A Class of Binary Sequences with Large Linear Complexity

Amparo Fúster Sabater
Security Information Institute, C.S.I.C.
144 Serrano, 28006 Madrid, Spain

amparo@iec.csic.es

Abstract

Sequence generators based on Linear Feedback Shift Registers (LFSRs) are very common procedures to generate pseudorandom sequences for multiple applications: computer simulation, circuit testing, error-correcting codes or cryptography (stream ciphers).

The encryption procedure in stream ciphers tries to imitate the mythic *one-time pad cipher* [1] that remains as the only known perfectly secure cipher. This encryption procedure is designed to generate from a short key a long sequence (*keystream sequence*) of seemingly random bits. Some of the most recent designs in stream ciphers can be found in [2]. Typically, a stream cipher consists of a keystream generator whose output sequence is bit-wise XORed with the plaintext (in emission) in order to obtain the ciphertext or with the ciphertext (in reception) in order to recover the original plaintext. References [3, 4] provide a solid introduction to the study of stream ciphers.

Most keystream generators are based on maximal-length LFSRs [6] whose output sequences or m-sequences are combined by means of nonlinear filters, nonlinear combinators, irregularly decimated generators, typical elements from block ciphers, etc to produce sequences of cryptographic application.

Desirable properties for such sequences can be enumerated as follows:

1. Long Period
2. Good statistical properties
3. Large Linear Complexity (LC).

One general technique for building a keystream generator is to use a nonlinear filter, i.e. a nonlinear function applied to the stages of a single maximal-length LFSR. That is the output sequence is generated as the image of a nonlinear Boolean function F in the LFSR stages. Period and statistical properties of the filtered sequences are characteristics deeply studied in the literature, see [7] and the references above mentioned. In addition, such sequences have to pass all 19 DIEHARD tests [8] to be accepted as cryptographic sequences.

Regarding the third requirement, linear complexity of a sequence is defined as the amount of known sequence necessary to reconstruct the entire sequence. In cryptographic terms, LC must be as large as possible in order to prevent the application of the Berlekamp-Massey algorithm [9]. A recommended value for LC is about half the sequence period. Although several contributions to the linear complexity of nonlinearly filtered sequences can be found in the literature [5], [10] or [11], the problem of determining the exact value of the linear complexity attained by any nonlinear filter is still open.

Now some basic notation is introduced:

Nonlinear filter. It is a Boolean function $F(x_0, x_1, \ldots, x_{L-1})$ in L variables of degree k. For a subset $A = \{a_0, a_1, \ldots, a_{r-1}\}$ of $\{0, 1, \ldots, L-1\}$ with $r \leq k$, the notation $x_A = x_{a_0}\, x_{a_1} \ldots x_{a_{r-1}}$ is used. The Boolean function can be written as:

$$F(x_0, x_1, \ldots, x_{L-1}) = \sum_A c_A\, x_A, \qquad (1)$$

where $c_A \in \{0, 1\}$ and the summation is taken over all subsets A of $\{0, 1, \ldots, L-1\}$.

Filtered sequence. The sequence $\{z_n\}$ is the keystream or output sequence of the nonlinear filter F applied to the L stages of the LFSR. The keystream bit z_n is computed by selecting bits from the m-sequence $\{s_n\}$ such that

$$z_n = F(s_n, s_{n+1}, \ldots, s_{n+L-1}). \qquad (2)$$

Equation (1) describes the Algebraic Normal Form (ANF) of a nonlinear filter F. That is the filter is represented as the sum of distinct products in the variables $(s_n, s_{n+1}, \ldots, s_{n+L-1})$.

The ANF representation of a nonlinear filter is unique. At the same time, a nonlinear filter $F(s_n, s_{n+1}, \dots, s_{n+L-1})$ can be represented in terms of a N-tuple of coefficients $(C_1, C_2, \dots, C_N)$ with $C_i \in GF(2^L)$ where each coefficient determines the starting point of its corresponding *characteristic sequence* and N denotes the number of cosets of weight $\leq k$, see [5].

In this work, a method of computing all the nonlinear filters of order k applied to a LFSR with linear complexity $LC \geq \binom{L}{k}$ (where L is the LFSR length) has been developed. The procedure is based on the concept of equivalence classes of nonlinear filters and on the handling of such filters from different classes.

Let G be the set of the kth-order nonlinear filters applied to a LFSR of length L. We are going to group the elements of G producing the filtered sequence $\{z_n\}$ or a shifted version of such a sequence. Therefore, two different nonlinear filters F_0, F_1 in the same equivalence class will produce shifted versions of the same filtered sequence.

After distinct operations on the nonlinear filters from different equivalence classes, the final result of this computing method is:

1. A set of N basic filters of the form $(0, 0, \dots, d_i, \dots, 0, 0)$ $(1 \leq i \leq N)$ with $d_i \in GF(2^L), d_i \neq 0$.
2. Their corresponding ANF representations.

The combination of all these basic filters with d_i $(1 \leq i \leq N)$ ranging in $GF(2^L)$ (with their corresponding ANF representations) gives rise to all the possible terms of order k that preserve the cosets of weight k. From such terms, all the nonlinear filters of order k with a guaranteed linear complexity $LC \geq \binom{L}{k}$ can be constructed. Recall that the construction method involves very simple operations:

- Sum operation: that is reduced to a sum of filters for the ANF representation or to a sum of elements of the extended field $GF(2^L)$ that expressed in binary representation is just the XOR logic operation.
- Shifting operation through an equivalence class: that means an increment by 1 in all the indexes in the ANF representation.

Consequently, the efficiency of the computation method is quite evident. In brief, we provide one with the complete class of nonlinear filters with $LC \geq \binom{L}{k}$ at the price of minimal computational operations.

No restriction is imposed on the parameters of the nonlinear filtering function. The method completes the families of nonlinear filters with guaranteed large LC given in [5].

Keywords

Pseudorandom sequences, linear complexity, nonlinear filter, cryptography

References

[1] N. Nagaraj, One-Time Pad as a nonlinear dynamical system. Commun Nonlinear Sci Numer Simulat 17 (2012) 4029-4036.

[2] eSTREAM, the ECRYPT Stream Cipher Project,The eSTREAM Portfolio in 2012, available at http://www.ecrypt.eu.org/documents/D.SYM.10-v1.pdf

[3] A.J. Menezes *et al.*, Handbook of Applied Cryptography, New York:CRC Press, 1997.

[4] C. Paar, J. Pelzl, Understanding Cryptography, Springer-Verlag, Berlin Heidelberg, 2010.

[5] R.A. Rueppel, Analysis and Design of Stream Ciphers, Springer, New York, 1986.

[6] S. Golomb, Shift-Register Sequences, Aegean Park Press, Laguna Hills, California, 1982.

[7] A. Fúster-Sabater *et al.*, Deterministic Computation of Pseudorandomness in Cryptographic Sequences. Proc. of ICCS 2009, Part I, LNCS, Vol. 5544, Springer-Verlag, 2009, pp. 621-630.

[8] A. Marsaglia, Test of DIEHARD, $http://stat.fsu.edu/pub/diehard/$, 1998.

[9] J.L. Massey, Shift-Register Synthesis and BCH Decoding. IEEE Trans. Information Theory, 15(1) (1969) 122-127.

[10] K. Limniotis, N. Kolokotronis, N. Kalouptsidis, On the Linear Complexity of Sequences Obtained by State Space Generators. IEEE Trans. Inform. Theory. 54 (2008) 1786-1793.

[11] S. Ronjom, C. Cid, Nonlinear Equivalence of Stream Ciphers. Proc. of Fast Software Encryption, FSE 2010, Seoul, Korea, LNCS, Vol. 6147, Springer-Verlag, 2010, pp. 40-54.