Information Processing Letters 112 (2012) 805–810

Contents lists available at SciVerse ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

Internal state recovery of keystream generator LILI-128 based on a novel weakness of the employed Boolean function

Miodrag J. Mihaljević [a,b,*], Sugata Gangopadhyay [c], Goutam Paul [d], Hideki Imai [e,f]

[a] Mathematical Institute, Serbian Academy of Sciences and Arts, Kneza Mihaila 36, Belgrade, Serbia
[b] Research Institute for Secure Systems, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan
[c] Department of Mathematics, Indian Institute of Technology, Roorkee 247 667, India
[d] Department of Computer Science and Engineering, Jadavpur University, Kolkata 700 032, India
[e] Faculty of Sciences and Engineering, Chuo University 1-13-27 Kasuga, Bunkyo-ku, Tokyo, 112-8551, Japan
[f] National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

ARTICLE INFO

Article history:
Received 9 March 2010
Received in revised form 25 March 2012
Accepted 23 July 2012
Available online 27 July 2012
Communicated by A. Tarlecki

Keywords:
Cryptography
Boolean functions
Stream ciphers
Internal state recovery
LILI-128 keystream generator

ABSTRACT

This paper proposes an algorithm for internal state recovery of the keystream generator LILI-128 and introduces a novel approach for cryptanalysis of certain stream ciphers which belong to the class of nonlinear filters. The proposed cryptanalysis is based on a cryptographic feature/weakness of the Boolean function employed as the nonlinear filter. It is shown that the developed algorithm is significantly more efficient than the previously reported ones against LILI-128 and can recover the internal state with time complexities of pre-processing and processing of the order of 2^{47} and 2^{35}, respectively, the space complexity of 2^{47}, and a sample of dimension approximately equal to 2^{46}. The developed cryptanalysis is also a practical confirmation on the significance of the so-called "non-normality" design criterion for Boolean functions.

1. Introduction

The LILI-128 keystream generator has been one of the submissions to the NESSIE project [5,11], and later on it became one of the typical algorithms for testing developed techniques for cryptanalysis. In this paper, we consider internal state recovery of LILI-128 with the following two goals: (i) to point out an illustration of a novel approach for cryptanalysis based on certain not yet considered features of Boolean functions, and (ii) to report an algorithm for internal state recovery of LILI-128 which is in a number of scenarios more efficient than the previously reported ones.

Description of LILI-128 can be found in [5], as well as in [4,7], and for the purpose of this paper it can be considered as follows. There are two subsystems in LILI-128: clock control and data generation ones. The clock control subsystem comprises of an *LFSR*, $LFSR_c$, of length 39 and a function, f_c, operating on the contents of the stages 12 and 20 of $LFSR_c$, denoted by x_{12} and x_{20}. The function f_c is defined as $f_c(x_{12}, x_{20}) = 2(x_{12}) + x_{20} + 1$. The sum is over the integers, therefore the output of f_c is $c(t) \in \{1, 2, 3, 4\}$. The data generation subsystem comprises of another *LFSR*, $LFSR_d$, of length 89 and a 10 variable Boolean function f_d. The function f_d is balanced with algebraic degree 3, nonlinearity 480 and correlation immunity of order 3. The algebraic immunity of this function is 4. The 10 inputs to f_d are taken from the positions $(0, 1, 3, 7, 12, 20, 30, 44, 65, 80)$. While generating the keystream, $LFSR_c$ is regularly clocked and the output of f_c is computed at each clock. The clock control subsystem advances the clock of $LFSR_d$ by the same value as the

* Corresponding author at: Mathematical Institute, Serbian Academy of Sciences and Arts, Kneza Mihaila 36, Belgrade, Serbia.
E-mail addresses: miodragm@mi.sanu.ac.rs (M.J. Mihaljević), gsugata@gmail.com (S. Gangopadhyay), goutam.paul@ieee.org (G. Paul).

0020-0190/$ – see front matter
http://dx.doi.org/10.1016/j.ipl.2012.07.013

output of the function f_c, that is by 1, 2, 3 or 4. Thus, on average $LFSR_d$ is advanced 2.5 clocks.

In the submission [5], the authors claimed that the complexity of any divide-and-conquer attack on LILI-128 is of at least 2^{112} operations. However, after the publication of the cipher, several more efficient attacks were successfully launched. Initially, two fast correlation attacks against LILI-128 have been reported in [8] and [7] and an improved one has been given in [9]. A time–memory trade-off based approach has been reported in [10]. The algebraic attacks have also been mounted against LILI-128, and the most efficient ones originate from in [4] and [6]. The state of the art regarding the algebraic attacks against LILI family has been reported in [12].

In this paper, we propose an attack on LILI-128 based on certain weakness of the filter function used in the keystream generation subsystem. The developed cryptanalysis is an alternative approach in comparison with the previously reported ones and provides additional options and suitability for certain attacking scenarios. The proposed approach follows the framework elements of time–memory–data trade-off (TMD-TO) over the decimated keystream [10] and the technique of BSW sampling [1], but it is substantially different from the both because it is based on certain feature of LILI-128 Boolean function for establishing an efficient enumeration of the special internal states and integrating it into the decimation based TMD-TO resulting in a significant gain.[1] The proposed algorithm is compared with the previously reported ones and its advantages are shown.

2. Theoretical framework for the cryptanalysis

2.1. A characteristic of the Boolean function

The ten-variable Boolean function f_d has the following algebraic normal form (ANF):

$$\begin{aligned}
&x_2 + x_3 + x_4 + x_5 + x_6x_7 + x_1x_8 + x_2x_8 + x_1x_9 + x_3x_9\\
&\quad + x_4x_{10} + x_6x_{10} + x_3x_7x_9 + x_4x_7x_9 + x_6x_7x_9 + x_3x_8x_9\\
&\quad + x_6x_8x_9 + x_4x_7x_{10} + x_5x_7x_{10} + x_6x_7x_{10} + x_3x_8x_{10}\\
&\quad + x_4x_8x_{10} + x_2x_9x_{10} + x_3x_9x_{10} + x_4x_9x_{10} + x_5x_9x_{10}\\
&\quad + x_3x_7x_8x_{10} + x_5x_7x_8x_{10} + x_2x_7x_9x_{10} + x_4x_7x_9x_{10}\\
&\quad + x_6x_7x_9x_{10} + x_1x_8x_9x_{10} + x_3x_8x_9x_{10} + x_4x_8x_9x_{10}\\
&\quad + x_6x_8x_9x_{10} + x_4x_6x_7x_9 + x_5x_6x_7x_9 + x_2x_7x_8x_9\\
&\quad + x_4x_7x_8x_9 + x_4x_6x_7x_9x_{10} + x_5x_6x_7x_9x_{10}\\
&\quad + x_3x_7x_8x_9x_{10} + x_4x_7x_8x_9x_{10} + x_4x_6x_7x_8x_9\\
&\quad + x_5x_6x_7x_8x_9 + x_4x_6x_7x_8x_9x_{10} + x_5x_6x_7x_8x_9x_{10}.
\end{aligned}$$

Our attack on LILI-128 is based on the observation that the function f_d is zero if $x_1 = x_2 = x_3 = x_4 = x_5 = x_6 = 0$, that is, $f_d(0,0,0,0,0,0,x_7,x_8,x_9,x_{10}) = 0$, for all $x_7, x_8, x_9, x_{10} \in \mathbb{F}_2$. Accordingly, the function $f_d(\cdot)$ is the $(k = 4)$-normal Boolean function of $n = 10$ variables (for more details on k-normal Boolean functions please refer to [2] and [3], for example). The inputs $x_1, x_2, x_3, x_4, x_5, x_6$ to the function f_d are obtained by tapping the positions $0, 1, 3, 7, 12, 20$ of $LFSR_d$, respectively.

The above is a particular example of the possibility that a Boolean function can be substantially modified (degraded) when a subset of its arguments take certain values. In the considered case, when certain variables are set to zero, the function is stuck at zero independently of all other variables.

2.2. A preliminary analysis

Let S be the transition matrix of $LFSR_d$. A sequence $\{c(t)\}_{t=0}^{m-1}$ of outputs of $LFSR_c$ is referred to as a *clocking sequence* of length m. Suppose that $\mathbf{X}_t = (X_0(t), \ldots, X_{88}(t))$ is the state of $LFSR_d$ at time t. Suppose $\mathbf{X}_0$ is the state of $LFSR_d$ after it is clocked according to the output $c(0)$. The subsequent states of $LFSR_d$ and the clocking sequence satisfy the following equations

$$\mathbf{X}_t = \mathbf{X}_{t-1}S^{c(t)}, \quad \text{for } t = 1, \ldots, m-1. \tag{1}$$

Let $S_j^{(\tau)}$ be the j-th column of the matrix S^τ, where τ is any integer. From (1), $\mathbf{X}_t = \mathbf{X}_0 S^{\beta_t} = \mathbf{X}_0(S_0^{(\beta_t)}, \ldots, S_j^{(\beta_t)}, \ldots, S_{88}^{(\beta_t)}) = (\mathbf{X}_0 S_0^{(\beta_t)}, \ldots, \mathbf{X}_0 S_j^{(\beta_t)}, \ldots, \mathbf{X}_0 S_{88}^{(\beta_t)})$, where $\beta_t = \sum_{i=1}^{t} c(i)$. At any time t, the inputs $(x_1, \ldots, x_{10})$ to the filter function f_d are as follows:

$$\begin{aligned}
&x_1 = X_0(t), \quad x_2 = X_1(t), \quad x_3 = X_3(t),\\
&x_4 = X_7(t), \quad x_5 = X_{12}(t), \quad x_6 = X_{20}(t),\\
&x_7 = X_{30}(t), \quad x_8 = X_{44}(t), \quad x_9 = X_{65}(t),\\
&x_{10} = X_{80}(t).
\end{aligned}$$

If $X_0(t) = X_1(t) = X_3(t) = X_7(t) = X_{12}(t) = X_{20}(t) = 0$, then the output of the function f_d is 0 irrespective of the values of $X_{30}(t)$, $X_{44}(t)$, $X_{65}(t)$ and $X_{80}(t)$.

On the other hand, note the following: when the index t is multiple of $2^{39} - 1$, we have (see [10], for example):

$$\begin{aligned}
&\big[X_0(t), X_1(t), \ldots, X_{88}(t)\big] = \mathbf{X}_0\mathbf{S}^{i(5\cdot 2^{38}-1)},\\
&\quad t = i\big(2^{39}-1\big),\ i = 1, 2, \ldots,
\end{aligned} \tag{2}$$

and accordingly the states $\mathbf{X}_t$, $t = i(2^{39} - 1)$, $i = 1, 2, \ldots$, are a function of $\mathbf{X}_0$ only (i.e. they depend only on the initial state $\mathbf{X}_0$ of $LFSR_d$).

Let $\mathcal{I}_0$ be the set of all states of $LFSR_d$ at certain time instance such that:

$$\begin{aligned}
&X_0(t) = X_1(t) = X_3(t) = X_7(t) = X_{12}(t) = X_{20}(t) = 0,\\
&\quad t = i\big(2^{39}-1\big),\ i = 0, \ldots, m-1,
\end{aligned} \tag{3}$$

and let a state belonging to $\mathcal{I}_0$ be considered as a realization of a vector random variable $\mathbf{x}$. The importance of the set $\mathcal{I}_0$ lies in the fact that if $\mathbf{x} \in \mathcal{I}_0$ is a state of $LFSR_d$ then the inputs $x_1, \ldots, x_6$ of the function f_d are 0 at times $t = i(2^{39} - 1)$, $i = 0, \ldots, m-1$, and they specify a system of $6m$ linear equations where unknowns are bits of the

[1] Note that the efficient enumeration of the suitable special states is the main issue regarding BSW sampling [1].