Efficient Prediction of Marsaglia–Zaman Random Number Generators

Eric Bach

Abstract—We show that the random number generator of Marsaglia and Zaman produces the successive digits of a rational b -adic number. (The b -adic number system generalizes p -adic numbers to an arbitrary integer base b .) Using continued fractions, we derive an efficient prediction algorithm for this generator.

Index Terms—Continued fractions, inductive inference, p -adic numbers, pseudorandom sequences.

I. INTRODUCTION

Marsaglia and Zaman [13] have suggested a new pseudorandom number generator with the following features. Elements of the pseudorandom sequence are base- b digits. Each digit in the sequence is a truncated sum or difference of two earlier digits, combined with the carry or borrow generated by the formation of the immediately preceding digit. (See Section III for a precise definition.) Our purpose is to make two observations:

- The successive pseudorandom digits are nothing other than digits of certain rational b -adic numbers.
- Because of this, there is an efficient algorithm to predict an unknown pseudorandom sequence of this type.

This shows that Marsaglia–Zaman generators are cryptographically insecure, in the sense of [2]. Although this does not affect their utility for Monte Carlo experiments, it does suggest that using them as a replacement for a random sequence in a one-time pad is ill-advised.

Our prediction algorithm relies on an integer analog to the continued fraction method of decoding Bose–Chaudhuri–Hocquenghem (BCH) codes [15], [17]. It is also related to a prediction algorithm discussed by Blum, Blum, and Shub [2] for their “ $1/p$ ” generator. We note that linear congruential generators are also easy to predict, but different techniques are required [4], [9].

II. THE b -ADIC NUMBER SYSTEM

In this section we review the system of b -adic numbers. This is standard material when b is prime, but we need to consider the general case (b possibly composite). A good reference is Mahler’s book [11].

Let $b \geq 2$ be an integer. A b -adic number is a sum of the type

$$\sum_{i=-k}^{\infty} x_i b^i$$

where the digits x_i satisfy $0 \leq x_i < b$. The set of b -adic numbers is equipped with a topology, in which a number is near 0 when it is divisible by a high power of b . Addition and multiplication are done by computing the formal result (i.e., treating b as an indeterminate), and then propagating carries.

When p is prime, the p -adic numbers form a field, which we will denote by $\mathbb{Q}_p$. This contains the subring $\mathbb{Z}_p$ of integral elements.

Manuscript received December 15, 1995; revised January 9, 1997. This research was supported in part by NSF under Grant CCR-9510244.

The author is with the Computer Sciences Department, University of Wisconsin, Madison, WI 53706 USA.

Publisher Item Identifier S 0018-9448(98)02703-5.

Otherwise, if $b = p_1^{e_1} \cdots p_r^{e_r}$, the ring of b -adic numbers splits

$$\mathbb{Q}_b \cong \mathbb{Q}_{p_1} \oplus \cdots \oplus \mathbb{Q}_{p_r}.$$

This has the field of rational numbers $\mathbb{Q}$ embedded on the diagonal. We also define

$$\mathbb{Z}_b \cong \mathbb{Z}_{p_1} \oplus \cdots \oplus \mathbb{Z}_{p_r}$$

so that a b -adic number is integral iff it is integral at p for every $p|b$. For any $m = p_1^{d_1} \cdots p_r^{d_r}$ we have a surjective homomorphism

$$\mathbb{Z}_b \rightarrow \mathbb{Z}/(m)$$

which allows us to speak of “ $x \bmod m$ ” when x is a b -adic integer. We will choose residue class representatives so as to make $0 \leq x \bmod m < m$.

We will write b -adic numbers so that the indices increase from right to left, for example

$$\cdots 34572.0 \tag{2.1}$$

indicates a 10-adic integer that is congruent to $34572 \bmod 10^5$. (Note that an element is integral iff it has no digits to the right of the “decimal point.”) The advantage of this representation is that the ordinary algorithms for addition, subtraction, and multiplication can be used essentially without modification. Readers familiar with computer arithmetic may wish to think of this as a “ b ’s complement” number system in which numbers extend infinitely far to the left.

III. ADD-WITH-CARRY GENERATORS

In this section we present the b -adic approach to Marsaglia–Zaman generators. We will concentrate on one of these, the so-called “add-with-carry” generator. The reader will not find it difficult to extend our results to similar generators based on subtraction and/or complementation.

In the Marsaglia–Zaman scheme, each pseudorandom number is a base- b digit, that is, an integer x_i in the range $0 \leq x_i < b$. There is also an auxiliary sequence of carries c_i , with $0 \leq c_i \leq 1$. The formation rules for these sequences are

$$x_n = x_{n-r} + x_{n-s} + c_{n-1} \bmod b \tag{3.1}$$

and

$$c_n = \left\lfloor \frac{x_{n-r} + x_{n-s} + c_{n-1}}{b} \right\rfloor. \tag{3.2}$$

Here, it is assumed that the lags r and s satisfy $0 < r < s$. The initial digits $x_0, x_1, \dots, x_{s-1}$ form the *seed*; to get more digits, we choose $c_{s-1} \in \{0, 1\}$ and use (3.1) and (3.2) for $n \geq s$. To summarize, we form x_n by adding two previously formed numbers, together with the carry generated by the computation of x_{n-1} . We now prove a) for this generator.

Theorem 1: Suppose the sequence $X = \{x_i\}_{i=0}^{\infty}$ of base- b digits is defined by (3.1) and (3.2), with $0 < r < s$. Then the b -adic number

$$x = \sum_{i=0}^{\infty} x_i b^i$$

is rational.

Proof: Defining x as above, we have

$$b^r x = \sum_{i=r}^{\infty} x_{i-r} b^i$$

and

$$b^s x = \sum_{i=s}^{\infty} x_{i-s} b^i.$$

Let $\alpha = x \bmod b^s$ and $\beta = b^r x \bmod b^s$. From (3.1) and (3.2), we get

$$x - \alpha = b^r x - \beta + b^s x + c_{s-1} b^s.$$

Therefore, $x - b^r x - b^s x$ is an integer z , satisfying $|z| < 2b^s$. Evidently, $1 - b^r - b^s$ is a unit $\bmod b$, so it is a unit in $\mathbf{Q}_b$. Then,

$$x = \frac{z}{1 - b^r - b^s}$$

so that x is rational. $\square$

Marsaglia observed that when $b^s + b^r - 1$ is prime, the period of X is the order of b modulo $b^s + b^r - 1$, except for two degenerate cases corresponding to $x = 0, -1$. He stated without proof [12, Sec. 4.2] that this also holds when $b^s + b^r - 1$ is composite, but this is false. (Let $b = 2, s = 3$ and $r = 1$. With the seed $(x_0, x_1, x_2) = (0, 1, 0)$, and initial carry $c_2 = 1$, we get a period of 2. But 2 has order $6 \bmod 3^3 + 2^2 - 1 = 9$.) The following result handles all cases.

Theorem 2: Define X and x as in Theorem 1. Let $x = c/d$, where c and $d \geq 1$ are relatively prime integers. Then d divides $b^s + b^r - 1$, and the period of X is the multiplicative order of $b \bmod d$.

Proof: The first assertion follows from the proof of Theorem 1. We may as well assume that $d > 1$, for otherwise the theorem is trivially true. Therefore, $c \neq 0$. Let ν be the order of b in $\mathbf{Z}/(d)^*$.

We now show that the period of x 's b -adic representation divides ν . First, assume that $0 < -c < d$. From the expansion

$$(1 - b^\nu)^{-1} = 1 + b^\nu + b^{2\nu} + \dots$$

we obtain the convergent series

$$\begin{aligned} \frac{c}{d} &= \dots + b^{3\nu} \left(-c \frac{b^\nu - 1}{d} \right) + b^{2\nu} \left(-c \frac{b^\nu - 1}{d} \right) \\ &\quad + b^\nu \left(-c \frac{b^\nu - 1}{d} \right) + \left(-c \frac{b^\nu - 1}{d} \right). \end{aligned}$$

Since $-c(b^\nu - 1)/d$ is a positive integer less than b^ν , in this case the b -adic expansion of c/d is periodic with period dividing ν . Since $d > 1$, c/d is not an integer, making its b -adic expansion nonzero infinitely often. Therefore, subtracting any positive integer will not change the period, and the period divides ν whenever $c/d < 0$. Finally, we obtain the expansion of $-c/d$ by replacing each digit x_i by $(b-1) - x_i$, to form the so-called complement, and adding 1. The complemented sequence has the same period as the original. It also has digits less than $b-1$ infinitely often, so the carry resulting from adding 1 will only propagate a finite distance. So the period divides ν for all c/d .

Conversely, any period μ is a multiple of ν . We see this by writing

$$\frac{c}{d} = (\dots + b^{2\mu} \delta + b^\mu \delta + \delta) b^\tau + \gamma = \frac{\delta}{1 - b^\mu} b^\tau + \gamma$$

where γ and δ are integers, satisfying $0 \leq \gamma < b^\tau$ and $0 \leq \delta < b^\mu$. Since c/d is in lowest terms, d divides $b^\mu - 1$. This means $b^\mu \equiv 1 \bmod d$, so ν divides μ . $\square$

We briefly mention some variants of the add-with-carry generator. Marsaglia and Zaman have also proposed using recurrences of the form

$$x_n = x_{n-r} - x_{n-s} - c_{n-1} \bmod b \quad (3.3)$$

(“subtract-with-borrow”) and

$$x_n = (b-1) - x_{n-r} - x_{n-s} - c_{n-1} \bmod b \quad (3.4)$$

(“complementary add-with-carry”). Here r and s are distinct positive integers, and $c_{n-1} \in \{0, 1\}$ is the “borrow” bit generated by the formation of x_{n-1} . Theorem 1 remains true for these generators, and Theorem 2 holds with the following modification: we must replace $b^s + b^r - 1$ by $b^s - b^r + 1$ (for (3.3) with $r < s$), by $b^r - b^s - 1$ (for (3.3) with $r > s$), or by $b^r + b^s + 1$ (for (3.4)). More generally, we could replace (3.1) and (3.2) by

$$x_n = \sum_{i=1}^s \alpha_i x_{n-i} + c_{n-1} \bmod b$$

and

$$c_n = \left\lfloor \frac{\sum_{i=1}^s \alpha_i x_{n-i} + c_{n-1}}{b} \right\rfloor$$

for suitable $\alpha_i \in \{0, \dots, b-1\}$. (One would probably restrict $|\alpha_i|$ to zero or powers of 2, for efficiency reasons.) We remark without proof that all these extensions generate b -adic digits of rational numbers, and our prediction methods apply to them.

IV. PREDICTION

Prediction of an add-with-carry sequence requires one to solve the following problem: given an initial segment of the b -adic representation of a rational number c/d , what are c and d ? Of course, for this to be meaningful, we must insist that c and d lie within prescribed bounds. Many authors have treated this problem in the context of exact arithmetic. (For this application, b is a prime power.) The most practical method is to generate candidate pairs $(\hat{c}, \hat{d})$, using the extended Euclidean algorithm. Dixon [6] proved that when $\hat{c}$ crosses a certain threshold, we have $\hat{c}/\hat{d} = c/d$. Krishnamurthy [10] and Kornerup and Gregory [8] proved that when c/d is in lowest terms, (c, d) is the unique candidate pair with both elements small.

We give a theorem below that combines the best features of both results. In particular, we only check the bound on c (as Dixon does), but obtain c and d exactly (as Krishnamurthy and Kornerup–Gregory do). We do not restrict b to be a prime power, as there is no need for it. We will, however, assume the fraction c/d is in lowest terms.

Let us first recall some facts about continued fractions, referring to [7] for proofs. Let u_0 and u_1 be positive integers with $u_0 > u_1$. Apply Euclid's process to these numbers, so that

$$u_0 = a_0 u_1 + u_2, \quad 0 \leq u_2 < u_1$$

$$u_1 = a_1 u_2 + u_3, \quad 0 \leq u_3 < u_2$$

$$u_2 = a_2 u_3 + u_4, \quad 0 \leq u_4 < u_3$$

...

$$u_{n-1} = a_{n-1} u_n.$$

We also put $u_{n+1} = 0$. Evidently $\{u_i\}$ is a decreasing sequence of nonnegative numbers. Using the recurrence

$$(q_{i+2} \ p_{i+2}) = (q_i \ p_i) - a_i (q_{i+1} \ p_{i+1})$$

with initial conditions

$$\begin{pmatrix} q_0 & p_0 \\ q_1 & p_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

the extended Euclidean algorithm determines integer sequences $\{p_i\}$

and $\{q_i\}$ satisfying

$$\begin{pmatrix} u_i \\ u_{i+1} \end{pmatrix} = \begin{pmatrix} q_i & p_i \\ q_{i+1} & p_{i+1} \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \end{pmatrix} \quad (4.1)$$

for $0 \leq i \leq n$. Both of these sequences alternate in sign and increase in absolute value. The matrix in (4.1) has determinant ± 1 , so

$$\begin{pmatrix} u_0 \\ u_1 \end{pmatrix} = \pm \begin{pmatrix} p_{i+1} & -p_i \\ -q_{i+1} & q_i \end{pmatrix} \begin{pmatrix} u_i \\ u_{i+1} \end{pmatrix}. \quad (4.2)$$

Using the above notation, we have the following result.

Theorem 3: Let $0 < k < m$, and $N = \sqrt{m/2}$ (this need not be an integer). Let $c \equiv dk \pmod{m}$, where c and d are relatively prime numbers such that $0 < c \leq N/2$ and $0 < |d| < N$. Apply the extended Euclidean algorithm to the inputs $u_0 = m$ and $u_1 = k$. There is a unique i for which $u_i \leq N < u_{i-1}$, and for this i , we have $u_i = c$ and $p_i = d$.

Proof: We first consider the condition on i and estimate p_i . Evidently, $u_0 > N$. There is an integer e for which $c = dk + em$; therefore, $u_n = \gcd(k, m)$ divides c . Therefore, $u_n \leq c < N$. Since $\{u_i\}$ decreases, there will be exactly one i with $u_i \leq N < u_{i-1}$. From the top row of (4.1), we see that

$$u_i \equiv p_i k \pmod{m}.$$

We note that $i \geq 1$, so $|p_i| \geq |p_1| > 0$. From the top row of (4.2), we find

$$u_0 = |p_i|u_{i-1} + |p_{i-1}|u_i \geq |p_i|u_{i-1}.$$

Therefore,

$$|p_i| \leq \frac{u_0}{u_{i-1}} < \frac{u_0}{u_{i-1}} = 2N.$$

We next show that $u_i/p_i = c/d$ as rational numbers. To prove this, start with

$$\begin{aligned} c &\equiv dk \pmod{m} \\ u_i &\equiv p_i k \pmod{m}. \end{aligned}$$

Cross-multiply by the coefficients of k and subtract to obtain

$$cp_i \equiv u_i d \pmod{m}.$$

Our bounds imply that $|cp_i|, |u_i d| < m/2$, so $cp_i = u_i d$ as integers, and we find $c/d = u_i/p_i$.

Finally, we use our assumption that c and d are coprime. We have

$$c = dk + em. \quad (4.3)$$

This implies that $\gcd(d, e)$ divides c ; it also divides d , so d and e are coprime. From (4.1), we have

$$u_i = q_i m + p_i k. \quad (4.4)$$

Solving (4.3) for c/d and (4.4) for u_i/p_i and setting these equal, we conclude that $q_i/p_i = e/d$. If $q_i = 0$, then $i = 1$. In this case, $e = 0$, so c/d is the positive integer k , making $p_1 = d = 1$ and $u_1 = c$. If $q_i \neq 0$, then q_i and p_i are coprime, so $|p_i| = |d|$. From $c > 0$ and $u_i/p_i = c/d$ we conclude $u_i = c$ and $p_i = d$. $\square$

We remark without proof that $\gcd(c, d) = 1$ implies that d is coprime to m . Conversely, if $c \equiv dk \pmod{m}$ with $g = \gcd(d, m) > 1$, we have $c/g \equiv (d/g)k$, with c/g coprime to d/g . If d is not a unit mod m , this need not hold. For example, mod 625 we have $5 \equiv 60 \cdot 448$, but $1 \not\equiv 12 \cdot 448$.

Nonconstructive versions of Theorem 3, which prove the existence of small c and d without giving any easy way to find them, go back at least to the work of Thue and Aubry in the early part of this century. For history and some related results see [3] and [16].

We briefly relate Theorem 3 to error-correcting codes. In decoding BCH codes, one gets a polynomial $k(X)$ and must solve the key equation

$$c(X) \equiv d(X)k(X) \pmod{X^{2t}}$$

for polynomials c and d , subject to $\deg c \leq t-1$ and $\deg d \leq t$. Thus the problem is to recover the rational function $c(X)/d(X)$ from its X -adic approximation $k(X)$, and this can be done with continued fractions. The polynomial analog to Theorem 3 was proved by Mills [14], and we have modeled our proof on the exposition of Blahut [1, p. 195]. Because our absolute value is not ultrametric, however, it was convenient to take $m = 2N^2$ rather than N^2 .

It follows from the proof of Theorem 3 that we will have $u_{i-1} > N$ and $u_i \leq N/2$. For this reason, it will suffice to test u_i against a crude approximation such as $2^{\lceil \log_2 N \rceil}$.

Theorem 3 suggests the following algorithm to predict a pseudo-random sequence generated by the add-with-carry method. We first assume that the base b and the largest lag s are known, but that the other lag r and the numerator z are unknown. From the proof of Theorem 1, we know that $|z|$ and $b^s + b^r - 1$ are less than $2b^s$. Therefore, taking $m = b^{2s+t}$ with $b^t > 32$ guarantees $|z| \leq N/2$ and $b^s + b^r - 1 < N$. We observe the first $2s+t$ elements of the sequence, and interpret these as base- b digits of an integer k . Since $1 - b^r - b^s$ is a unit mod m , there are coprime integers c, d such that

$$\frac{c}{d} = \frac{z}{1 - b^r - b^s} \equiv k \pmod{m}.$$

With the aid of Theorem 3, we can obtain c and d . If $d = 1 - b^r - b^s$, we have recovered the generator. (Intuitively, this seems likely, as the probability is about 61% that two random numbers will be coprime.) If not, we still have enough information to predict the rest of the sequence, as it is the b -adic representation of c/d .

Using the bound of Collins [5] for the bit complexity of Euclid's algorithm, it can be shown that our prediction process uses $O((s \log b)^2)$ bit operations, if ordinary arithmetic is employed. We observe that $O(s \log b)$ bits are necessary to specify the sequence in general, so that the time to predict the generator varies as the square of the length of its description. We also note that the prediction algorithm uses $O(s \log b)$ bits of input, which is optimal up to a constant factor.

In practice, it is not necessary to know s . We can simply try $s = 2, 4, 8, 16, \dots$ in order until a value of s at most double the correct one is reached. The time for prediction remains $O(s^2(\log b)^2)$.

For the add-with-carry sequence, one can find the generator in $O(s^3(\log b)^2)$ steps simply by trying all possible r and s , and seeing which ones work. In defense of our method, we point out that it is direct (no guesswork needed), and that it is equally fast when applied to more general sequences, such as those defined by

$$x_n = x_{n-r} + x_{n-s} + x_{n-t} + c_{n-1}.$$

Discovering b when it is not provided for us poses more of a problem. In practice, a cryptanalyst would expect that $b = 2^w$, where w is the word size of the computer used to generate the sequence. The rationale for this is that most computers provide addition-with-carry and subtraction-with-borrow as primitive operations.

A heuristic argument based on the coupon collection problem suggests that $O(b \log b)$ samples should be sufficient to determine b . (With high probability, all possible digits appear in the sample, so we estimate b to be one larger than the largest digit seen so far.) This is useful if b is small, but it is not polynomial in the length of b .

If b is large, more work appears necessary. In the add-with-carry generator, there are only $s-2$ choices for r with $0 < r < s$. For each of these, we can compute

$$x_n - x_{n-r} - x_{n-s} - c \quad (4.5)$$

for various values of n and $c \in \{0, 1\}$. If the x_i 's are well distributed, we can expect (4.5) to be a small multiple of b , certainly no more than $3b$ in absolute value. By removing small prime factors from these numbers, we obtain a candidate for b . Heuristically, then, we can easily get $O(s)$ tentative values for b ; one of these will be correct. But the complexity of this procedure quickly gets out of hand for more complicated generators.

Trying to make the arguments of the last two paragraphs rigorous leads to the following problem, which seems to be open: Given the first few b -adic digits of a rational number x , where both b and x are unknown, quickly find x .

V. EXAMPLES

In this section, we will give some examples of the prediction method. We will use a sample generator based on the Fibonacci sequence, where $b = 10$ and

$$x_n = x_{n-1} + x_{n-2} + c_{n-1} \bmod 10.$$

Here is an initial segment of the sequence

$$n = 0 \ 1 \ 2 \ 3 \ \dots$$

$$x_n = 1 \ 2 \ 3 \ 5 \ 8 \ 3 \ 2 \ 6 \ 8 \ 4 \ 3 \ 8 \ \dots$$

$$c_n = 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ \dots$$

As b -adic numbers, we have

$$\dots 834862385321.0 = x,$$

$$\dots 348623853210.0 = bx,$$

$$\dots 486238532100.0 = b^2x,$$

$$\text{so } \dots 00000000011.0 = (1 - b - b^2)x.$$

(To check this, verify that the last three rows sum to the first.) Thus $z = 11$, and $1 - b - b^2 = -109$, so x is the 10-adic representation of $-11/109$.

For prediction, we must compute the values u_i and p_i . This is conveniently done by defining row vectors

$$R_0 = (u_0, 0), R_1 = (u_1, 1)$$

and

$$R_{i+2} = R_i - a_i R_{i+1}$$

for $i \geq 0$. Then we have $R_i = (u_i, p_i)$ for $0 \leq i \leq n+1$.

To predict the "Fibonacci" sequence above, we take $s = 2$ and that $m = 10^7$. Taking $k = 2385321$, the first seven digits of x , the extended Euclidean algorithm applied to m , and k produces the quotient sequence 4, 5, 5, 8340, 11. So we get the following table:

i	u_i	p_i
0	10000000	0
1	2385321	1
2	458716	-4
3	91741	21
4	11	-109
5	1	909081
6	0	-10000000

To the nearest integer $N = 2236$, so $u_i \leq N$ for the first time when $i = 4$. From this row of the table, we read off $x = 11/(-109)$.

Here is another example, taken from the middle of the "Fibonacci" sequence

$$\dots 3486238.0 = y,$$

$$\dots 4862380.0 = by,$$

$$\dots 8623800.0 = b^2y,$$

$$\text{so } \dots 0000058.0 = (1 - b - b^2)y.$$

Here the seed is 38, but the initial carry is 1. We apply the Euclidean algorithm to $m = 10000000$ and $k = 3486238$, and get the quotient sequence 2, 1, 6, 1, 1, 1, 1, 1581, 5, 1, 4. Therefore, the table is

i	u_i	p_i
0	10000000	0
1	3486238	1
2	3027524	-2
3	458714	3
4	275240	-20
5	183474	23
6	91766	-43
7	91708	66
8	58	-109
9	10	172395
10	8	-862084
11	2	1034479
12	0	-5000000

From $i = 8$, we get $y = 58/(-109)$.

Both quotient sequences in these examples contain atypically large entries, and this deserves an explanation. Suppose we are trying to recover c/d from its representation mod $m = 2N^2$, and $c \leq N/2^k$. Since $u_i = c$, we will have $u_{i-1}/u_i > 2^k$, so $a_i \geq 2^k$. If c is small, then k will be large, forcing a large value of the i th quotient.

ACKNOWLEDGMENT

The author wishes to thank the referee for a carefully written report.

REFERENCES

- [1] R. E. Blahut, *Theory and Practice of Error Control Codes*. Reading, MA: Addison-Wesley, 1983.
- [2] L. Blum, M. Blum, and M. Shub, "A simple unpredictable pseudo-random number generator," *SIAM J. Comp.*, vol. 15, pp. 364-383, 1986.
- [3] A. Brauer and R. L. Reynolds, "On a theorem of Aubry-Thue," *Canadian J. Math.*, vol. 3, pp. 367-374, 1951.
- [4] J. Boyar, "Inferring sequences produced by pseudo-random number generators," *J. Assoc. Comput. Mach.*, vol. 36, pp. 129-141, 1989.
- [5] G. Collins, "Computing time analyses for some arithmetic and algebraic algorithms," in *Proc. 1968 Summer Institute on Symbolic Mathematical Computation* (IBM Federal Systems Center, 1969), pp. 195-231.
- [6] J. D. Dixon, "Exact solution of linear equations using P -adic expansions," *Numer. Math.*, vol. 40, pp. 137-141, 1982.
- [7] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*. 5th ed. Oxford, U.K.: Oxford Univ. Press, 1979.
- [8] P. Kornerup and R. T. Gregory, "Mapping integers and Hensel codes onto Farey fractions," *BIT*, vol. 23, pp. 9-20, 1983.
- [9] H. Krawczyk, "How to predict congruential generators," *J. Algorithms*, vol. 13, pp. 527-545, 1992.
- [10] E. V. Krishnamurthy, "On the conversion of Hensel codes to Farey rationals," *IEEE Trans. Comp.*, vol. C-32, pp. 331-337, 1983.
- [11] K. Mahler, *p -adic Numbers and Their Functions*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1981.

- [12] G. Marsaglia, "The mathematics of random number generators," *Proc. Symp. Applied Math.*, vol. 46, pp. 73–90, 1982.
- [13] G. Marsaglia and A. Zaman, "A new class of random number generators," *Ann. Appl. Prob.*, vol. 1, pp. 462–480, 1991.
- [14] W. H. Mills, "Continued fractions and linear recurrences," *Math. Comp.*, vol. 29, pp. 173–180, 1975.
- [15] Y. Sugiyama, M. Kasahara, S. Hirasawa, and T. Namekawa, "A method for solving key equation for decoding Goppa codes," *Inform. Contr.*, vol. 27, pp. 87–99, 1975.
- [16] A. Thue, "Et par antydninger til en taltheoretisk metode," *Forhandlinger i Videnskabs-selskabet i Christiania*, no. 7, 1902. Reprinted with English abstract in *Selected Mathematical Papers of Axel Thue*, T. Nagell et al., Eds., 1977, pp. 57–75.
- [17] L. R. Welch and R. A. Scholtz, "Continued fractions and Berlekamp's algorithm," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 19–27, 1979.

Periods of Interleaved and Nonuniformly Decimated Sequences

Jovan Dj. Golić

Abstract—Periods of interleaved and nonuniformly decimated integer sequences are analyzed. A characterization of the period of an interleaved sequence in terms of the constituent periodic integer sequences is first derived. This is then used to generalize the result of Blakley and Purdy on the period of a decimated integer sequence obtained by a periodic nonuniform decimation. The technique developed and results may be interesting for analyzing the period of various pseudorandom sequences, especially in stream cipher and spread-spectrum applications.

Index Terms—Cryptography, interleaving, nonuniform decimation, period, pseudorandom sequences.

I. INTRODUCTION

Pseudorandom sequences generated by finite-state machines have numerous applications in many areas of computer science and electrical engineering such as, for example, stream cipher cryptography, spread-spectrum communications, radar ranging, and random number generation. A basic characteristic of these sequences is their period, which is often required to be large enough, depending on a particular application. Sequences are usually obtained from some simple-structured elementary recurring sequences combined by a suitable operation, for example, based on nonlinear switching functions with or without memory, nonuniform decimation, or interleaving. Interleaving is especially interesting for high-speed parallel generation of pseudorandom sequences, whereas nonuniform decimation is a common way of achieving large sequence complexities in stream cipher and spread-spectrum applications. Deriving the period of a pseudorandom sequence is generally a difficult algebraic problem which seems to be tractable only for relatively simple sequences and under special constraints.

In this correspondence we study the periods of interleaved and nonuniformly decimated integer sequences. The fundamental result

on the period of nonuniformly decimated integer sequences is established by Blakley and Purdy [1]. Our objective is to obtain a more general result and thus show that the period can be guaranteed without the constraint assumed in [1] which presents a practical limitation in the design of pseudorandom sequence generators, especially those based on clock-controlled shift registers, see [7]. To this end, in Section II, we first obtain a basic characterization of the period of interleaved integer sequences which may also be useful for determining the period of various types of pseudorandom sequences. In Section III, we then use this characterization to derive the period of nonuniformly decimated integer sequences and thus generalize the result from [1]. In Section IV, we give some practically useful conditions for the maximum period to be achieved which extend the results on the period of clock-controlled shift registers [2], [7]. Conclusions are presented in Section V.

II. INTERLEAVED SEQUENCES

Consider a set of K periodic integer sequences $a_i = \{a_i(t)\}_{t=0}^{\infty}$ with periods $P_i, 0 \leq i \leq K-1$, respectively. It is well known that the interleaved sequence b is defined by

$$b(i + Kt) = a_i(t), \quad 0 \leq i \leq K-1, t \geq 0. \quad (1)$$

It is clear that b is periodic with period $P_b | PK$, where

$$P = \text{l.c.m. } (P_i: 0 \leq i \leq K-1).$$

Our objective is to characterize P_b more precisely. We first give an elementary lower bound on P_b in terms of P and K (for a special case, see [3]).

Lemma 1: The period P_b of the interleaved sequence b has the form

$$P_b = P(P_b, K) \quad (2)$$

which is equivalent to

$$P_b = PK'n, \quad n | K'' \quad (3)$$

where $K = K'K''$, $(K'', P) = 1$, and every prime factor of K' divides P (here and throughout $(\cdot, \cdot)$ denotes the greatest common divisor).

Proof: We first prove (2). Since a_i can be obtained from the decimation by K of a translate by i of b , see (1), we get that $P_i | \text{l.c.m. } (P_b, K)/K, 0 \leq i \leq K-1$, and hence $P | P_b/(P_b, K)$. On the other hand, from $P_b | PK$, which is a direct consequence of (1), we have $P_b/(P_b, K) | PK/(P_b, K)$ and hence $P_b/(P_b, K) | P$, because $(P_b/(P_b, K), K/(P_b, K)) = 1$. Therefore, $P = P_b/(P_b, K)$.

We now prove that (2) is equivalent to (3). From (2) it follows that $(P, K/(P_b, K)) = 1$ which by definition of K' and K'' implies that $K' | (P_b, K)$. Consequently, (2) implies (3). Conversely, (3) results in $(P_b, K) = K'n$ and hence (2) as well. $\square$

Lemma 1 provides a characterization of P_b only in terms of $P_i, 0 \leq i \leq K-1$, and K . In order to specify P_b exactly we must in addition involve the sequences $a_i, 0 \leq i \leq K-1$, themselves. Let $P_b = Pk, k | K$, and $K = km$. By Lemma 1 it follows that $(m, P) = 1$, since $K' | k$ and $(K'', P) = 1$. Our basic result on the period of interleaved sequences is given by the following theorem. As a consequence, one may derive the periods of various special classes of interleaved sequences [9] for spread-spectrum or stream-cipher applications.

Manuscript received July 10, 1995; revised March 10, 1997. Part of this work was done while the author was with the Information Security Research Centre, Queensland University of Technology, Brisbane, Qld. 4001, Australia.

The author is with the School of Electrical Engineering, University of Belgrade, 11001 Belgrade, Yugoslavia.

Publisher Item Identifier S 0018-9448(98)02704-7.