Available online at www.elixirpublishers.com (Elixir International Journal)

Discrete Mathematics

Elixir Dis. Math. 99 (2016) 43171-43174

ISSN: 2229-712X

Linear Complexity of Pseudorandom Binary Sequences from de Bruijn Graphs

Mufutau Babatunde Akinwande
Independent Higher Education Professional, Ottawa, Canada.

ARTICLE INFO

Article history:
Received: 17 August 2016;
Received in revised form:
3 October 2016;
Accepted: 13 October 2016;

Keywords
Binary sequences,
Periodic sequences,
Correlation,
Linear complexity,
De Bruijn graphs.

ABSTRACT
Linear complexity is a vital complexity measure and pseudorandom sequences with good correlation properties, large linear complexity, and balance statistics are widely used in modern communication and cryptology. This paper study the linear complexity of an inverse paths of a binary de Bruijn cycle by presenting set of pseudorandom binary sequences from de Bruijn graphs. And, it is shown that such sequences have large linear complexity.

1. Introduction

A good illustration method of a finite sequence, which is used to measure the randomness, was developed in the 1970s. Lempel and Ziv in [1] justified using the shortest length of Linear Feedback Shift Register (LFSR) models to measure the unpredictability of the sequences. Since LFSRs are easy to implement by hardware and fast to process, they were often recommended to be the pseudorandom sequences generators. Mainly, because of the adoption of LFSRs in stream ciphers, cryptographers and mathematicians used thorough mathematical theory to analyze their security. Also because LFSRs are linear devices, the linear complexity is an important concept to determine the security levels of stream ciphers (for other complexity measurements such as higher-order complexity, 2-adic complexity measures and complexity measures based on pattern counting, one could refer to [2]). And the linear complexity profile is also a good tool to measure the randomness of generated sequences.

The randomness and complexity properties of these sequences are vital in some applications where security is an issue. Large linear complexity (also known as linear span) of the sequence is required to prevent it from being reconstructed from a subset of the sequence, for example, using the Berlekamp-Massey algorithm. With the exception of [3], the sequences in these sets are obtained from sequences of powers of primitive elements in fields of characteristic 2 by algebraic manipulations. As a consequence, the sequences all have periods of the form $2^n - 1$ (or, in the case of sequences obtained from the duals of nonprimitive BCH codes, period dividing $2^n - 1$) and the sizes of the sets are rather restricted. In literature, comprehensive research has been performed on how to generate sequences with these desired properties, some representative examples can be found in [4] and references therein.

A detailed characterization of a class of homomorphisms between de Bruijn digraphs of different orders with a property *D* that can be used to construct de Bruijn cycles recursively. For two positive integers *n* and *k*, property *D* allows a recursive construction of de Bruijn cycles that the inverse of a factor in a lower order de Bruijn graph, $B_n(q)$, is a factor in a higher order de Bruijn graph, $B_{n+k}(q)$, of the same alphabet, *q*, which generalized a well-known binary construction of Lempel is discussed in [5]. There are three standard test batteries for empirically estimating random number quality: Diehard, STS, and Rabbit (in TestU01) test suites to test the randomness properties of binary sequences. It is shown in [6] that sequences generated recursively from de Bruijn graphs pass these test suites. It is also found that such sequences exhibit good autocorrelation and cross correlation properties [7].

In this correspondence, we discuss and illustrates set of pseudorandom binary sequences from de Bruijn graphs. Linear complexity of the binary sequences so generated is determined using Massey - Berlekamp algorithm [8] and results are analyzed. It is shown that such sequences exhibit large linear complexity which is desirable characteristics of random sequences required for key sequences in stream cipher systems. Section 2 contains a brief description of de Bruijn graphs we need, and reviews a novel method to generates pseudorandom binary sequences, which relies on *D*-homomorphism between de Bruijn digraphs of different orders. Section 3 describes necessary concepts and measures used to quantify linear complexity of the generated binary sequences. Section 4 analyzes and gives results. Finally, Section 5 gives conclusion.

2. Pseudorandom Binary Sequences from de Bruijn Graphs

The main graphical tool used in the study of de Bruijn sequences are *de Bruijn digraphs*. Besides its use in the context of the de Bruijn sequences, they are also used as models for transportation networks, DNA algorithms, and computer networks to mention a few. The properties of de Bruijn digraphs are well discussed in [9, 10].

E-mail address: mboakinwande@yahoo.com

A homomorphism *H* between two digraphs G_1 and G_2 is a function that preserves the structure of the digraph. That is, if (x_1, x_2) is an edge in G_1 then (Hx_1, Hx_2) is also an edge in G_2. For two positive integers *n* and *k*, [5] characterizes such homomorphisms and describes a family of homomorphisms from $B_{n+k}(q)$ to $B_n(q)$ whose inverse assigns to an arbitrary vertex disjoint path in $B_n(q)$ a constant number, q^k, of non-overlapping preimage paths in $B_{n+k}(q)$. We will say that such a homomorphism enjoys property *D* or simply is a *D*-homomorphism. The following theorem is proved in [7] and they characterize *D*-homomorphisms between binary de Bruijn digraphs.

Theorem 2.1

A necessary and sufficient condition for a homomorphism $D_{n,k}$ from $B_{n+k}(2)$ to $B_n(2)$ to have property *D* is that $d_k(x_1,...,x_{k+1}) = x_1 + h(x_2,...,x_k) + x_{k+1}$, where $h(x_2,...,x_k)$ is any Boolean function of $k-1$ variables.

The method relies on *D*-homomorphisms between de Bruijn digraphs of different orders that were defined above. Thus, we treat the backbone generator as a cycle in a binary de Bruijn digraph $B_n(2)$. The inverse of the backbone generator by a *D*-homomorphism makes a large number of inverse sequences that all have the same size as the original cycle, where the former are regarded as paths in the higher order de Bruijn digraph. The method is based on mapping one generalized shift register sequence to many distinct sequences, using the inverse of a well designed homomorphism between two de Bruijn digraphs of different orders developed in [11]. The significance of the produced sequences is that they are all of the same length as the original, and no sequence of consecutive numbers of a certain prescribed length is common to any two sequences produced.

3. Linear Complexity

Complexity measures of sequences are very much useful in the security analysis of stream ciphers and other applications. Periodic sequences should satisfy certain criteria for the suitability as keystream sequences in stream ciphers. One condition for suitability is that it should be very hard to reproduce the entire keystream from the knowledge of a portion of it. Another criteria is that the sequence must belong to a large class of sequences possessing similar behavior in a suitable sense.

The linear complexity of a sequence is not only a measure of unpredictability for suitability in cryptographic applications, but also of interest in information theory. It is an important complexity measure in the system theoretic approach to stream ciphers. The linear complexity *L(S)* of an ultimately periodic sequence *S* defined over a finite field of prime order, F_q, is defined as the length of the shortest LFSR that generates the sequence *S*. In other words, it is the least order of a linear recurrence relation over F_q which generates *S*.

The linear complexity of a finite sequence is determined using Massey-Berlekamp algorithm [8] which is described as follows:

Algorithm

Input:

A binary sequence $S = s_0, s_1, s_2,..., s_{n-1}$ of length n.

Output:

The linear complexity $L(S)$ of S, $0 \le L(S) \le n$.

Step 1:

Initialization. $C(D) \leftarrow 1, \quad L \leftarrow 0, \quad m \leftarrow -1, \quad B(D) \leftarrow 1, \quad N \leftarrow 0.$

Step 2:

While ($N < n$) do the following:

Compute the next discrepancy *d*:

$$d \leftarrow (s_N + \sum_{i=1}^{L} c_i s_{N-i}) \bmod 2.$$

If $d = 1$ then do the following:

$$T(D) \leftarrow C(D), \quad C(D) \leftarrow C(D) + B(D).D^{N-m}.$$

If $L \le N/2$ then $L \leftarrow N+1-L, \quad m \leftarrow N, \quad B(D) \leftarrow T(D).$

$$N \leftarrow N+1.$$

Return (*L*).

Let $S = s_1, s_2, s_3,....$ be an arbitrary sequence of elements defined over F_q, for any integer n, $1 \le n \le |S|$, the nth linear complexity $L_n(S)$ of *S* is the length of the shortest LFSR that generates the first *n* terms of *S*. In terms of linear recurrence relation, it is the least order of a linear recurrence relation over F_q that generates *S*. Then $0 \le L_n(S) \le n$ and $L_n(S) \le L_{n+1}(S)$.

Thus, we can define the linear complexity of an ultimately periodic sequence S in terms of $L_n(S)$ as $L(S) = Sup_{n \ge 1} L_n(S)$.

Some properties of linear complexity of binary sequences are given in the following remark [8].

Remark

Let *S* and *T* be binary sequences. Then

a. For any $n \geq 1$, the linear complexity of the subsequence S_n satisfies $0 \leq L(S_n) \leq n$.

b. $L(S_n) = 0$ if and only if S_n is the zero sequence of length *n*.

c. $L(S_n) = n$ if and only if $S_n = 0,0,0,...,0,1$.

d. If the sequence S is periodic with period *N*, then $L(S) \leq N$.

e. $L(S \oplus T) \leq L(S) + L(T)$, where $S \oplus T$ denotes the bitwise XOR of the sequences *S* and *T*.

To study the statistical performance of the linear complexity of produced sequences, we calculate the Mean, Standard deviation, and Variance of linear complexity. Some of the basic concepts and definitions about Mean, Standard deviation, and Variance we introduce in this section can be found in [12, 13].

Mean

The mean of a numerical variable is computed as the sum of all of the observations divided by the number of observations. Let $L_1, L_2, L_3, \text{K}, L_n$ be *n* linear complexity values of *n* sequence of the same length. Thus, the mean value or the average, denoted by μ, of linear complexity is defined as follows:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} L_i$$

The mean is a common way to measure the center of a distribution of data.

Standard deviation and Variance

The mean describe the center of a data set, but the variability in the data is also important. Thus, we introduce two measures of variability: the variance and the standard deviation. Variance of a random variable L, denoted by σ^2, is defined as follows:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(L_i - \mu)^2$$

The variance is roughly the average squared distance from the mean. The standard deviation is the square root of the variance. The standard deviation is useful when considering how close the data are to the mean. A low standard deviation shows that the values are very close to the mean, whereas high standard deviation shows that the values are spread out over a large range of values, around the mean.

4. Analysis and Results

We now consider the parallel inverse images obtained in [6] as the six different backbone sequences each for sequence of lengths of 64bits, 128bits, 256bits, 512bits and 1024bits, i.e., they are of orders 6, 7, 8, 9, and 10 respectively. The linear complexities of the produced six sequences are computed using an Online Calculator of Berlekamp-Massey Algorithm developed by Bo Zhu [14].

The statistical performance of linear complexity is obtained by finding its mean, standard deviation, and variance after the linear complexities of six sequences are calculated. The results are given in Table 4.1.

Table 4.1. Linear Complexity of the produced binary sequences from de Bruijn graphs.

Sequence	Order 6	Order 7	Order 8	Order 9	Order 10
1	32	64	128	256	512
2	32	65	128	257	513
3	34	62	129	256	514
4	33	66	128	256	512
5	32	64	129	254	514
6	31	63	128	256	513
Mean	32.33333	64	128.3333	255.8333	513
Variance	1.066667	2	0.266667	0.966667	0.8
Standard Deviation	1.032796	1.414214	0.516398	0.983192	0.894427

We observed that the mean linear complexities for sequences of order 6, 7, 8, 9, and 10 are 32.3333, 64, 128.3333, 255.8333, and 513 respectively. And, the corresponding variances are 1.0667, 2, 0.2667, 0.9667, and 0.8000 respectively. While the corresponding standard deviations are given as 1.0328, 1.4142, 0.5164, 0.9832, and 0.8944 respectively.

Clearly, the above results show that the linear complexity value is increasing with the order of the sequence and tends to *n/2*, where *n* is the sequence length. While, the values of variance and standard deviation are low.

Now, we use Chebychev inequality to obtain bounds on the probability of linear complexities of the sequences. The following theorem is proved in [12] and is related to the idea that the variance of a random variable is a measure of how spread out its distribution is.

Theorem 4.1 (Chebyshev Inequality).

Let *L* be a random variable with mean $\mu = E(L)$ and variance $\sigma^2 = Var(L)$. Then for every number *k* > 0,

$\Pr(|L-\mu| \geq k) \leq \frac{\sigma^2}{k^2}$, i.e., $\Pr(|L-\mu| \geq k)$ is the probability that L lies outside the range $(\mu-k) \leq L \leq (\mu+k)$. This probability is always less than or equal to $\frac{\sigma^2}{k^2}$.

Conversely, the above inequality can be restated as follows: For every number $k > 0$, $\Pr(|L-\mu| \leq k) \geq 1-\frac{\sigma^2}{k^2}$, i.e., $\Pr(|L-\mu| \leq k)$ is the probability that L lies inside the range $(\mu-k) \leq L \leq (\mu+k)$. This probability is always greater than or equal to $1-\frac{\sigma^2}{k^2}$.

If we let $k = 9$, then the bound on the probability of linear complexity L lying inside the range $(\mu-9) \leq L \leq (\mu+9)$ for all the orders are given in Table 4.2.

Table 4.2. Statistical results for the produced binary sequences of orders 6, 7, 8, 9 ,and 10 with corresponding lengths of 64bits, 128bits, 256bits, 512bits and 1024bits respectively.

	Order 6	**Order 7**	**Order 8**	**Order 9**	**Order 10**
Mean, μ	32.33333	64	128.3333	255.8333	513
Variance, σ^2	1.066667	2	0.266667	0.966667	0.8
$\Pr(\|L-\mu\| \leq k) \geq 1-\frac{\sigma^2}{k^2} = 1-\frac{\sigma^2}{81}$	0.9868	0.9753	0.9967	0.9881	0.9901

5. Conclusions

The set of pseudorandom binary sequences obtained from de Bruijn graphs which is based on mapping one generalized shift register sequence to many distinct sequences by using the inverse of a well designed homomorphism between two de Bruijn digraphs of different orders results in sequences with large linear complexity determined using Massey- Berlekamp algorithm.

The statistical behaviour of the linear complexity of produced sequences was study by calculating the Mean, Standard deviation, and Variance of linear complexity. Six sequences for each orders 6, 7, 8, 9, and 10 with corresponding sequence lengths of 64 -, 128 -, 256 -, 512 -, and 1024 - bits respectively are investigated and analyzed.

We observed from Table 4.2 that the probability that the linear complexity L differ by mean value by 9, i.e., the probability that linear complexity L is lying inside the range (256 ± 9) is always greater than or equal to 0.9967.

References

[1] A. Lempel, and J. Ziv, "On the complexity of finite sequences," IEEE Transactions on Information Theory, vol. 22, no. 1, pp. 75-81, 1976.

[2] H. Niederreiter, "Some computable complexity measures for binary sequences," Sequences and Their Applications (C. Ding, T. Helleseth, and H. Niederreiter, eds.), pp. 67-78, Springer, London, 1999.

[3] P. Udaya, and M. U. Siddiqi, "Optimal biphase sequences with large linear complexity derived from sequences over Z_4," IEEETrans.Inform.Theory,42:206 – 216, 1996.

[4] S. W. Golomb, and G. Gong, "Signal Designs for Good Correlation: for wireless communications, cryptography and radar applications," Cambridge University Press, 2005.

[5] M. B. O. Akinwande, "Characterization of de Bruijn graphs homomorphisms," African Journal of Mathematics and Computer Science Research, Vol. 4(10), pp. 300-307, 2011

[6] M. B. O. Akinwande, "Homomorphisms of Nonbinary de Bruijn Graphs with Applications," Ph.D. dissertation, Clarkson University, New York, USA, 2010

[7] M. B. Akinwande, "Correlation of Pseudorandom Binary Sequences from de Bruijn Graphs," Global Journal of Mathematical Sciences: Theory and Practical, Volume 4, Number 2,pp. 121–132, 2012

[8] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, Handbook of Applied Cryptography, CRC Press, 2001

[9] R. D. Bryant and H. Fredricksen, "Covering the de Bruijn graph," Discrete Mathematics, 89, 2; 133 - 148, 1991

[10] C. Lu, J. Xu and K. Zhang, "On (d; 2)-dominating numbers of binary undirected de Bruijn graphs," Discrete Application Mathematics, 105(1-3), 137 - 145, 2000

[11] A. Alhakim and M. Akinwande, "A recursive construction of nonbinary de Bruijn sequences," Des. Codes Cryptogr., 60:155–169, 2011

[12] M. H. DeGroot, and M.J. Schervish, Probability and statistics, 4th edition, Addison-Wesley, 2012

[13] D. M. Diez, C.D. Barr, and M. Cetinkaya-Rundel, OpenIntro Statistics, 3rd edition, 2015. Available at openintro.org

[14] An Online Calculator of Berlekamp-Massey Algorithm by Bo Zhu. Available at http://bma.bozhu.me/.