Periodicity and Distribution Properties of Combined FCSR Sequences

Mark Goresky[1,⋆] and Andrew Klapper[2,⋆⋆]

[1] Institute for Advanced Study, Princeton NJ
www.math.ias.edu/~goresky
[2] Dept. of Computer Science, University of Kentucky, Lexington KY
www.cs.uky.edu/~klapper

Abstract. This is a study of some of the elementary statistical properties of the bitwise exclusive or of two maximum period feedback with carry shift register sequences. We obtain conditions under which the resulting sequences has the maximum possible period, and we obtain bounds on the variation in the distribution of blocks of a fixed length. This may lead to improved design of stream ciphers using FCSRs.

Keywords: Feedback with carry shift register, pseudorandom sequence, stream cipher.

1 Introduction

The *summation combiner* [8] is a stream cipher in which two binary m-sequences are combined using addition-with-carry. This cipher attracted considerable attention during the 1980's because it was fast, simple to construct in hardware, and the linear span of the resulting combined sequence was known to approach its period, which is approximately the product of the periods of the constituent sequences.

The security of the summation combiner was called into question following the introduction of feedback-with-carry shift registers, or FCSRs [4], [5], and the associated rational approximation algorithm [5]. This is because the *2-adic complexity* of the output of the summation combiner is no more than the sum of the 2-adic complexities of the constituent sequences. Nevertheless, the summation combiner remains an interesting and difficult to analyze procedure for generating pseudorandom sequences and many basic questions concerning this combiner have never been satisfactorily addressed.

One might just as well consider the reverse procedure, and combine two binary FCSR sequences using binary addition ("XOR"). Sequences of this type are just as difficult to analyze, which perhaps explains why they have been largely ignored despite having been suggested ten years ago [5], [9].

⋆ Partially supported by DARPA grant no. HR0011-04-1-0031.
⋆⋆ Partially supported by N.S.F. grant no. CCF-0514660.

G. Gong et al. (Eds.): SETA 2006, LNCS 4086, pp. 334–341, 2006.

Recall that a binary ℓ-sequence is a maximal length FCSR sequence [4] of 0's and 1's. Such a sequence is obtained whenever the *connection integer* $q \geq 3$ is a prime number such that 2 is a primitive root modulo q. The period of such an ℓ-sequence is $q-1$ and it is known to have a number of desirable statistical properties, one of which is that the number of occurrences of any given block $f = (f_0, f_1, \cdots, f_{k-1})$ of size k differs at most by one, as f ranges over all 2^k possibilities [4].

In this paper we consider "combining" two distinct ℓ-sequences $\mathbf{a} = (a_0, a_1, \cdots)$ and $\mathbf{b} = (b_0, b_1, \cdots)$ using addition modulo 2 (or "XOR", denoted $\oplus$) to obtain a sequence $\mathbf{c} = (c_0, c_1, \ldots)$ with $c_j = a_j \oplus b_j$. Suppose $\mathbf{a}$ is the ℓ-sequence that is generated by an FCSR with connection polynomial q and that $\mathbf{b}$ is the ℓ-sequence that is generated by an FCSR with connection polynomial r. We are interested in the resulting sequence $\mathbf{c}$, perhaps as a possible constituent in a stream cipher — there is experimental evidence (not reported on in this paper) that the 2-adic complexity is close to half its period.

We first show that the combined sequence $\mathbf{c}$ will have maximal period if one of the periods, say, $q-1$ is divisible by 4, if the other period, $r-1$ is not divisible by 4, and if no odd prime divides both.

We also consider the distribution properties of these sequences. That is, we bound the number of occurrences of each block of size k within such a sequence. We are able to show that by careful choice of the constituent sequences it is possible to guarantee good distribution properties for the resulting combined sequence. The precise statement is given in Theorem 3.

2 Recollections on Binary FCSR Sequences

Let $q > 2$ be a prime number, the *connection integer*. Let $s = \mathrm{ord}_q(2)$ be the smallest integer such that $2^s \equiv 1 \pmod{q}$ or equivalently, such that q divides $2^s - 1$.

For any integer h, with $0 \leq h < q$, the base-2 expansion of the fraction h/q will be periodic with (minimal) period s. It is a *binary sequence*, meaning that its symbols are taken from the alphabet $\Sigma = \mathbf{Z}/(2)$. These sequences have been studied since the time of Gauss [3], [2] (p. 163). The reverse of this sequence is known as an *FCSR sequence* [4], [5] since it is the output sequence of a *feedback with carry shift register* with connection integer q, with cell contents taken from $\mathbf{Z}/(2)$, and with initial loading that depends on h, cf. [5]. This FCSR sequence can also be described as the 2-adic expansion of the fraction $-h/q$. To be explicit, let $0 \leq h \leq q$ and suppose the 2-adic expansion

$$-\frac{h}{q} = a_0 + a_1 2 + a_2 2^2 + \cdots \tag{1}$$

(with $a_i \in \{0, 1\}$) is periodic with period s. Then the sequence $\mathbf{a} = a_0, a_1, \cdots$ is an FCSR sequence. Its reverse is the base 2 expansion of the fraction h/q :

$$\frac{h}{q} = \frac{a_{s-1}}{2} + \frac{a_{s-2}}{2^2} + \cdots + \frac{a_0}{2^{s-1}} + \frac{a_{s-1}}{2^s} + \cdots \tag{2}$$

as may easily be seen by summing the geometric series in (1) and (2).

The period s of such a sequence satisfies $0 \le s \le q-1$. The period is maximal ($s = q-1$) if and only if 2 is a *primitive root* modulo q, meaning that the distinct powers 2^j modulo q, account for all the nonzero elements in $\mathbf{Z}/(q)$. In this case the base 2 expansion of h/q is known as a $1/q$ *sequence* [1] or as a Barrows-Mandelbaum codeword [7]. Its reverse, the corresponding FCSR sequence, is known as a (binary) *ℓ-sequence*.

It is also known [5] that there exists $B \in \mathbf{Z}/(q)$ (the choice of which depends on the value of h) such that

$$a_j = B2^{-j} \pmod q \pmod 2 \tag{3}$$

for all j, meaning that first $B2^{-j} \in \mathbf{Z}/(q)$ is computed; this number is represented as an integer between 0 and $q-1$, and it is then reduced modulo 2. The $q-1$ possible different non-zero choices of $B \in \mathbf{Z}/(q)$ give cyclic shifts of the resulting sequence $\mathbf{a}$, and this accounts for all the binary ℓ-sequences with connection integer q. The following fact was observed over a hundred years ago [2, p. 163].

Lemma 1. *Let $\mathbf{a} = a_0, a_1, a_2, \cdots$ be the binary ℓ-sequence corresponding to the fraction $-h/q$ where 2 is primitive modulo the (odd) prime q, and where $0 < h < q$. Then*

$$a_{j+\frac{q-1}{2}} \equiv q - a_j \equiv q_0 - a_j \pmod 2,$$

where $q_0 = q \pmod 2$. In other words, within any period of the ℓ-sequence $\mathbf{a}$, the second half is the complement of the first half, [6].

Proof. Since 2 is primitive mod q, we have: $2^{q-1} \equiv 1 \pmod q$ hence $2^{\frac{q-1}{2}} \equiv -1 \pmod q$ so $2^{-\frac{q-1}{2}} \equiv -1 \pmod q$. It suffices to prove the lemma for any single shift of the sequence $\mathbf{a}$. Accordingly, we may take $B = 1$ in equation (3), then calculate

$$\begin{aligned} a_{j+\frac{q-1}{2}} &\equiv -2^{-j} \pmod q \pmod 2 \\ &\equiv (q - 2^{-j}) \pmod q \pmod 2 \end{aligned}$$

If $A_j \in \{1, 2, \cdots, q-1\}$ is the positive integer representation of the number $2^{-j} \pmod q \in \mathbf{Z}/(q)$ then $0 < q - A_j < q$ so $q - A_j$ is the positive integer representation of the number $q - 2^{-j} \pmod q \in \mathbf{Z}/(q)$. Therefore, reducing this equation modulo 2 gives

$$a_{j+\frac{q-1}{2}} \equiv q_0 - a_j \pmod 2$$

where $q_0 = q \pmod 2 \in \mathbf{Z}/(2)$.

3 Period

In this section we describe a very general criterion which guarantees that the period of a sequence $\mathbf{c}$ obtained by "combining" two periodic sequences $\mathbf{a}, \mathbf{b}$ is the least common multiple of the periods of $\mathbf{a}$ and $\mathbf{b}$. It would surprise us to find that this theorem is unknown, but we are not aware of its having appeared in print.

Let Σ be an alphabet (i.e., a finite set). Let $\odot$ be a binary operation on Σ. That is, $\odot : \Sigma \times \Sigma \to \Sigma$. We write $a \odot b$ for the value of $\odot$ at (a, b).

Definition 1. *The operation $\odot$ is* cancellative *if for all $a, b, c \in \Sigma$, if $a \odot b = a \odot c$, then $b = c$.*

Theorem 1. *Let $\mathbf{a} = (a_0, a_1, \cdots)$ be a periodic sequence of (minimal) period n with each $a_i \in \Sigma$, and let $\mathbf{b} = (b_0, b_1, \cdots)$ be a periodic sequence of (minimal) period m with each $b_i \in \Sigma$. Let $\mathbf{c} = (c_0, c_1, \cdots)$ be the sequence with $c_i = a_i \odot b_i$ for each i. Suppose that for every prime r, the largest power of r that divides n is not equal to the largest power of r that divides m. Then $\mathbf{c}$ is periodic and the period of $\mathbf{c}$ is the least common multiple of n and m.*

Proof. It is straightforward to see that $\mathbf{c}$ is periodic and its (least) period divides the least common multiple of n and m. Let t denote the (least) period of $\mathbf{c}$. Suppose that $t < \mathrm{lcm}(n, m)$. Then there is some prime r so that t divides $\mathrm{lcm}(n, m)/r$. In particular, $\mathbf{c}$ has $\mathrm{lcm}(n, m)/r$ as a period.

Suppose that the largest power of r dividing n is r^e and the largest power of r dividing m is r^f. By symmetry we may assume that $e < f$. Thus the largest power of r dividing $\mathrm{lcm}(n, m)/r$ is r^{f-1}, so n divides $\mathrm{lcm}(n, m)/r$ and m does not divide $\mathrm{lcm}(n, m)/r$. For every i we have

$$\begin{aligned} a_i \odot b_i &= c_i \\ &= c_{i+\mathrm{lcm}(n,m)/r} \\ &= a_{i+\mathrm{lcm}(n,m)/r} \odot b_{i+\mathrm{lcm}(n,m)/r} \\ &= a_i \odot b_{i+\mathrm{lcm}(n,m)/r}. \end{aligned}$$

By the cancellative property of $\odot$, it follows that for every i,

$$b_i = b_{i+\mathrm{lcm}(n,m)/r}.$$

But this contradicts the fact that $\mathrm{lcm}(n, m)/r$ is not a multiple of the minimal period of $\mathbf{b}$, and thus proves the theorem. □

Corollary 1. *Let $\mathbf{a} = (a_0, a_1, \cdots)$, $\mathbf{b} = (b_0, b_1, \cdots)$ be binary ℓ-sequences with connection integers q and r respectively. Suppose that 4 divides $q - 1$ but does not divide $r - 1$ and that no odd prime divides both $q - 1$ and $r - 1$ (so that $\gcd(q - 1, r - 1) = 2$). Then the sequence $\mathbf{c} = \mathbf{a} \oplus \mathbf{b} \pmod p$ obtained by taking the termwise sum, modulo 2, of $\mathbf{a}$ and $\mathbf{b}$ has period $(q - 1)(r - 1)/2$.*

4 Distributions

By an *occurrence* of a block $e = (e_0, \cdots, e_{k-1})$ in a sequence $\mathbf{a}$ of period n we mean an index i, $0 \le i < n$ so that $a_i = e_0, a_{i+1} = e_1, \cdots, a_{i+k-1} = e_{k-1}$. Recall the following result of [1] (Theorem 1). See also [5].

Theorem 2. *Let $\mathbf{a} = (a_0, a_1, \cdots)$ be a binary ℓ-sequence with connection integer q. Then the number of occurrences of any block $e = (e_0, e_2, \cdots, e_{k-1})$ of size k in $\mathbf{a}$ varies at most by 1 as the block e varies over all 2^k possibilities. That is, there is an integer w so that every block of length k occurs either w times or $w+1$ times in $\mathbf{a}$. The number of blocks of length k that occur $w+1$ times is $q - 1 \pmod{2^k}$, and the number of blocks of length k that occur w times is $2^k - (q - 1 \pmod{2^k})$.*

Proof. The first statement is explicitly given in [1] Theorem 1 (for the corresponding $1/q$ sequence). The second statement follows immediately: let Q be the number of blocks of length k that occur $w+1$ times in $\mathbf{a}$. Then

$$\begin{aligned} q - 1 &= Q(w+1) + (2^k - Q)w \\ &= 2^k w + Q. \end{aligned}$$

It follows that $Q = q - 1 \pmod{2^k}$, as claimed. □

Throughout the remainder of this section we fix prime numbers q and r such that 2 is a primitive root modulo q and also modulo r. Let $\mathbf{a} = (a_0, a_1, \cdots)$ and $\mathbf{b} = (b_0, b_1, \cdots)$ be binary ℓ-sequences with connection integers q and r respectively, (and thus periods $q-1$ and $r-1$ respectively). We will further assume that 4 divides $q-1$, and that 4 does not divide $r-1$, so that $\gcd(q-1, r-1) = 2$. Let $\mathbf{c} = \mathbf{a} \oplus \mathbf{b}$ be the sequence obtained as sum, modulo 2 (or the exclusive or) of these two sequences: $c_i = a_i \oplus b_i \pmod 2$. According to Corollary 1, the period of the sequence $\mathbf{c}$ is maximal, and is equal to $(q-1)(r-1)/2$.

Lemma 2. *Let $0 \le i < q-1$ and $0 \le j < r-1$. Then in a full period of $\mathbf{c}$, a_i is combined with b_j if and only if j and i have the same parity. That is, there are integers k and l with $i + k(q-1) = j + l(r-1)$ if and only if $i \equiv j \pmod 2$.*

Proof. This is an application of the Euclidean theorem. The integer 2 is the greatest common divisor $q-1$ and $r-1$. The integers i and j have the same parity if and only if $i - j$ is a multiple of 2, which by the Euclidean theorem is equivalent to the existence of k and l. □

Lemma 3. *Within any single period, the second half of the sequence $\mathbf{c} = \mathbf{a} \oplus \mathbf{b}$ is the complement of the first half.*

Proof. The second half of a period of the sequence $\mathbf{a}$ is the complement of the first half and the same is true for the sequence $\mathbf{b}$. Let $T = (q-1)(r-1)/2$ be the period of $\mathbf{c}$. Then

$$\frac{T}{2} = \frac{q-1}{2} \cdot \frac{r-1}{2} = \frac{q-1}{2} \cdot \text{ odd } = \frac{r-1}{2} \cdot \text{ even.}$$

Therefore $a_{j+T/2} = \overline{a}_j$ and $b_{j+T/2} = b_j$ whenever $0 \le j < T/2$. Here, $\overline{a}_j$ denotes the complement of $a_j \in \mathbb{Z}/(2)$. Hence, for these values of j,

$$c_{j+T/2} = \overline{a}_j \oplus b_j = \overline{c}_j$$

which proves the lemma. □

Theorem 3. *Fix $k \ge 0$. Let $Q = q - 1 \pmod{2^k}$ and let $R = r - 1 \pmod{2^k}$. Define*

$$s = \frac{\min(Q, R) - \max(0, Q + R - 2^k)}{2}.$$

Then the number of occurrences of a block $e = (e_0, e_2, \cdots, e_{k-1})$ of size k in the sequence $\mathbf{c} = \mathbf{a} \oplus \mathbf{b}$ varies at most by s as the block e varies over all 2^k possibilities.

Proof. Let $\mathbf{b}^{(1)} = (b_1, b_2, \cdots)$ be the shift of the sequence $\mathbf{b}$ by one. Then we claim that the sequence

$$\mathbf{d} = \mathbf{a} \oplus \mathbf{b}^{(1)}$$

is a shift of the sequence $\mathbf{c} = \mathbf{a} \oplus \mathbf{b}$.

To prove this claim, note that because $(r-3)/2$ is even and $\gcd(r-1, q-1) = 2$, there exist integers ℓ and m such that

$$\frac{r-3}{2} = m(q-1) - \ell(r-1).$$

That is,

$$m(q-1) = \frac{r-3}{2} + \ell(r-1).$$

Therefore, for all j,

$$\begin{aligned} d_{j+m(q-1)} &= a_{j+m(q-1)} \oplus b^{(1)}_{j+\frac{r-3}{2}+\ell(r-1)} \\ &= a_{j+m(q-1)} \oplus b_{j+\frac{r-1}{2}+\ell(r-1)} \\ &= a_j \oplus b_{j+\frac{r-1}{2}} \\ &= a_j \oplus \overline{b}_j \\ &= \overline{c}_j. \end{aligned}$$

since $\mathbf{d}$ is obtained by shifting $\mathbf{b}$ by one before adding it to $\mathbf{a}$. By Lemma 3 the sequence $\mathbf{c}$ is a shift of its complement, so $\mathbf{d}$ is also a shift of $\mathbf{c}$.

Therefore, if we count the occurrences of each block of a fixed length k in both $\mathbf{c}$ and $\mathbf{d}$, then for each block we will have exactly twice the number of occurrences of that block in $\mathbf{c}$. However, in the construction of these two sequences, each occurrence of each block of length k in $\mathbf{a}$ is matched with each occurrence of each block of length k in $\mathbf{b}$. Thus to count the occurrences of a block e of length k in $\mathbf{c}$, we want to count the number of pairs (i, j) where i is an occurrence of

a block f in $\mathbf{a}$, j is an occurrence of a block g in $\mathbf{b}$, and $f \oplus g = e$. That is, $g = f \oplus e$. Thus we sum over all blocks f of length k the number of occurrences of f in $\mathbf{a}$ times the number of occurrences of $f \oplus e$ in $\mathbf{b}$.

Let w denote the minimum number of occurrences of a block of length k in $\mathbf{a}$, so that by Theorem 2 every possible block of length k occurs either w or $w+1$ times. Similarly, let z denote the minimum number of occurrences of a block of length k in $\mathbf{b}$, so that every possible block of length k occurs either z or $z+1$ times. For a fixed block e of length k, as we have seen, the occurrences of a block f of length k in $\mathbf{a}$ are matched with the occurrences of block $e \oplus f$ in $\mathbf{b}$. There are four possibilities:

1. f occurs w times in $\mathbf{a}$ and $e \oplus f$ occurs z times in $\mathbf{b}$;
2. f occurs $w+1$ times in $\mathbf{a}$ and $e \oplus f$ occurs z times in $\mathbf{b}$;
3. f occurs w times in $\mathbf{a}$ and $e \oplus f$ occurs $z+1$ times in $\mathbf{b}$;
4. f occurs $w+1$ times in $\mathbf{a}$ and $e \oplus f$ occurs $z+1$ times in $\mathbf{b}$.

Let Y_i denote the number of fs in case i above, $i = 1, 2, 3, 4$. Then the number of occurrences of e in $\mathbf{c}$ is

$$N_e = \frac{wzY_1 + (w+1)zY_2 + w(z+1)Y_3 + (w+1)(z+1)Y_4}{2}. \tag{4}$$

We have $Y_2 + Y_4 = Q$ since cases (2) and (4) together account for all the blocks f that occur $w+1$ times in $\mathbf{a}$. Similarly, $Y_3 + Y_4 = R$, and $Y_1 + Y_2 + Y_3 + Y_4 = 2^k$. Thus $Y_1 = 2^k - Q - R + Y_4$, $Y_2 = Q - Y_4$, and $Y_3 = R - Y_4$. Therefore, substituting these values into (4) gives

$$N_e = \frac{wz2^k + zQ + wR + Y_4}{2}.$$

It follows that the possible variation in N_e is one half the possible variation in Y_4. By the definition of Y_4 we have $Y_4 \le \min(Q, R)$ and $Y_4 \ge 0$. Also, $Y_2 \le 2^k - R$, so that $Y_4 = Q - Y_2 \ge Q + R - 2^k$. It follows that the possible variation in Y_4 for various e is at most

$$\min(Q, R) - \max(0, Q + R - 2^k).$$

The theorem follows immediately from this. □

Corollary 2. *The sequence* $\mathbf{c}$ *is balanced and the distribution of consecutive pairs in* $\mathbf{c}$ *is uniform.*

Proof. Balance follows from the case of Theorem 3 when $k = 1$. The uniform distribution of pairs follows from Theorem 3 with $k = 2$. In both cases the bound s in the theorem equals zero. □

It follows from Theorem 3 that the sequence $\mathbf{c} = \mathbf{a} \oplus \mathbf{b}$ is highly uniform if $\min(Q, R) - \max(0, Q + R - 2^k)$ is small for all small k.

A small amount of experimental evidence indicates that this bound is very close to optimal, in the sense that there are blocks of length k whose numbers of occurrences differ by almost $\min(Q, R) - \max(0, Q + R - 2^k)/2$. Further experimentation is planned.

5 Conclusions

It is apparent from these results how to look for pairs of ℓ-sequences whose exclusive ors have large period and for small k have near uniform distribution of blocks of length k. This situation is an improvement over the situations for many sequence generators that have been proposed previously as components of stream ciphers – in many cases the period has not even been computed. On the basis of experimentation we believe that our exclusive or sequences have other good properties such as large 2-adic complexity. Before they are used as components in stream cipher construction, however, we need to test them with the NIST test suite and examine their resistance to other attacks such as correlation attacks and algebraic attacks.

References

1. L. Blum, M. Blum, and M. Shub, A simple unpredictable pseudorandom number generator, SIAM J. Comput. **15** (1986), 364-383.
2. L. E. Dickson, *History of the Theory of Numbers*, vol. 1, Chelsea, New York, 1950.
3. C. F. Gauss, *Disquisitiones Arithmeticae*, Leipzig, 1801, English translation, Yale, New haven, 1966.
4. A. Klapper and M. Goresky, 2-adic shift registers, in *Fast Software Encryption: Proceedings of 1993 Cambridge Algorithms Workshop*, Lecture Notes in Computer Science **809**, Springer Verlag, 1994, 174-178.
5. A. Klapper and M. Goresky, Feedback Shift Registers, Combiners with Memory, and 2-Adic Span, *Journal of Cryptology* **10** (1997), 111-147.
6. A. Klapper and M. Goresky, Arithmetic crosscorrelation of feedback with carry shift registers, *IEEE Trans. Info. Theory* **43** (1997), 1342-1345.
7. D. Mandelbaum, Arithmetic codes with large distance, *IEEE Trans. Info. Theory* **IT-13** (1967), 237-242.
8. R. Rueppel, *Analysis and Design of Stream Ciphers.* Springer Verlag, New York, 1986.
9. B. Schneier, *Applied Cryptography.* John Wiley & Sons, New York, 1996.