Aperiodic Linear Complexities of de Bruijn Sequences

Richard T.C.Kwok, M.Sc., and Maurice Beale, B.Sc., Ph.D.,
Electrical Engineering Department,
University of Manchester,
Manchester, M13 9PL,
England.

Extended Abstract.

Binary de Bruijn sequences of period 2^n bits have the property that all 2^n distinct n-tuples occur once per period. To generate such a sequence with an n-stage shift-register requires the use of nonlinear feedback. These properties suggest that de Bruijn sequences may be useful in stream ciphers. However, any binary sequence can be generated using a linear-feedback shift register (LFSR) of sufficient length. Thus, the *linear complexity* of a sequence, defined as the length of the shortest LFSR which generates it, is often used as a measure of the unpredictability of the sequence. This is a useful measure, since a well-known algorithm[1] can be used to successfully predict all bits of any sequence with linear complexity C from a knowledge of 2C bits. As an example, an m-sequence of period 2^n -1 has linear complexity C=n, which clearly indicates that m-sequences are highly predictable.

Now, the widely used definition of linear complexity stated above is open to different interpretations. We distinguish here between the *periodic linear complexity* (PLC) - the length of the shortest LFSR which generates the given sequence and then repeats it cyclically - and the *aperiodic linear complexity* (ALC) - the length of the shortest LFSR which generates the given sequence followed by any arbitrary sequence of bits. This distinction is not made in the literature on de Bruijn sequences, but it has important practical consequences. In a stream cipher, it is clearly undesirable for keystream sequences to be allowed to repeat. Consequently, no more than P bits (where P is the sequence period) will ever be used, which implies that it is the ALC, not the PLC, which is of real concern. Unfortunately, all of the published results on the linear complexity of de Bruijn sequences (e.g [2] and [3]) relate only to the PLC not the ALC. The research described here goes some way towards addressing this imbalance.

Having decided that the ALC is the most useful measure of unpredictability, we note however that a large value of the ALC of an entire sequence (one period) is not , by itself, a sufficient condition for high randomness. We also require the ALC of all sub-sequences of the given sequence to be as large as possible. For a given sequence of P bits, we are therefore interested in the ALC of the first k bits of the sequence, as a function of k ($1 \le k \le P$). The importance of this function was identified by Rueppel [4], who referred to it as the linear complexity profile (LCP) of the sequence. Note that, in general, the LCP of a sequence depends on the starting point

S. Goldwasser (Ed.): Advances in Cryptology - CRYPTO '88, LNCS 403, pp. 479-482, 1990.

within the sequence. Thus, if we consider all P cyclic shifts of any given sequence of P bits, some cyclic shifts may have notably better LCPs than others.

Although statistical results on the LCP of random binary sequences have been obtained [4], these authors are unaware of any published results on the LCPs of any class of finite deterministic sequences. The LCPs of de Bruijn sequences were therefore investigated and the results of this study are summarized below. For comparison, it is interesting to note that the expected value of the LCP of a random binary sequence (with equiprobable ones and zeros), as a function of the sub-sequence length k, is given by [4] :-

$$E[C] = k/2 + [(4+R_2(k)]/18 - 2^{-k}(k/3 + 2/9), \quad \ldots\ldots(1)$$

which rapidly approaches k/2 as k increases. (Here, $R_2(k)$ denotes the remainder when k is divided by 2). On the basis of this result and other observations, Rueppel[4] proposed that a "good random sequence" for cryptographic use should have a LCP which closely, but irregularly, follows the k/2 line. Also note that a linear complexity of k/2 for a sub-sequence of length k is sufficient to foil an attack based on the Berlekamp-Massey algorithm[1].

Now consider the aperiodic LCP of a de Bruijn sequence of period 2^n. As noted earlier, the LCP depends on the cyclic shift of the sequence under consideration. However, if we take the average value of the LCP of a de Bruijn sequence over all 2^n cyclic shifts, it is readily seen that for sub-sequence lengths $k \leq n$, equal numbers of all 2^k possible sub-sequences have been included in the averaging process. For such k, the average LCP is therefore identical to the expected value for random sequences, since the latter is also an ensemble average over all choices of sub-sequence of length k. Hence, we have the following :

Theorem : For $k \leq n$, the average LCP of any de Bruijn sequence of length 2^n, over all cyclic shifts, is identical to the average LCP for random sequences given by eqn (1).

For $k > n$, when all cyclic shifts of a fixed de Bruijn sequence are considered, only a subset of 2^n of the possible 2^k k-bit sub-sequences occur. Which of the 2^k possible sub-sequences occur depends on the de Bruijn sequence being considered. For this case, it has proved difficult to derive analytical results concerning the LCP. However, extensive numerical investigations have been carried out on the sets of de Bruijn sequences generated by Fredricksen's 'cross-join' algorithm[5]. Although this algorithm (in common with all other practical algorithms) generates only subsets of de Bruijn sequences, these subsets are large, containing 2^{2n-5} or 2^{2n-6} sequences of length 2^n, for odd and even n, respectively. Furthermore, an implementation in the form of a programmable nonlinear-feedback shift-register can be derived from this algorithm[6], making it an attractive choice for applications. The LCP investigations were carried out over all de Bruijn sequences generated by this algorithm for all $n \leq 12$.

Consider again the average LCP over all cyclic shifts of a given de Bruijn sequence. As a measure of the non-randomness of a sequence, we can take the difference between this average LCP and the ensemble average for random sequences in

eqn(1). For each of the de Bruijn sequences investigated, the fluctuations themselves appear to be random, and show no tendency to increase or decrease as a function of the sub-sequence length k. Typical results for de Bruijn sequences of length 512 and 4096 are shown in Figs.1 and 2. Of course the fluctuations are identically zero for $k \leq n$, as predicted by the previous Theorem, although this is clearly visible only in Fig.1 due to the scale. It is also apparent that the magnitude of the fluctuations decreases, albeit slowly, as the de Bruijn sequence length is increased.

Now, although the average LCP over all cyclic shifts of a sequence is of some interest for comparison with the ensemble average in eqn(1), an issue of greater practical concern is the LCP behaviour of fixed cyclic shifts of a sequence. In particular, one would like to know if there are any 'bad' cyclic shifts of a de Bruijn sequence which ought to be avoided. The results of our investigations suggest that this question can be answered in the negative, at least for all the de Bruijn sequences generated by Fridricksen's 'cross-join' algorithm for all $n \leq 12$. The LCPs of these de Bruijn sequences are relatively insensitive to the choice of cyclic shift and all appear to satisfy Rueppel's criterion for closely, but irregularly, following the k/2 line. A typical example of the LCP of a de Bruijn sequence of length 256 is shown in Fig.3. In this case, the cyclic shift chosen was that beginning with the all-zeros n-tuple. The steps in the LCP are all of the order of n or less in magnitude; indeed, the inevitable step associated with the all-zeros n-tuple is the largest present. This result appears to hold in general.

To illustrate the insensitivity of the LCP to the choice of cyclic shift of a de Bruijn sequence, Fig.4 shows typical results for the average, maximum and minimum values of the LCP, over all cyclic shifts of a 512-bit de Bruijn sequence. As expected, the average LCP is indistinguishable from the k/2 line. An interesting feature, which appears to hold in general, is that the maxima and minima of the LCP show a remarkable symmetry about its average value. More importantly, the peak deviation from the average (and, in effect, from the k/2 line) is small relative to the sequence length, and shows no tendency to increase or decrease as a function of the sub-sequence length k.

References

[1] Massey, J.L : " Shift-Register synthesis and BCH decoding". *IEEE Trans on IT*, Vol. IT-15,Jan,1969, pp.122-127.

[2] Etzion, T. & Lempel,A : "On the distribution of de Bruijn sequences of given complexity". *IEEE Trans on IT*, Vol.IT-30, no.4, July 1984, pp.611-614.

[3] Chan, A.H , Games, R.A & Key, E.L : " On the complexities of de Bruijn sequences". *J.Comb. Theory*, (Ser A), Vol.3, No.3, Nov.1982, pp.223-246.

[4] Rueppel, R.A: " New Approaches to Stream Ciphers". D.Sc. dissertation, No.ETH-7714, Swiss Federal Institute of Technology, Zurich, 1984.

[5] Fredricksen. H.M : " A survey of full length non-linear shift register cycle algorithms". *SIAM Review*, Vol.24, Apr.1982, pp.195-221.

[6] Beale, M., Cochrane.S.D & Lau ,S.M.S : "A programmable de Bruijn sequence generator for stream ciphers". *Proc.IEE Int. Conf.on Secure Communication Systems*, Oct. 1986.pp.69-73.

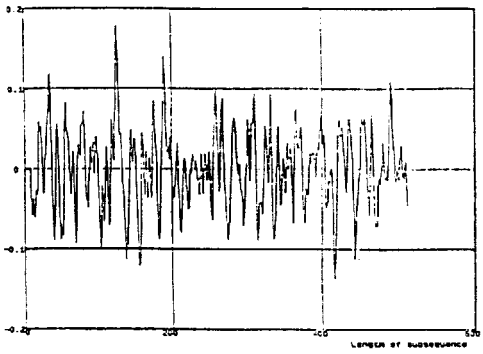

Fig.1 : Fluctuations of the average LCP (over all cyclic shifts) of a 512-bit de Bruijn sequence from the expected value for random sequences.

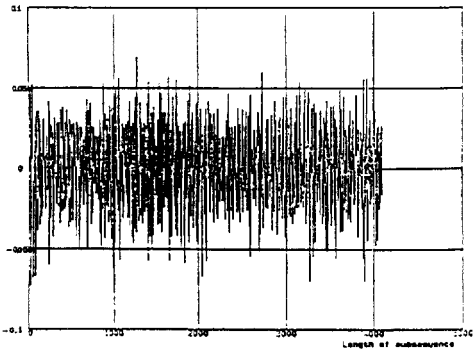

Fig.2 : As Fig.1, but for a de Bruijn sequence of length 4096.

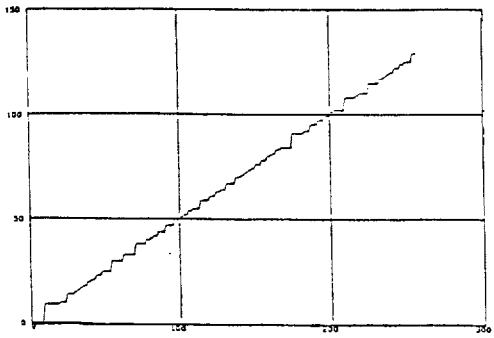

Fig.3: LCP of a de Bruijn sequence of length 256

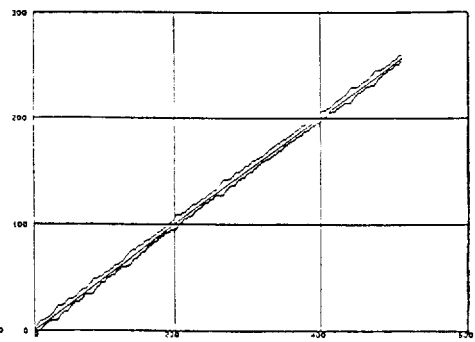

Fig.4: The average, maximum and minimum values of the LCP of a 512-bit de Bruijn sequence, over all cyclic shifts.