Non-Uniqueness and Radius of Cyclic Unary NFAs

Michael Domaratzki*
School of Computing, Queen's University,
Kingston, ON K7L 3N6 Canada
email: domaratz@cs.queensu.ca

Keith Ellul, Jeffrey Shallit†and Ming-Wei Wang
School of Computer Science, University of Waterloo,
Waterloo, ON N2L 3G1 Canada
email:{kbellul@alumni,shallit@graceland,m2wang@math}.uwaterloo.ca

Abstract

In this paper we study some properties of cyclic unary languages. We find a connection between the uniqueness of the minimal NFA for a cyclic unary language and a certain Diophantine equation first studied by Sylvester.

We also obtain some results on the radius of unary languages. We show that the nondeterministic radius of a cyclic unary regular language L is not necessarily obtained by any of the minimal NFAs for L. We also give a class of examples which demonstrates that the nondeterministic radius of a regular language cannot necessarily even be approximated by the minimal radius of its minimal NFAs.

1 Introduction

State complexity of regular languages, that is, the study of the size of minimal finite automata, both deterministic and nondeterministic, has received much attention recently. We refer the reader to the recent survey of Yu [17] for a survey of deterministic state complexity. For recent work on nondeterministic state complexity, we refer the interested reader to Holzer and Kutrib [10] or Ellul [9].

Unary languages are often of particular interest in the study of state complexity, due to their relation to many number-theoretic results, as well as their difference from the general, non-unary case. Cyclic unary languages were investigated by Jiang *et al.* [14] in their study of the complexity of minimizing nondeterministic finite automata (NFAs).

Motivated by the work of Jiang *et al.*, in this paper, we study some properties of cyclic unary regular languages and NFAs. We find a connection between the Diophantine equation

$$1 = \sum_{i=1}^{k} \frac{1}{n_i} + \frac{1}{\prod_{i=1}^{k} n_i},$$

*Work supported in part by an NSERC PGS-B graduate scholarship.
†Research supported by NSERC.

with $1 < n_1 < n_2 < \cdots < n_k$, and the uniqueness of minimal cyclic unary NFA. This Diophantine equation has been studied since at least 1880, when it was investigated by Sylvester [16].

We also investigate the relation between the size of a minimal NFA and the nondeterministic radius of regular languages. We show that the nondeterministic radius of a regular language L is not necessarily attained by a minimal NFA for L. In the deterministic case, Ellul showed that the analogous result does hold, i.e., that the deterministic radius of a regular language L is the radius of the minimal deterministic finite automaton (DFA) [9]. This eliminates a potential algorithm for computing the nondeterministic radius of a regular language, which remains an open problem.

2 Preliminaries

Let 2^X denote the power set of X: $2^X = \{Y : Y \subseteq X\}$. Let $\mathbb{N}$ denote the set of natural numbers $\mathbb{N} = \{0, 1, \dots\}$ and $\mathbb{N}^+$ denote the set of non-zero natural numbers: $\mathbb{N}^+ = \mathbb{N} - \{0\}$.

In this paper we work with languages over the unary alphabet $\Sigma = \{a\}$. We assume the reader is familiar with the concepts of and notation for deterministic finite automata (DFAs) and nondeterministic finite automata as described, for example, in Hopcroft and Ullman [12]. In particular, for us a nondeterministic finite automaton (NFA) is a 5-tuple $M = (Q, \Sigma, \delta, q_0, F)$ such that Q is a finite set of states, $\Sigma = \{a\}$ is the unary input alphabet, $\delta : Q \times \Sigma \to 2^Q$ is the transition function, $q_0 \in Q$ is the start state and $F \subseteq Q$ is the set of final states. An NFA is a DFA if $|\delta(q, a)| = 1$ for all $q \in Q$, $a \in \Sigma$.

A minimal NFA (DFA) for a given regular language is an NFA (DFA) that recognizes the language with a minimal number of states. It is known that the minimal DFA is unique up to the names of the states, but minimal NFAs are not necessarily unique.

Let L be a regular language. Then $\mathrm{nsc}(L)$, the *nondeterministic state complexity* of L, is the number of states in any minimal NFA accepting L. Similarly $\mathrm{sc}(L)$ is the *deterministic state complexity* of L, that is, the number of states in the minimal DFA accepting L.

A unary language $L \subseteq \Sigma^*$ is *cyclic* if there exists an integer n such that $a^i \in L \iff a^{i+n} \in L$ for all $i \geq 0$. In this case, we say that L is *n-cyclic*.

We say that a set $\{n_1, n_2, \dots, n_k\} \subseteq \mathbb{N}$, is *division-free* if for all $1 \leq i < j \leq k$, we have $n_i \not| n_j$ and $n_j \not| n_i$. In the literature, division-free sets are sometimes called primitive sets.

For any finite set $S = \{s_1, s_2, \dots, s_n\} \subseteq \mathbb{N}^+$, we define $\mathrm{lcm}(S) = \mathrm{lcm}(s_1, \dots, s_n)$. We let $[n]$ denote the set $\{1, 2, 3, \dots, n\}$ for any $n \in \mathbb{N}^+$.

3 Results on Minimal Unary NFAs

In this section, we give some new and known results on minimal unary NFAs that will be applied to give the non-uniqueness of minimal cyclic unary NFAs.

The minimal DFA of a cyclic language is a simple loop. However, the following result, quoted essentially verbatim from Jiang *et al.* [14, Theorem 2.1], gives the structure of minimal NFAs for cyclic languages:

Theorem 3.1 *Let L be an m-cyclic language. Then L has a minimal NFA in one of the following two forms:*

(a) M consists of a single directed cycle having k states where $k \mid m$.

(b) M consists of two or more pairwise disjoint cycles, each reached by a transition from the start state. The start state does not belong to any cycle. There are no other transitions in M, and each cycle length is a divisor of m.

We say that a unary NFA is in *tail-less Chrobak Normal Form* (t-CNF) if it is in form (b) of Theorem 3.1 (see Chrobak [8] for a definition of Chrobak Normal Form). We will require the following three technical lemmas on deterministic and nondeterministic state complexity.

Lemma 3.2 *Let $m_1, m_2, \ldots, m_k \in \mathbb{N}^+$. Then for any NFA M in t-CNF with loops of size $m_1, \ldots, m_k$, there exists a DFA with $\operatorname{lcm}(m_1, m_2, \ldots, m_k)$ states accepting the same language.*

Proof. Let L be the language accepted by the NFA $M = (Q, \{a\}, \delta, q_0, F)$ with loops of size $m_1, m_2, \ldots, m_k$. Let $m = \operatorname{lcm}(m_1, m_2, \ldots, m_k)$. Let $L[i]$ be the set of natural numbers $\leq m_i$ corresponding to lengths of strings accepted by the i-th loop of M (of size m_i). Then we let $M' = ([m], \{a\}, \delta, m, F')$ be the DFA given by $\delta(i, a) = i + 1 \pmod m$ and

$$F' = \bigcup_{i=1}^{k} \{j \ : \ \exists \ell \in L[i] \text{ such that } j \equiv 0 \pmod \ell\}.$$

It is easy to establish that $L(M') = L$. ■

Lemma 3.3 *Let $k \geq 2$, $\{m_1, m_2, \ldots, m_k\} \subseteq \mathbb{N}^+$ be a division-free set and $L = \bigcup_{i=1}^{k} (a^{m_i})^*$. Then $\operatorname{sc}(L) = \operatorname{lcm}(m_1, m_2, \ldots, m_k)$.*

Proof. Assume that there exists a DFA accepting $L = \cup_{i=1}^{k} (a^{m_i})^*$ with less than $\operatorname{lcm}(m_1, \ldots, m_k)$ states. Let ℓ be chosen with $1 \leq \ell < \operatorname{lcm}(m_1, \ldots, m_k)$ so that there exists $M = ([\ell], \{a\}, \delta, \ell, F)$ accepting L for some F and δ given by $\delta(j, a) = j + 1 \pmod \ell$ for all $j \in [\ell]$.

Let $i \in [k]$ be chosen so that m_i does not divide ℓ. Let $g = \gcd(m_i, \ell)$. Consider the Diophantine equation $xm_i = g + y\ell$, with x, y variables. Let $[x_0, y_0]$ be a solution to this equation with $x_0, y_0 \geq 0$; such solutions can easily be seen to exist regardless of m_i and ℓ. Thus, we have that $x_0 m_i \equiv g \pmod \ell$. This implies that $a^g \in L$. Therefore, there exists $j \in [k]$ such that $a^g \in (a^{m_j})^*$. Thus, $m_j \mid g$. As $g \mid m_i$ by our choice of g, we have that $m_j \mid m_i$. If $i \neq j$, then we have a contradiction to our assumption that $\{m_1, \ldots, m_k\}$ is division-free. Otherwise, $m_i = m_j = g$ and $m_i \mid \ell$, which contradicts our choice of i. Thus, $\operatorname{sc}(L) \geq \operatorname{lcm}(m_1, \ldots, m_k)$. ■

Lemma 3.4 *Let $n_1, n_2, \ldots, n_k \in \mathbb{N}^+$, and $L = \bigcup_{i=1}^{k} (a^{n_i})^*$. Let M be a unary NFA in t-CNF accepting L. Assume that M has loops of size $c_1, c_2, \ldots, c_\ell$, with the state we arrive at on input a^k in the loop of size c_i labeled $[i,k]$ for all $1 \leq i \leq \ell$ and $1 \leq k \leq c_i$. Then for each c_i and each final state $[i,r]$ of c_i there exists an n_j such that $n_j \mid \gcd(c_i, r)$. In particular, $n_j \mid c_i$ and $n_j \mid r$.*

Proof. Fix an index i with $1 \leq i \leq \ell$. Assume that $n_k \geq n_j$ for all $1 \leq j \leq k$.

Suppose that the loop with length c_i has a final state $[i,r]$, that is, suppose the loop accepts $a^r(a^{c_i})^*$.

Then let $d = \gcd(r, c_i)$, $r' = r/d$, and $c' = c_i/d$. By Dirichlet's Theorem, there are infinitely many primes in the arithmetic progression $\{r' + mc' \ : \ m \geq 0\}$. Choose an $m \geq 0$ such that $p = r' + mc'$ is a prime with $p > n_k$. Since $d \cdot p = r + mc_i$, a^{dp} is accepted by the loop of length c_i, and there must be q and n_j such that $d \cdot p = q \cdot n_j$. Since $p > n_j$, we must have $p \mid q$. Thus $n_j \mid d$. This proves the lemma. ■

Theorem 3.5 *Let $k \geq 2$. Let $m_1, m_2, \ldots, m_k \in \mathbb{N}^+$, and $L = \bigcup_{i=1}^{k} (a^{m_i})^*$. Then $\mathrm{nsc}(L) = 1 + \sum_{i=1}^{k} m_i$ if and only if the following conditions hold:*

(a) $1 + \sum_{i=1}^{k} m_i \leq \mathrm{lcm}(m_1, m_2, \ldots, m_k)$, *and*

(b) for all proper subsets $S \subset \{m_1, m_2 \ldots, m_k\}$ with $|S| \geq 2$, we have $\sum_{m \in S} m \leq \mathrm{lcm}(S)$.

Remark: For $k = 2$, Theorem 3.5 was proven under slightly weaker conditions by Holzer and Kutrib [11, Thm. 4].

We call the condition that $\sum_{m \in S} m \leq \mathrm{lcm}(S)$ for all $S \subset \{m_1, \ldots, m_k\}$ the *local-lcm property.*

Proof.
($\Rightarrow$): Assume that $\mathrm{nsc}(L) = 1 + \sum_{i=1}^{k} m_i$. Let $m = \mathrm{lcm}(m_1, m_2, \ldots, m_k)$. Then L is accepted by an m-state DFA by Lemma 3.2. Thus, $\mathrm{nsc}(L) \leq m$.

Suppose that condition (a) does not hold. Then since $\mathrm{nsc}(L) \leq m$, it is not true that $\mathrm{nsc}(L) = 1 + \sum_{i=1}^{k} m_i$.

On the other hand, suppose that condition (b) does not hold. Let S be a subset of $\{m_1, \ldots, m_k\}$ with $\sum_{m \in S} m > \mathrm{lcm}(S)$. Let $m' = \mathrm{lcm}(S)$, and $L_S = \bigcup_{m \in S} (a^m)^*$. By Lemma 3.2, there is a DFA with m' states accepting L_S, so that $\mathrm{nsc}(L_S) \leq m'$. Now, $L = L_S \cup (\bigcup_{m \notin S} (a^m)^*)$. Since we can accept each $(a^m)^*$ by a loop of size m, the nondeterministic state complexity of L is at most $1 + m' + \sum_{m \notin S} m$ (by the standard union construction; see, e.g., Holzer and Kutrib [11, Thm. 3]) which is strictly less than $1 + \sum_{i=1}^{k} m_i$, since $m' < \sum_{m \in S} m$.
($\Leftarrow$): Assume that (a) and (b) do hold, but $\mathrm{nsc}(L) \neq 1 + \sum_{i=1}^{k} m_i$. By the standard union construction, we know that $\mathrm{nsc}(L) \leq 1 + \sum_{i=1}^{k} m_i$. Thus, it must be that $\mathrm{nsc}(L) < 1 + \sum_{i=1}^{k} m_i$. Thus $\mathrm{nsc}(L) < \mathrm{lcm}(m_1, m_2, \ldots, m_k)$.

First, suppose that M is a DFA. We show that $\{m_1, \ldots, m_k\}$ is division-free. Assume not. Therefore, there exist m_i, m_j such that $m_i \mid m_j$. Then $\mathrm{lcm}(m_i, m_j) = \max(m_i, m_j) < m_i + m_j$. This contradicts our assumptions on $\{m_1, \ldots, m_k\}$. Thus,

we can apply Lemma 3.3, and M has $\text{lcm}(m_1, m_2, \ldots, m_k)$ states. Thus, we may assume that M is in t-CNF. Let M be any NFA in t-CNF, with loops $C_1, C_2, \ldots, C_\ell$ of sizes $c_1, c_2, \ldots, c_\ell$ respectively. For all $1 \leq i \leq \ell$ and $1 \leq r \leq c_i$, let the state in loop C_i we arrive at on input a^r be labeled by $[i, r]$. Assume that there exists a C_i such that $c_i \notin \{m_1, m_2, \ldots, m_k\}$. Let $S \subseteq \{m_1, m_2, \ldots, m_k\}$ be the set of m_j such that $m_j \in S$ if and only if $m_j \mid c_i$. Thus, $c_i \geq \text{lcm}(S)$.

Consider first the case where $|S| = 1$. Then let $m_j \mid c_i$. By Lemma 3.4, for each final state $[i, r]$ of C_i, $m_j \mid r$ (since no other m_i with $1 \leq i \leq k$ divides c_i). Thus, we can replace the final state $[i, r]$ in loop C_i with a loop of size m_j, with only one final state (the state we arrive at on input a^{m_j}), accepting $(a^{m_j})^*$. Since this accepts all the strings accepted by C_i, and no strings that are not in L, the resulting NFA accepts L, and has $c_i - m_j \geq 0$ fewer states.

Now, let $|S| \geq 2$. For each final state $[i, r]$ of loop C_i, we have that, for some $m_j \in S$, $m_j \mid r + mc_i$ for all $m \geq 0$. Thus,

$$a^r(a^{c_i})^* \subseteq (a^{m_j})^*. \tag{1}$$

Thus, we take M and we replace the loop C_i with a loop of size m_j for each $m_j \in S$. Further, we assign final states so that the loop of size m_j accepts $(a^{m_j})^*$. By (1), this does not cause M to accept any fewer strings. Further, as $(a^{m_j})^* \subseteq L(M) = L$, the replacement of C_i does not cause M to accept any more strings. Thus, we have an NFA with $c_i - \sum_{m \in S} m \geq \text{lcm}(S) - \sum_{m \in S} m \geq 0$ fewer states accepting the same language.

In this way, we can always replace each loop of size $c_i \notin \{m_1, \ldots, m_k\}$ with loops of sizes taken from $\{m_1, \ldots, m_k\}$, without increasing the size of M. Thus, for some $R \subseteq \{m_1, m_2, \ldots, m_k\}$, we have that $\text{nsc}(L) = 1 + \sum_{m \in R} m$.

We now claim that this $R = \{m_1, m_2 \ldots, m_k\}$. If $R \neq \{m_1, \ldots, m_k\}$, then we have a t-CNF NFA M for L with loop sizes R for some proper subset R of $\{m_1, \ldots, m_k\}$. Using the standard union construction, we can construct a DFA for L with $\text{lcm}(R)$ states. As the above process creates M that accepts L, we must have that $\text{lcm}(R) \geq \text{lcm}(m_1, m_2, \ldots, m_k)$, by Lemma 3.3. Thus,

$$\text{lcm}(R) \geq 1 + \sum_{i=1}^{k} m_i > \sum_{m \in R} m.$$

This contradicts condition (b). Thus, we must have $R = \{m_1, m_2, \ldots, m_k\}$, and $\text{nsc}(L) = 1 + \sum_{i=1}^{k} m_i$. ■

4 Non-Uniqueness of Minimal Unary NFAs

In their paper on minimal unary NFAs, Jiang *et al.* note that minimal NFAs are not unique. However, the example they provide is not a unary NFA (it accepts the language $L = (0+1)^*1$ [14, Fig. 3]). This raises the following question: are minimal unary NFAs unique? It is easy to see that Figure 1 shows two minimal NFAs for the language $L = a + a^n$ for any $n \geq 2$.

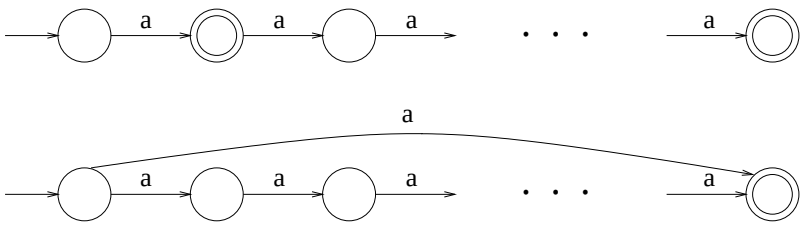

Figure 1: Two minimal NFAs accepting $L = a + a^n$.

However, the subclass of cyclic unary languages, to which Theorem 3.1 applies, is not so simple. We first note that trivial examples of cyclic unary languages possessing non-unique minimal NFAs exist. For instance, consider any language L which has a minimal NFA $M = (Q, \{a\}, \delta, q_0, F)$ in t-CNF in which one loop of size n has two final states. Thus, there exist $0 \leq n_1, n_2 < n$ such that $(a^n)^* a^{n_1} \cup (a^n)^* a^{n_2}$ is the language accepted by the loop of size n. Consider replacing the loop of size n with another of size n with only one final state, and two transitions from the initial state q_0 to the loop of size n. These two transitions leaving q_0 can be assigned to ensure that the loop of size n accepts exactly $(a^n)^* a^{n_1} \cup (a^n) a^{n_2}$. The resulting NFA will also be minimal if n is odd or $n_1 \not\equiv n_2 \pmod{n/2}$. An example is given in Figure 2 (additional loops in both automata are not depicted).

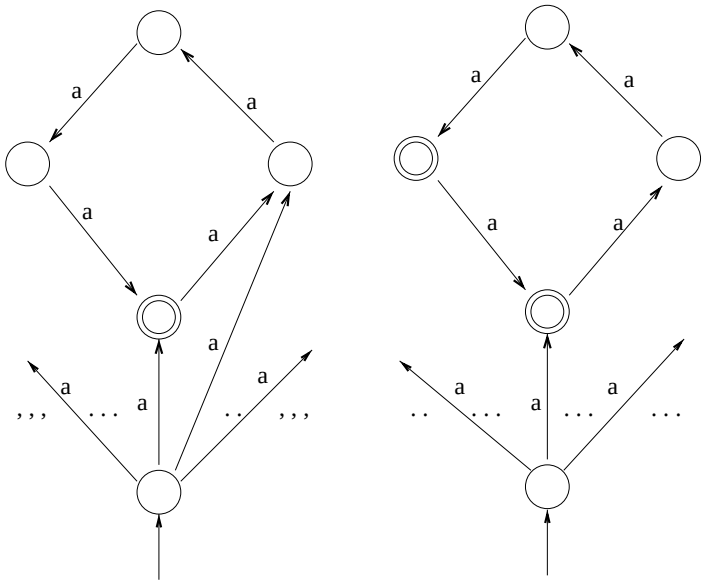

Figure 2: Two equivalent loops accepting $(a + a^4)(a^4)^*$.

However, we view these as trivial examples of non-unique minimal NFAs, since the set of loop sizes in both NFAs are equal. Thus, in what follows, we seek cyclic unary languages with non-unique minimal NFAs such that either (a) one NFA is in t-CNF and one is a DFA or (b) both NFAs are in t-CNF, but have different sets of loop sizes. This leads us to study connections to a Diophantine equation that has many applications and has been studied for over 120 years.

4.1 The Diophantine equation $\sum 1/n_i + 1/\prod n_i = 1$

The study of the equation

$$1 = \sum_{i=1}^{k} \frac{1}{n_i} + \frac{1}{\prod_{i=1}^{k} n_i} \tag{2}$$

for $1 < n_1 < n_2 < \cdots < n_k$ has a long history. Sylvester studied it in 1880 [16], and it has connections to complex surface singularities and perfect graphs [2, 3, 4, 5, 7, 13].

Consider any solution $1 < n_1 < n_2 < \cdots < n_k$ to (2). Note that, clearing denominators, (2) implies (cf., Brenton and Drucker [4, Eq. (2)])

$$\prod_{i \neq j} n_i \equiv -1 \pmod{n_j}. \tag{3}$$

Thus, we have $\gcd(n_i, n_j) = 1$ for all $1 \leq i < j \leq k$. Let $n = \prod_{i=1}^{k} n_i$. Then we have

$$n = \sum_{i=1}^{k} \frac{n}{n_i} + 1.$$

Letting $m_i = \frac{n}{n_i}$ for all $1 \leq i \leq k$, we get

$$n = \sum_{i=1}^{k} m_i + 1 \tag{4}$$

and

$$\operatorname{lcm}(m_1, m_2, \ldots, m_k) = n. \tag{5}$$

4.2 Non-Uniqueness of Unary Cyclic NFAs

We establish the solutions to (2) always yield a cyclic unary language whose minimal NFA is not unique. First, we observe that any $m_1, m_2, \ldots, m_k$ derived from (2) satisfy the local-lcm property.

Lemma 4.1 *Let $\{m_1, m_2, \ldots, m_k\} \subseteq \mathbb{N}$ satisfy (4) and (5), derived from a set $\{n_1, n_2, \ldots, n_k\} \subseteq \mathbb{N}$ satisfying (2). Then for any proper subset $S \subset \{m_1, \ldots, m_k\}$ with $|S| \geq 2$, we have*

$$\sum_{x \in S} x \leq \operatorname{lcm}(S).$$

Proof. Let $n = \operatorname{lcm}(m_1, m_2, \ldots, m_k)$. Let $m_i, m_j \in S$ for $i \neq j$. Then $m_i = n/n_i$ and $m_j = n/n_j$. By the fact that $\gcd(n_i, n_j) = 1$, we have that

$$\operatorname{lcm}(m_i, m_j) = n.$$

Thus, as $\operatorname{lcm}(m_1, m_2, \ldots, m_k) = n$ as well, we must have that $\operatorname{lcm}(S) = n$. Thus, by (4),

$$\sum_{m \in S} m \leq \sum_{i=1}^{k} m_i \leq n = \operatorname{lcm}(S).$$

■

Theorem 4.2 *Let $\{m_1, m_2, \ldots, m_k\} \subseteq \mathbb{N}$ satisfy (4) and (5), derived from a solution to (2). Then there exists a unary n-cyclic regular language L such that the minimal NFA for L is not unique.*

Proof. Let $L = \bigcup_{i=1}^{k} (a^{m_i})^*$. Then by Lemma 4.1 and Theorem 3.5, there exists a minimal NFA M_1 for L with $1 + \sum_{i=1}^{k} m_i$ states. By Lemma 3.2, there exists a DFA M_2 of size $\mathrm{lcm}(m_1, m_2, \ldots, m_k)$ states recognizing the same language as M_1. Since the number of states in M_1 and M_2 are equal, they are both minimal NFAs for L. They are clearly non-isomorphic. ■

Corollary 4.3 *There are infinitely many $n \in \mathbb{N}^+$ such that there exists a unary n-language whose minimal NFA is not unique.*

Proof. This corollary is due to a result of Sylvester [16], which gives an infinite sequence of numbers n_i, for which $\{n_1, n_2, \ldots, n_k\}$ satisfies (2) for all $k \geq 2$. By the construction of Theorem 4.2, each of the sets $\{n_1, n_2, \ldots, n_k\}$ yields a language whose minimal NFA is not unique. The infinite sequence is given by the recurrence [16]

$$\begin{aligned} n_1 &= 2 \\ n_k &= n_{k-1}^2 - n_{k-1} + 1 \end{aligned} \tag{6}$$

which gives the sequence $2, 3, 7, 43, 1807, \ldots$. (sequence A000058 in Sloane [15]). This yields the following sequence of sizes of non-unique cyclic unary NFAs: 6, 42, 1806, 3263442, ... (i.e., one less than the terms of A000058 beginning from 7). ■

4.3 Non-unique Minimal NFAs which use Nondeterminism

Consider a solution to the equation

$$1 = \frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_k} \tag{7}$$

where $1 < n_1 < n_2 < \cdots < n_k$ and for all $1 \leq i < j \leq k$, n_i does not divide n_j. Further, let $t \in \mathbb{N}^+$ be chosen so that t does not divide n_i and n_i does not divide t for all $1 \leq i \leq k$.

Let $n = \mathrm{lcm}(n_1, n_2, \ldots, n_k)$. Then clearing fractions, we get that

$$n = \sum_{i=1}^{k} \frac{n}{n_i}.$$

Let $m_i = \frac{n}{n_i}$. Then $\mathrm{lcm}(m_1, m_2, \ldots, m_k) = n$. Thus,

$$\sum_{i=1}^{k} m_i = \mathrm{lcm}(m_1, m_2, \ldots, m_k).$$

Thus, if we have a solution to (7), then we can construct two NFAs, neither of which is a DFA, of the same size accepting the same language. To do so, we construct one

NFA with loops of size $t, m_1, m_2, \ldots, m_k$, and one with loops of size t and n. Both use $t + n + 1$ states.

Solutions to (7) are given by Burshtein [6] and Barbeau [1]. The solution of Burshtein gives

$$n = 237001175614269189185623840224094316345178 0,$$

$k = 79$ and $t = 8$. The solution of Barbeau gives

$$n = 92240748796496554021780901360901994476034912459320579235643337 0,$$

$k = 101$ and $t = 4$.

Unfortunately, however, neither Burshtein's nor Barbeau's solution satisfy the local-lcm property of Theorem 3.5. Consider that $14, 21, 35 \in \{n_1, n_2, \ldots, n_k\}$ for both solutions. Then we have that $n/14, n/21, n/35 \in \{m_1, m_2, \ldots, m_k\}$ for both solutions. As $n = \text{lcm}(m_1, m_2, \ldots, m_k)$,

$$\text{lcm}(\frac{n}{14}, \frac{n}{21}, \frac{n}{35}) = \frac{n}{7}$$

and further

$$\frac{n}{14} + \frac{n}{21} + \frac{n}{35} = \frac{n}{7}(\frac{1}{2} + \frac{1}{3} + \frac{1}{5}) > \frac{n}{7}.$$

To illustrate this point, for the Burshtein example, we note that we can replace the loops of size

$$\begin{aligned} m_{70} &= 67714621604076911195892525778312661812908 \\ m_{75} &= 11285770267346151865982087629718776968818 0 \\ m_{77} &= 16928655401019227798973131444578165453227 0 \end{aligned}$$

by a loop the size of their lcm,

$$33857310802038455597946262889156330906454 0$$

which is less than the sum of these three loop sizes,

$$34985887828773070784544471652128208603335 8.$$

The following remains open:

Open Problem 4.4 *Does there exist a cyclic unary regular language L with two minimal NFAs, both in t-CNF?*

5 Nondeterministic Radius

We now turn to the automata-theoretic notion of the *radius* of a DFA, NFA or regular language.

If $M = (Q, \Sigma, \delta, q_0, F)$ is an initially connected NFA, for all $q \in Q$, define

$$\text{depth}(q) = \min_{x \in \Sigma^*} \{|x| \ : \ q \in \delta(q_0, x)\}.$$

The radius of M, denoted $\mathrm{rad}(M)$, is $\max\{\mathrm{depth}(q) \ : \ q \in Q\}$.

Given a regular language L, its deterministic radius (denoted $\mathrm{rad}(L)$) is the minimal radius of any DFA accepting it. The nondeterministic radius of L (denoted $\mathrm{nrad}(L)$) is the minimal radius of any NFA accepting it.

Ellul [9, Thm. 44] has established the following result about (deterministic) radius:

Theorem 5.1 *Let L be a regular language, and M be the minimal DFA for L. Then* $\mathrm{rad}(L) = \mathrm{rad}(M)$.

Our goal in this section is to establish that the result does not hold if we replace deterministic radius with nondeterministic radius in Theorem 5.1. We will require the following result:

Theorem 5.2 *Let $k \geq 2$, $m_1, m_2, \ldots, m_k \in \mathbb{N}^+$, and $L = \bigcup_{i=1}^{k} (a^{m_i})^*$. Suppose the following conditions hold:*

(a) $1 + \sum_{i=1}^{k} m_i > \mathrm{lcm}(m_1, m_2, \ldots, m_k)$, *and*

(b) for all proper subsets $S \subset \{m_1, m_2 \ldots, m_k\}$ with $|S| \geq 2$, we have $\sum_{m \in S} m \leq \mathrm{lcm}(S)$.

Then we have that

(i) $\mathrm{nsc}(L) = \mathrm{lcm}(m_1, m_2, \ldots, m_k)$, *and*

(ii) the minimal NFA for L is unique.

Note that condition (b) is the same as Theorem 3.5, while condition (a) is reversed from Theorem 3.5 (further, Theorem 3.5 gives both necessary and sufficient conditions).

Proof. The proof is adapted from the proof of Theorem 3.5. We can again assume, contrary to what we want to prove, that there exists a NFA M in t-CNF which accepts L with less than $\mathrm{lcm}(m_1, m_2, \ldots, m_k)$ states. We then apply the same construction to M to arrive at an NFA with $1 + \sum_{i=1}^{k} m_i$ states. However, in this case, this is more than $\mathrm{lcm}(m_1, m_2, \ldots, m_k)$, by (a). Thus, M is not minimal. This establishes the result. ■

We now give our result:

Theorem 5.3 *For infinitely many n, there exists a cyclic unary regular language L_n with $nsc(L_n) = n$, but for any minimal NFA M accepting L_n,* $\mathrm{rad}(M) > \mathrm{nrad}(L_n)$.

Proof. Let n_i for $i \geq 1$ be defined by (6). Let n_i' for $i \geq 1$ be defined by $n_i' = n_i - 2$. Note that we have that $n_k' = (\prod_{i=1}^{k-1} n_i) - 1$ for all $k \geq 2$.

Let $k \geq 3$. Let $n = \mathrm{lcm}(n_1, \ldots, n_{k-1}, n_k')$ and define $m_i = n/n_i$ for $1 \leq i \leq k-1$ and $m_k = n/n_k'$. Let $L_n = \bigcup_{i=1}^{k} (a^{m_i})^*$. Then we claim that L_n satisfies the conditions of the theorem.

First, we claim that $\gcd(n_i, n_j) = 1$ for all $1 \leq i < j \leq k-1$ and $\gcd(n_i, n'_k) = 1$ for all $1 \leq i \leq k-1$. The first claim follows by (3). For the second, as $n_k = \prod_{i=1}^{k-1} n_i + 1$, we have that $n'_k = \prod_{i=1}^{k-1} n_i - 1$, and $n'_k \equiv -1 \pmod{n_j}$ for all $1 \leq j \leq k-1$. From these facts, it follows that $\gcd(m_i, m_j) = 1$ for all $1 \leq i < j \leq k$, and $\mathrm{lcm}(m_1, \ldots, m_k) = n$.

To establish (a), note that as $n_1, \ldots, n_{k-1}$ are defined by (6), then by (2), we have that

$$\sum_{i=1}^{k-1} \frac{1}{n_i} + \frac{1}{\prod_{i=1}^{k-1} n_i} = 1. \tag{8}$$

By definition, $n'_k < \prod_{i=1}^{k} n_i$. Thus, $\sum_{i=1}^{k-1} \frac{1}{n_i} + \frac{1}{n'_k} > \sum_{i=1}^{k-1} \frac{1}{n_i} + \frac{1}{\prod_{i=1}^{k-1} n_i} = 1$. Consider now

$$1 + \sum_{i=1}^{k} m_i > \sum_{i=1}^{k} m_i = n \left(\left(\sum_{i=1}^{k-1} \frac{1}{n_i} \right) + \frac{1}{n'_k} \right) > n = \mathrm{lcm}(m_1, \ldots, m_k).$$

Thus, (a) is satisfied.

We now turn to condition (b). For all $S \subset \{m_1, \ldots, m_k\}$ with $|S| \geq 2$, $\mathrm{lcm}(S) = n$. Thus, it suffices to show that $\sum_{m \in S} m \leq n$ for all $S \subset \{m_1, \ldots, m_k\}$ with $|S| \geq 2$. Note that if $S \neq \{m_1, \ldots, m_k\}$,

$$\sum_{m \in S} m \leq \sum_{i=1}^{k-1} m_i \leq n \left(\sum_{i=1}^{k-1} \frac{1}{n_i} \right) \leq n.$$

The last inequality is by (8). Thus, (b) holds as well. Therefore, $\mathrm{nsc}(L_n) = n$ by Theorem 5.2, and the minimal NFA for L_n is unique, and is a DFA. Thus $\mathrm{rad}(M) = n$ for all minimal NFAs M accepting L_n.

However, consider the t-CNF NFA M accepting L_n with loop sizes $m_1, \ldots, m_k$. The radius of this NFA is $m_k = n/2$. Thus, $\mathrm{nrad}(L_n) \leq n/2$. This establishes the theorem. ■

As an example, for $k = 5$,

$$(n_1, n_2, n_3, n_4, n'_5) = (2, 3, 7, 41, 1805),$$

and $n = \mathrm{lcm}(2, 3, 7, 41, 1805) = 3108210$. Finally,

$$(m_1, m_2, m_3, m_4, m_5) = (1554105, 1036070, 444030, 75810, 1722).$$

The nondeterministic radius of $L_n = \cup_{i=1}^{5} (a^{m_i})^*$ is thus at most 1554105, which is achieved by a t-CNF NFA of size 3111738. However, $nsc(L_n) = 3108210$.

We can extend Theorem 5.3 as follows. Let p_i be the i-th prime with $p_1 = 2$. Let $\varpi : \mathbb{N}^+ \to \mathbb{N}^+$ be the function defined by

$$\varpi(n) = \min\{m \in \mathbb{N}^+ \ : \ \sum_{i=n}^{m} \frac{1}{p_i} > 1\}.$$

As the sum $\sum_{i \geq 1} 1/p_i$ diverges, $\varpi(n)$ exists for all $n \geq 1$; further $\varpi(n) > n$. The first few values of $\varpi(n)$ for $n \geq 1$ are 3, 10, 29, 69, 148, 258, 430, 658, 985, 1401 (A092325

in Sloane [15]). Let $n \geq 1$, and $P_n = \prod_{i=n}^{\varpi(n)} p_i$. For $n \geq 1$, let $\{m_n, \ldots, m_{\varpi(n)}\}$ be defined by $m_i = P_n/p_i$ for all $n \leq i \leq \varpi(n)$.

Then we can verify that we can apply Theorem 5.2 to $\{m_n, \ldots, m_{\varpi(n)}\}$ for all $n \geq 1$. Thus, if $L_n = \cup_{i=n}^{\varpi(n)} (a^{m_i})^*$, $\mathrm{nsc}(L_n) = P_n$. However, $\mathrm{nrad}(L_n) \leq P_n/p_n$.

Given a regular language L, let nrad–min(L) be the minimal radius of any *minimal* NFA accepting L. Then we have established the following result:

Theorem 5.4 *There exists a family of cyclic unary regular languages* $\{L_n\}_{n=1}^{\infty}$ *such that* $nsc(L_n) > n$ *and*

$$\lim_{n\to\infty} \frac{\mathrm{nrad}(L_n)}{\text{nrad–min}(L_n)} = 0.$$

We recall the following open problem raised by Ellul [9, p. 112]:

Open Problem 5.5 *Given a regular language* L*, is the quantity* nrad(L) *computable?*

We note that in the class of examples we have given for Theorem 5.3, the state complexity of the t-CNF recognizing L_n (with smaller radius) grows unboundedly in the size of the minimal NFA for L_n. Further, the example given for Theorem 5.4 shows that the size of nrad–min(L) grows unboundedly compared to nrad(L). Thus, we cannot hope to compute nrad(L) by even checking all NFAs within a constant size of the minimal NFA for L, or to approximate nrad(L) by calculating nrad–min(L).

6 Conclusion

We have given infinitely many cyclic unary languages whose minimal NFA is not unique. We have also noted that there exist infinitely many unary regular languages whose nondeterministic radius is not given by the radius of any minimal NFA.

References

[1] Barbeau, E. Expressing one as a sum of distinct reciprocals: Comments and bibliography. *Eureka 3* (1977), 178–181.

[2] Brenton, L., and Bruner, R. On recursive solutions of a unit fraction equation. *J. Austral. Math. Soc. (A) 57* (1994), 341–356.

[3] Brenton, L., and Drucker, D. Perfect graphs and complex surface singularities with perfect local fundamental group. *Tôhoku Math. J. 41* (1989), 507–525.

[4] Brenton, L., and Drucker, D. On the number of solutions of $\sum_{j=1}^{s}(1/x_j) + 1/(x_1 \cdots x_s) = 1$. *J. Num. Th. 44* (1993), 25–29.

[5] Brenton, L., and Hill, R. On the Diophantine equation $1 = \sum 1/n_i + 1/\prod n_i$ and a class of homologically trivial complex surface singularities. *Pacific J. Math. 133*, 1 (1988), 41–67.

[6] Burshtein, N. On distinct unit fractions whose sum equals 1. *Disc. Math. 5* (1973), 201–206.

[7] Butske, W., Jaje, L., and Mayernik, D. On the equation $\sum_{p|N} \frac{1}{p} + \frac{1}{N} = 1$, pseudoperfect numbers and perfectly weighted graphs. *Math. Comp. 69*, 229 (2000), 407–420.

[8] Chrobak, M. Finite automata and unary languages. *Theor. Comput. Sci. 47* (1986), 149–158.

[9] Ellul, K. Descriptional complexity measures of regular languages. M.Math thesis, University of Waterloo, 2002.

[10] Holzer, M., and Kutrib, M. Nondeterministic descriptional complexity of regular languages. *Intl. J. Found. Comput. Sci. 14* (2003), 1087–1102.

[11] Holzer, M., and Kutrib, M. Unary language operations and their nondeterministic state complexity. In *Developments in Language Theory: 6th International Conference* (2003), M. Ito and M. Toyama, Eds., vol. 2450 of *Lecture Notes in Computer Science*, pp. 162–172.

[12] Hopcroft, J. E., and Ullman, J. D. *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley, 1979.

[13] Janák, J., and Skula, L. On the integers x_i for which $x_i | x_1 \cdots x_{i-1} x_{i+1} \cdots x_n + 1$ holds. *Math. Slovaca 28*, 3 (1978), 305–310.

[14] Jiang, T., McDowell, E., and Ravikumar, B. The structure and complexity of minimal NFA's over a unary alphabet. *Intl. J. Found. Comp. Sci. 2*, 2 (1991), 163–182.

[15] Sloane, N. The on-line encyclopedia of integer sequences. *Published electronically at http://www.research.att.com/∼njas/sequences* (2004).

[16] Sylvester, J. On a point in the theory of vulgar fractions. *Amer. J. Math. 3* (1880), 322–335,388–389.

[17] Yu, S. State complexity of finite and infinite regular languages. *Bull. Eur. Assoc. Theor. Comput. Sci. 76* (2002), 142–152.