The Asymptotic Behavior of N-Adic Complexity

Andrew Klapper*

Abstract

We study the asymptotic behavior of stream cipher security measures associated with classes of sequence generators such as linear feedback shift registers and feedback with carry shift registers. For nonperiodic sequences we consider normalized measures and study the set of accumulation points for a fixed sequence. We see that the the set of accumulation points is always a closed subinterval of $[0, 1]$. For binary or ternary FCSRs we see that this interval is of the form $[B, 1 - B]$, a result that is an analog of an earlier result by Dai, Jiang, Imamura, and Gong for LFSRs.

Keywords: Sequences, N-adic complexity, Stream ciphers, shift registers.

AMS Classification: 94A05, 94A55, 94A60.

1 Introduction

The purpose of this paper is to study the asymptotic behavior of security or randomness measures for infinite sequences. The kinds of measures we are interested in arise in the following manner. There is a class $\mathcal{G}$ of finite state devices that generate infinite sequences over some alphabet Σ, such that every eventually periodic sequence is generated by at least one element of $\mathcal{G}$. We also assume there is a notion of the size of a generator in $\mathcal{G}$, a positive real number. In general this measure should be close to the number, n, of elements of Σ needed to represent a state of the generator. Typically "close" means differing from n by at most $O(\log(n))$. Examples of such classes of generators include the *linear feedback shift registers* (LFSRs), where the size of an LFSR is the number of cells,

*Dept. of Computer Science, 779A Anderson Hall, University of Kentucky, Lexington, KY, 40506-0046. www.cs.uky.edu/∼klapper. Part of this work was carried out while the author was visiting the Fields Institute at the University of Toronto. This material is based upon work supported by the National Science Foundation under Grant No. CCF-0514660. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation..

and *feedback with carry shift registers* (FCSRs) [6], where the size of an FCSR is the log of the connection number (the log with base equal to the size of the output alphabet).

We denote by $\lambda^{\mathcal{G}}(S)$ the minimum size of a generator in $\mathcal{G}$ that outputs the eventually periodic sequence S. In many cases there is an algorithm that efficiently finds the minimum size generator of S given a prefix of S whose length is a small (typically linear) function of $\lambda^{\mathcal{G}}(S)$. We call this a $\mathcal{G}$-*synthesis algorithm*. Examples include the Berlekamp-Massey algorithm for LFSRs [9] and the 2-adic rational approximation algorithm for FCSRs [6]. When a $\mathcal{G}$-synthesis algorithm exists, the quantity $\lambda^{\mathcal{G}}(S)$ is a measure of the security of S.

For sequences that are not eventually periodic, such measures are undefined. However, we can apply the measure to the various prefixes of S and try to understand the asymptotic behavior. For $n > 0$, let $\lambda_n^{\mathcal{G}}(S)$ denote the minimum size of a generator from $\mathcal{G}$ that outputs the first n symbols of S as its first n outputs. The sequence of numbers $(\lambda_n^{\mathcal{G}}(S) : n = 1, 2, \cdots)$ is called the $\mathcal{G}$-*complexity profile of* S. For a sequence that is not eventually periodic, the limit of the $\lambda_n^{\mathcal{G}}(S)$ is infinite, so we normalize the measure by letting $\delta_n^{\mathcal{G}}(S) = \lambda_n^{\mathcal{G}}(S)/n$. For the typical measures we are interested in we have

$$\lambda_n^{\mathcal{G}}(S) \leq n + O(\log(n)),$$

so that

$$0 \leq \delta_n^{G}(S) \leq 1 + o(n).$$

In general the $\delta_n^{\mathcal{G}}(S)$ do not have a single limit, but rather have a set $T(S)$ of accumulation points. Our goal is to determine what sets of accumulation points are possible.

It is immediate for such a measure $\lambda_n^{\mathcal{G}}(S)$ that

$$\lambda_{n+1}^{\mathcal{G}}(S) \geq \lambda_n^{\mathcal{G}}(S)$$

for all $n \geq 1$, so that

$$\delta_{n+1}^{\mathcal{G}}(S) \geq \frac{n}{n+1}\delta_n^{\mathcal{G}}(S) \geq \delta_n^{\mathcal{G}}(S) - \frac{1}{n+1}. \tag{1}$$

This allows us to show that the set of accumulation points is a closed interval.

Theorem 1.1 *Let* $\{\lambda_n : n = 1, 2, \infty\}$ *be a sequence of integers,* $1 \leq \lambda_n \leq n$*, satisfying* $\lambda_n \leq \lambda_{n+1}$ *for all* $n = 1, 2, \cdots$*. Let* $\delta_n = \lambda_n/n \in [0, 1]$*. Then the set* T *of accumulation points of the* δ_n *is a closed interval* $[B, C]$*.*

Proof: First note that under these hypotheses, for any n we have

$$\begin{aligned}\delta_n - \delta_{n+1} &= \frac{\lambda_n}{n} - \frac{\lambda_{n+1}}{n+1} \\ &\leq \lambda_n(\frac{1}{n} - \frac{1}{n+1}) \\ &= \frac{\lambda_n}{n(n+1)} \\ &< \frac{1}{n}.\end{aligned}$$

Let B and C be the least and largest accumulation points of T, respectively. Then there are sequences of integers $n_1 < n_2 < \cdots$ and $m_1 < m_2 < \cdots$ so that

$$\lim_{i\to\infty} \delta_{n_i} = B \text{ and } \lim_{i\to\infty} \delta_{m_i} = C.$$

By deleting some integers, we may assume that $m_1 < n_1 < m_2 < n_2 < \cdots$ and that

$$|B - \delta_{n_i}| > |B - \delta_{n_{i+1}}|$$

and

$$|C - \delta_{m_i}| > |C - \delta_{m_{i+1}}|$$

for all i. Let $B < D < C$. Take i sufficiently large that

$$\delta_{n_i} < D \text{ and } \delta_{m_i} > D.$$

Let k_i be the largest index between m_i and n_i so that

$$\delta_{k_i} \geq D.$$

Then

$$\delta_{k_i+1} < D$$

and

$$\delta_{k_i} - \delta_{k_i+1} \leq 1/k_i$$

by equation (1). It follows that

$$\lim_{i\to\infty} \delta_{k_i} = D.$$

That is, every real number in the interval $[B, C]$ is an accumulation point, which proves the theorem. □

Dai, Jiang, Imamura, and Gong studied this problem in the case when $\mathcal{G}$ is the set of LFSRs over $\mathbb{F}_2$, the finite field with two elements, and $\lambda^{\mathcal{G}}$ is linear complexity [2]. They showed that in this case $T(S)$ is an interval of the form $[B, 1-B]$, with $0 \leq B \leq 1/2$.

They also showed that for every such B there are sequences S with $T(S) = [B, 1-B]$. Dai, Imamura, and Yang, and Feng and Dai have also studied this problem for vector valued non-periodic sequences [1, 3]. In this setting, however, there are much more limited results. They showed that there is a number associated with the generalized continued fraction expansion of a multisequence that is a lower bound for the maximum accumulation point and an upper bound for the minimum accumulation point. One might ask whether $T(S)$ is a closed interval centered at this number. Theorem 1.1 shows that the set is a closed interval but leaves open the remainder of this question.

The primary goal of this paper is to do a similar analysis for N-adic complexity, where N is an integer greater than 1. In fact we show that if N is a power of 2 or 3, then exactly the same theorem holds, although we use different techniques. For more general N we obtain weaker results.

Consider the alphabet $\Sigma = \{0, 1, \cdots, N-1\}$. Recall that N-adic complexity is the security measure for N-ary sequences associated with the set $\mathcal{G}$ of N-ary feedback with carry shift registers (FCSRs) [6, 7, 8]. Such a register (with a particular initial state) can be identified with a rational number a/b with $\gcd(b, N) = 1$ in much the same way that a linear feedback shift register can be identified with a rational function $a(x)/b(x)$. The output sequence is then the N-adic expansion of a/b,

$$\frac{a}{b} = \sum_{i=0}^{\infty} s_i N^i. \tag{2}$$

We denote the set

$$\left\{ \sum_{i=0}^{\infty} s_i N^i : s_i \in \{0, 1, \cdots, N-1\} \right\}$$

of N-adic numbers by $\mathbb{Z}_N$. It is well-known that $\mathbb{Z}_N$ is an algebraic ring. If $\Phi(a, b) = \max(|a|, |b|)$, then the size of the register (measured by the number of N-ary memory cells needed) differs from $\lambda(a, b) = \log_N(\Phi(a, b))$ by at most a constant times $\log_N(\lambda(a, b))$. Thus we take $\lambda(a, b)$ as our security measure. That is, we let $\Phi_n(S)$ denote the least $\Phi(a, b)$ so that

$$\frac{a}{b} \equiv \sum_{i=0}^{\infty} s_i N^i \pmod{N^n}$$

and let $\lambda_n(S) = \log_N(\Phi_n(S))$. Let $\delta_n(S) = \lambda_n(S)/n$. Also, we let $\Phi(S)$ denote the least $\Phi(a, b)$ so that equation (2) holds, if such a pair a, b exists. We let $\Phi(S) = \infty$ otherwise. Let $\lambda(S) = \log_N(\Phi(S))$, the *$N$-adic complexity of S*. The set of numbers $\{\lambda_n(S) : n = 1, 2, \cdots\}$ is sometimes called the *N-adic complexity profile of S*. Note that $\Phi_n(S) \leq N^n - 1$ since we can take

$$a = \sum_{i=0}^{n-1} s_i N^i, \quad b = 1.$$

Thus $\lambda_n(S) < n$ and $0 \leq \delta_n(S) < 1$. It is known that there is an effective register synthesis algorithm for the N-adic generalization of feedback with carry shift registers, even if N is not a prime [8], so N-adic complexity is an important security measure for N-ary sequences.

Dai, et al. [2] analyzed the asymptotic linear complexity by using known results relating linear complexity to continued fraction expansions of power series. Although there is a relationship between integer continued fractions and N-adic complexity, the techniques used by Dai, et al. in the linear complexity case cannot apparently be used to analyze the asymptotic properties of the N-adic complexity. Thus we use entirely different methods here. As it turns out, these methods can also be used in the case of linear complexity (even more simply than for N-adic complexity, in fact) and result in much simpler proofs than were given by Dai, et al.

2 A Basic Lemma

Our goal is to show that the set $T(S)$ of accumulation points of the normalized N-adic complexity profile of an ultimately nonperiodic sequence S is a closed interval centered at $1/2$. In fact we see in the next section that this holds when $N \leq 4$, and in general when the least accumulation point is at most $\log_N(2)$. In this section we develop the technical tools needed to prove this.

Our goal is to understand the asymptotic behavior of $\delta_n(S)$. If S is ultimately periodic, then $\Phi_n(S)$ is constant for n sufficiently large, so the limit of the $\delta_n(S)$ exists and is zero. Therefore from here on in this section we assume that S is not ultimately periodic.

We need a lemma to bound $\delta_{n+1}(S)$ in terms of $\delta_n(S)$.

Lemma 2.1 *Suppose that $\Phi_{n+1}(S) > \Phi_n(S)$. Then*

1. *For all $N \geq 2$ we have*

$$\frac{N^n}{2\Phi_n(S)} \leq \Phi_{n+1}(S) \leq \frac{N\Phi_n(S)}{2} + \frac{N^{n+1}}{2\Phi_n(S)}.$$

2. *For all $N \geq 2$ we have*

$$\frac{n-1}{n+1} - \frac{n}{n+1}\delta_n(S) \leq \delta_{n+1}(S).$$

3. *For all $N \geq 2$ and $\epsilon > 0$, if n is sufficiently large and $\delta_n(S) > \max(1/2 + \epsilon, 1 - \log_N(2/(1+\epsilon)))$, then $\delta_{n+1}(S) < \delta_n(S)$.*

Proof: Let

$$\frac{a}{b} \equiv \sum_{i=0}^{\infty} s_i N^i \pmod{N^n} \tag{3}$$

with $\gcd(b, N) = 1$ and $\Phi(a, b) = \Phi_n(S)$. By the assumption that $\Phi_{n+1}(S) > \Phi_n(S)$, equation (3) does not hold modulo N^{n+1}. Thus for some integer v with v not divisible by N we have

$$\frac{a}{b} \equiv vN^n + \sum_{i=0}^{\infty} s_i N^i \pmod{N^{n+1}}.$$

Suppose also that

$$\frac{c}{d} \equiv \sum_{i=0}^{\infty} s_i N^i \pmod{N^{n+1}}$$

with $\gcd(d, N) = 1$ and $\Phi(c, d) = \Phi_{n+1}(S)$. Then

$$\frac{a}{b} \equiv \frac{c}{d} + vN^n \pmod{N^{n+1}}.$$

It follows that $ad - bc \equiv bdvN^n \pmod{N^{n+1}}$. Since v is nonzero, we have

$$N^n \leq |ad - bc| \leq 2\Phi_n(S)\Phi_{n+1}(S).$$

This implies the lower bound in the first assertion. The lower bound on $\delta_{n+1}(S)$ in the second assertion follows by taking logarithms and dividing by $n + 1$.

To obtain an upper bound on $\Phi_{n+1}(S)$ we construct a "pretty good" rational approximation modulo N^{n+1}. Then $\Phi_{n+1}(S)$ is upper bounded by the value of Φ on this approximation. Note that a and b are relatively prime: if they weren't, then we could factor out a common factor and reduce $\Phi(a, b)$.

First assume that $0 < |b| < |a|$. Let u be an integer so that $u \equiv bv \pmod{N}$ and $|u| \leq N/2$. Since aN and b are relatively prime, there exist integers e and f so that

$$aNe - bf = uN^n - a. \tag{4}$$

Let $g = 1 + Ne$, so that $g \equiv 1 \pmod{N}$ and

$$ag - bf = uN^n. \tag{5}$$

It then follows that $ag - bf = uN^n \equiv bvgN^n \pmod{N^{n+1}}$, so that

$$\frac{a}{b} - \frac{f}{g} \equiv vN^n \pmod{N^{n+1}}.$$

Thus

$$\frac{f}{g} \equiv \sum_{i=0}^{\infty} s_i N^i \pmod{N^{n+1}},$$

and $\gcd(g, N) = 1$, so $\Phi(f, g)$ is an upper bound for $\Phi_{n+1}(S)$. In fact there are many choices for (f, g) satisfying equation (5). By the relative primality of aN and b, the solutions to equation (5) with $g \equiv 1 \pmod{N}$ are exactly the pairs $(f, g) = (f_0, g_0) + (raN, rbN)$ where (f_0, g_0) is a fixed solution and $r \in \mathbb{Z}$. In particular, we can take $|f| \leq |a|N/2 = N\Phi_n(S)/2$. We then have $ag = bf + uN^n$ so that

$$\begin{aligned} |g| &\leq \frac{|bf|}{|a|} + \frac{|u|N^n}{|a|} \\ &\leq \frac{N|b|}{2} + \frac{|u|N^n}{|a|} \\ &< \frac{N\Phi_n(S)}{2} + \frac{N^{n+1}}{2\Phi_n(S)}. \end{aligned}$$

Therefore

$$\Phi_{n+1}(S) \leq \frac{N\Phi_n(S)}{2} + \frac{N^{n+1}}{2\Phi_n(S)}. \tag{6}$$

This proves the upper bound in the first assertion when $|b| < |a|$.

Now let $0 < |a| < |b|$. As in the previous case there are integers $g = 1 + Ne$ and f with $ag - bf = uN^n$. By adding a multiple of (bN, aN) to the pair (g, f), we may assume that $|g| \leq |b|N/2$. It follows that

$$|f| < \frac{N\Phi_n(S)}{2} + \frac{N^{n+1}}{2\Phi_n(S)}.$$

Finally we prove the third assertion. Let $\epsilon > 0$ and suppose that $\delta_n(S) > \max(1/2 + \epsilon, 1 - \log_N(2/(1+\epsilon)))$. Take n large enough that

$$\frac{1}{2} + \epsilon > \frac{1}{2} + \frac{\log(\epsilon^{-1})}{2n}. \tag{7}$$

From $\delta_n(S) > (1/2) + \epsilon$ and equation (7) it then follows that

$$\frac{N^{n+1}}{2\Phi_n(S)} < \epsilon \frac{N\Phi_n(S)}{2}.$$

Also, from $\delta_n(S) > 1 - \log_N(2/(1+\epsilon))$ it follows that

$$(1+\epsilon)\frac{N\Phi_n(S)}{2} < \Phi_n(S)^{(n+1)/n}.$$

It then follows that

$$\begin{aligned} \Phi_{n+1}(S) &\leq \frac{N\Phi_n(S)}{2} + \frac{N^{n+1}}{2\Phi_n(S)} \\ &\leq (1+\epsilon)\frac{N\Phi_n(S)}{2} \\ &< \Phi_n(S)^{(n+1)/n}. \end{aligned}$$

Taking logarithms and dividing by $n+1$ then gives $\delta_{n+1}(S) < \delta_n(S)$ as desired. □

This will suffice to characterize sets $[B, C]$ of accumulation points of normalized N-adic complexities of sequences when

$$1 - B \geq \lim_{\epsilon \to 0} \max\left(\frac{1}{2} + \epsilon, 1 - \log_N\left(\frac{2}{1+\epsilon}\right)\right) = \max\left(\frac{1}{2}, 1 - \log_N(2)\right).$$

This is equivalent to having $B \leq \min(1/2, \log_N(2))$. If $N \leq 4$, then $1 - \log_N(2) \leq 1/2$, so this suffices to characterize all sets of accumulation points. For larger N the characterization is incomplete.

3 Sets of Accumulation Points

Let S be an ultimately non-periodic N-ary sequence. In this section we show that in many cases the set of accumulation points $T(S)$ satisfies $T(S) = [B, 1-B]$ for some B. Let $T(S) = [B, C]$. Let $m_1, m_2, \cdots$ be a sequence of indices such that $B = \lim_{n\to\infty} \delta_{m_n}(S)$. If $\lambda_{n+1}(S) = \lambda_n(S)$, then $\delta_{n+1}(S) < \delta_n(S)$. If we replace m_n by the next index j so that $\lambda_j(S) < \lambda_{j+1}(S)$, then the resulting sequence will have a limit $D \leq B$. Since B is the minimal accumulation point of the $\delta_i(S)$, $D = B$. Therefore we may assume that $\lambda_{m_n}(S) < \lambda_{m_n+1}(S)$.

Lemma 3.1 *Let $N \geq 2$ and $B < 1/2$. Then*

$$\lim_{n\to\infty} \delta_{m_n+1} = 1 - B.$$

Proof: Let $\epsilon > 0$. Take n large enough that $m_n \geq 4/\epsilon$ and $|B - \delta_{m_n}(S)| < \min(\epsilon/2, (1-2B)/4)$. Then $\delta_{m_n}(S) < 1/2$ and by Lemma 2.1.2

$$\begin{aligned} 1 - B - \delta_{m_n+1}(S) &\leq 1 - B - \left(\frac{m_n - 1}{m_n + 1} - \frac{m_n}{m_n+1}\delta_{m_n}(S)\right) \\ &= (\delta_{m_n}(S) - B) + \frac{2 - \delta_{m_n}(S)}{m_n + 1} \\ &\leq \epsilon. \end{aligned}$$

Also, $\delta_{m_n}(S) < 1/2$ implies that $\Phi_{m_n}(S) < N^{m_n}/\Phi_{m_n}(S)$, so by Lemma 2.1.1 we have $\Phi_{m_n+1}(S) < N^{m_n+1}/\Phi_{m_n}(S)$. Thus $\lambda_{m_n+1}(S) < m_n + 1 - \lambda_{m_n}(S)$, so

$$\begin{aligned} \delta_{m_n+1}(S) - (1 - B) &\leq \frac{m_n + 1}{m_n + 1} - \frac{m_n}{m_n+1}\delta_{m_n}(S) - (1 - B) \\ &< (B - \delta_{m_n}(S)) + \frac{1}{m_n+1}\delta_{m_n}(S) \\ &\leq \epsilon. \end{aligned}$$

Thus $|1 - B - \delta_{m_n+1}(S)| < \epsilon$ for n sufficiently large, proving the lemma. □

Corollary 3.2 *In general* $1/2 \leq C$.

Proof: By Lemma 3.1, if $B < 1/2$, then $1 - B > 1/2$ is an accumulation point. If $B \geq 1/2$, then $C \geq B \geq 1/2$. In either case $C \geq 1/2$. □

We can now prove our main result.

Theorem 3.3 *Let S be an N-ary sequence and suppose that the set of accumulation points of the set of $\delta_n(S)$ is the interval $[B, C]$. Then $B \leq \max(1/2, 1 - \log_N(2))$. If $B < \log_N(2)$ then $C = 1 - B$.*

Proof: There is a sequence of integers $\ell_1 < \ell_2 < \cdots$ such that $\lim_{n\to\infty} \delta_{\ell_n}(S) = C$ and we can assume that $|C - \delta_{\ell_n}(S)| > |C - \delta_{\ell_{n+1}}(S)|$ for all n. By possibly deleting some of the ℓ_n and m_n, we can assume that $m_n < \ell_n < m_{n+1}$ for all $n \geq 1$. For n sufficiently large we have $\delta_{m_n} < \delta_{\ell_n}$, so we can assume this holds for all $n \geq 1$. Thus there is an $\ell \leq \ell_n$ so that $\delta_{\ell-1}(S) < \delta_\ell(S)$. If we replace ℓ_n by the largest such ℓ, then we still have a sequence whose limit is C. So we can assume that $\delta_{\ell_n-1}(S) < \delta_{\ell_n}(S)$ for all n. In particular, $\Phi_{\ell_n-1}(S) < \Phi_{\ell_n}(S)$. Then by Lemma 2.1, part 3, for every $\epsilon > 0$ there is an n so that

$$\delta_{\ell_n-1}(S) < \max(1/2 + \epsilon, 1 - \log_N(2/(1+\epsilon))).$$

This implies that there is an accumulation point of the $\delta_n(S)$ that is less than or equal to $\max(1/2, 1 - \log_N(2))$, so $B \leq \max(1/2, 1 - \log_N(2)) = 1 - \min(1/2, \log_N(2))$. This proves the first statement.

To prove the second statement, let us assume to the contrary that $1 - B < C$ and that $B \leq \log_N(2)$. Thus $C > 1 - \log_N(2)$. Also $B \leq 1/2$ (since when $N = 2$ or 3, $\max(1/2, 1 - \log_N(2)) = 1/2$ and when $N \geq 4$, $\log_N(2) \leq 1/2$), so $1/2 \leq 1 - B < C$. By part (1) of Lemma 2.1,

$$\Phi_{n+1}(S) \leq \frac{N\Phi_n(S)}{2} + \frac{N^{n+1}}{2\Phi_n(S)} \leq \max(N\Phi_n(S), N^{n+1}/\Phi_n(S)).$$

Thus

$$\delta_{\ell_n}(S) \leq \max\left(\frac{1}{\ell_n} + \frac{\ell_n - 1}{\ell_n}\delta_{\ell_n-1}(S), 1 - \frac{1}{\ell_n}\delta_{\ell_n-1}(S)\right). \tag{8}$$

Suppose that $\delta_{\ell_n-1}(S) \leq 1/2$. Then the right hand side of equation (8) equals the second term, so

$$\delta_{\ell_n-1}(S) \leq \frac{\ell_n}{\ell_n - 1} - \frac{\ell_n}{\ell_n - 1}\delta_{\ell_n}(S).$$

If this occurs for infinitely many n, then the set $\{\delta_{\ell_n-1}(S) : n \geq 1\}$ has an accumulation point less than or equal to

$$\lim_{n\to\infty} \frac{\ell_n}{\ell_n - 1} - \frac{\ell_n}{\ell_n - 1}\delta_{\ell_n}(S) = 1 - \lim_{n\to\infty} \delta_{\ell_n}(S) = 1 - C < B.$$

This is a contradiction, so (by possibly deleting finitely many ℓ_ns) we may assume that $\delta_{\ell_n-1}(S) > 1/2$ for every n. Thus the right hand side of equation (8) equals the first term, and

$$C = \lim_{n\to\infty} \delta_{\ell_n}(S) \leq \lim_{n\to\infty} \frac{1}{\ell_n} + \frac{\ell_n - 1}{\ell_n}\delta_{\ell_n-1}(S) = \lim_{n\to\infty} \delta_{\ell_n-1}(S).$$

But C is the maximum accumulation point of the $\delta_i(S)$, so in fact $\lim_{n\to\infty} \delta_{\ell_n-1}(S) = C$.

Again using part (3) of Lemma 2.1 and taking limits, we see that $C \leq \max(1/2, 1 - \log_N(2))$, which is a contradiction. □

Corollary 3.4 *Let $N = 2$, 3, or 4. Let S be an eventually non-periodic N-ary sequence. Then $T(S) = [B, 1 - B]$ for some real number B.*

Proof: For these values of N we have $\max(1/2, 1 - \log_N(2)) = 1/2$ and $\log_N(2) \geq 1/2$, so the first assertion of Theorem 3.3 says that $B \leq 1/2$ for all such S and the second assertion then says $C = 1 - B$ for all such S. □

Now fix a positive integer k. Consider a sequence $S = s_0, s_1, \cdots$ with each $s_i \in \{0, 1, \cdots, N-1\}$. For each i, let

$$t_i = \sum_{j=0}^{k-1} s_{ki+j} N^j \in \{0, 1, \cdots, N^k - 1\}$$

and let $T = t_0, t_1, \cdots$. Then the function $\Gamma : \mathbb{Z}_N \to \mathbb{Z}_{N^k}$ defined by

$$\Gamma\left(\sum_{i=0}^{\infty} s_i N^i\right) = \sum_{i=0}^{\infty} t_i (N^k)^i$$

is a ring isomorphism. By abuse of notation we also write $\Gamma(S) = T$.

Theorem 3.5 *Let $S = s_0, s_1, \cdots$ with each $s_i \in \{0, 1, \cdots, N-1\}$. Then the set of accumulation points of $\{\delta_n(S)\}$ is identical to the set of accumulation points of $\{\delta_n(\Gamma(S))\}$ (where we use N^k-adic complexity to define $\delta_n(\Gamma(S))$).*

Proof: First observe that if R is a ring, then there is a unique ring homomorphism from $\mathbb{Z}$ into R, defined by mapping 1 to 1. This is called the *canonical map* from $\mathbb{Z}$ to R. If U is any subring of the rational numbers, then there is at most one ring homomorphism from U to R, since any such homomorphism is still determined by mapping 1 to 1. In fact such a homomorphism exists if and only if the image of every invertible integer in U under the canonical map is a unit in R.

When $R = \mathbb{Z}_N$, the maximal subring of the rational numbers that maps to R is $U = \{a/b : \gcd(N, b) = 1\}$. In his case the map is an injection. Since $\gcd(N, b) = 1$

if and only if $\gcd(N^k, b) = 1$, U is also the maximal subring of the rational numbers that maps to $\mathbb{Z}_{N^k}$. By the uniqueness of these maps, they must be identified under the identification of $\mathbb{Z}_N$ and $\mathbb{Z}_{N^k}$.

Now suppose that D is an accumulation point of the $\delta_n(S)$, say

$$D = \lim_{i\to\infty} \delta_{n_i}(S).$$

There is some $j \in \{0, \cdots, k-1\}$ so that $\{n_i \equiv j \pmod k\}$ is infinite. Thus D is a limit of $\delta_n(S)$s with all n congruent to j.

Let $\alpha = \sum_{i=0}^{\infty} s_i N^i$. Suppose that k divides n. Then

$$\alpha \pmod{N^n} = \Gamma(\alpha) \pmod{(N^k)^{n/k}}.$$

Thus $\Phi_n(S)$ — the minimal $\Phi(a, b)$ among rational approximations of α modulo N^n — is the same as $\Phi_{n/k}(\Gamma(S))$ — the minimal $\Phi(a, b)$ among rational approximations of $\Gamma(\alpha)$ modulo $(N^k)^{n/k} = N^n$. Thus $\lambda_n(S) = k\lambda_{n/k}(\Gamma(S))$, and

$$\delta_n(S) = \frac{\lambda_n(S)}{n} = \frac{k\lambda_{n/k}(\Gamma(S))}{n} = \frac{\lambda_{n/k}(\Gamma(S))}{n/k} = \delta_{n/k}(\Gamma(S)).$$

Thus the set of accumulation points of the $\delta_n(S)$ that are limits of $\delta_n(S)$s with k dividing n coincides exactly with the set of accumulation points of the $\delta_n(\Gamma(S))$. Thus it remains only to show that if $1 \le j \le k-1$, then every accumulation point of $\{\delta_n(S) : n \equiv j \pmod k\}$ is also an accumulation point of $\{\delta_n(\Gamma(S))\}$.

Let S^j denote the shift of S by j positions, $S^j = s_j, s_{j+1}, \cdots = s_0^j, s_1^j, \cdots$. We claim that the set of accumulation points of $\{\delta_n(S) : n \equiv j \pmod k\}$ equals the set of accumulation points of $\{\delta_n(S^j) : k|n\}$. Indeed, letting $\alpha^{(j)} = \sum_{i=0}^{\infty} s_i^j N^i$, we have $\alpha = r_j + N^j\alpha^j$ with $r_j \in \mathbb{Z}$, $0 \le r_j < N^j$. Thus if a and b are integers so that $a/b \equiv \alpha \pmod{N^n}$ and $\gcd(b, N) = 1$, then $a - br_j = N^j c$ for some integer c and $c/b \equiv \alpha^{(j)} \pmod{N^{n-j}}$. We have $|c| \le 2\max(|a|, |b|)$, so $\Phi_{n-1}(S^j) \le 2\Phi_n(S)$. Therefore

$$\delta_{n-j}(S^j) \le \frac{n}{n-j}\delta_n(S) + \frac{\log_N(2)}{n-j}.$$

Conversely, suppose that a and b are integers so that $a/b \equiv \alpha^j \pmod{N^{n-j}}$. Then $\alpha \equiv r_j + N^j\alpha^j \equiv (r_j b + N^j a)/b \pmod{N^n}$. We have $|r_j b + N^j a| \le 2N^j \max(|a|, |b|)$, so $\Phi_n(S) \le 2N^j\Phi_{n-j}(S^j)$. Therefore

$$\frac{n}{n-j}\delta_n(S) - \frac{1 + \log_N(2)}{n-j} \le \delta_{n-j}(S^j).$$

Thus the set of accumulation points of $\{\delta_n(S) : n \equiv j \pmod k\}$ equals the set of accumulation points of $\{\delta_n(S^j) : k|n\}$, as claimed.

Let $\Gamma(S^j) = v_0, v_1, \cdots$, with $v_i \in \{0, 1, \cdots, N^2-1\}$. Note that $\Gamma(S^j)$ is not in general equal to $\Gamma(S)^j$. The same argument as above shows that the set of accumulation points of $\{\delta_n(S^j) : k|n\}$ equals the set of accumulation points of $\{\delta_n(\Gamma(S^j))\}$. Let

$$\beta = \sum_{i=0}^{\infty} t_i (N^k)^i \text{ and } \gamma = \sum_{i=0}^{\infty} v_i (N^k)^i.$$

Then $\beta \equiv r_j + N^j \gamma \pmod{(N^k)^n}$ for any $n \geq 1$. Thus if $a/b \equiv \beta \pmod{(N^k)^n}$ with $\gcd(b, N) = 1$, then $a - br_j = N^j c$ for some integer c, and

$$\begin{aligned} \frac{c}{b} &\equiv \gamma \pmod{N^{-j}(N^k)^n} \\ &\equiv \gamma \pmod{(N^k)^{n-1}}. \end{aligned}$$

Therefore $\Phi_{n-j}\Gamma(S^j)) \leq 2\Phi_n(\Gamma(S))$, so

$$\delta_{n-j}(\Gamma(S^j)) \leq (n/(n-j))\delta_n(\Gamma(S)) + \log_{N^k}(2)/(n-j).$$

Similarly, if $a/b \equiv \gamma \pmod{(N^k)^n}$, then

$$(N^j a + r_j b)/b \equiv \beta \pmod{(N^k)^n}.$$

Thus

$$\Phi_n(\Gamma(S)) \leq 2N^j \Phi_n(\Gamma(S^j)),$$

so that

$$\delta_n(\Gamma(S)) \leq \delta_n(\Gamma(S^j)) + \log_{N^k}(2N^j)/n.$$

It follows that any infinite sequence of $\delta_n(\Gamma(S^j))$s is upper bounded by one infinite sequence of $\delta_n(\Gamma(S))$s and lower bounded by another infinite sequence of $\delta_n(\Gamma(S))$s. Hence every accumulation point of the $\{\delta_n(\Gamma(S^j))\}$ is lower bounded and upper bounded by accumulation points of the $\{\delta_n(\Gamma(S))\}$. But from Theorem 1.1 we know that the set of accumulation points of the $\{\delta_n(\Gamma(S))\}$ is a closed interval, so every accumulation point of the $\{\delta_n(\Gamma(S^j))\}$ is an accumulation point of the $\{\delta_n(\Gamma(S))\}$. In particular, the accumulation points of $\{\delta_n(S) : n \equiv j \pmod{k}\}$ are all accumulation points of $\{\delta_n(\Gamma(S))\}$. This completes the proof of the theorem. □

Corollary 3.6 *Let N be a power of 2 or 3. Let S be an eventually non-periodic N-ary sequence. Then $T(S) = [B, 1-B]$ for some real number B.*

Proof: The proof is immediate from Corollary 3.4 and Theorem 3.5. □

4 All Balanced Intervals Occur as $T(S)$s

We denote by β the function that associates the real number B with the sequence S, where $T(S) = [B, C]$. That is,

$$\beta : \{S = s_0, s_1, \cdots, s_i \in \{0, 1, \cdots, N-1\}\} \to [0, \max(1/2, 1 - \log_N(2))]$$

and $\beta(S)$ is the least accumulation points of the set of N-adic complexities of prefixes of S. In this section we see that the image of β contains $[0, 1/2]$. In particular, if N is a power of 2 or 3, then β is surjective.

Theorem 4.1 *For every $N \geq 2$ and every $B \in [0, 1/2]$ there is an N-ary sequence S with $T(S) = [B, 1 - B]$.*

Proof: Let $0 \leq B < 1/2$. We build S with $\beta(S) = B$ in stages. Suppose that we have chosen $n_1 \leq n_2 \leq \cdots \leq n_r \in \mathbb{Z}^+$ and $S^{n_r} = s_0, \cdots, s_{n_r - 1} \in \{0, 1, \cdots, N-1\}$ so that

$$|\delta_{n_i}(S) - B| \leq \frac{1}{n_i},$$

$\delta_{n_i}(S) \leq B < \delta_j(S)$ for $n_i < j < n_{i+1}$, and

$$n_1 > \max\left(\frac{B+1}{1-2B}, \frac{2\log_N(2)}{1-2B}\right). \tag{9}$$

Choose s_{n_r} so that $\Phi_{n_r+1}(S) \neq \Phi_{n_r}(S)$. Some care must be taken here to see that this is in fact possible. Suppose that a/b is the rational approximation to $\sum_{i=0}^{r-1} s_i N^i$ modulo N^{n_r} so that $\Phi(a, b)$ is minimal. Suppose also that $a/b \equiv \sum_{i=0}^{n_r - 1} s_i N^i + s'_{n_r} N^{n_r} \pmod{N^{n_r+1}}$. We choose $s_{n_r} \neq s'_{n_r}$. Then the bound in equation (9) and the fact that $\Phi(a, b) < N^{nB}$ ensure that there are not integers c and d so that $c/d \equiv \sum_{i=0}^{n_r} s_i N^i \pmod{N^{n_r+1}}$ with $\Phi(a, b) = \Phi(c, d)$. That is, the N-adic complexity profile must increase at this point.

Now by Lemma 2.1, part (2),

$$\begin{aligned} \delta_{n_r+1}(S) &> \frac{n_r - 1}{n_r + 1} - \frac{n_r}{n_r + 1}\delta_{n_r}(S) \\ &\geq \frac{n_r - 1}{n_r + 1} - \frac{n_r}{n_r + 1}B \\ &> B, \end{aligned}$$

where the last inequality follows from equation (9). Also, as in the proof of Lemma 3.1, the limit of the δ_{n_r+1} is $1 - B$. Choose n_{r+1} and $s_{n_r+1}, \cdots, s_{n_{r+1}-1}$ so that $\Phi_{n_r+1}(S) = \cdots = \Phi_{n_{r+1}}(S)$ and $\delta_{n_{r+1}}(S) \leq B < \delta_{n_{r+1}-1}(S)$. This is possible since the $\delta_j(S)$ are

decreasing with limit 0 if the $\Phi_j(S)$ are unchanged. Moreover, for any j such that $\Phi_{j+1}(S) = \Phi_j(S)$, we have $\lambda_{j+1}(S) = \lambda_j(S)$ and

$$\begin{aligned}\delta_j(S) - \delta_{j+1}(S) &= \frac{\lambda_j(S)}{j} - \frac{\lambda_j(S)}{j+1} \\ &= \frac{\lambda_j(S)}{j(j+1)} \\ &\leq \frac{1}{j+1}.\end{aligned}$$

It follows that $B - 1/n_{r+1} < \delta_{n_{r+1}}(S) < B$. Thus $B = \lim_{i\to\infty} \delta_{n_i}(S)$. It also follows that B is the least accumulation point of the $\delta_j(S)$. Also, as in the proof of Lemma 3.1, the limit of the δ_{n_r+1} is $1 - B$ and this is the maximum accumulation point.

Finally, let $B = 1/2$. We construct S a term at a time as follows. If $\delta_r(S) < 1/2 - 2/r$, then choose s_r so that $\Phi_r(S) \neq \Phi_{r+1}(S)$. As before, we have made $\delta_r(S)$ small enough that $\Phi_r(S) \neq \Phi_{r+1}(S)$ if we choose the "wrong" s_r. Otherwise choose s_r so that $\Phi_r(S) = \Phi_{r+1}(S)$. We claim that $\lim_{r\to\infty} \delta_r(S) = 1/2$, so that $1/2$ is the only accumulation point.

Let r be an arbitrary index. If $\delta_r(S) \geq 1/2 - 2/r$, then $\delta_{r+1}(S) \geq 1/2 - 3/r$, which guarantees that the least accumulation point is at least $1/2$. If $\delta_r(S) < 1/2 - 2/r$, then

$$\begin{aligned}\delta_{r+1}(S) &> \frac{r-1}{r+1} - \frac{r}{r+1}\delta_r(S) \\ &> \frac{r-1}{r+1} - \frac{r}{r+1}(\frac{1}{2} - \frac{2}{r}) \\ &= \frac{1}{2} + \frac{1}{2(r+1)}.\end{aligned}$$

On the other hand, suppose $\delta_r(S) > 1/2 + 1/r$. Then

$$\delta_{r+1}(S) = \frac{r}{r+1}\delta_r(S) > \frac{1}{2} + \frac{1}{2(r+1)}.$$

If we apply our constructions and $\delta_j(S) > 1/2 - 2/j$ for $j = r, r+1, \cdots, k$ for some k, then

$$\delta_k(S) = \frac{r}{k}\delta_r(S).$$

Thus after finitely many steps we reach a k for which $\delta_k(S) < 1/2 - 2/k$. (In fact the first k for which $\delta_k(S) \leq 1/2$ is at most $k = 2(r+2)$.) We have shown that $\delta_r(S) \in [1/2 - 1/r, 1/2]$ for infinitely many r, and that $\delta_r(S) > 1/2 - 1/r$ for all r. Thus $B = 1/2$ is the least accumulation point. Again, as in the proof of Lemma 3.1, the limit of the δ_{r+1} for which $\Phi_r(S) \neq \Phi_{r+1}(S)$ is $1 - 1/2 = 1/2$ and this is the maximum accumulation point. This completes the proof. □

Note that if we modify this construction so that the the N-adic complexity changes when $\delta_{n_r}(S) < \delta_{n_r-1}(S) < B < \delta_{n_r-2}(S)$, then B is still the least accumulation point. In fact, at each phase we can either use this method or the one in the proof to determine when to change the N-adic complexity. Since there are infinitely many phases, this gives uncountably many sequences for which B is the least accumulation point.

Corollary 4.2 *Every balanced interval $[B, 1-B]$ occurs uncountably often as a $T(S)$ with S an N-ary sequence.*

5 Conclusions

We have characterized the sets of accumulation points of the normalized N-adic complexities of non-periodic sequences. This gives us a fuller understanding of the properties of these important security measures. It provides another in a growing list of ways that feedback with carry shift registers are similar to linear feedback shift registers. There are also practical implications to this and Dai, et al.'s work on linear complexity. Suppose a stream cipher uses an infinite non-periodic sequence S as a keystream, and suppose that the set of accumulation points of the normalized N-adic or linear complexity is $[B, 1-B]$. Now imagine a cryptanalyst who has observed a prefix of S and wants to predict the next symbol. If the normalized complexity up to this point is close to B, the next symbol is likely to change the complexity so the normalized complexity increases. Likewise, if the normalized complexity up to this point is close to $1-B$, then the next symbol is likely to leave the complexity unchanged so the normalized complexity decreases. In this sense sequences for which the set of accumulation points is $[0, 1]$ are the most random sequences.

The methods used in this paper to analyze N-adic complexity can also be applied to linear complexity, eliminating the need for continued fractions. In fact in the case of linear complexity the proofs are significantly simpler than the ones in this paper and than the ones originally given by Dai, et al. The pivotal fact is a lemma due to Massey [9] that says that if $\lambda'_n(S)$ is the linear complexity of the length n prefix of S, then $\lambda'_{n+1}(S) = \max(\lambda'_n(S), n+1-\lambda'_n(S))$. Thus we know more precisely how the linear complexity changes. Modifying the proofs in this paper using this fact is straightforward. Morever, it is straightforward to use this to prove the analogous theorems for the linear complexity of sequences over arbitrary finite fields. Since this was not observed by Dai, et al., we include it as a theorem without proof.

Theorem 5.1 *Let S be a sequence over the finite field F. Let $\delta'_n(S) = \lambda'_n(S)/n$ be the normalized linear complexity of S. Then the set of accumulation points of the set of $\delta'_n(S)$ is a closed interval $[B, 1-B]$. For every B, there are uncountably many sequences over F for which the set of accumulation points is $[B, 1-B]$.*

Various questions remain. If N is not a power of 2 or 3, and $\beta(S) > \log_N(2)$, is $T(S) = [\beta(S), 1 - \beta(S)]$? Is there any (perhaps measure theoretic) sense in which some least accumulation points are more likely than others? In the case of linear complexity over a field $\mathbb{F}_q$, Niederreiter showed that $T(S) = [1/2, 1/2]$ with probability 1 (where the set of infinite sequences is endowed with the infinite product measure arising from the uniform measure on $\mathbb{F}_q$) [10]. We do not have such a result for N-adic complexity. For a fixed S, what can be said about the distribution of the $\delta_n(S)$ in $[B, 1 - B]$? If the distribution is non-uniform, this might lead to better prediction methods than those outlined in the first paragraph of this section.

Finally, we have treated, in varying detail, a number of generalizations of feedback with carry shift registers and linear feedback shift registers [4, 5, 7]. Do the same results hold in these settings?

References

[1] Z. Dai, K. Imamura, and J. Yang, *Asymptotic behavior of normalized linear complexity of multi-sequences*, Sequences and Their Applications - SETA 2004, Springer-Verlag Lecture Notes in Computer Science, **3486** (2005), 126-142.

[2] Z. Dai, S. Jiang, K. Imamura, and G. Gong, *Asymptotic behavior of normalized linear complexity of ultimately non-periodic sequences*, IEEE Trans. Info. Theory, **50** 2911-2915.

[3] X. Feng and X. Dai, *The expected value of the normalized linear complexity of 2-dimensional binary sequences*, Sequences and Their Applications - SETA 2004, Springer-Verlag Lecture Notes in Computer Science, **3486** (2005), 113-128.

[4] M. Goresky and A. Klapper, *Polynomial pseudo-noise sequences based on algebraic feedback shift registers*, IEEE Trans. Info. Theory, **53** (2006), 1649-1662.

[5] A. Klapper, *Distribution properties of d-FCSR sequences*, Journal of Complexity **20**, (2004), 305-317.

[6] A. Klapper and M. Goresky, *Feedback Shift Registers, 2-Adic Span, and Combiners with Memory*, J. Cryptology, **10** (1997), 111-147.

[7] A. Klapper and J. Xu, *Algebraic feedback shift registers*, Theoretical Comp. Sci., **226** (1999), 61-93.

[8] A. Klapper and J. Xu, *Register synthesis for algebraic feedback shift registers based on non-primes*, Designs, Codes, and Cryptography, **31** (2004), 227-25.

[9] J. Massey, *Shift-register synthesis and BCH decoding*, IEEE Trans. Infor. Theory, **IT-15** (1969), 122-127.

[10] H. Niederreiter, *The probabilistic theory of linear complexity*, in C.G. Gunther, ed., Advances in Cryptology EUROCRYPT 88, Springer-Verlag Lecture Notes in Computer Science, **330** (1988), 191209.