On the Quadratic Span of Binary Sequences

A.M. Youssef and G. Gong
Center for Applied Cryptographic Research
Department of Combinatorics and Optimization
University of Waterloo, Waterloo, ON N2L 3G1
{a2youssef, ggong}@cacr.math.uwaterloo.ca

Abstract

The length of the shortest FSR that generates a sequence is called the span of the sequence. If the feedback function is linear, then the Berlekamp-Massey algorithm can be used to efficiently determine the the length of the shortest linear FSR that generate the sequence and its associated linear feedback function. However, for a general nonlinear feedback function, determining the span and an associated feedback function efficiently is difficult because of the nonlinearities involved. Because of its tractability, most of the current research has focused on studying the linear span of a sequence. However, a sequence with a large linear span may be generated by a much shorter feedback shift register with nonlinear feedback function. In this paper we study the quadratic span of binary sequences. We prove that (i) If the quadratic span of the sequence $s_0, s_1, \cdots, s_{n-1}$ is $> n/2$, then the quadratic span of the sequence $s_0, s_1, \cdots, s_n$ remains unchanged. Based on our experimental results, we conjecture the following: (ii) Let $N_n(q)$ be the number of binary sequences of length n and quadratic span $q > n/2$. Then $N_n(q)$ is a function of the difference $(n-q)$ only, i.e., $N_n(q) = N_{n+i}(q+i)$. (iii) For moderately large n, the expected value of the quadratic span of a randomly selected sequence of length n is given by $E(q_n) \approx \sqrt{2n}$.

Keywords: stream ciphers, shift registers, quadratic span, quadratic span profile

1. Introduction

Stream ciphers are an important class of encryption algorithms. By contrast to block ciphers which tend to simultaneously encrypt groups of characters of plaintext messages using fixed transformation, stream ciphers encrypt individual characters of the plaintext message one at a time using an encryption transformation which varies with time. Feedback shift registers (FSR) are attractive option for implementing stream ciphers because of their simple hardware circuity. There are different number of ways to introduce nonlinearity into sequences generated by shift registers [3]. One approach is to use nonlinear feedback function. The limitation of this approach is that it is theoretically difficult, if not virtually impossible, to predict the period of these sequences. A second approach is to use a nonlinear function to combine several sequences. The period of the combined sequence can be theoretically predicted in terms of the period of the individual component sequences. A third approach is to use a nonlinear function to combine the individual stages of a maximum length linear FSR. Another approach is to use clock-controlled shift registers in which shift registers are clocked in a quasi-random manner under the control of another shift register [1]. Irrespective of the approach used to generate the stream cipher sequence, an important criterion to measure the complexity of the resulting stream is the sequence span, which is the length of the shortest FSR that can generate the sequence. Most of the current research has been focused on studying the linear span [2]. However, a sequence with very large linear span may be generated by a much shorter FSR if the feedback function is allowed to have some nonlinear terms. In [4] [5] the authors studied the quadratic span of de Bruijn sequences. However the behavior of the quadratic span for a general finite length sequences or periodic sequences remains an open problem. In this paper, we study the quadratic span distribution and the quadratic span profile for randomly selected sequences.

2. Calculating the quadratic span

The m^{th}-span of a sequence can be determined by iteratively solving certain structured systems of Boolean linear equations. The Berlekamp-Massey algorithm [7] reduces the complexity of solving the systems involved in the linear case. It is an open problem whether the special structure of the matrix in the quadratic case can be utilized to reduce the complexity of calculating the quadratic span. An m-stage FSR with a feedback function $f : GF(2)^m \rightarrow GF(2)$ generates a sequence $s_0, s_1, \cdots, s_i, \cdots, s_{n-1}$ where $s_0, s_1, \cdots, s_{m-1}$ corresponds to the initial state of the shift register and $s_{i+m} = f(s_i, s_{i+1}, \cdots, s_{i+m-1}), i \geq 0$. The function f is called a quadratic function iff it can be represented as

$$f(x_0, x_1, \cdots, x_{q-1}) = \sum_{0 \leq i < q} a_i x_i + \sum_{0 \leq i < j < q} a_{i,j} x_i x_j.$$

A sequence is said to have quadratic span q if it can be generated by a q-stage FSR, but not with a $(q-1)$-stage FSR. To compute a quadratic feedback function of a q-stage FSR from a given sequence of n terms, we solve the following system of linear equation in the unknowns $a_i, 0 \leq i < q$ and $a_{i,j}, 0 \leq i < j < q$:

$$\begin{aligned} s_q &= f(s_0, s_1, \cdots, s_{q-1}) \\ s_{q+1} &= f(s_1, s_2, \cdots, s_q) \\ &\vdots \\ s_{n-1} &= f(s_{n-q-1}, s_{n-q}, \cdots, s_{n-2}). \end{aligned} \tag{1}$$

If the above set of equations has a solution, then the quadratic span of the sequence S is $\leq q$; otherwise it is $> q$. For example, if the sequence $(s_0, s_1, \cdots, s_7)$ has a quadratic span $n = 3$ then the following set of equations

$$\begin{pmatrix} s_0 & s_1 & s_2 & s_{0,1} & s_{0,2} & s_{1,2} \\ s_1 & s_2 & s_3 & s_{1,2} & s_{1,3} & s_{2,3} \\ s_2 & s_3 & s_4 & s_{2,3} & s_{2,4} & s_{3,4} \\ s_3 & s_4 & s_5 & s_{3,4} & s_{3,5} & s_{4,5} \\ s_4 & s_5 & s_6 & s_{4,5} & s_{4,6} & s_{5,6} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_{0,1} \\ a_{0,2} \\ a_{1,2} \end{pmatrix} = \begin{pmatrix} s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \end{pmatrix}$$

will have at least one solution while the set of equations

$$\begin{pmatrix} s_0 & s_1 & s_{0,1} \\ s_1 & s_2 & s_{1,2} \\ s_2 & s_3 & s_{2,3} \\ s_3 & s_4 & s_{3,4} \\ s_4 & s_5 & s_{4,5} \\ s_5 & s_6 & s_{5,6} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_{1,2} \end{pmatrix} = \begin{pmatrix} s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \end{pmatrix}$$

will not have any solution. So, in general, to calculate the quadratic span of a given sequence of length n, we start by assuming that the quadratic span q is equal to $n-1$ and form the set of $(n-q) \times (\frac{q(q+1)}{2} + 1)$ linear equations in (1). If these equations have a solution, we decrement q by one and iterate until we form a system of equations that doesn't have a solution. The quadratic span will equal to q of the previous step. One might use some simple search techniques (such as binary search) to speed up the calculations of the quadratic span. However, solving the set of linear equations for each iteration still an expensive operation for large matrix dimensions. Studying the behavior of the quadratic span jump profile is crucial in finding an efficient algorithm to determine the quadratic span.

3. Main Results

The following theorem provides a partial answer for the open problem of determining the jump in the quadratic span profile.

Theorem 1 *Let $s_0, s_1, \cdots, s_{N-1}$ be a binary sequence of length N where $N > 0$. Let q_n be a quadratic span of $s_0, s_1, \cdots, s_{n-1}$, $0 < n < N$. If $q_n > n/2$, then $q_{n+1} = q_n$.*

In order to prove this result, we need some results on linear algebra. In the followings let $\mathbf{s}^n = s_0, s_1, \cdots, s_{n-1}$, $0 < n < N$. Let $QS(\mathbf{s}^n)$ and $LS(\mathbf{s}^n)$ represent the quadratic span and the linear span of $\mathbf{s}^n$, respectively. We define the matrix $M(n-q, q)$ as the coefficient matrix associated with the systems of linear equations (2). I.e.,

$$M(n-q,q) = \begin{pmatrix} s_0 & s_1 & \cdots & s_{q-1} & s_0 s_1 & \cdots & s_0 s_{q-1} & \cdots s_{q-2} s_{q-1} \\ s_1 & s_2 & \cdots & s_q & s_1 s_2 & \cdots & s_1 s_q & \cdots s_{q-1} s_q \\ \cdots \\ s_{n-q-1} & s_{n-q} & \cdots & s_{n-2} & s_{n-q-1} s_{n-q} & \cdots & s_{n-q} s_{n n-2} & \cdots s_{n-3} s_{n-2} \end{pmatrix}$$

We write the first q columns as the matrix $A(n-q, q)$ (which is a Hankel matrix [8]) and the rest of columns as the matrix $\Delta(n-q, q)$. So we have

$$M(n-q,q) = (A(n-q,q), \Delta(n-q,q)).$$

Fact 1 *The quadratic span of $\mathbf{s}^n$, $0 < n < N$, is equal to q if and only if*

$$Rank(M(n-q,q)) = Rank((A(n-q,q+1), \Delta(n-q,q))$$

and q is the smallest number satisfies the above identity.

Lemma 1 *Let $\mathbf{s}^n$ have the quadratic span $q_n = q$, $0 < n < N$. If $q > n/2$, then*

$$Rank(M(n+1-q,q)) = 1 + Rank(M(n-q,q)).$$

Proof: We only need to prove that the last row of $M(n+1-q,q)$ is not a combination of the first $n-q$ rows of $M(n+1-q,q)$. If it is not true, we write $\alpha_0, \alpha_1, \cdots, \alpha_{n-q}$ as the row vectors of $M(n+1-q,q)$, then there exist $k_i \in \{0,1\}$ which are not all zeros such that $\alpha_t = \sum_{i=0}^{t-1} k_i \alpha_i$ where $t = n-q$. Therefore, we have

$$s_t = \sum_{i=0}^{t-1} k_i s_i,\, s_{t+1} = \sum_{i=0}^{t-1} k_i s_{i+1}, \cdots, s_{n-1} = \sum_{i=0}^{t-1} k_i s_{q-1+i}.$$

From the above identities, the linear span of $\mathbf{s}^n$ is less than or equal to t. Since $q > n/2$, then $t = n-q < n/2$. Thus

$$LS(\mathbf{s}^n) \le t < n/2.$$

But on the other hand, we have

$$LS(\mathbf{s}^n) \ge QS(\mathbf{s}^n) = q > n/2$$

which is a contradiction. Thus the last row of $M(n+1-q,q)$ is not a combination of the first $n-q$ rows of $M(n+1-q,q)$ which proves the Lemma.

Fact 2 *If $QS(\mathbf{s^n}) = q$ and*

$$Rank(A(n+1-q,q), \Delta(n-q,q)) = 1 + Rank(A(n-q,q), \Delta(n-q,q)), \tag{2}$$

then

$$Rank(A(n+1-q,q), \Delta(n-q,q)) = Rank(A(n+1-q,q+1), \Delta(n-q,q)). \tag{3}$$

Proof of Theorem 1. Let f_n be a quadratic Boolean function in $q_n = q$ variables which generates the sequence $\mathbf{s}^n = s_0, s_1, \cdots, s_{n-1}$. If f_n still generates $s_0, s_1, \cdots, s_{n-1}, s_n$, then $q_{n+1} = q_n$. If f_n does not generate $\mathbf{s}^{n+1}$, according Lemma 1, we have (2). From Fact 2, we get (3). Applying Fact 1, we obtain that $q_{n+1} = q = q_n$.

Remark 1 *Since the proof of theorem 1 doesn't use any information about the matrix $\Delta(n-q,q)$, then it follows that the result in theorem 1 applies to all higher order spans. Note that this result has been proved for linear span in [7] and the maximum order span in [6]. Therefore this result is true for all spans.*

In the rest of this section we give two conjectures supported by experimental results. Table 1 shows the number of sequences of length n and quadratic span q. Based on the results in this table, we have the following conjecture

Conjecture 1 *Let $N_n(q)$ be the number of binary sequences of length n and quadratic span $q > n/2$. Then $N_n(q)$ is a function of the difference $(n-q)$ only, i.e., $N_n(q) = N_{n+i}(q+i)$.*

Numerical value for the function $N_n(q)$ is given Table 2. By noting that the maximum period of the sequence generated by q-stage shift register is 2^q and that $q + q(q+1)/2$ bits of such a sequence will uniquely identify it, then $N_n(q)$ is a function of q only for $n \geq \max(2^q, q + q(q+1)/2)$.

q	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
n																									
24	1	2	11	105	1950	106129	3655217	5744240	3733790	1933617	899744	400789	17401	74384	31376	13048	5336	2140	838	314	114	40	13	4	1
23	1	2	11	105	1950	105608	2210847	2783812	1723099	866093	395794	173678	74384	31376	13048	5336	2140	838	314	114	40	13	4	1	
22	1	2	11	105	1950	104252	1280704	1336361	787000	384134	172208	74352	31376	13048	5336	2140	838	314	114	40	13	4	1		
21	1	2	11	105	1950	100836	708419	634921	355219	168457	74007	31376	13048	5336	2140	838	314	114	40	13	4	1			
20	1	2	11	105	1950	92128	372081	297899	158291	72913	31347	13048	5336	2140	838	314	114	40	13	4	1				
19	1	2	11	105	1950	71832	190409	137442	69588	31100	13048	5336	2140	838	314	114	40	13	4	1					
18	1	2	11	105	1950	50712	95169	62250	30122	13022	5336	2140	838	314	114	40	13	4	1						
17	1	2	11	105	1950	33430	46265	27711	12797	5336	2140	838	314	114	40	13	4	1							
16	1	2	11	105	1950	20740	21867	12083	5313	2140	838	314	114	40	13	4	1								
15	1	2	11	105	1946	12001	10100	5138	2140	838	314	114	40	13	4	1									
14	1	2	11	105	1874	6407	4540	2120	838	314	114	40	13	4	1										
13	1	2	11	105	1502	3276	1971	838	314	114	40	13	4	1											
12	1	2	11	105	1085	1585	821	314	114	40	13	4	1												
11	1	2	11	105	721	722	31	114	40	13	4	1													
10	1	2	11	105	433	300	114	40	13	4	1														
9	1	2	11	105	221	114	40	13	4	1															
8	1	2	11	81	103	40	13	4	1																
7	1	2	11	56	40	13	4	1																	
6	1	2	11	32	13	4	1																		
5	1	2	11	13	4	1																			
4	1	2	8	4	1																				
3	1	2	4	1																					
2	1	2	1																						
1	1	1																							

Table 1. Quadratic Span Distribution for $n \leq 24$

$n-q$	0	1	2	3	4	5	6	7	8	9	10	11
$N_n(q)$	1	4	13	40	114	314	838	2140	5336	13048	31376	74384

Table 2. $N_n(q)$ for $q \geq n/2$

Experimental average value of the quadratic span of randomly selected sequences of length $100 \geq n \geq 1$ is given in Figure 1. Based on the results of this experiment, we have the following conjecture:

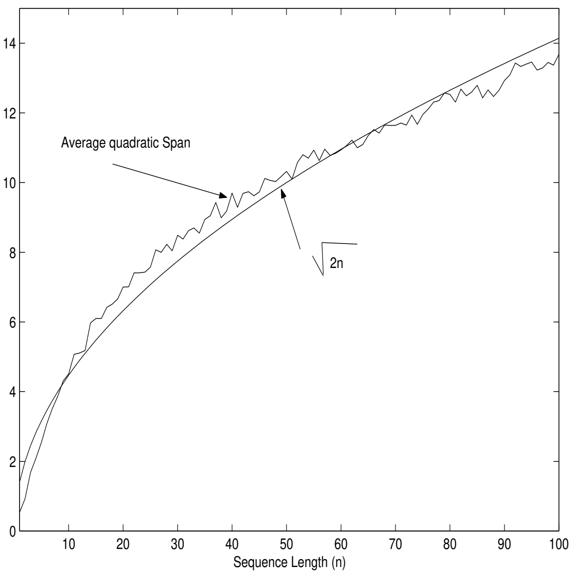

Figure 1. Average value of the quadratic span

Conjecture 2 *For moderately large n, the expected value of the quadratic span of a random sequence of length n is given by $E(q_n) \approx \sqrt{2n}$.*

For moderately large n, the expected value of the linear span of a random sequence of length n is given by $\approx n/2$ [2]. The expected value of the maximum order span is given by $\approx 2log_2(n)$ [6].

References

[1] S.W. Golomb, *Shift Register Sequences*, Aegean Park Press. Laguna Hills, California. 1982.

[2] R.A. Rueppel, *Analysis and Design of Stream Ciphers*, Springer Verlag, Berlin. 1986.

[3] D. Gollmann and W. G. Chambers, *Clock-controlled shift registers: a review*, IEEE Journal on Selected Areas in communications. Vol. 7. no. 4. pp. 525-533. May, 1989.

[4] A. H. Chan and R. A. Games, *On the Quadratic Span of Periodic Sequences*, Proceedings of Crypto'89. LNCS 435. pp. 82-89. 1989.

[5] A. H. Chan, R. A. Games and J. Rushanan, *On quadratic M-Sequences*, Proceedings of Fast Software Encryption. LNCS 809. pp. 166-173. 1994.

[6] C. J. A. Jansen and D. E. Boekee, *The shortest feedback shift register that generate a given sequence*, Proceedings of Crypto'89. LNCS 435. pp. 90-99. 1989.

[7] J. M. Massey, *Shift register synthesis and BCH decoding*, IEEE transactions on Information Theory. Vol. 15. no. 1. pp. 122-127. January, 1969.

[8] K.Imamura and W. Yoshida, *A simple derivation of Berlekamp-Massey algorithm and some applications*, IEEE transactions on Information Theory. Vol. 33. no. 1. pp. 146-150. January, 1987.