

2-Adic Shift Registers

Andrew Klapper* Mark Goresky†

February 23, 1994

Pseudorandom sequences, with a variety of statistical properties (such as high linear span, low autocorrelation and pairwise cross-correlation values, and high pairwise hamming distance) are important in many areas of communications and computing (such as cryptography, spread spectrum communications, error correcting codes, and Monte Carlo integration). Binary sequences, such as m -sequences, more general nonlinear feedback shift register sequences, and summation combiner sequences, have been widely studied by many researchers. Linear feedback shift register hardware can be used to relate certain of these sequences (such as m -sequences) to error correcting codes (such as first order Reed-Muller codes).

In this paper a new type of feedback register, feedback with carry shift registers (or FCSRs), will be presented. These relatively simple devices can be used to relate summation combiner sequences, arithmetic codes, and $1/q$ sequences. We describe an algebraic framework, based on algebra over the 2-adic numbers, in which the sequences generated by FCSRs can be analyzed, in much the same way that algebra over finite fields can be used to analyze LFSR sequences. As a consequence of this analysis, we present a method for cracking the summation combiner [9] which has been suggested for generating cryptographically secure binary sequences. In general, one must consider this "2-adic span" as a measure of security along with ordinary linear span. At the same time, FCSRs are a new, general, and therefore exciting, mechanism for generating sequences with enough structure for analysis. Many of the methods of nonlinearization that have been applied to linear feedback shift registers (LFSRs) can be applied to FCSRs, and some of these possibilities are described here. Hopefully, they will result in sequences with greater cryptologic security.

The many threads that are brought together by our analysis have analogues in the theory of LFSRs. In an LFSR, certain register cells are "tapped", their contents are added modulo 2 (using exclusive OR gates) and the sum is returned to the first cell of the shift register. Any periodic binary sequence may be realized as the output sequence from some LFSR with appropriate taps. Recall some of the well known concepts and consequences which are derived from this point of view.

1. The size of the smallest LFSR which generates a given periodic binary sequence is an important measure of the cryptographic security of the sequence. It is called the linear span of the binary sequence. The Berlekamp-Massey algorithm gives an efficient procedure for the construction of this smallest equivalent LFSR.
2. If two periodic binary pseudorandom sequences are combined by adding their terms modulo 2 (using exclusive OR gates) then the linear span of the resulting sequence is no more than the sum of the linear spans of the original two sequences.
3. The r taps q_1, q_2, \dots, q_r on the cells of an r -stage LFSR correspond to a *connection polynomial* $q(X) = q_r X^r + q_{r-1} X^{r-1} + \dots + q_1 X - 1$ with coefficients q_i in $\mathbf{Z}/(2)$. The period (and many other properties) of the output sequence can be analyzed using Galois the-

*Dept. of Computer Science, 915 POT, University of Kentucky, Lexington, KY 40506-0027, USA, klapper@cs.uky.edu. Project sponsored by the Natural Sciences and Engineering Research Council under Operating Grant OGP0121648 and the National Security Agency under Grant Number MDA904-91-H-0012. The United States Government is authorized to reproduce and distribute reprints notwithstanding any copyright notation hereon.

†Dept. of Mathematics, Northeastern University, Boston, MA 02115, USA, goresky@neu.edu

ory which interprets the roots of this polynomial as elements of the Galois field $GF(2^r)$. Moreover, every LFSR sequence \mathbf{a} with irreducible connectin polynomial can be written $a_i = \text{Trace}_1^r(A\gamma^i)$, where $A, \gamma \in GF(2^n)$.

4. An m -sequence is a LFSR sequence of maximum possible period $T = 2^r - 1$ (where the shift register has r stages). M -sequences have important distribution and correlation properties. It is a well known (but amazing) fact that the m -sequences are exactly those sequences generated by LFSRs whose taps correspond to *primitive* connection polynomials. Moreover, a LFSR sequence is an m -sequence if and only if the element γ in point (3) is a primitive element of $GF(2^n)$.
5. The $2^n - 1$ cyclic permutations of a single period of an m -sequence form the nonzero codewords of a (“punctured”) first order cyclic Reed-Muller code. These codes are of fundamental importance in coding theory and are prototypes of the general “finite geometry” codes.
6. Shift register analysis may be applied to nonbinary pseudorandom sequences. For example, linearly recurrent sequences $\mathbf{a} = a_0, a_1, \dots$ with entries $a_i \in \mathbf{Z}/(p)$ (where p is a large prime number) are used in the generation of pseudorandom numbers for Monte Carlo integration.

Feedback-with-carry shift registers (FCSRs) can be thought of as LFSRs with ordinary addition in place of addition modulo 2, and auxiliary memory for storing the carry. The contents (0 or 1) of the tapped cells of the shift register are added *as integers* to the current contents of the memory to form a sum, Σ . The parity bit ($\Sigma \pmod 2$) of Σ is fed back into the first cell, and the higher order bits ($\lfloor \Sigma/2 \rfloor$) are retained for the new value of the memory. See Figure 1.

Any periodic binary sequence may be generated by such a FCSR, and hence one may mimic the developments (1) through (7) above using FCSRs instead of LFSRs.

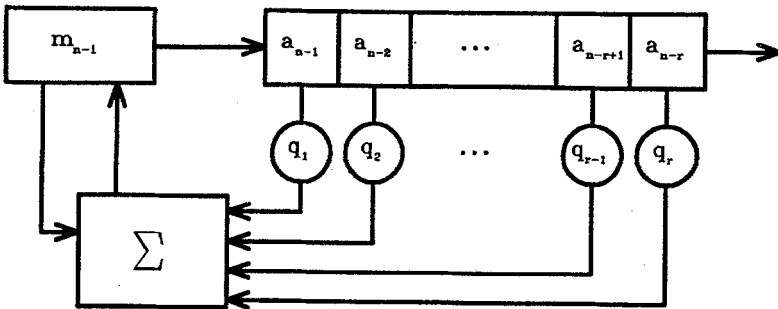


Figure 1: Feedback with Carry Shift Register

- 1'. **Definition** The size of the smallest FCSR which generates the periodic part of an eventually periodic sequence \mathbf{a} is called the 2-adic span of the sequence \mathbf{a} .

There is an analog (due to Mandelbaum [7]) of the Berlekamp-Massey algorithm. For any periodic binary sequence, this algorithm may be used to construct the smallest FCSR which generates the sequence.

- 2'. If two periodic binary sequences are added *with carry* operation then the 2-adic span of the resulting sequence is no more than the sum of the 2-adic spans of the original sequences. This fact and (1') above can be used to provide a cryptographic attack on the “summation combiners” described in [9].

- 3'. The taps q_1, q_2, \dots, q_r on the cells of a FCSR correspond to an integer $q = q_r 2^r + q_{r-1} 2^{r-1} + \dots + q_1 2 - 1$ which is called the *connection integer*. The period and other properties of the binary sequence are determined from number theoretic properties of this integer. Moreover, any FCSR sequence \underline{a} can be written $a_i = (A\gamma^i \pmod{q}) \pmod{2}$, where $A \in \mathbf{Z}$ and $\gamma = 2^{-1} \pmod{q}$.
- 4'. An ℓ -sequence is a FCSR sequence with maximum possible period $T = q - 1$ (where q is the connection integer of the FCSR). Such a sequence is the cyclic shift of the sequence formed by *reversing* the period of the binary expansion of the fraction $1/q$ and have been studied since the time of Gauss ([2, 3, 5]). These ℓ -sequences have remarkable distribution and correlation properties and are generated by connection integers q for which 2 is a *primitive root* modulo q , or, equivalently, if the number $\gamma = 2^{-1}$ in point (3') is a primitive root modulo q .
- 5'. The $q - 1$ cyclic permutations of a single period of a given ℓ sequence form the nonzero codewords of a Barrows-Mandelbaum [1, 6] cyclic arithmetic code.
- 6'. Two special cases of FCSR sequences with entries in $\mathbf{Z}/(p)$ (where p is a large prime number) have recently been proposed [8] for use in Monte Carlo problems.

Although they are not new, points (1'), (4'), (5') and (6') are diverse ideas that come together from the parallel study of LFSR sequences and FCSR sequences. The result is an interesting mathematical framework in which the analysis of sequences generated by a FCSR may be carried out. Suppose that $\mathbf{a} = \{a_0, a_1, a_2, \dots\}$ is an infinite periodic binary sequence. It is customary to associate to such a sequence the generating function (or Z-transform),

$$A(X) = \sum_{i=0}^{\infty} a_i X^i$$

which is considered as an element of the ring $\mathbf{Z}/(2)[[X]]$ of formal power series with coefficients in the integers modulo 2. As the following well known theorem [4] shows, this ring is the natural algebraic structure for the analysis of linear feedback shift registers:

Fact 1 *If the sequence \mathbf{a} is periodic, then its generating function is equal to a quotient of two polynomials, $A(X) = r(X)/q(X) \in \mathbf{Z}/(2)[[X]]$ and the denominator $q(X)$ is the connection polynomial for a linear feedback shift register which generates the sequence \mathbf{a} . (The numerator $r(X)$ determines the initial loading of the shift register).*

Many operations on pseudorandom sequences are conveniently described using operations in the ring $\mathbf{Z}/(2)[[X]]$. For example, suppose that $\mathbf{b} = \{b_0, b_1, b_2, \dots\}$ is another eventually periodic binary sequence, with generating function $B(X) = \sum_{i=0}^{\infty} b_i X^i = s(X)/t(X)$. Then the bit-wise exclusive-or sum of these two sequences, $\mathbf{c} = \{a_0 + b_0 \pmod{2}, a_1 + b_1 \pmod{2}, \dots\}$ has generating function

$$C(X) = A(X) + B(X) = \frac{r(X)t(X) + q(X)s(X)}{q(X)t(X)}.$$

Now suppose instead that the sequence \mathbf{c}' is obtained by adding the sequences \mathbf{a} and \mathbf{b} with *carry operation*, in other words,

$$\begin{aligned} c'_0 &= (a_0 + b_0) \pmod{2}, & m_1 &= (a_0 + b_0) \div 2 \\ c'_1 &= (a_1 + b_1 + m_1) \pmod{2}, & m_2 &= (a_1 + b_1 + m_1) \div 2 \end{aligned} \quad (1)$$

and so on. (Here, m_j is the bit carried from stage $j - 1$ to stage j , and \div denotes integer division, i.e. with remainder neglected.)

We model this addition with carry operation by associating to the infinite binary sequence \mathbf{a} the formal power series $\alpha = a_0 + a_1 2 + a_2 2^2 + \dots = \sum_{i=0}^{\infty} a_i 2^i$, and by associating to the infinite binary sequence \mathbf{b} the formal power series $\beta = \sum_{i=0}^{\infty} b_i 2^i$.

Such a power series does not converge in the usual sense but it can nevertheless be interpreted as defining an element in the ring \mathbf{Z}_2 of 2-adic integers. This ring consists of all formal power series $\sum_{i=0}^{\infty} s_i 2^i$ with $s_i \in \{0, 1\}$, and has been studied extensively by mathematicians for many years. The main difference between the two rings $\mathbf{Z}/(2)[[x]]$ and \mathbf{Z}_2 is that addition in \mathbf{Z}_2 is performed by “carrying” overflow bits to higher order terms, so that $2^i + 2^i = 2^{i+1}$. It follows that the formal power series $\gamma = c'_0 + c'_1 2 + c'_2 2^2 + \dots$ associated to the sum-with-carry sequence \mathbf{c}' is given by addition, $\gamma = \alpha + \beta \in \mathbf{Z}_2$ in the ring of 2-adic integers. Furthermore, the 2-adic integers are well suited to the study of feedback shift registers with carry, as the following result shows:

Fact 2 *If the sequence \mathbf{a} is periodic then the associated 2-adic integer is a quotient of two ordinary integers,*

$$\alpha = \sum_{i=0}^{\infty} a_i 2^i = \frac{-p}{q} \in \mathbf{Z}_2 \quad (2)$$

with $0 \leq p < q$. The denominator q will be odd. Furthermore, q is the connection integer for a FCSR which generates the sequence \mathbf{a} .

This result is the “with carry” analog of Fact 1 above and it explains exactly how to generate any periodic binary sequence using a FCSR: express the corresponding 2-adic number as a fraction p/q (which may involve computations in the ring \mathbf{Z}_2) then write $q = -1 + q_1 2 + q_2 2^2 + \dots + q_r 2^r$. The bits q_1, q_2, \dots, q_r specify which taps to incorporate in the FCSR. (The numerator p determines the initial loading of the shift register.) The 2-adic span of the sequence \mathbf{a} is therefore one less than the number of bits in the binary expansion of q , that is, $\lfloor \log_2(q+1) \rfloor$. (Actually one additional $\log_2(wt(q+1))$ bits of memory are also needed in the construction of the FCSR, where $wt(q+1)$ is the number of nonzero taps in the shift register.) The period of the periodic tail of the sequence is given by $T = \text{ord}_q(2)$, the smallest integer T such that $2^T - 1$ is divisible by q .

A specific consequence of this is that the sequences produced by summation combiners are vulnerable to attack. If \mathbf{a} and \mathbf{b} are two m -sequences, the summation combiner adds them *exactly as if they were the sequences of coefficients of 2-adic numbers* to produce \mathbf{c} . If \mathbf{a} represents p/q , and \mathbf{b} represents p'/q' , then their sum using the summation combiner represents $(p'q + pq')/(qq')$. This sequence can thus be output by a FCSR of size at most $\lfloor \log(qq' + 1) \rfloor \leq \lfloor \log(q+1) \rfloor + \lfloor \log(q'+1) \rfloor$. In other words, the 2-adic span of \mathbf{c} is at most the sum of the 2-adic spans of \mathbf{a} and \mathbf{b} , which is at most the sum of the periods of \mathbf{a} and \mathbf{b} . In contrast, the linear span of \mathbf{c} is approximately the product of the periods of \mathbf{a} and \mathbf{b} (cf. [9]). This means that \mathbf{c} is far more vulnerable to Mandelbaum’s variant of the Berlekamp-Massey algorithm than it is to the usual Berlekamp-Massey algorithm. Moreover, the more sequences are added, the worse things get.

There are many remaining questions concerning FCSRs. Virtually all questions of LFSRs that have been asked can be asked of FCSRs. These include questions concerning 2-adic span profile, how sequences can be generated with large 2-adic span (as well as large linear span), and the distributional properties of FCSR sequences. Various generalizations of FCSRs are possible. For example, the 2-adic numbers can be replaced by a totally ramified extension of the 2-adics (this corresponds to an addition with a carry that jumps several positions), giving rise to another type of binary sequence generator. The 2-adics can also be replaced by an unramified extension of the 2-adics, giving rise to sequences over finite fields of characteristic two. It may be possible to apply nonlinear operations to such sequences to produce binary sequences with large 2-adic span.

References

- [1] J. T. BARROWS, JR., A new method for constructing multiple error correcting linear residue codes, Rep. R-277, Coordinated Sci. Lab., Univ. of Illinois, Urbana, 1966.
- [2] L. BLUM, M. BLUM, AND M. SHUB, A simple unpredictable pseudo-random number generator, *Siam J. Comput.* vol. 15, 1986 pp. 364-383.
- [3] C. F. GAUSS, *Disquisitiones Arithmeticae*, 1801; reprinted in English translation by Yale Univ. Press, New Haven, CT. 1966.
- [4] S. GOLOMB *Shift Register Sequences*. Aegean Park Press, Laguna Hills CA, 1982.
- [5] D. KNUTH, *The Art of Computer Programming, Vol 2. Seminumerical Algorithms*. Addison-Wesley, Reading MA, 1981.
- [6] D. MANDELBAUM, Arithmetic codes with large distance. *IEEE Trans. Info. Theory*, vol. IT-13, 1967 pp. 237-242.
- [7] D. MANDELBAUM, An approach to an arithmetic analog of Berlekamp's algorithm. *IEEE Trans. Info. Theory*, vol. IT-30, 1984 pp. 758-762.
- [8] G. MARSAGLIA AND A. ZAMAN, A new class of random number generators, *Annals of Applied Probability*. vol. 1, 1991 pp. 462-480.
- [9] R. RUEPPEL *Analysis and Design of Stream Ciphers*. Springer Verlag, New York, 1986.