Two Alerts for Design of Certain Stream Ciphers: Trapped LFSR and Weak Resilient Function over GF(q)

Paul Camion[1], Miodrag J. Mihaljević[2,3], and Hideki Imai[4]

[1] Centre National de la Recherche Scientifique, Universite Pierre et Marie Curie, UMR 7090, Combinatoire, 175 rue du Chevaleret, 75013 Paris, France
paul.camion@wanadoo.fr
[2] SONY Corporation, 6-7-35 Kitashinagawa, Shinagawa-ku, Tokyo, 141-0001 Japan
[3] Mathematical Institute, Serbian Academy of Sciences and Arts, Kneza Mihaila 35, 11001 Belgrade, Yugoslavia
miodragm@turing.mi.sanu.ac.yu
[4] University of Tokyo, Institute of Industrial Science, 4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505, Japan
imai@iis.u-tokyo.ac.jp

Abstract. This paper points out: (i) a possibility for *malicious* selection of the LFSRs feedback polynomials in order to install a trap-door for the cryptanalysis; and (ii) a weakness of the construction of the resilient functions over GF(q) proposed at CRYPTO'96. Two corresponding methods for cryptanalysis are proposed. It is shown that although certain keystream generators over GF(q) are resistant against correlation and linear complexity based attacks, they are vulnerable by some novel attacks. The efficiency of these attacks depends on characteristics of the employed LFSRs and resilient functions. The developed attacks imply that LFSRs with certain characteristic polynomials and certain resilient functions are inappropriate as the building components for nonlinear combination generators and related schemes. They imply certain design criteria for employment of LFSRs and resilient functions over GF(q) in the nonlinear combination keystream generators and related schemes.

Keywords: linear feedback shift registers over GF(q), keystream generators, nonlinear combination generator, resilient functions, cryptanalysis.

1 Introduction

A number of the published keystream generators are based on linear feedback shift registers (LFSRs) assuming that parts of the secret key are used to load the LFSRs initial states (see [12], for example). Particularly, LFSRs over GF(q) appear as the interesting building blocks for certain keystream generators.

The unpredictability request, which is one of the main cryptographic requirements, implies that the linearity inherent in LFSRs should not be "visible" in the generator output. One general technique for destroying the linearity is to use

K. Nyberg and H. Heys (Eds.): SAC 2002, LNCS 2595, pp. 196–213, 2003.

several LFSRs which run in parallel, and to generate the keystream as a nonlinear function of the outputs of the component LFSRs. Such keystream generators are called nonlinear *combination generators*. In this paper we consider nonlinear combination generators where the only unknown elements to a cryptanalyst are the initial states of the LFSRs.

It has been shown that nonlinearity which prevent the linear complexity based attacks (see [11], for example) can not provide resistance against the correlation attack initially proposed in [19], improved by developing fast correlation attack method in [13], and further improved in a number of papers including the following most recent ones [5], [6], [14], [9], [16], [17] and [7]. A very important issue related to the combination function is that a trade-off exists between the correlation-immunity order and the nonlinearity order (see [20], for example).

In a general case the combination generators can be constructed over GF(q), $q > 2$. This assumes employment of LFSRs over GF(q), as well as combining functions over GF(q). Certain results related to this topic are reported in [8], [2], [3], [10], and [4], for example.

A particular class of the functions relevant for the combination generator are the *resilient functions*. f is t-resilient function over GF(q) if f is t-th order correlation-immune over GF(q) and balanced. A (n, m, t)-resilient function is an n-input m-output function f with the property that it runs through every possible output m-tuple an equal number of times when t arbitrary inputs are fixed and the remaining $n-t$ inputs runs through all q^{n-t} input tuples. Resilient functions are particularly appropriate for combining the outputs of linear feedback shift registers since such combination leads to pseudo-random generator which resists certain correlation attack. Knowing that a high correlation-immunity order is not sufficient for ensuring the security of the resulting generator the second important parameter, the nonlinearity order of the resilient function should be taken into account, as well, to ensure that linear-complexity attack can not be employed.

Motivation for the work.
Main intention of this paper is to point out certain issues which should be taken into account related to design of the nonlinear combination or filter like keystream generators over GF(q): Otherwise these schemes, although resistant on the all reported attacks could be vulnerable by the developed specialized attacks. Recall that the previously reported results on the considered topic were focused on the correlation and linear complexity issues. Accordingly, it appears interesting to consider if certain schemes, which are resistant on the correlation attacks and the linear-complexity attacks, are at the same time vulnerable by some other attacks. A goal of this paper is to point out the novel approaches for cryptanalysis of certain nonlinear combination keystream generators over GF(q), and to employ it in a context of the design criteria.

Our results.
This paper points out: (i) a possibility for *malicious* selection of the LFSRs feedback polynomials in order to install a trap-door for the cryptanalysis; (ii) a

weakness of the construction of the resilient functions over GF(q) proposed at CRYPTO'96.

Particularly, this paper presents two methods for cryptanalysis of certain nonlinear combination generators over GF(q). It is shown that the property of "resilience" is not sufficient for security, and a few concrete examples are given. Although, the previous is not entirely unexpected, the paper gives a precise support to the heuristic, and a particular importance of the results is that they show a way in which some previously proposed constructions can fail.

Certain characteristics of LFSRs over GF(q) based on the matrix characterization are discussed related to the behaviour of the different powers of the LFSR state transition matrix and it is pointed out the possibility of choosing the feedback polynomial, among all primitive polynomials of a given degree over $GF(q)$, which has a behaviour very far from the expected one, yielding the way for construction of a keystream generator with the trapped LFSRs which have the feedback polynomial with a hidden trap-door for a cryptanalysis. These characteristic are used for developing the algorithm (Algorithm A) for cryptanalysis of certain nonlinear combination keystream generators. These generators show all characteristics of a secure one but are however trapped and breakable. (As a trapped cryptographic component we assume the component which shows the characteristics of a secure one but however hides a previously unconsidered property that allows a cryptanalysis.) The derived characteristics and the developed algorithm for cryptanalysis imply a design criterion for employment of LFSRs over GF(q) in certain nonlinear combination keystream generators. The criterion requires check of the patterns distribution in a sequence of successive powers of the LFSR state transition matrix. Violation of the criterion can result in employment of an LFSR which is the trapped one so that the keystream is breakable when its very short output segment is available.

The second proposed methods for cryptanalysis (Algorithm B) points out weaknesses in the construction of the resilient functions over GF(q) with optimal nonlinearity order reported in [2]-[3]. It is shown that this method for construction opens a door for attacking the nonlinear combination generator which employs such resilient function based on the method for cryptanalysis proposed in this paper. The developed attack shows how a particular failure of the avalanche property of a function over GF(q) can be employed for the cryptanalysis, and accordingly the attack implies a particular condition necessary for the good avalanche characteristic of the combining function over GF(q).

The developed algorithms for cryptanalysis directly imply the design alerts and restrictions on LFSRs characteristic polynomials and resilient functions which are appropriate for nonlinear combination generators over GF(q) and the related schemes.

Organization of the paper.
Section 2 considers matrix characterization and certain characteristics of LFSRs over GF(q). The related security implications and a design criterion are pointed out in Section 3. The first algorithm for cryptanalysis, as well as an illustrative

example of cryptanalysis based on the proposed algorithm are given in Section 4. The second developed algorithm for cryptanalysis and its consequences, as well as an illustrative example, are discussed in Section 5. Finally, the conclusions are given in Section 6.

2 Matrix Characterization and Certain Characteristics of LFSRs over GF(q)

An LFSR can be considered as a linear finite state machine. Let us recall how a linear finite state machine is a realization or an implementation of certain linear operators. The characteristic polynomial or feedback polynomial of the LFSR is

$$b(u) = 1 + b_1 u + \ldots + b_L u^L \tag{1}$$

and the recursion implemented by the LFSR is then

$$X_{L+t} = -b_1 X_{L+t-1} - \ldots - b_L X_t = b_1 X_{L+t-1} + \ldots + b_L X_t, \tag{2}$$

since q here is a power of 2. The reader is referred to [18], for example, for more details.
When the LFSR feedback polynomial being given by (1), then the **state transition q-ary matrix A** can be written as:

$$\mathbf{A} = \begin{bmatrix} b_1 & b_2 & b_3 & \ldots & b_L \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & . \\ . & . & . & \ldots & . \\ 0 & & & \ldots 1 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \\ . \\ \mathbf{A}_L \end{bmatrix}, \tag{3}$$

where each $\mathbf{A}_i$, $i = 1, 2, ..., L$, represents a $1 \times L$ matrix (a row-vector). Having denoted by $\mathbf{X}_0$ the vector $[X_{L-1}, ..., X_0]$ representing the initial contents or initial state and by $\mathbf{X}_t = [X_{L+t-1}, ..., X_t]$ the L-dimensional vector over GF(q) representing the LFSR state after t clocks, then in the matrix form (2) writes

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1}^T = \mathbf{A}^t \mathbf{X}_0^T, \ t = 1, 2, \ldots, \tag{4}$$

where $\mathbf{A}^t$ is the t-th power over GF(q) of the $L \times L$ state transition binary matrix $\mathbf{A}$ and $\mathbf{X}^T$ denotes the transpose of the L-dimensional vector $\mathbf{X}$. Accordingly, a state of a length-L LFSR after t clocks is given by the matrix-vector product over GF(q) in relations (4).

Relations (4) show that powers of the matrix $\mathbf{A}$ determine algebraic replica of the LFSR initial state.

The next two parts of this section show two important characteristics relevant for specification of a design criterion for a nonlinear combination keystream generator over GF(q). At the first we show the behaviour of different powers

of the matrix $\mathbf{A}$ assuming a random primitive characteristic polynomial. In the next, we point out the possibility of choosing the characteristic polynomial, among all primitive polynomials of a given degree over $GF(q)$, which has a behaviour very far from the expected one, yielding a way for construction of the trapped keystream generator. These issues are the origins for specification of a design criterion for employment of LFSRs over GF(q) as the building blocks for keystream generators.

2.1 The Expected Behaviour

Proposition 2.1 *In a sequence of N different powers of the matrix $\mathbf{A}$, corresponding to an L-length LFSR with primitive characteristic polynomial, we can expect that Nq^{-L+B} of the vectors $\mathbf{A}_1^{(\cdot)}$, take a specific pattern of values (say all zero pattern) in certain $L-B$ coordinates, assuming arbitrary values in the remained B coordinates, $B < L$.*

Sketch of the Proof. Recall that an LFSR with a primitive characteristic polynomial of order L generates the sequences $\{X_i\}$ of period equal to $q^L - 1$, and note that since $X_{L+i} = \mathbf{A}_1^{(i)}\mathbf{X}_0$, where $\mathbf{X}_0$ is a vector constant, the sequences $\{X_i\}$ and $\{\mathbf{A}_1^{(i)}\}$ have the same period. Accordingly, the sequence $\mathbf{A}_1^{(i)}$, $i = 1, 2, ..., q^L - 1$, consists of all possible $q^L - 1$ different L-dimensional vectors over GF(q), except the all-zero vector. As a result, in a statistical model, any vector $\mathbf{A}_1^{(i)}$ could be considered as a realization of a random L-dimensional vector source of L-dimensional patterns over GF(q), and accordingly, we have the proposition statement.

Proposition 2.1 directly implies the following corollary.

Corollary 2.1. *In a sequence of N elements generated by an L-length LFSR over GF(q), we can expect that Nq^{-L+B} symbols depend only on certain $B < L$ symbols of the LFSR initial state assuming that the LFSR characteristic polynomial is a primitive one.*

As an illustration we consider the following particular example over GF(2^3). Let the coefficients of the primitive characteristic polynomial be $b_1 = 1$, $b_2 = 1$, $b_3 = \alpha$, $b_4 = 0$, $b_5 = 1 + \alpha^2$, where α is a root of $x^3 + x + 1$ over $GF(2)$. Here $L = 5$ and according to Corollary 2.1, for $B = 2$ and $N = 4000$ we can expect $4000/512 = 7.81$ occurrences of the i-th power $\mathbf{A}^i$ of the transition q-ary matrix $\mathbf{A}$ such that the first row of $\mathbf{A}^i$ has a 0 entry in any fixed three positions. When the first row of $\mathbf{A}$ is $[1, 1, \alpha, 0, 1 + \alpha^2]$, the first row of $\mathbf{A}^i$ for $i = 25$, 161, 197, 1007, 1510, 2565 and 3910, respectively, has the following forms:
$[0, 0, 0, 1 + \alpha^2, \alpha]$, $[0, 0, 0, \alpha^2, \alpha]$, $[0, 0, 0, 1, \alpha]$, $[0, 0, 0, \alpha, \alpha + \alpha^2]$, $[0, 0, 0, 1 + \alpha + \alpha^2, 1 + \alpha]$, $[0, 0, 0, \alpha^2, \alpha^2]$, and $[0, 0, 0, 1 + \alpha, 1]$.

Accordingly, the LFSR output sequence symbols on the positions $i = 25$, 161, 197, 1007, 1510, 2565 and 3910, depend only on two symbols of the LFSR initial state.

2.2 The Probability of a High Deviation from the Expected Behaviour

Proposition 2.2 *In a set of R different LFSRs over GF(q) of length L and primitive characteristic polynomial, the probability P that an LFSR exits such that at least M from its first N output symbols depends only on certain $B < L$ elements of the LFSR initial state is given by the following:*

$$P = 1 - [\sum_{m=0}^{M-1} P(m)]^R \,, \tag{5}$$

where

$$P(m) = \binom{q^B}{m} \binom{q^L - 1 - q^B}{N - m} / \binom{q^L - 1}{N} \,. \tag{6}$$

Sketch of the Proof. We consider the set of all primitive polynomial of degree L over $GF(q)$ and the corresponding roots over $GF(q^L)$. Any power α^i of such a root α can be expressed as $\alpha^i = \sum_{j=0}^{L-1} a_j \alpha^j$. It can be directly shown that the probability that exactly m patterns, such that $a_j = 0$ for any j out of certain set of B indices, appear in the first N powers of α is given by (6). On the other hand, $Pr = \sum_{i=0}^{M-1} P(i)$ is the probability that at most $M - 1$ powers of α with the specified pattern appear in the first N powers of α. Accordingly, the probability for the desired root is $1 - Pr^R$ where R is the number of primitive polynomials of degree L over $GF(q)$ that we consider.

The statement of Proposition 2.2 is illustrated by the following example. We consider the set of all primitive polynomial of degree $L = 6$ over $GF(q = 2^7)$ and the corresponding roots over $GF(q^L)$. Any power α^i of such a root α can be expressed as $\alpha^i = \sum_{j=0}^{5} a_j \alpha^j$. We are interested of the probability to have in the first N powers of the primitive root a subset I of elements such that $a_5 = a_4 = 0$ in the expression of $\alpha^\ell, \ell \in I$. The length of the generator output sequence is assumed to be N, which will here range from 10 to 50, maybe somewhat larger but not more than 100. Then $P = \sum_{i=0}^{2} P(i)$ is the probability that at most two powers of α with the specified pattern appear in the first N powers of α. The probability for the desired root is $1 - P^{R_N}$ where R_N is the number of primitive polynomials of degree 6 over $GF(q)$ that we need to go through for $1 - P^{R_N}$ to be at least 99%. Here the total number R of primitive polynomials is the number of integers prime to $128^6 - 1$ divided by 6. Then $R = 404620054272 \simeq 4 \cdot 10^{11}$. Since $P \simeq .999999995952$, then for $N = 50$ we need R_N to be $1.2\ 10^9$ and the probability for the desired root is $1 - P^{R_N} \simeq .992$. For $N = 10$ we need R_N to be $1.9 \cdot 10^{11}$ and the probability for the desired root is $1 - P^{R_N} \simeq .99$.

3 Security Implications and Design Criterion

Proposition 2.1 and Corollary 2.1 open a door for a part-by-part recovery of a LFSR initial state employed in certain keystream generators. The critical observation is that if you can identify some output locations that depend on only a subset of the key, then you can speed-up exhaustive keysearch by guessing only that subset of the key.

Proposition 2.2 opens a door for a *malicious* selection of the characteristic polynomial so that the LFSR initial state recovery, in certain keystream generators, can be performed assuming availability of very short keystream output sequence. This malicious selection of the LFSRs feedback polynomials yields a possibility to install a trap-door for the cryptanalysis in certain keystream generators.

Accordingly, this section proposes the following design criterion for employment of LFSRs over GF(q) in order to prevent the cryptanalysis using only a very short segment of the output keystream sequence.

– Check the patterns distribution of the first N powers of the LFSR state transition matrix, and do not employ the considered candidate if the distribution significantly deviates from the expected one given by Proposition 2.1.

Violation of the proposed rule for selection of the LFSR characteristic polynomial can result in a scheme vulnerable based on very short segments from the keystream generator.

4 An Algorithm for Cryptanalysis and Its Characteristics

As an illustration of the statements given in the previous section, this section proposes a particular approach for the cryptanalysis.

We assume a nonlinear combination keystream generator over GF(q) which consists of two LFSRs and a 1-resilient nonlinear function $f(\cdot)$ which inverse is $f^{-1}(\cdot)$. The lengths of the LFSRs, $L1$ and $L2$, are co-prime, $L1 < L2$, and the secret key determines only the LFSRs initial states. Accordingly the considered keystream generator is resistant on the correlation attack and the linear complexity attack.

4.1 Underlying Ideas

The developed technique for cryptanalysis is based on the **divide and conquer** approach. This approach is widely used in cryptanalysis, and particularly it has been employed for initial state reconstruction of an LFSR as it is reported in [1], [6] and [9], for example, but these techniques are inappropriate in the cases under our consideration due to the following. The technique reported in [1] requires that the LFSR feedback taps are highly concentrated, and so it is inefficient or can not work in the general cases under our consideration. The techniques reported in [6], [9], [14], [17] and [7], are based on certain sums of the LFSR

output symbols so that the relevant inversions can not be performed. In this paper we propose a novel divide and conquer approach which is suitable for the problem under our consideration.

Two main underlying ideas for the cryptanalytic method are the following:
- recovering of complete secret key by independent recovering of its certain parts;
- employment of a technique based on the linear finite state machine model of an LFSR over GF(q) for autonomous reconstruction of the secret key parts using appropriate subsequence of an LFSR output sequence which depend on a part of the LFSR initial state only.

Characterization of an LFSR over GF(q) by the state transition matrix, and the characteristics of a sequence of powers of the state transition matrix, derived in Section 2, are the main origins for construction of the algorithm for the cryptanalysis.

4.2 Algorithm A

– *INPUT.*
Output sequence from the keystream generator, $\{Z_i\}_{i=1}^{N}$, and parameter ℓ.
– *PREPROCESSING.*
Denote by $\mathbf{Aj}$ the state transition matrix of LFSRj, $j = 1, 2$, and by $\mathbf{Aj}_1^i$ the first row of the i-th power $\mathbf{Aj}^i$ of $\mathbf{Aj}$. Let $P_{(\ell)}$ be a subset of $\{1, 2, ..., L1\}$, and let N be large enough for having a set $S_{(\ell)}$ of indices with the size $L2-\ell$ such that:
- the first row of $\mathbf{A1}^i$, $i \in S_{(\ell)}$, has nonzero components only for the indices which belong to $P_{(\ell)}$;
- the vectors $\mathbf{A2}_1^i$, $i \in S_{(\ell)}$ are linearly independent.

– *PROCESSING.*
1. Suppose previously unconsidered hypothesis about the initial state part $\mathbf{X1}_{(0,\ell)}$ which components belong to the set $P_{(\ell)}$, and generate $L2 - \ell$ output elements of $LFSR1$, $X1_i$, $i \in S_\ell$ according to the following:
$$X1_i = \mathbf{A1}_1^i (\mathbf{X1}_{(0,\ell)})^T \ , \quad i \in S_{(\ell)} \ .$$
2. For each $(X1_i, Z_i)$, $i \in \{P_{(\ell)}, S_{(\ell)}\}$, calculate
$$X2_i = f^{-1}(X1_i, Z_i) \ , \ i \in \{P_{(\ell)}, S_{(\ell)}\} \ ,$$
and solve the system
$$\mathbf{A2}_1^i (\mathbf{X2}_0)^T = X2_i \ , \ i \in S_{(\ell)} \ ,$$
for $\mathbf{X2}_0$ where it is a candidate for the initial state of LFSR2.
3. Based on the current candidate for $\mathbf{X2}_0$ and $\{Z_i\}$, recover complete candidate for $\mathbf{X1}_0$.
4. For the established LFSRs states calculate the generator output sequence $\tilde{Z}_i$, $i = 1, 2, ..., L1 + L2$.
If $\tilde{Z}_i = Z_i$, $i = 1, 2, ..., L1 + L2$, accept current hypothesis of the LFSR1 and LFSR2, as the true one; otherwise go to the step 1.

– *OUTPUT.*
Secret key recovered from the reconstructed initial states.

4.3 Complexity and Required Input

The structure of Algorithm A directly implies the following statement.

Corollary 4.1. *The time complexity of Algorithm A is proportional to* q^{ℓ}, $\ell < L1$, *and it is smaller for a factor equal to* $2^{L1-\ell}$ *than a brute force attack,* $L2 > L1 > \ell$.

The required sample for the cryptanalysis strongly depends on the characteristic polynomial of employed LFSRs.
The algorithm preprocessing phase requires enough long output sequence from the generator to make possible collection of the suitable powers of the state transition matrix. Also note that, according to Proposition 2.2, the required length can be extremely small if an LFSR with certain characteristic polynomial is employed.
Table 1 gives a numerical illustration of the algorithm time complexity, according to Corollary 4.1, and assuming operations over GF(2^3).

Table 1. Time complexity and required imput sample of the algorithm for cryptanalysis as the functions of the algorithm parameters L_1, L_2 and method of the characteristic polynomial selection assuming the combination generator over GF(2^3) with two LFSRs and 1-resilient function.

secret key length (in bits)	LFSR lengths $L2$, $L1$ and the parameter ℓ (over GF(8))	order of time complexity	order of required sample dependent on selection of the characteristic polynomial	
			random	malicious
63	11, 10, 5	8^5	8^5	100
96	17, 15, 8	8^8	8^8	100
156	27, 25, 15	8^{15}	8^{15}	1000
183	31, 30, 15	8^{15}	8^{15}	1000

4.4 Illustrative Example

Key Ideas

Our aim is to conceive a nonlinear combination keystream generator over GF(q) with two LFSRs which shows all characteristics of a secure one but is however trapped and breakable. The key statements are as follows.

1. The nonlinear combining function is 1-resilient with maximal non-linearity order, and it combines LFSR1 and LFSR2 over GF(2^7) of lengths 6 and 7,

respectively. We have chosen $GF(2^7)$ in order to produce ASCII symbols which are bytes as generator output symbols.

2. We first choose the feedback polynomial from the

$$\phi(2^{42}-1)/6 = 404620054272 \simeq 4 \cdot 10^{11}$$

ones with the following property: Among the first 50 powers of the corresponding primitive root α a set I of at least 3 have the same pattern with 4 nonzero components and 2 zero components in its expression with 6 components over the ground field. It is established that the probability is larger than 99.2% so that a primitive polynomial for $LFSR1$ exists among $1.2 \cdot 10^9$ of them with the desired property. What we have in mind is to complete the crytanalysis with the knowledge of only 50 bytes of the generator output sequence. It is still feasible to find a primitive polynomial with the desired property such that among the first 10 powers of the corresponding primitive root α a set I of at least 3 have the same pattern with 4 nonzero components and 2 zero components in its expression with 6 components over the ground field. Here the probability of finding such a polynomial still is close to 99% but at the condition however to be able to go through $1.9 \cdot 10^{11}$ primitive polynomials, and this would allow to complete the cryptanalysis with only 10 bytes of the generator output sequence, corresponds to an attack with a known plaintext of only 10 characters.

3. Let I be the set of three powers of the transition matrix **A1** of LFSR1 for which the first row of $\mathbf{A1}^{\ell}$, $\ell \in I$ has 0 in the first two positions. For LFSR2 we need to choose independently a feedback polynomial such that the first rows of $\mathbf{A2}^{\ell}$, $\ell \in I$ together with the first four unit vectors of length 6 forms of a set of linearly independent vectors. This reduces to considering the probability that a 3×3 matrix over $GF(128)$ be invertible. That probability is $\simeq .99126$. Thus, such a feedback polynomial can be easily found.

Example of a Vulnerable 1-Resilient Function

We give an explicit form of the inversion of f when it is given as in [2, Proof of Proposition 14] by

$$f(x,y) = (x^{q-2} + y^{\frac{q-2}{2}} + 1)^{q-5}, \tag{7}$$

where $q > 4$ is a power of 2. This is an example of a 1-resilient function with optimal nonlinearity order over a finite field GF(q), q a power of 2.

It can be shown that the following statement holds.

Proposition 1. *Over the finite field $GF(q)$, $q > 2$, q a power of 2, if $z = f(x,y) = (x^{q-2} + y^{\frac{q-2}{2}} + 1)^{q-5}$, then $y = (z^{q-2} + x^{4(q-2)} + 1)^{\frac{q-2}{2}}$.*

Details of the Attack

The LFSRi state transition q-ary matrix is denoted by $\mathbf{Ai}$, $i = 1, 2$. The unknown initial state of $LFSR1$ is $[x_5, x_4, x_3, x_2, x_1, x_0]$ and that of $LFSR2$ is $[y_6, y_5, y_4, y_3, y_2, y_1, y_0]$. We recall that we have the feedback polynomial of LFSR1 such that there is a set I of size 3 of integers and that the first row of $\mathbf{A1}^\ell,\ \ell \in I$ has 0 for its first two entries. Our aim is to reduce the number of candidates to be checked for the whole initial state: $[x_5, x_4, x_3, x_2, x_1, x_0], [y_6, y_5, y_4, y_3, y_2, y_1, y_0]$ to at most q^4. By considering the $128^4 = 2^{28}$ possible partial initial states $[x_3, x_2, x_1, x_0]$ of LFSR1, we are able to recover the complete initial states of LFSR1 and LFSR2 as follows. We first have the set of equations

$$y_\ell = g(x_\ell, z_\ell), \ell = \ 3, 2, 1, 0. \tag{8}$$

By those equations every choice of $[x_3, x_2, x_1, x_0]$ forces the partial initialization $[y_3, y_2, y_1, y_0]$. By our assumption on the matrices $\mathbf{A1}^\ell,\ \ell \in I$, we have that

$$x_\ell = \mathbf{A1}^\ell [0, 0, x_3, x_2, x_1, x_0]^T,\ \ell \in I \tag{9}$$

We are then able to solve for y_ℓ the equations

$$y_\ell = g(x_\ell, z_\ell), \ell \in I \tag{10}$$

Let us denote by $[e_1], [e_2], [e_3], [e_4]$ the row vectors $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0)$, $(0, 0, 1, 0, \ldots, 0)$, $(0, 0, 0, 1, 0, \ldots, 0)$, respectively. We then denote by $[\mathbf{A2}^\ell],\ \ell \in I$ the first row of the matrix $\mathbf{A2}^\ell,\ \ell \in I$.

We finally have the set of equations

$$[e_{\ell+1}][y_6, y_5, y_4, y_3, y_2, y_1, y_0]^T = y_\ell,\ \ell = 0, 1, 2, 3. \tag{11}$$

and

$$\mathbf{A2}^\ell [y_6, y_5, y_4, y_3, y_2, y_1, y_0]^T = y_\ell,\ \ell \in I \tag{12}$$

We have chosen the feedback polynomial of LFSR2 in order that the vectors $[e_\ell],\ \ell = 1, 2, 3, 4$, and $\mathbf{A2}^\ell,\ \ell \in I$ are linearly independent. We thus can solve (11) and (12) for $y_6, y_5, y_4, y_3, y_2, y_1, y_0$. The complete initial state of LFSR2 is then entirely recovered though only the assumed partial initial state $[x_3, x_2, x_1, x_0]$ of LFSR1. Now we can solve

$$x_\ell = h(y_\ell),\ \ell = 4, 5.$$

What we have just seen is that exploiting the property of the function $f(x_\ell, y_\ell) = z_\ell$ and the fact that we can choose the feedback polynomial, first for LFSR1 and then independently for LFSR2, so that every partial initialization x_3, x_2, x_1, x_0 of LFSR1 completely determines the whole initialization of both LFSR. We are then able to produce the keystream sequence corresponding to that initialization which eventually will match the keystream sequence used. Cryptanalysis will be achieved in an average of $\frac{q^4}{2} = 2^{27}$ steps with negligible memory space, whereas a straightforward brute force attack requires $\frac{q^7}{2} = 2^{48}$ steps.

5 A Weak Resilient Function over GF(q)

This section points out vulnerability of certain resilient functions over GF(q). The developed method for cryptanalysis applies when some input variables x of the combining function over GF(q) only appear as x^{q-1} in the algebraic normal form. A particular class of the functions which have certain optimal characteristics but fulfill this condition, as well, is proposed at CRYPTO'96, [2].

5.1 Keystream Generator Under Consideration

We assume that the nonlinear combination generator consists of the following:
• n LFSRs over GF(q): LFSR1, LFSR2, ..., LFSRn, with lengths L_1, L_2, ..., L_n, respectively, and known characteristic polynomials;
• 1-resilient function of n variables with the optimal nonlinearity order constructed as it is proposed in [2] employing the following Proposition 2 and (7);
• and that the secret key determine the LFSRs initial states only.

Proposition 2. *(Proposition 15, [2]) Let $q \neq 2$ or $t \neq n-1$. Let f_1, $f_2 : GF(q)^n \to GF(q)$ be two t-resilient functions with optimal nonlinearity, such that $degree(f_1 - f_2) = degree(f_1)$. Then $g : GF(q)^{n+1} \to GF(q)$ defined by*

$$g(x_1, \ldots, x_{n+1}) = x_{n+1}^{q-1} f_1(x_1, \ldots, x_n) + (1 - x_{n+1}^{q-1}) f_2(x_1, \ldots, x_n) \qquad (13)$$

is a t-resilient function with optimal nonlinearity.

Finally, note the following. By Proposition 14 of [2] there exists a t-resilient function $f : GF(q)^{t+1} \to GF(q)$ with optimal nonlinearity order. Applying Proposition 2 with $f_1 = f$ and $f_2 = \alpha f$, where $\alpha \in GF(q) \setminus \{0, 1\}$ leads to a t-resilient function with $t+2$ variables and optimal nonlinearity order. If we iterate this construction $n + t - 2$ times, we obtain a t-resilient function with n variables and optimal nonlinearity order. Siegenthaler [20] proved this result in Boolean case.

5.2 Underlying Ideas for Cryptanalysis

The algebraic normal form of the considered t-resilient function with optimal nonlinearity order has a monomial (with highest possible degree) that should have t variables with degree $q-2$ and all other variables with degree $q-1$. This is naturally brought to the attention of the cryptanalyst since a variable x raised to the power $q-1$ takes the value 1 for every nonzero x.
The following statement points out an underlying result relevant for the cryptanalysis.

Proposition 3. *The probability $P(n,t)$ of at least t consecutive ones in a length n string of zeros and ones with independent probabilities p for zero and $1-p=q$ for one is given by*

$$P(n,t) = 0 \text{ for } n < t,\ q^t \text{ for } n = t \text{ and}$$

$$P(n,t) = q^t + \sum_{i=2}^{n-t+1} pq^t(1 - P(i-2,t)) \text{ for } n > t.$$

Sketch of the Proof. The probability of at least t consecutive *ones* starting in the first position is q^t. Next the probability of at least t consecutive *ones* appearing for the first time with the first one in the i-th position, $i > 1$ is $pq^t - P(i-2,t)pq^t$. Such a string has a *zero* in the $(i-1)$-th position and there is no t consecutive ones in the first $i-2$ positions.

5.3 Algorithm for Cryptanalysis

Algorithm B

- *INPUT.*
  Characteristic polynomials of the employed n LFSRs of lengths $L1, L2, ..., Ln$, assuming (without loosing generality) that $L1 < L2 < ... < Ln$, the combining function, assuming that the outputs from LFSR3-LFSRn appear only as x^{q-1} in the algebraic normal form of the combining function, and an output sequence from the keystream generator, $\{Z_i\}_{i=1}^{I}$,
- *PROCESSING.*
  Joint recovering of the LFSR1-LFSR2 initial states, and than sequential recovering of the LFSR3-LFSRn initial states.
  1. Recovering of the LFSR1-LFSR2 initial states.
     For each output sequence index $i = 1, 2, ..., I1$, where $I1 < I - (L1 + L2)$, do the following:
     (a) Assume that $L1 + L2$ successive output bits from each of LFSR3 to LFSRn are nonzero, implying that the keystream generator output depends only on the output sequences from LFSR1 and LFSR2;
     (b) Employ the following procedure for recovering of the initial states of LFSR1 and LFSR2.
        i. Suppose a previously unconsidered initial state $X_1, X_2, ..., X_{L1}$, of $L1$-length LFSR1 which generates the sequence X_{L1+1}, X_{L1+2}, $..., X_K$, $K < I$.
        ii. Using the sequences X_{L1+1}, X_{L1+2}, ..., X_{L1+L2} and $\{Z_i\}$ calculate $Y_k = f^{-1}(X_k, Z_k)$, $k = L1 + 1, L1 + 2, ..., L1 + L2$, where f^{-1} denotes inversion of f, and generate Y_{L2+1}, Y_{L2+2}, ..., Y_K.
        iii. Calculate $\tilde{Z}_k = f(X_k, Y_k)$, $k = L2+1, L2+2, ..., K$, where $K \leq I$ provides an unique solution in the next step test.
        If $\tilde{Z}_k = Z_k$, $k = L2 + 1, L2 + 2, ..., K$, go to Step iv; otherwise go to Step i.

iv. Memorize the recovered initial states of LFSR1 and LFSR2. (i.e. the corresponding secret key parts), and go to the next step.

2. Recovering of the LFSR3-LFSRn initial states based on the recovered initial states of LFSR1-LFSR2.
   For $j = 3, 4, ..., n$, based on the recovered initial states LFSR1-LFSR$(j-1)$, reconstruct LFSRj employing the following:
   (a) assume previously unconsidered initial state of LFSRj
   (b) generate the output sequence $\{\tilde{Z}_i\}$ based on the recovered initial states of LFSR1-LFSRj, assuming that the outputs of LFSR$(j+1)$-LFSRn are always nonzero
   (c) Calculate $d = \sum_{i=1}^{I} \delta(\tilde{Z}_i, Z_i)$ where $\delta(\tilde{Z}_i, Z_i) = 0$ for $\tilde{Z}_i = Z_i$ and $\delta(\tilde{Z}_i, Z_i) = 1$ for $\tilde{Z}_i \neq Z_i$, assuming that the algorithm parameter I provides the relevant separability.
   (d) if d is smaller than a certain threshold accept the current hypothesis as the true one; otherwise go to Step 2.(a).

– *OUTPUT.*
Secret key recovered from the reconstructed initial states.

5.4 Required Sample and Complexity of Algorithm B

According to Algorithm B structure, it can be directly shown that the following is valid.

When available keystream output sequence is $O(256^2(\frac{256}{255})^{n-3})$, assuming $n \geq 3$, the complexity of attack is $O(q^{L1} + \sum_{j=3}^{n} q^{Lj})$.

Recall that a direct exhaustive search over all possible secret keys has complexity $O(q^{\sum_{j=1}^{n} Lj})$.

5.5 Illustrative Example: A Nonlinear Combination Generator with Five LFSRs

We assume that the generator under cryptanalysis consists of:
- 5 LFSRs, LFSR1, LFSR2, LFSR3, LFSR4, LFSR5, of respective lengths $L1 = 7$, $L2 = 11$, $L3 = 5$, $L4 = 9$, $L5 = 8$, over GF(256) with primitive characteristic polynomials;
- A known 1-resilient function of 5 variables $(x_1, x_2, x_3, x_4, x_5)$, corresponding to the LFSR1, LFSR2, LFSR3, LFSR4, LFSR5. It has been constructed according to the following: a 1−resilient function $f_2(x_1, x_2)$ defined by (7) is first obtained. Then three more variables x_3, x_4, x_5 are introduced in three steps with the recurrence (13).

Also, for the first step, we assume that 174 bytes of keystream are at our disposal. Under this condition, the attack succeeds with probability 0.97. If enough long sequence is available, the reconstruction can be completed.

The attack consists of the following two distinct parts:
• Recovering the initial states of LFSR1 and LFSR2 with probability 0.97, based on a search over no more than 8^7 hypothesis;

• Assuming that a long sequence of the generator output (or very long sequence of the ciphertext) is available, recovering the initial states of LFSR3, LFSR4 and LFSR5 based on three successive independent search which employ testing of 8^5, 8^9 and 8^8 hypothesis, respectively.

Starting steps of the algorithm for complete cryptanalysis are the following:
- Assume that the values of x_3, x_4, x_5 all are nonzero for 18 successive bytes of the keystream by taking those bytes starting from the first, the second, ..., and the 156-th (156=174-18) byte.
- For each of the previous assumptions that the values of x_3, x_4, x_5 are nonzero for 18 successive bytes perform the cryptanalysis on what is now essentially function $f(x_1, x_2)$ defined by (7).

Considering relation (13), we see that the probability of a successful attack is that of having simultaneously 18 ones in three strings of length 174 of *zeros* and *ones*, where each *zero* appears with independent probability $p = \frac{1}{256}$ and *one* with probability $q = \frac{255}{256}$.
When the initial states of LFSR1 and LFSR2 are recovered, long segments of the keystream sequence can be generated, since, for any t, the probability that the output byte of the generator at time t is equal to $f_2(x_1, x_2)$ is $(\frac{255}{256})^3 = 0.9883$. If a sufficiently long generator output (or even ciphertext only) is at disposal it is possible to recover the initial state of LFSR3, LFSR4 and LFSR5 using the following approach.
- Suppose previously unconsidered initial state of LFSR3. If the assumption is correct, than, for any t, the probability that the generator output byte at time t is defined by $f_2(x_1, x_2)$ is equal to $(\frac{255}{256})^2 = 0.9922$. Since 9922 is 0.39% greater than 9883, the generated sequence will be more similar to the given one, so that a discrimination between incorrect and correct hypothesis is possible.
- The analog techniques can be employed for the initial states recovering of LFSR4 and LFSR5.

6 Conclusions

This paper yields two alerts relevant for nonlinear combination keystream generators over GF(q) and the related schemes, and it points out: (i) a possibility for *malicious* selection of the LFSRs feedback polynomials in order to install a trap-door for the cryptanalysis; (ii) a weakness of the construction of the resilient functions over GF(q) proposed in [2]-[3]. On the other hand the results are novel general elements for design of keystream generators over GF(q). More precisely, the results of this paper can be summarized as follows.

Certain characteristics of LFSRs over GF(q) based on the matrix characterization are discussed related to the behaviour of the different powers of the LFSR state transition matrix and it is pointed out the possibility of choosing the feedback polynomial, among all primitive polynomials of a given degree over $GF(q)$, which has a behaviour very far from the expected one, yielding the way for construction of the trapped keystream generator. These characteristic are used for developing the algorithm, Algorithm A, for cryptanalysis of certain nonlinear

combination keystream generators. These generators show all characteristics of a secure one but are however trapped and breakable. The derived characteristics and the developed algorithm for cryptanalysis imply a design criterion for employment of LFSRs over GF(q) in certain nonlinear combination keystream generators. The criterion requires checking of the patterns distribution in a sequence of successive powers of the LFSR state transition matrix. Violation of the criterion can result in employment of an LFSR which is the trapped one so that the keystream is breakable when a very short output segment is available.

The developed Algorithm A is an illustration of significance of the proposed criterion, and an efficient algorithm for cryptanalysis of the nonlinear combination generator with two LFSRs and 1-resilient function. The critical observation is that if you can identify some output locations that depend on only a subset of the key, then you can speed-up exhaustive keysearch by guessing only that subset of the key. The developed algorithm has divide-and-conquer nature, and recovering of complete secret key is obtained by independent reconstruction of its parts employing a technique based on the linear finite state machine model of an LFSR over GF(q), and identification of certain subsequence of an LFSR output sequence. Complexity of the cryptanalysis and required length of the keystream sequence are discussed. When two LFSRs have co-prime lengths $L_1 < L_2$ and the primitive characteristic polynomials, the complexity of attack is $O(q^\ell)$, $\ell < L_1$. Particularly, it is pointed out that for certain LFSRs characteristic polynomials, the attack can be performed on very short generator output segments, and such LFSRs are called trapped LFSRs.

The second proposed methods for cryptanalysis, Algorithm B, points out weaknesses in the construction of the resilient functions over GF(q) with optimal nonlinearity order reported in [2]-[3]. It is shown that this method for construction opens a door for attacking the nonlinear combination generator which employs such resilient function based on the method for cryptanalysis proposed in this paper. The developed attack shows how a particular failure of the avalanche property of a function over GF(q) can be employed for the cryptanalysis, and accordingly the attack implies a particular condition necessary for the good avalanche characteristic of the combining function over GF(q).

Algorithm B, the second algorithm for cryptanalysis proposed in this paper, applies when some input variables x of the combining function appear only as x^{q-1} in the algebraic normal form. In that case the function is very close to a function of less input variables (since x^{q-1} takes the value 1 for all non-zero values of x). In [2] a family of resilient functions achieving Siegenthaler's bound was constructed, and these functions have this particular algebraic property: any such n-variable t-resilient function over GF(q) is at low distance from a t-variable function. The second developed algorithm points out that the family of resilient functions proposed in [2] (CRYPTO'96) is not suitable for cryptographic applications even if these functions have the highest possible degree. Most notably, the second algorithm implies a new criterion on the algebraic normal form of the combining function which is specific to q-ary case. In other words, the proposed attack shows how a particular failure of the avalanche property of a function

over GF(q) can be employed for the cryptanalysis, and accordingly the attack implies a particular condition necessary for the good avalanche characteristic of the combining function over GF(q).

When n LFSRs over GF(q) have lengths $L_1 < L_2 < ... < L_n$, and the outputs from LFSR3-LFSRn appear as x^{q-1} in the algebraic normal form, the following is shown: When available keystream output sequence is $O(256^2(\frac{256}{255})^{n-3})$, assuming $n \geq 3$, the complexity of attack is $O(q^{L_1} + \sum_{j=3}^{n} q^{L_j})$, while a direct exhaustive search over all possible secret keys has complexity $O(q^{\sum_{j=1}^{n} L_j})$. In certain cases, and particularly if the LFSRs are the trapped ones, the first algorithm for cryptanalysis can be combined with the second one.

The developed algorithms for cryptanalysis directly imply the design alerts and restrictions on LFSRs characteristic polynomials and resilient functions which are appropriate for nonlinear combination generators over GF(q) and the related schemes. Accordingly, the results of this paper are focused toward some particular problems and, at the same time, they are general guidelines for construction of certain stream ciphers.

Finally note that archiving hidden and nontrivial weaknesses of certain building blocks for keystream generators is an important issue relevant for construction of the secure schemes, as well as for the security evaluation of the proposed ones.

References

1. R. J. Anderson, "A faster attack on certain stream ciphers", *Electronics Letters*, vol. 29, pp. 1322-1323, 22nd July 1993.
2. P. Camion and A. Canteaut, "Generalization of Siegenthaler inequality and Schnorr-Vaudenay multipermutations", Advance in Cryptology - CRYPTO'96, *Lecture Notes in Computer Science*, vol. 1109, pp. 372-386, 1996.
3. P. Camion and A. Canteaut, "Correlation-immune and resilient functions over a finite alphabet and their applications in cryptography", *Design, Codes and Cryptography*, vol 16, pp.103- 116, 1999.
4. P. Camion, M. J. Mihaljević and H. Imai, "On employment of LFSRs over GF(q) in certain stream ciphers", *IEEE Int. Symp. Inform. Theory - ISIT2002*, Lausanne, Switzerland, July 2002, Proceedings, p. 210.
5. A. Canteaut and M. Trabbia, "Improved fast correlation attacks using parity-check equations of weight 4 and 5", Advances in Cryptology - EUROCRYPT 2000, *Lecture Notes in Computer Science*, vol. 1807, pp. 573-588, 2000.
6. V.V. Chepyzhov, T. Johansson and B. Smeets, "A simple algorithm for fast correlation attacks on stream ciphers", Fast Software Encryption 2000, *Lecture Notes in Computer Science*, vol. 1978, pp. 180-195, 2001.
7. P. Chose, A. Joux and M. Mitton, "Fast correlation attacks: An algorithmic point of view", Advances in Cryptology - EUROCRYPT 2002, *Lecture Notes in Computer Science*, vol. 2332, pp. 209-221, 2002.
8. J. Dj. Golić, "On linear complexity of functions of periodic GF(q) sequences", *IEEE Trans. Inform. Theory*, vol. 35, pp. 69-75, Jan. 1989.
9. T. Johansson and F. Jonsson, "Fast correlation attacks through reconstruction of linear polynomials", Advances in Cryptology - CRYPTO 2000, *Lecture Notes in Computer Science*, vol. 1880, pp. 300-315, 2000.

10. F. Jonsson and T. Johansson, "Correlation attacks on stream ciphers over GF(2^n)", *2001 IEEE Int. Symp. Inform. Theory - ISIT2001*, Washington DC, June 2001, Proceedings, p. 140.
11. J. L. Massey, "Shift-register synthesis and BCH decoding", *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 122-127, 1969.
12. A. Menezes, P.C. van Oorschot and S.A. Vanstone, *Handbook of Applied Cryptography*. Boca Raton: CRC Press, 1997.
13. W. Meier and O. Staffelbach, "Fast correlation attacks on certain stream ciphers," *Journal of Cryptology*, vol. 1, pp. 159-176, 1989.
14. M. J. Mihaljević, M. P. C. Fossorier and H. Imai, "A low-complexity and high-performance algorithm for the fast correlation attack", Fast Software Encryption - FSE 2000, *Lecture Notes in Computer Science*, vol. 1978, pp. 196-212, 2001.
15. M. J. Mihaljević and J. Golić, "A method for convergence analysis of iterative probabilistic decoding", *IEEE Trans. Inform. Theory*, vol. 46, pp. 2206-2211, Sept. 2000.
16. M. J. Mihaljević, M. P. C. Fossorier and H. Imai, "On decoding techniques for cryptanalysis of certain encryption algorithms", *IEICE Trans. Fundamentals*, vol. E84-A, pp. 919-930, April 2001.
17. M. J. Mihaljević, M.P.C. Fossorier and H. Imai, "Fast correlation attack algorithm with the list decoding and an application", Fast Software Encryption - FSE 2001, *Lecture Notes in Computer Science*, vol 2355, pp. 196-210, 2002.
18. R. A. Rueppel, *Analysis and Design of Stream Ciphers*. Berlin: Springer-Verlag, 1986.
19. T. Siegenthaler, "Decrypting a class of stream ciphers using ciphertext only", *IEEE Trans. Comput.*, vol. C-34, pp. 81-85, 1985.
20. T. Siegenthaler, "Correlation-immunity of nonlinear combining functions for cryptographic applications", *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 776-780, 1984.