On the Quadratic Spans of Periodic Sequences[1]

Agnes Hui Chan[2]
Richard A. Games
The MITRE Corporation, Bedford MA 01730

1. Introduction

Random binary sequences are required in many applications of modern communication systems and in designing reliable circuits. However, truly random sequences are often associated with extremely high costs, and are therefore infeasible to use. Deterministically generated sequences that pass certain statistical tests suggested by random sequences are often used instead and are referred to as *pseudorandom* sequences. In applications involving, for instance, secure or spread spectrum communications, it is essential that these pseudorandom sequences be unpredictable. This paper addresses the problem of predicting the terms of a pseudorandom sequence from some initial portion of the sequence. A good introduction to the issues involved in this area can be found in [7].

Sequences that are generated deterministically by a finite-state machine must ultimately be periodic, and as such, can be generated by a simple feedback shift register (FSR) whose length is long enough to contain the terms of the sequence up to the point where they start to repeat. This pure cycling FSR has a single tap at the point corresponding to the initial point of the periodic part of the sequence, which will be the first stage of the shift register only if there is no initial acyclic part.

If the period is extremely long, this pure cycling FSR is impractical, but there will usually be much shorter FSRs that can be used to generate the sequence. These shorter FSRs will have more general feedback functions, Boolean functions defined on the states of the register. The length of a shortest FSR that generates the sequence is called the *span* of the sequence. Determining the span and an associated Boolean feedback function is difficult because of the nonlinearities involved.

Because of its tractability, most attention has been focused on determining the *linear span* of a sequence—the length of the shortest linear FSR that generates the sequence. If the linear span of a sequence is small, then the feedback function that defines the linear FSR can be determined easily using the Berlekamp-Massey algorithm [6]. Once the feedback function is determined, the remainder of the sequence can be easily generated.

A sequence with very large linear span may be generated by a much shorter FSR if nonlinear terms are allowed in the feedback function. The case of additional quadratic terms $s_i s_{i+j}$ is considered in this paper, since it is the most computationally tractable nonlinear case. The *quadratic span* of a periodic sequence is defined to be the length of the shortest length quadratic FSR that generates the sequence. For example, the periodic sequence $s = \overline{000101101011110}00101101\ldots$ of period 15 has linear span 14, but this sequence is generated by a quadratic FSR of length 4, and in this case, the span and the quadratic span both equal 4.

[1] This work was supported by the United States Air Force's Electronic Systems Division under Contract F19628-86-C-0001 and the MITRE-Sponsored Research Program.

[2] A. H. Chan is also with the College of Computer Science, Northeastern University, Boston, MA 02115.

G. Brassard (Ed.): Advances in Cryptology - CRYPTO '89, LNCS 435, pp. 82-89, 1990.

The complexity of binary sequences formed using nonlinear functions has been studied previously in [3], [5], [7], and [8]. These authors have considered sequences obtained by applying nonlinear feedforward functions to the contents of one or more linear FSR's. Their results concern only the linear spans of these nonlinear feedforward sequences. The quadratic spans of this class of sequences have not been analyzed.

The quadratic span of a sequence can be determined by solving certain structured systems of linear equations. Determining the linear span of a sequence also involves solving certain structured systems of linear equations. The Berlekamp-Massey algorithm reduces the complexity of solving the systems involved in the linear case. Updating procedures similar to those used in the Berlekamp-Massey algorithm can only sometimes be applied to the quadratic case. We present an algorithm based on Gaussian elimination for calculating the quadratic span. It is still an open problem whether the special structure of the matrix in the quadratic case can be used to decrease the complexity of this algorithm. We do present a useful result concerning the increase in the quadratic span when an existing linear FSR fails to generate a particular term in a sequence. This is a partial generalization to the quadratic case of theorem 1 of [6].

In addition to obtaining a more efficient algorithm, another important result would be to assess further the predictability of a particular class of sequences by determining their quadratic spans. In this paper, we focus on de Bruijn sequences. The de Bruijn sequences of span n are the sequences of maximum period 2^n generated by nonlinear FSR's of length n. There have been a number of new algorithms proposed for generating large numbers of de Bruijn sequences of span n. The problem of determining the quadratic spans of de Bruijn sequences of span n was introduced in [2]. The linear span of a de Bruijn sequence of span n is greater than half of its period [1], and it has been observed empirically (through $n = 6$) that the vast majority of de Bruijn sequences have linear spans nearly equal to the upper bound of one less than the period.

For the case of quadratic spans of de Bruijn sequences, the situation appears to be quite different. In this paper, we determine a new upper bound on the quadratic span of de Bruijn sequences of span n. We show that the quadratic spans of de Bruijn sequences of span n are bounded above by $2^n - 1 - \binom{n}{2}$, and show that this bound is achieved by de Bruijn sequences obtained from adding the 0 state to maximum-period linear sequences of period $2^n - 1$. It is easy to see that a lower bound for this case is $n + 1$, but we conjecture that a lower bound for $n > 3$ is in fact $n + 2$.

The distributions of quadratic spans for the de Bruijn sequences through span 6 are calculated by computer, correcting and extending the results of [2]. The situation for $n = 6$ is displayed graphically in figure 1, where the linear span distribution may be seen to be concentrated near the linear span upper bound of 63, while the quadratic span distribution is concentrated around 11. The problem of determining the quadratic span distribution of de Bruijn sequences remains open.

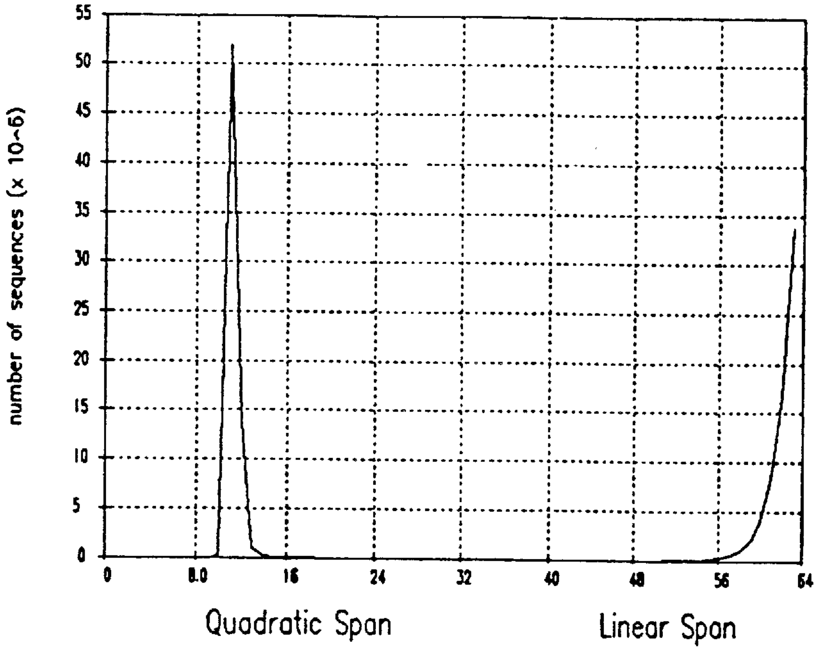

Figure 1. Linear and Quadratic Span Distribution for de Bruijn Sequences of Span 6

2. Definitions and Main Results

In this section, we state without proof the main results of this paper. We first give the necessary definitions. All the sequences we consider have terms in GF(2). Most of the equations and expressions discussed will be over GF(2) unless otherwise stated.

An n-stage FSR with feedback function f: $\mathrm{GF}(2)^n \rightarrow \mathrm{GF}(2)$ generates a sequence $\mathbf{s} = (s_0, s_1, \ldots, s_i, \ldots)$ where $s_0, s_1, \ldots, s_{n-1}$ corresponds to the initial loading and

$$s_{i+n} = f(s_i, s_{i+1}, \ldots, s_{i+n-1}), \quad i \geq 0 \tag{1}$$

The function f in (1) is *linear* if there exist $a_0, a_1, \ldots, a_{n-1}$ in GF(2) such that for all $(x_0, x_1, \ldots, x_{n-1}) \in \mathrm{GF}(2)^n$,

$$f(x_0, x_1, \ldots, x_{n-1}) = \sum_{j=0}^{n-1} a_j x_j.$$

The function f is *quadratic* if there exists $a_{j,k}$, $0 \leq j \leq k \leq n-1$ in GF(2) such that for all $(x_0, x_1, \ldots, x_{n-1}) \in \mathrm{GF}(2)^n$,

$$f(x_0, x_1, \ldots, x_{n-1}) = \sum_{j=0}^{n-1} \sum_{k=j}^{n-1} a_{j,k} x_j x_k. \tag{2}$$

Note that $x_j x_j = x_j$ in GF(2); for simplicity we sometimes denote $a_{j,j}$ by a_j. Note this definition implies that $f(0) = 0$. A more general quadratic feedback function could include a constant term. As in [2], we consider the case of no constant term. The generalized results for the constant term case will be reported in the full paper. Higher-order feedback functions are defined in a similar fashion.

A sequence generated by an n-stage FSR, but not an $(n-1)$-stage FSR, is said to have *span n*. *Linear span* and *quadratic span* are defined similarly for FSR's with linear and quadratic feedback functions. The span, quadratic span, and linear span of a sequence s will be denoted respectively by $sp(\mathrm{s})$, $q(\mathrm{s})$, and $l(\mathrm{s})$. By definition,

$$sp(\mathrm{s}) \leq q(\mathrm{s}) \leq l(\mathrm{s}).$$

The above notions of span apply to finite sequences as well. Given a sequence $\mathrm{s} = (s_0, s_1, \ldots, s_i, \ldots)$, we denote any finite subsequence of length N starting at s_i by s_i^N, that is,

$$\mathrm{s}_i^N = (s_i, s_{i+1}, \ldots, s_{i+N-1}).$$

We write s^N for s_0^N. A FSR generates s^N if (1) is satisfied for $0 \leq i \leq N-n-1$, i.e. if the terms through s_{N-1} can be successfully generated. Note that any FSR with N or more stages generates s^N, since the entire sequence can be loaded in the N initial stages of the register.

Let E denote the sequence shift operator; that is, $E\mathrm{s}$ denotes the sequence with the i-th term $(E\mathrm{s})_i = s_{i+1}$. We define $(E^j \circ E^k)\mathrm{s}$ to be the sequence with i-th term given by $s_{i+j}s_{i+k}$. For a quadratic feedback function (2), equation (1) can be expressed in terms of the shift operator as

$$(E^n + \sum_{j=0}^{n-1} \sum_{k=j}^{n-1} a_{j,k} E^j \circ E^k)\mathrm{s} = 0.$$

To compute a quadratic feedback function (2) of an n-stage FSR from a given sequence of N terms, a system of linear equations in the unknowns $a_{j,k}$, $0 \leq j \leq k \leq n-1$, can be formed:

$$\begin{aligned} f(s_0, s_1, \ldots, s_{n-1}) &= s_n \\ f(s_1, s_2, \ldots, s_n) &= s_{n+1} \\ &\vdots \\ f(s_{N-n-1}, s_{N-n}, \ldots, s_{N-2}) &= s_{N-1}. \end{aligned} \tag{3}$$

If (3) has a solution, then $q(\mathrm{s}^N) \leq n$; otherwise $q(\mathrm{s}^N) > n$.

We define the matrix $M(N, n)$ as the coefficient matrix associated with the system of linear equations in (3). This matrix has $N - n$ rows and $n(n+1)/2$ columns indexed by the variables $a_{j,k}$. Actually, the column associated with $a_{j,k}$ is given by $((E^j \circ E^k)\mathrm{s})^{N-n}$. We use a particular ordering of the variables $a_{j,k}$ that simplifies the notation when the number of stages n is increased. For example, when $N = 8$ and $n = 3$, we write (3) as

$$\begin{pmatrix} s_0 & s_1 & s_0s_1 & s_2 & s_0s_2 & s_1s_2 \\ s_1 & s_2 & s_1s_2 & s_3 & s_1s_3 & s_2s_3 \\ s_2 & s_3 & s_2s_3 & s_4 & s_2s_4 & s_3s_4 \\ s_3 & s_4 & s_3s_4 & s_5 & s_3s_5 & s_4s_5 \\ s_4 & s_5 & s_4s_5 & s_6 & s_4s_6 & s_5s_6 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_{0,1} \\ a_2 \\ a_{0,2} \\ a_{1,2} \end{pmatrix} = \begin{pmatrix} s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \end{pmatrix}.$$

Our analysis of quadratic spans involves examining the properties of $M(N, n)$, including its rank. A similar point of view for the linear span case can be found in [4].

Given a sequence s, if an FSR generates s^N, but does not generate s^{N+1}, then we say a *discrepancy* occurs at s_N. The first proposition states when the occurrence of a discrepancy results in an increase in the quadratic span. It uses standard results from linear algebra and forms the basis of an algorithm, which uses Gaussian elimination to solve the system (3), for computing the quadratic span.

PROPOSITION 1. *If a quadratic FSR of length n generates* s^N *but not* s^{N+1}, *then no quadratic FSR of length n generates* s^{N+1} *if and only if* $\mathrm{rank}(M(N+1, n)) = \mathrm{rank}(M(N, n))$.

An important issue is predicting the increment in the quadratic span when proposition 1 implies there has to be some increase. In the case of linear span, this increment is determined by theorem 1 of [6]. The precise increment for the quadratic case is an open question. However, the increment can be determined in terms of the rank of the matrix $M(N, L)$ if the quadratic FSR that generates s^N is in fact a linear FSR of length L. The following theorem gives the increment for this situation and can be viewed as a partial generalization to the quadratic case of theorem 1 in [6].

THEOREM 2. *If some linear FSR of length L generates the sequence* $\mathrm{s}^N = (s_0, s_1, \ldots, s_{N-1})$, *but no quadratic FSR of length L generates the sequence* $\mathrm{s}^{N+1} = (s_0, s_1, \ldots, s_N)$, *then any quadratic FSR that generates the latter sequence has length Q satisfying*

$$Q \geq N + 1 - \mathrm{rank}(M(N, L)).$$

Sequences with period 2^n generated by a nonlinear FSR of length n are called *de Bruijn sequences* of span n. We derive upper and lower bounds on the quadratic spans of de Bruijn sequences of span n. In these results, we consider matrices $M(2^n + q, q)$, which have 2^n rows, so that each column contains a period of the sequence. It then follows that the quadratic span of a de Bruijn sequence s is given by the smallest number q with the property that $(E^q\mathrm{s})^{2^n}$ is a linear combination of the columns of $M(2^n + q, q)$. The next proposition follows from the observation that the matrix $M(2^n + q, q)$ has rank at most 2^n and $\mathrm{rank}(M(2^n + n, n)) = \binom{n}{2} + n$.

PROPOSITION 3. *If* s *is a de Bruijn sequence of span* n, *then*

$$q(\mathrm{s}) \leq 2^n - \binom{n}{2}.$$

To improve the upper bound to $2^n - \binom{n}{2} - 1$ requires quite a bit of work. The idea is to show that $M(2^n + q, q)$ contains at least one more linearly independent column. We do this by proving the following lemma.

LEMMA 4. *Let* s *be a de Bruijn sequence of span* n, $n \geq 3$. *The column* $((E^0 \circ E^n)\mathrm{s})^{2^n}$ *in the matrix* $M(2^n + n + 1, n + 1)$ *is linearly independent of all the columns in* $M(2^n + n, n)$ *and the column* $(E^n \mathrm{s})^{2^n}$.

Lemma 4 is used to establish the improved upper bound stated below.

THEOREM 5. *If* s *is a de Bruijn sequence of span* $n, n \geq 3$, *then*

$$q(\mathrm{s}) \leq 2^n - \binom{n}{2} - 1.$$

We show that the upper bound given in theorem 5 is best possible by showing it is attained by the class of de Bruijn sequences formed by adding a zero term to the *m-sequences* — maximum-period sequences of period $2^n - 1$ generated by linear FSR's of length n. The proof uses theorem 2 to establish the increase in quadratic span when the discrepancy due to the added term is encountered.

THEOREM 6. *Let* s *be a de Bruijn sequence of span* n *obtained from an m-sequence of span* n *by adding the zero* n*-tuple. Then*

$$q(\mathrm{s}) = 2^n - \binom{n}{2} - 1.$$

For example, 000111101011001 is one period of an m-sequence of span 4. Then 0001111010110010 is a de Bruijn sequence of span 4 with maximum quadratic span 9.

It is easy to see that a lower bound on the quadratic span of a de Bruijn sequence of span n is $n + 1$. From our experimental results, we conjecture the following:

CONJECTURE. *The quadratic span of a de Bruijn sequence of span* n *is at least* $n + 2$, *for* $n > 3$.

Finally, we report the results of computer runs that determined the quadratic spans of the de Bruijn sequences of span n, n = 3, 4, 5, and 6. Our results for n = 4 and 5 correct the previously published results [2] for these cases. In each case, the $2^{2^{n-1}-n}$ de Bruijn sequences of span n were generated by considering all possible feedback functions. Each time a de Bruijn sequence was found, its linear and quadratic spans were determined and the results tallied. The linear spans were computed and compared with the results of [1] as a check.

The case n=6, with its 2^{26} de Bruijn sequences, presented the only real computational burden. In this case, we used several SUN workstations running in parallel, both at the MITRE Corporation and at Northeastern University to complete the run. The quadratic span distributions are listed in figure 2.

N = 3	
Quadratic Span	Number of Sequences
4	2

N = 4	
Quadratic Span	Number of Sequences
6	4
7	8
8	2
9	2

N = 5	
Quadratic Span	Number of Sequences
7	28
8	753
9	877
10	263
11	86
12	11
13	4
14	2
15	0
16	0
17	8
18	0
19	4
20	6
21	6

N = 6	
Quadratic Span	Number of Sequences
8	7
9	356
10	137869
11	5190500
12	1373661
13	962534
14	228087
15	75812
16	31376
17	12362
18	8919
19	5248
20	2220
21	960
22	529
23	223
24	223
25	100
26	72
27	36
28	26
29	44
30	72
31	22
32	6
33	6
34	12
35	8
36	6
37	4
38	2
39	6
40	2
41	2
42	0
43	0
44	8
45	0
46	12
47	6
48	6

LIST OF REFERENCES

1. A. H. Chan, R. A. Games and E. L. Key, "On the Complexity of de Bruijn Sequences," *J. Comb. Theory (A)* **33-3** (1982), 233–246.
2. H. Fredricksen, "A Survey of Full Length Nonlinear Shift Register Cycle Algorithms," *SIAM Review* **24** (1982), 195–221.
3. E. J. Groth, "Generation of Binary Sequences with Controllable Complexity," *IEEE Trans. on Inform. Theory* **IT-17** (1971), 288–296.
4. K. Imamura and W. Yoshida, "A Simple Derivation of the Berlekamp-Massey Algorithm and some Applications," *IEEE Trans. on Inform. Theory* **IT-33** (1987), 146–150.
5. E. L. Key, "An Analysis of the Structure and Complexity of Nonlinear Binary Sequence Generators," *IEEE Trans. on Inform. Theory* **IT-22** (1976), 732–736.
6. J. L. Massey, "Shift-Register Synthesis and BCH Decoding," *IEEE Trans. on Inform. Theory* **IT-15** (1969), 122–127.
7. R. A. Rueppel, *New Approaches to Stream Ciphers*, Swiss Federal Institute of Technology, Zurich, Switzerland: Ph.D. Thesis, 1984.
8. R. A. Rueppel and Staffelbach, "Linear Recurring Sequences With Maximum Complexity," *IEEE Trans. on Inform. Theory* **IT-33** (1987), 126–131.