

[8] H. Molland, "Improved linear consistency attack on irregular clocked keystream generators", in Proceedings of Fast Software Encryption (FSE 2004), Lecture Notes in Computer Science LNCS 3017, pp. 109–126, Springer–Verlag, 2004.

[9] H. Molland, T. Helleseth, "An improved correlation attack against irregular clocked and filtered keystream generators", in Proceedings of CRYPTO 2004, Lecture Notes in Computer Science LNCS 3152, pp. 373-389, Springer–Verlag, 2004.

[10] B. Oommen, "Constrained String Editing", Inform. Sci., vol. 40, no. 9, pp. 267–284, 1986.

[11] S. Petrović, A. Fúster, "Clock control sequence reconstruction in the ciphertext only attack scenario", in Proceedings ICICS 2004, Lecture Notes in Computer Science LNCS 3269, pp. 427–439, Springer–Verlag, 2004.

[12] D. Sankoff, J. Kruskal, "Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison", Addison-Wesley, 1983.

[13] K. Zeng, C. Yang, and T. Rao, "On the linear consistency test (LCT) in cryptanalysis with applications", in Advances in Cryptology, Proceedings of CRYPTO '89, Lecture Notes in Computer Science LNCS 435, pp. 164-174, Springer–Verlag, 1990.

[14] E. Zenner, "On the efficiency of the clock control guessing attack", in Proceedings of ICISC 2002, Lecture Notes in Computer Science LNCS 2587, pp. 200-212, Springer–Verlag, 2002.