A Hadamard matrix-based algorithm to evaluate the strength of binary sequences*

Amparo Fúster-Sabater[1], Verónica Requena[2], and Sara D. Cardell[3]

[1] Instituto de Tecnologías Físicas y de la Información, C.S.I.C., Madrid, Spain
amparo@iec.csic.es
[2] Departament de Matemàtiques, Universitat d'Alacant, Alacant, Spain
vrequena@ua.es
[3] Centro de Matemática, Computação e Cognição, Universidade Federal do ABC (UFABC), Santo André-SP Brazil
s.cardell@ufabc.edu.br

Abstract. Nowadays, a wide range of critical services relies on Internet of Things (IoT) devices. Nevertheless, they often lack proper security, becoming the gateway to attack the whole system. IoT security protocols are often based on stream ciphers, where pseudo-random number generators (PRNGs) are an essential part of them. In this work, we introduce a novel algorithm based on Hadamard matrices to evaluate the strength (unpredictability) of binary sequences, a key part of the IoT security stack. A comparative study with other algorithms that compute the same parameter is also presented.

Keywords: Hadamard matrix · unpredictability · PRNG · IoT.

1 Introduction

Nowadays, diverse critical services such as smart-grid, e-health, e-govern or industrial automation depend on an IoT infrastructure. At any rate, as the services around IoT grow dramatically so do the security risks [3]. Low-cost IoT devices, currently characterized by their resource constrains in processing power, memory, size and energy consumption, are also characterized by their minimum security. Combining lack of security with network dependability, they become the perfect gateway to compromise the whole network. This is the reason why 5G related research [8] or specific calls such as that of NIST for lightweight cryptography primitives [10], are addressing this concerning topic. In brief, lightweight cryptography as well as stream ciphers (binary sequence generators) are the key stones for designing security protocols.

* This work was supported in part by the Spanish State Research Agency (AEI) of the Ministry of Science and Innovation (MICINN), project P2QProMeTe (PID2020-112586RB-I00/AEI/ 10.13039/501100011033), co-funded by the European Regional Development Fund (ERDF, EU).

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_19

In this work, we propose a novel algorithm based on the fractal structure of the Hadamard matrices that analyses the unpredictability (linear complexity) of binary sequences with application in cryptography. Finally, we discuss a comparison among different algorithms that measure the same feature.

2 Preliminaries

Basic notation and concepts are now introduced.

Binary sequences: let $\{u_n\}_{n\geq 0} = \{u_0, u_1, u_2, \ldots\}$ be a binary sequence with $u_n \in \mathbb{F}_2$. Here we will just consider binary sequences with period a power of 2.

Linear Feedback Shift Register(LFSR): an LFSR is an electronic device with L interconnected memory cells (stages) with binary content. Maximum-length LFSRs generate PN-sequences with period $T = 2^L - 1$, see [5].

Linear Complexity: the LC of a sequence measures its unpredictability and is related with the amount of sequence we need in order to reconstruct the whole sequence. In cryptographic applications, LC must be as large as possible.

Generalized sequences: they are a family of binary sequences $\{s_n\}_{n\geq 0}$ obtained by means of the self-decimation of a PN-sequence, see [6]. The period of any generalized sequence is a divisor of 2^{L-1} and the linear complexity satisfies $2^{L-2} < LC \leq 2^{L-1} - (L-2)$ as proved in [4].

Binomial sequences: a binomial sequence $\left\{\binom{n}{k}\right\}_{n\geq 0}$, k being an integer, is a binary sequence

$$\left\{\binom{n}{k}\right\}_{n\geq 0} = \left\{\binom{0}{k}, \binom{1}{k}, \binom{2}{k}, \ldots\right\}_{mod\,2} \tag{1}$$

whose terms are the binomial numbers $\binom{n}{k}$ reduced mod 2. The sequence given in (1) is the ***k*-th binomial sequence**. Additional characteristic and properties of such sequences can be found in [1].

3 Binomial representation of binary sequences

Every binary sequence $\{u_n\}$ whose period $T = 2^t$ is a power of 2 can be written as a linear combination of binomial sequences [1],[2]:

$$\{u_n\} = \sum_{i=0}^{2^t-1} c_i \left\{\binom{n}{i}\right\}, \tag{2}$$

where t is a non-negative integer, $\left\{\binom{n}{i}\right\}$ is the i-th binomial sequence and the $c_i$are binary coefficients. The previous equation is the binomial representation of the sequence $\{u_n\}$. Moreover, $imax$ is an integer $(0 \leq imax \leq 2^t - 1)$ such that the coefficient c_{imax} of the binomial representation satisfies that $c_{imax} \neq 0$ while $c_i = 0$ for all index i in the range $(imax < i \leq 2^t - 1)$.

The coefficient c_{imax} and the binomial representation provide information about two fundamental parameters of the sequence: the period T of $\{u_n\}$ is that

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_19

of the binomial sequence $\left\{\binom{n}{imax}\right\}$ [1, Proposition 3] while the linear complexity LC of $\{u_n\}$ is that of the binomial sequence $\left\{\binom{n}{imax}\right\}$ [1, Corollary 14], that is $LC = imax + 1$. The **binomial matrix** H_{2^t} is a binary Hadamard matrix [7] of size $2^t \times 2^t$ constructed as:

$$H_{2^t} = \begin{bmatrix} H_{2^{t-1}} & H_{2^{t-1}} \\ 0_{2^{t-1}} & H_{2^{t-1}} \end{bmatrix},$$

being $0_{2^{t-1}}$ the $2^{t-1} \times 2^{t-1}$ null-matrix. Indeed, H_{2^t} exhibits the typical structure of a Hadamard matrix: three identical blocks plus the null-block. In turn, each block $H_{2^{t-1}}$ is a Hadamard matrix too. In addition, any binomial matrix H_{2^t} can be easily constructed from the binomial sequences such as follows: (a) its rows correspond to the first 2^t bits of the first 2^t binomial sequences and (b) its columns correspond to shifted versions of the first 2^t binomial sequences starting each of them in its first 1.

We write the binomial matrix $H_{2^4} = H_{16}$ that will be a basic structure for the construction of other binomial matrices of higher dimensions.

$$H_{16} = H_{2^4} = \begin{bmatrix} H_{2^3} & H_{2^3} \\ 0_{2^3} & H_{2^3} \end{bmatrix},$$

where

$$H_8 = H_{2^3} = \begin{bmatrix} 1&1&1&1&1&1&1&1 \\ 0&1&0&1&0&1&0&1 \\ 0&0&1&1&0&0&1&1 \\ 0&0&0&1&0&0&0&1 \\ 0&0&0&0&1&1&1&1 \\ 0&0&0&0&0&1&0&1 \\ 0&0&0&0&0&0&1&1 \\ 0&0&0&0&0&0&0&1 \end{bmatrix}.$$

Due to the particular structure of the binomial sequences, we can reformulate the binomial representation of $\{u_n\}$ given in (2) and convert it into a matrix equation including the binomial matrix H_{2^t}. That is:

$$(c_0, c_1, \ldots, c_{2^t-1}) = (u_0, u_1, \ldots, u_{2^t-1}) \cdot H_{2^t} \bmod 2, \tag{3}$$

where $(u_0, u_1, \ldots, u_{2^t-1})$ corresponds to the 2^t successive terms of the sequence $\{u_n\}$ and $(c_0, c_1, \ldots, c_{2^t-1})$ are the binary coefficients of the equation (2).

4 An algorithm to compute the *LC* of binary sequences

The equation (3) is the core of a new algorithm that computes the LC of a sequence. In fact, if the binomial matrix is written in terms of its column vectors $H_{2^t} = (\boldsymbol{h_0}, \boldsymbol{h_1}, \ldots, \boldsymbol{h_{2^t-1}})$, then the coefficients c_i are easily calculated as:

$$c_i = (u_0, u_1, \ldots, u_{2^t-1}) \cdot \boldsymbol{h_i} \qquad (0 \leq i \leq 2^t - 1). \tag{4}$$

The computation starts with the coefficient c_{2^t-1} and proceeds in reverse order until the first coefficient $c_i \neq 0$ is reached. In that case, $imax = i$ and LC is

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_19

Algorithm 1: Computation of the LC of a given sequence

Input: *seq*: sequence of period 2^t, H_t: the $(2^t \times 2^t)$ binomial matrix.
  $imax = 0$; $i = length(seq) - 1$;
  while $i \geq 0$ **do**
    $c_i = (u_0, u_1, \ldots, u_{2^t-1}) \cdot \boldsymbol{h_i}$;
    if $c_i \neq 0$ **then**
      $imax = i$;
      Break;
    endif
    $i = i - 1$;
  endwhile
Output: $LC = imax + 1$: Linear complexity of the sequence.

easily computed as $LC = imax + 1$. Algorithm 1 illustrates such a computation. Now, two basic ideas can be drawn:

1. The computation of LC is reduced to products modulo 2 of binary vectors. Clearly, its computational complexity is lower than that of other algorithms found in the literature, see sub-section 4.2.
2. If the column $\boldsymbol{h_{imax}}$ has many $0's$ and only a few $1's$, then only a few terms of the sequence $\{u_n\}$ will be required to compute its LC.

The previous algorithm is particularly useful when we analyse sequences whose LC is upper bounded by a maximum value LC_{max}. In that case, the computation of coefficients is simplified as $c_i = 0$ for every coefficient in the range $(imax < i \leq 2^t - 1)$. Thus, the Algorithm 1 starts with the index $i = imax = LC_{max} - 1$ computing directly the coefficient c_{imax}.

4.1 Application of the algorithm to generalized sequences

The generalized sequences $\{s_n\}_{n\geq 0}$ are ideal candidates for the application of Algorithm 1. In fact, their T is a power of 2 and their LC is upper bounded. Therefore, the Algorithm 1 starts with $i = imax = 2^{L-1} - (L-1)$ and computes the value of c_{imax}. If the coefficient $c_{imax} \neq 0$, then the complexity of the generalized sequence is $LC = 2^{L-1} - (L-2)$, otherwise LC will take a lower value. The column $\boldsymbol{h_{imax}}$ corresponds to the $(L-1)$-th column of the matrix $H_{2^{L-1}}$ read from right to left. As far as L increases by 1, the column $\boldsymbol{h_{imax}}$ is shifted one position to the left. Next, we will apply these Hadamard matrices to the computation of LC for different values of L.

For L in the range $2 \leq L \leq 17$: we fix H_{16} as our reference matrix, called 16-box and depicted in Table 1. The successive matrices $H_{2^{L-1}}$ in this range are made up of 16-boxes. Then, we divide the period $T = 2^{L-1}$ of the generalized sequence by 16 to determine the number of 16-boxes included in its binomial matrix $H_{2^{L-1}}$. Next, we count the number of $1's$ in the column $\boldsymbol{h_{imax}}$ of the 16-box and, finally, we multiply this number by the number of 16-boxes in $H_{2^{L-1}}$ to get the total number of $1's$ in the general column $\boldsymbol{h_{imax}}$.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_19

Table 1. 16-box to analyse generalized sequences with $2 \leq L \leq 17$

	H_8	H_8
	0_8	H_8
$L =$	**17** ... **14** **13** ... **10**	**9** ... **6** **5** ... **2**

For L in the range $18 \leq L \leq 33$: we now use a 32-box as shown in Table 2, where H_{16} is the 16-box. Next, we divide the period T of $\{s_n\}$ by 32 and analyse the number of $1's$ in the successive columns $\boldsymbol{h_{imax}}$ of the 32-box.

Table 2. 32-box to analyse generalized sequences with $18 \leq L \leq 33$

	H_{16}	H_{16}
	0_{16}	H_{16}
$L =$	**33** ... **26** **25** ... **18**	17 ... 10 9 ... 2

For L in the range $34 \leq L \leq 65$: we now use a 64-box as shown in Table 3, where H_{32} is the 32-box. Now we divide the period T of the sequence by 64 and analyse the $1's$ of the successive columns of the 64-box.

For L in the range $66 \leq L \leq 129$: the study is similar to that of the previous intervals but now we would use a 128-box made up of 64-boxes according to the typical structure of a Hadamard-matrix. We would divide the period T of the sequence by 128 and would analyse the number of $1's$ in the successive columns of the 128-box.

In all this analysis, the least suitable cases are generalized sequences whose columns $\boldsymbol{h_{imax}}$ in $H_{2^{L-1}}$ correspond to binomial sequences $\left\{\binom{n}{2^m}\right\}$ as half their digits are $1's$. Consequently, $T/2$ digits of $\{s_n\}$ are needed to compute its LC. Conversely, the most suitable cases are generalized sequences whose columns $\boldsymbol{h_{imax}}$ in $H_{2^{L-1}}$ correspond to binomial sequences $\left\{\binom{n}{2^m-1}\right\}$ where only $T/2^m$ of their digits are $1's$. Consequently, $T/2^m$ digits of $\{s_n\}$ are enough to compute its LC.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_19

Table 3. 64-box to analyse generalized sequences with $34 \leq L \leq 65$

	H_{16}	H_{16}	H_{16}	H_{16}
	0_{16}	H_{16}	0_{16}	H_{16}
	0_{32}		H_{16}	H_{16}
			0_{16}	H_{16}
$L =$	**65 ... 50**	**49 ... 34**	33 ... 18	17 ... 2

Remark 1 *With only four boxes (16, 32, 64 and 128-boxes), we have easily got values of L in the range $L > 128$, which is the real cryptographic range.*

Moreover, we realize that Algorithm 1 will never require the knowledge of the whole sequence $\{s_n\}$ to compute its LC as there will always be at least a null-block in the binomial matrix. Consequently, the amount of sequence required is much less than that of other algorithms, see next sub-section.

4.2 Comparison with other algorithms

We compare the proposal here developed with other algorithms computing LC. The comparison is summarized in Table 4, where l is the sequence length, r the number of binomial sequences in the binomial decomposition and 5G denotes algorithm focussed on 5G Technologies.

5 Conclusions

In this work, a new algorithm, the Hadamard matrix-based algorithm to compute the linear complexity of binary sequences, has been introduced. It exhibits a better performance (lower computational complexity and sequence requirements) than that of similar algorithms computing LC. This is a big step in the study of binary sequences with period a power of 2 as well as it makes easier to detect flaws such as predictability in this kind of sequences. Moreover, the binomial decomposition as a way to extract information from a given sequence is an innovative tool and it is left for future work its application to other features, e.g. auto-correlation, balancedness, compression, of binary sequences.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_19

Table 4. Comparison between proposed and existing algorithms to calculate LC

Authors	Merits	Demerits	O(.)	Length	5G
Berlekamp et al. [9] (1969)	For sequences of any length.	High requirements in length	$O(l^2)$	$2 \times l$	**N**
Cardell et al. [1] (2019)	Improvement in sequence length requirements.	Applicable to seq. with length a power of 2. Sequential.	$O(r \times l)$	$l - log\ l$	**N**
Martin et al. [7] (2020)	Improvement in sequence length requirements. It outperforms previous algorithms. Concurrent.	Applicable to seq. with length a power of 2	$O(l)$	$l - log\ l$	**Y**
Fúster et al. (This work) (2022)	Great improvement in sequence length requirements. It outperforms all the previous algorithms. Concurrent.	Applicable to seq. with length a power of 2	$O(l/2)$	$l/2$	**Y**

References

1. Sara D. Cardell and Amparo Fúster-Sabater. Binomial Representation of Cryptographic Binary Sequences and Its Relation to Cellular Automata. *Complexity*, **1**, 1–13 (2019) doi: 10.1155/2019/2108014
2. Sara D. Cardell and Amparo Fúster-Sabater. *Cryptography with Shrinking Generators*. Springer Briefs in Mathematics, Springer International Publishing (2019) doi: 10.1007/978-3-030-12850-0
3. W.L. Chin, W. Li and H.H. Chen. Energy big data security threats in IoT-based smart grid communications. *IEEE Commun. Mag.*, **2017**(55), 70–75 (2017)
4. Amparo Fúster-Sabater and Sara D. Cardell. Linear complexity of generalized sequences by comparison of PN-sequences. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. (RACSAM)*, **114**(4), 79–97 (2020)
5. Solomon W. Golomb. *Shift Register-Sequences*. 2nd edn. Aegean Park Press, Laguna Hill, California (1982)
6. Yupu Hu and Guozhen Xiao. Generalized self-shrinking generator. *IEEE Trans Inform Theory*, **50**(4), 714–719 (2004)
7. Jose L. Martin-Navarro and Amparo Fúster-Sabater. Folding-BSD Algorithm for Binary Sequence Decomposition. *Computers*, **9**(4), 100 (2020) doi: 10.3390/computers9040100
8. C.X. Mavromoustakis, G. Mastorakis and J.M. Batalla. *Internet of Things (IoT) in 5G Mobile Technologies*, Volume 8. Springer, Berlin/Heidelberg, Germany (2016)
9. James L. Massey. Shift-register synthesis and BCH decoding. *IEEE Trans Inform Theory*, **15**(1), 122–127 (1969)
10. NIST Lightweight Cryptography Project. Available online: https://csrc.nist.gov/Projects/Lightweight-Cryptography (accessed 28 March 2022).

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_19