Small NFA's for Cofinite Unary Languages

William Gasarch
University of Maryland

Erik Metz
University of Maryland

Zan Xu
University of Maryland

Yuang Shen
University of Maryland

Sam Zbarsky
Princeton University

Abstract

For all n there is a DFA for $\{a^i : i \neq n\}$ of size $n+2$; however there is no smaller DFA. What about NFA's? We show that there is an NFA for $\{a^i : i \neq n\}$ of size $\sqrt{n} + \widetilde{O}(1)$. We also find small NFA's for many other cofinite unary sets. How small can we go? We show that any NFA for $\{a^i : i \neq n\}$ must have at least $\sqrt{n}$ states.

1 Introduction

Consider the language

$$\mathrm{MN}(n) = \{a^i : i \neq n\}.$$

(MN stands for *Missing Number*.)

It is easy to show that (1) there is a DFA for $\mathrm{MN}(n)$ with $n+2$ states, and (2) any DFA for $\mathrm{MN}(n)$ has at least $n+2$ states. What about an NFA for $\mathrm{MN}(n)$? We show that there is an NFA for $\mathrm{MN}(n)$ that has substantially fewer than n states. We also obtain small NFA's for many other cofinite unary languages.

Melvin Fitting, Dov Gabbay, Massoud Pourmahdian, Adrian Rezuş, Ali Sadegh Daghighi (eds.)
« Research Trends in Contemporary Logic » College Publications, London 2020

Notation 1.1. N is $\{0, 1, 2, \ldots\}$ (that is, we include 0).

Def 1.2. If $A \subseteq \mathsf{N}$ then

$$\text{MN}(A) = \{a^i : i \notin A\}.$$

We will only use this definition when A is finite. We will write $\text{MN}(a, b, c)$ instead of the formally correct $\text{MN}(\{a, b, c\})$.

Notation 1.3. If f and g are functions then, informally, $f \le \widetilde{O}(g)$ means that f is less than g if we ignore polylog factors. Formally it means that

$$(\exists n_0)(\exists c)(\forall n \ge n_0)[f(n) \le c(\log n)^c g(n)].$$

1. In Section 3 we show that (1) there is an NFA for MN(100) on 29 states, and (2) for all n there is an NFA for $\text{MN}(n)$ with $\le n^{1/2} + \widetilde{O}(1)$ states.
2. In Section 4 we show that (1) there is an NFA for MN(998, 999, 1000) on 104 states, (2) for any $A \subseteq \{998, 999, 1000\}$ there is an NFA for $\text{MN}(A)$ on 104 states, (3) for all n, for all $0 < \delta < 1$ there is an NFA for $\text{MN}(n - n^\delta, \ldots, n)$ on $5n^{\max\{1/2,\delta\}} + \widetilde{O}(1)$ states, and (4) for any $A \subseteq \{n - n^\delta, \ldots, n\}$ there is an NFA for $\text{MN}(A)$ on $5n^{\max\{1/2,\delta\}} + \widetilde{O}(1)$ states.
3. In Section 5 we show that, for all n, for all $0 < \alpha < 1$ such that $\alpha n \in \mathsf{N}$, there is an NFA for $\text{MN}(\alpha n, n)$ on $2n^{1/2}\ln(n) + \widetilde{O}(1)$ states.
4. In Section 6 we prove a general theorem about unary sets with big gaps. We obtain the following corollary: for all $0 < \delta < 1$ there is an NFA for $\text{MN}(n^\delta, n)$ on $n^{1/2} + n^\delta + \widetilde{O}(1)$ states.
5. In Section 7 we show that any NFA for $\text{MN}(n)$ requires at least $n^{1/2}$ states.
6. In Section 8 we discuss our empirical results.
7. In Section 9 we state open problems.

Def 1.4. A set X has a *small NFA* if there is an NFA that accepts it that is much smaller than any DFA for it. We do not define the term *much smaller* rigorously. However, all of our results are about small NFA's.

All of our general results are asymptotic; however, we will present empirical evidence that indicates the results hold for small n as well.

2 Needed Lemma

The following problem is attributed to Frobenius:

Given a set of relatively prime positive integers $\{a_1, \ldots, a_m\}$ *find the set* $\{\sum_{i=1}^{n} a_i x_i : x_1, \ldots, x_m \in \mathsf{N}\}$.

It is known that this set is always cofinite. The $m = 2$ case was solved by James Joseph Sylvester in 1884:

Lemma 2.1. *Let* $c, d \in \mathsf{N}$ *be relatively prime.*

1. *For all* $i \geq cd - c - d + 1$ *there exists* $x, y \in \mathsf{N}$ *such that* $i = cx + dy$.
2. *There is no* $x, y \in \mathsf{N}$ *such that* $cd - c - d = cx + dy$.
3. *There is no* $x, y, C, D \in \mathsf{N}$ *such that* $cd - c - d - Cc - Dd = cx + dy$. *(If there was then* $cd - c - d = (C + x)c + (D + y)d$.*) We use this part in Section 4.*

3 Small NFA's for MN(100) and MN(n)

3.1 Small NFA for MN(100)

Theorem 3.1.

1. *For all* $i \geq 96$ *there exists* $x, y \in \mathsf{N}$ *such that* $i = 13x + 9y$.
2. *There does not exist* $x, y \in \mathsf{N}$ *such that* $95 = 13x + 9y$.
3. *For all* $i \geq 101$ *there exists* $x, y \in \mathsf{N}$ *such that* $i = 13x + 9y + 5$.
4. *There does not exist* $x, y \in \mathsf{N}$ *such that* $100 = 13x + 9y + 5$.
5. *There exists an NFA* M *such that the following are true:*
   (a) *For all* $i \geq 101$, M *accepts* a^i.
   (b) M *rejects* a^{100}.
   (c) *We have no comment on the behavior of* M *on other* a^i.
   (d) M *has 13 states.*
6. *There exists an NFA on 29 states that accepts* MN(100).

Proof. 1,2) These follow from Lemma 2.1, though they can be proven directly by an easy induction.
3,4) These follow from Parts 1 and 2
5) The NFA is constructed as follows: (also see Figure 1, the caption will be explained later).

- M has states $0, \ldots, 12$, 0 is the start state, and 5 is the only final state. For $0 \le j \le 12$, $\delta(j, a) = j + 1 \pmod{13}$. ($\delta$ is not fully defined yet.)
- If we go no further then M accepts $\{a^{13x+5} : x \in \mathsf{N}\}$.
- We put in an e-transition from state 5 to state 9. Now M accepts

$$\{a^{13x+9y+5} : x, y \in \mathsf{N}\}.$$

(The $9y$ is not because the e-transition went to state 9. It is because the distance from state 9 back to state 5 is 9.)

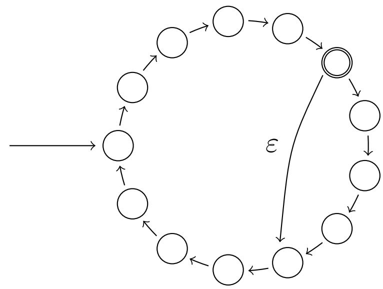

Figure 1: LOOP(9,13,5) Case 1

By Parts 3,4 M satisfies 5a and 5b. M clearly has 13 states, so it satisfies 5d.

6) Let $Q = \{3, 5, 7\}$. Note that $3 \times 5 \times 7 = 105 > 100$. For each $p \in Q$ let M_p be the DFA that accepts $\{a^i : i \not\equiv 100 \pmod{p}\}$.

The NFA is constructed as follows: (see also Figure 2)

1. The NFA M is part of our new NFA. We create a new start state, and then put an e-transition from this new state to M's original start state. Note that M (a) accepts all a^i with $i \ge 101$ (it also accepts other strings), (b) rejects a^{100}, and (c) has 13 states.
2. For each $p \in Q$ put an e-transition from our new start state to the start state of M_p. Note that M_p (a) accepts all a^i with $i \not\equiv 100 \pmod{p}$, (b) rejects a^{100}, and (c) has p states.

Clearly the NFA has $13 + 3 + 5 + 7 + 1 = 29$ states and rejects a^{100}. We show that it accepts everything else.

Let a^i be rejected by this NFA.

- Since the M part rejects a^i, $i \leq 100$ (note, hence $i \leq 3 \times 5 \times 7 = 105$).
- Since the M_3 part rejects a^i, $i \equiv 100 \pmod 3$
- Since the M_5 part rejects a^i, $i \equiv 100 \pmod 5$
- Since the M_7 part rejects a^i, $i \equiv 100 \pmod 7$

By the Chinese Remainder Theorem there is a unique number $0 \leq z \leq 3 \times 5 \times 7 = 105$ such that, for every $p \in \{3, 5, 7\}$, $z \equiv 100 \pmod p$. Since both i and 100 satisfies these criteria, $i = n$. □

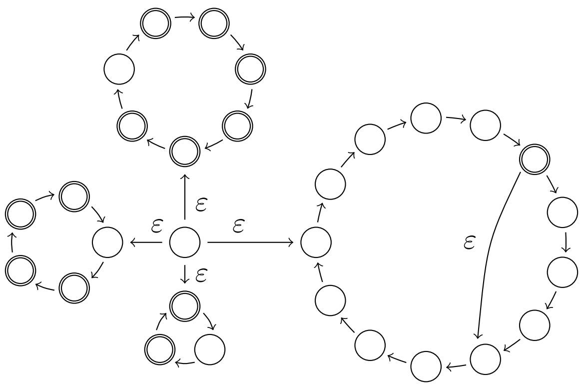

Figure 2: NFA for MN(100)

3.2 Small NFA for MN(n)

We generalize the construction of a small NFA for MN(100) to get a small NFA for MN(n).

Def 3.2. Let $c, d, e \in \mathsf{N}$ be such that $c < d$ and c, d are relatively prime. LOOP(c, d, e) is the NFA defined as follows. There are two cases.
Case 1: $e \leq d - 1$.

1. The NFA has states $0, \ldots, d - 1$, with 0 as the start state and e as the only final state. For $0 \leq j \leq d - 1$, $\delta(j, a) = j + 1 \pmod d$.
2. So far this NFA accepts $\{a^{dx+e} : x \in \mathsf{N}\}$.

3. We put in an e-transition from state e to state $e - c \pmod{d}$. Note that the distance from state $e - c \pmod{d}$ to state e is c. Now the NFA accepts

$$\{a^{cx+dy+e} : x, y \in \mathsf{N}\}.$$

4. This NFA has d states.

Note that Figure 1 is LOOP$(9, 13, 5)$ which is an example of a Case 1 LOOP.

Case 2: $e \geq d$

1. The NFA has states $s_0, s_1, \ldots, s_{e-d+2}$ such that s_0 is the start state. For $0 \leq j \leq e - d + 1$, $\delta(s_j, a) = s_{j+1}$.
2. The NFA has states $0, \ldots, d - 1$, with $d - 1$ as the only final state. For $0 \leq j \leq d - 1$, $\delta(j, a) = j + 1 \pmod{d}$. The state 0 is identical to the state s_{e-d+2}.
3. In total there are $(e - d + 1) + (d - 1) = e$ transitions to get to the final state the first time, after which each loop of length d brings you back to the same state, so the NFA accepts $\{a^{dx+e} : x \in \mathsf{N}\}$.
4. We put in an e-transition from state $d - 1$ to state $d - c - 1$. Note that the distance from state $d - c - 1$ to state $d - 1$ is c. Now the NFA accepts

$$\{a^{cx+dy+e} : x, y \in \mathsf{N}\}.$$

5. This NFA has $e + 1$ states.

Figure 3 is LOOP$(9, 13, 17)$ which is an example of a Case 2 LOOP.

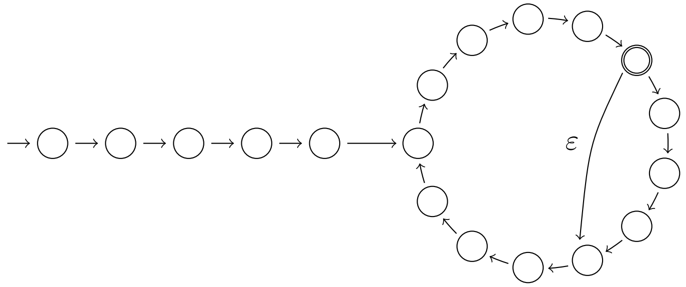

Figure 3: NFA LOOP(9,13,17) Case 2

The following is clear:

Lemma 3.3. *Let $c, d, e \in \mathsf{N}$ be such that $e, c < d$ and c, d are relatively prime.*

1. $\mathrm{LOOP}(c, d, e)$ *accepts* $\{a^i : i \geq cd - c - d + e + 1\}$
2. $\mathrm{LOOP}(c, d, e)$ *rejects* $\{a^{cd-c-d+e}\}$.
3. $\mathrm{LOOP}(c, d, e)$ *rejects* $\{a^{cd-Cc-Dd+e} : C, D \in \mathsf{N}\}$ *(since if $a^{cd-Cc-Dd+e}$ can reach the accept state then by adding C c's and D d's the NFA gets back to the accept state). We will use this part in Section 4.*
4. *If we used Case 1 then* $\mathrm{LOOP}(c, d, e)$ *has d states.*
5. *If we used Case 2 then* $\mathrm{LOOP}(c, d, e)$ *has $e + 1$ states.*

Note 3.4. Below, we use o(1) to denote a number that may be positive or negative, but that goes to 0 as our variable of interest (N in Lemma 3.5), goes to infinity. This may be non-standard.

Lemma 3.5. *Let $N \in \mathsf{N}$. Let Q_N be the set of the first N primes.*

1. $\prod_{p \in Q_N} p \sim e^{((1+o(1))N \log N}$. *(This is well known.)*
2. $\sum_{p \in Q_N} p \sim O(N^2 \log N) = \widetilde{O}(N^2)$. *(For references and more precise estimates see Axler [1].)*
3. *Let $n \in \mathsf{N}$. The product of the first $\Omega(\log n)$ primes is $\geq n$. The sum of the first $O(\log n)$ primes is $\leq O((\log n)^2 \log \log n) \leq \widetilde{O}(1)$. (This follows from parts 1 and 2.)*
4. $\prod_{p \leq N,\ p \text{ prime}} p \sim e^{(1+o(1))N}$. *(This is well known.)*

Theorem 3.6. *Let $n \in \mathsf{N}$.*

1. *There exists an NFA M such that the following are true:*
   (a) *For all $i \geq n + 2\left\lceil n^{1/2} \right\rceil$, M accepts a^i.*
   (b) *M rejects a^n.*
   (c) *We have no comment on the behavior of M on other a^i.*
   (d) *M has $\leq n^{1/2} + O(1)$ states.*
2. *There exists an NFA on $\leq n^{1/2} + \widetilde{O}(1)$ states that accepts* $\mathrm{MN}(n)$.

Proof. 1) Let $c = \lceil n^{1/2} \rceil + 1$ and $e = n + 1 \pmod{c}$. Note that $e \le c$. Let M be LOOP$(c, c+1, e)$. Note that

$$c(c+1) - c - c - 1 + e = c^2 - c - 1 + e \le c^2 - c - 1 + c = c^2 - 1$$

By Lemma 3.3 M accepts a^i where $i \ge c^2 - 1 + 1 = c^2 \ge n + 2\lceil n^{1/2} \rceil$.

We show that M rejects a^n. Assume, by way of contradiction, that M accepts a^n. Then there exists $x, y \ge 0$ such that

$$cx + (c+1)y + e = n$$

Take this equation mod c. Then

$$0x + 1 \times y + (n+1) \equiv n \pmod{c}$$

$$y + 1 \equiv 0 \pmod{c}$$

$$y \equiv -1 \pmod{c}.$$

Since $y \ge 0$, $y \ge c - 1$. Hence

$$n = cx + (c+1)y + e \ge (c+1)(c-1) = c^2 - 1 = (\lceil n^{1/2} \rceil + 1)^2 - 1 = n + 2n^{1/2} - 1.$$

This is a contradiction.

Since $e \le c$, M has $c + 1 = n^{1/2} + O(1)$ states.

2) By Lemma 3.5.3 there is a set of primes Q such that

- $\prod_{p \in Q} p \ge n + 2\lceil n^{1/2} \rceil$.
- $\sum_{p \in Q} p \le \widetilde{O}(1)$.

For each $p \in Q$ let M_p be the DFA that accepts $\{a^i : i \not\equiv n \pmod{p}\}$.

The NFA is constructed as follows:

1. The NFA M is part of our NFA. We create a new start state, and then put an e-transition from this new state to M's original start state. Note that (1) M accepts a^i for $i \ge n + 2\lceil n^{1/2} \rceil$ (it also accepts other strings), (2) M rejects a^n, and (3) M has $\le n^{1/2} + O(1)$ states.

2. For each $p \in Q$ there is an e-transition from our new start state to the start state of M_p. Note that (1) M_p accepts a^i if $i \not\equiv n \pmod{p}$, (2) M_p rejects a^n, and (3) M_p has p states.

Clearly the NFA has $\le n^{1/2} + \sum_{p\in Q} p \le n^{1/2} + \widetilde{O}(1)$ states and rejects a^n. We show that it accepts everything else.

Let a^i be rejected by this NFA.

- Since the M part rejects a^i, $i \le n + 2\left\lceil n^{1/2} \right\rceil$ (note, hence $i \le \prod_{p\in Q} p$).
- For each $p \in Q$, since the M_p part rejects a^i, $i \equiv n \pmod{p}$.

By the Chinese Remainder Theorem there is a unique number $0 \le z < \prod_{p\in Q} p \ge n + 2\left\lceil n^{1/2} \right\rceil$ such that, for every $p \in Q$, $z \equiv n \pmod{p}$. Since both i and n satisfy the criteria, $i = n$. □

4 Small NFA's for $\mathrm{MN}(998, 999, 1000)$ and $\mathrm{MN}(A)$

4.1 Small NFA for $\mathrm{MN}(998, 999, 1000)$ and $\mathrm{MN}(998, 1000)$

Theorem 4.1.

1. *There exists an NFA M such that the following are true:*
   (a) *For all $i \ge 1067$, M accepts a^i.*
   (b) *For all $i \in \{998, 999, 1000\}$ M rejects a^i.*
   (c) *We have no comment on the behavior of M for other a^i's.*
   (d) *M has 34 states.*
2. *There exists an NFA with 104 states that accepts* $\mathrm{MN}(998, 999, 1000)$.
3. *There exists an NFA with 104 states that accepts* $\mathrm{MN}(A)$ *where* $A \subseteq \{998, 999, 1000\}$. *(For $A = \emptyset$ this is trivial.)*

Proof. 1) Let M be $\mathrm{LOOP}(33, 34, 11)$. From Lemma 3.3 we know the following:

a) For all $i \ge 1067$, M accepts a^i.

b) For all $C, D \in \mathsf{N}$, M rejects $a^{1066-33C-34D}$ which we write as $a^{1066-33(C+D)-D}$.

We set (C, D) carefully to obtain, using item b, strings that M rejects.

- If $(C, D) = (2, 0)$ then we get $1066 - 33 \times 2 - 0 = 1000$
- If $(C, D) = (1, 1)$ then we get $1066 - 33 \times 2 - 1 = 999$
- If $(C, D) = (0, 2)$ then we get $1066 - 33 \times 2 - 2 = 998$

Clearly M has 34 states.

2) We will once again use primes and mods. We can't use mod 2 or mod 3 since then one of a^{998}, a^{999}, a^{1000} will be accepted.

We can use any mod from 5 up. We need another trick, as you will see.

Let $Q = \{5, 7, 11\}$. Note that $3 \times 5 \times 7 \times 11 = 1155 > 1066$ (That is not a typo. We really do mean to multiply by 3. We chose 3 because $|\{998, 999, 1000\}| = 3$. We chose $\{5, 7, 11\}$ since none of them divide 3 and the product $3 \times 5 \times 7 \times 11 > 1066$. The fact that 3 is a prime is not important.)

For each $p \in Q$ let M_{3p} be the DFA that accepts

$$\{a^i : i \not\equiv 998, 999, 1000 \pmod{3p}\}$$

Note that $\text{LCM}(3 \times 5, 3 \times 7, 3 \times 11) = 3 \times 5 \times 7 \times 11 = 105 > 100$.

The NFA is constructed as follows:

1. The NFA M is part of our new NFA. We create a new start state, and then put an e-transition from this new state to M's original start state. Note that M (1) accepts all a^i with $i \geq 1067$ (it also accepts other strings), (2) rejects any of a^i with $i \in \{998, 999, 1000\}$, and (3) has 34 states.

2. For each $p \in Q$, there is an e-transition from our new start state to the start state of M_{3p}. Note that M_{3p} (1) accepts a^i when $i \not\equiv 998, 999, 1000 \pmod{3p}$, (2) rejects any a^i with $i \in \{998, 999, 1000\}$, and (3) has $3p$ states.

This NFA has $34 + 3(5 + 7 + 11) + 1 = 104$ states and rejects any a^i with $i \notin \{998, 999, 1000\}$. We show that it accepts everything else.

Let a^i be rejected by this NFA.

- Since the M part rejects a^i, $i \leq 1066$ (note, hence $i \leq 3 \times 5 \times 7 \times 11 = 1155$).
- For all $p \in Q$, since the M_{3p} part rejects a^i, there exists $x \in \{998, 999, 1000\}$ such that $i \equiv x \pmod{3p}$.

We cannot use the Chinese Remainder Theorem (yet) since it is possible that, say $i \equiv 998 \pmod{3 \times 7}$ but $i \equiv 1000 \pmod{3 \times 11}$. We need that i is equivalent to the same x with all of those mods.

Let $i \equiv x \pmod 3$ where $x \in \{998, 999, 1000\}$. Note that x is unique. Let $p \in Q$. Let $y \in \{998, 999, 1000\}$ be such that $i \equiv y \pmod{3p}$. Note that y is unique since $3 < 3p$.

We show that $x = y$.

Since $i \equiv x \pmod 3$ there exists $a \in \mathsf{Z}$ such that

$$\text{Eq 1 } i = x + 3a.$$

Since $i \equiv y \pmod{3p}$ there a $b \in \mathsf{Z}$ such that

$$\text{Eq 2 } i = y + 3pb.$$

By subtracting Eq 2 from Eq 1 we get

$$x - y = 3pb - 3a \equiv 0 \pmod 3$$

Since $x, y \in \{998, 999, 1000\}$ and $x \equiv y \pmod 3$, $x = y$. To recap we now have that there exists $x \in \{998, 999, 1000\}$ such that, for all $p \in Q$, $i \equiv x \pmod{3p}$.

By the Chinese Remainder Theorem there is a unique number $0 \le z \le \mathrm{LCM}(3 \times 5, 3 \times 7, 3 \times 11) = 1155$ such that, for all $p \in Q$, $z \equiv x \pmod{3p}$. Since both i and x satisfy those criteria, $i = x$.

3) We look at $\mathrm{MN}(998, 1000)$ as an example. The construction is similar to the one for $\mathrm{MN}(998, 1000)$ except that, at the end, use the DFA for $\{a^i : i \not\equiv 998, 1000 \pmod{3p}\}$ instead of $\{a^i : i \not\equiv 998, 999, 1000 \pmod{3p}\}$. The other cases are similar. □

Theorem 4.2. *Let $0 < \delta < 1$. Let $n \in \mathsf{N}$. (We will assume $n^\delta \in \mathsf{N}$ and leave it to the reader to adjust the statement and the proof for when $n^\delta \notin \mathsf{N}$.) Assume $n = c^2 + f$ where $0 \le f \le 2c$.*

1. *There exists an NFA M such that the following are true:*
   - (a) *For all $i \ge n + n^{1/2+\delta} + n^{2\delta} + 1$, M accepts a^i.*
   - (b) *For all $i \in \{n - n^\delta, n - n^\delta + 1, \ldots, n\}$, M rejects a^i.*
   - (c) *We have no comment on the behavior of M for other a^i's.*
   - (d) *M has $\le 5n^{\max\{1/2,\delta\}} + O(1)$ states.*
2. *There exists an NFA on*

$$\le 5n^{\max\{1/2,\delta\}} + \widetilde{O}(1) \text{ states}$$

that accepts $\mathrm{MN}(n - n^\delta, n - \delta + 1, \ldots, n)$.

3. *Let* $A \subseteq \{n - n^\delta, \ldots, n\}$. *There exists an NFA on*

$$5n^{\max\{1/2,\delta\}} + \widetilde{O}(1) \text{ states}$$

that accepts MN(A). *(For* $A = \emptyset$ *this is trivial.)*

Proof. 1) Let M be the NFA LOOP$(c+k, c+k+1, f+1+x(k))$ where we determine k and $x(k)$ later.
Claim:

1. If M rejects $a^{n+n^\delta(c+k)}$ then, for $i = n - n^\delta, \ldots, n$, M rejects a^i.
2. If $x(k) = n^\delta(c+k) - k^2 - 2ck + c + k$ then M rejects $a^{n+n^\delta(c+k)}$.
3. If $x(k) = n^\delta(c+k) - k^2 - 2ck + c + k$ then, for $i = n - n^\delta, \ldots, n$, M rejects a^i (this follows from parts 1 and 2).

Proof of Claim:
1) Assume M rejects $n + n^\delta(c+k)$. Then it also rejects everything of the form

$$n + n^\delta(c+k) - (c+k)C - (c+k+1)D = n + n^\delta(c+k) - (C+D)(c+k) - D$$

(since otherwise M would accept $n + n^\delta(c+k) - (c+k)C - (c+k+1)D + (c+k)C + (c+k+1)D = n + n^\delta(c+k)$).

We set (C, D) as follows:

- If $(C, D) = (n^\delta, 0)$ then we get $n + n^\delta(c+k) - n^\delta(c+k) - 0 = n$.
- If $(C, D) = (n^\delta - 1, 1)$ then we get $n + n^\delta(c+k) - n^\delta(c+k) - 1 = n - 1$.
- $\vdots$
- If $(C, D) = (0, n^\delta)$ then we get $n + n^\delta(c+k) - n^\delta(c+k) - n^\delta = n - n^\delta$.

2) For $k \in \mathsf{N}$ we need $x(k) \in \mathsf{N}$ such that LOOP$(c+k, c+k+1, f+1+x(k))$ rejects $n + n^\delta(c+k)$. Note that this NFA rejects

$$(c+k)(c+k+1) - (c+k) - (c+k+1) + f + 1 + x(k)$$

$$= c^2 + k^2 + 2ck + c + k - 2c - 2k - 1 + f + 1 + x(k)$$

$$= n + k^2 + 2ck - c - k + x(k)$$

Hence we find $x(k)$ via:

$$n + k^2 + 2ck - c - k + x(k) = n + n^\delta(c + k)$$

or equivalently

$$x(k) = n^\delta(c + k) - k^2 - 2ck + c + k$$

End of Proof of Claim

We choose k such that the max of $\{c + k + 1, f + 1 + x(k)\}$ is small. We look at what happens to $x(k)$ for $k \in \{0, \ldots, n^\delta\}$. We consider only when $n \geq 9$, with smaller n being expressed within the $O(1)$ term. Note that

- $x(0) = n^\delta c + c > 0$.
- $x(n^\delta) = n^{2\delta} + n^\delta c - n^{2\delta} - 2cn^\delta + c + n^\delta = c + n^\delta - cn^\delta < 0$ (since $n \geq 9$).
- there exists k_o such that $x(k_o) \geq 0$ and $x(k_o + 1) \leq 0$ (this follows from the first two points).

Note that

$$x(k_o) \leq x(k_o) - x(k_o + 1) \leq |-2c + n^\delta - 2k_o| \leq 2c + n^\delta.$$

Let $M = \mathrm{LOOP}(c+k_o, c+k_o, f+1+x(k_o))$. Since $c \leq n^{1/2}$, $f \leq 2n^{1/2}$, $k_o \leq n^\delta$, and $x(k_o) \leq 2c + n^\delta \leq 3n^{\max\{1/2,\delta\}}$. $e = x(k_o) + f + 1$ has $\leq 5n^{\max\{1/2,\delta\}} + O(1)$ states, while $c + k_o + 1$ has $\leq 3n^{\max\{1/2,\delta\}} + O(1)$ states, so M must have $\leq 5n^{\max\{1/2,\delta\}} + O(1)$ states overall.

M satisfies conditions of what to reject and how many states it has. We now consider what it accepts. Note that $x(k)$ was chosen so that the largest number M (with $k = k_o$) rejects is $a^{n+n^\delta(c+k_o)}$. We need to estimate this.

$$n + n^\delta(c + k_o) \leq n + n^\delta(n^{1/2} + n^\delta) \leq n + n^{1/2+\delta} + n^{2\delta}.$$

By Lemma 3.3 M accepts what it should.

2) To simplify the algebra we just use that the NFA in Part 1 accepts $\{a^i : i \geq n^2\}$.

By Lemma 3.5 there is a set of primes Q' such that (1) $\prod_{p\in Q'} p \geq n^2$, (2) $\sum_{p\in Q'} p \leq \widetilde{O}(1)$. We form Q as follows: (1) remove from Q' all of the primes that divide n^δ, (2) add in the smallest primes possible that do not divide n^δ so that $n^\delta \prod_{p\in Q} p \geq n^2$.

One can show that $\sum_{p\in Q} p \leq O(\sum_{p\in Q'} p) \leq \widetilde{O}(1)$. Hence we have a set Q such that (1) $n^\delta \prod_{p\in Q} p \geq n^2$ (2) $\sum_{p\in Q} p \leq \widetilde{O}(1)$, and (3) for all $p \in Q$, p does not divide n^δ. For each $p \in Q$ let $M_{n^\delta p}$ be the DFA that accepts $\{a^i : i \not\equiv n \pmod{n^\delta p}\}$.

Note that the LCM$\{n^\delta p : p \in Q\} = n^\delta \prod_{p \in Q} p \geq n$.

The NFA is constructed as follows:

1. The NFA M is part of our new NFA . We create a new start state, and then put an e-transition from this new state to M's original start state. Note that M (1) accepts all a^i with $i \geq n^2$ (it also accepts other strings), (2) rejects any of a^i with $i \in \{n - n^\delta, \ldots, n\}$, and (3) has $\leq 5n^{\max\{1/2,\delta\}} + O(1)$ states.

2. For each $p \in Q$ put an e-transition from our new start state to the start state of $M_{n^\delta p}$. Note that $M_{n^\delta p}$ (1) accepts a^i with $i \not\equiv n - n^\delta, \ldots, n \pmod{n^\delta p}$, (2) rejects any of a^i with $i \in \{n - n^\delta, \ldots, n\}$, and (3) has $n^\delta p$ states.

This NFA has $5n^{\max\{1/2,\delta\}} + \widetilde{O}(1)$ states and rejects any a^i with $i \in \{n - n^\delta, \ldots, n\}$. We show that it accepts everything else.

Let a^i be rejected by this NFA.

- Since the M part rejects a^i, $i \leq n^2$ (note, hence $i \leq n^\delta \prod_{p \in Q} p$).

- For each $p \in Q$, since the $M_{n^\delta p}$ part rejects a^i, there exists $x \in \{n - n^\delta, \ldots, n\}$ such that $i \equiv x \pmod{n^\delta p}$.

We cannot use the Chinese Remainder Theorem (yet) since it is possible that, say $i \equiv 95 \pmod{n^\delta \times 7}$ but $i \equiv 92 \pmod{n^\delta \times 11}$. We need that a^i is equivalent to the same $n^\delta p$ with all those mods.

Let $i \equiv x \pmod{n^\delta}$ where $x \in \{n - n^\delta, \ldots, n\}$. Note that x is unique. Let $n^\delta p \in n^\delta Q$. Let $y \in \{n - n^\delta, \ldots, n\}$ be such that $i \equiv y \pmod{n)^\delta p}$. Note that y is unique since $n^\delta < n^\delta p$. We show that $x = y$.

Since $i \equiv x \pmod{n^\delta}$ there exists $a \in \mathsf{Z}$ such that:

$$\text{Eq 1 } i = x + n^\delta a.$$

Since $i \equiv y \pmod{n^\delta}$ there exists a $b \in \mathsf{Z}$ such that

$$\text{Eq 2 } i = y + n^\delta pb.$$

By subtracting Eq 2 from Eq 1 we get

$$x - y = n^\delta pb - n^\delta a \equiv 0 \pmod{n^\delta}$$

Since $x, y \in \{n-\delta, \ldots, n\}$ and $x \equiv y \pmod{n^\delta}$, $x = y$. To recap we now have that there exists $x \in \{n-\delta, \ldots, n\}$ such that, for all $n^\delta p \in Q$, $i \equiv x \pmod{n^\delta p}$.

By the Chinese Remainder Theorem there is a unique number $0 \le z \le \mathrm{LCM}\{n^\delta p : p \in Q\} \ge n$ such that, for all $p \in Q$, $z \equiv x \pmod{n^\delta p}$. Since both i and x satisfy those criteria, $i = x$.

3) This is an easy modification of Part 2 which we leave to the reader. □

5 Small NFA's for $\mathrm{MN}(\alpha n, n)$

Lemma 5.1. *Let $x, x', y, y', c \in \mathsf{N}$ with $c \ge 1$ be such that the following hold.*

1. *$c(x - x') + (c+1)(y - y') = 0$.*
2. *$|x - x'| \le c$.*

Then $x = x'$ and $y = y'$.

Proof. Since $c+1$ divides $c(x-x')$ and $c+1$ is rel prime to c we have that $c+1$ divides $x - x'$. Since $|x - x'| \le c$, $x = x'$. Hence $y = y'$. □

Theorem 5.2. *Let $n \in \mathsf{N}$ and $0 < \alpha < 1$ be such that $\alpha n \in \mathsf{N}$.*

1. *There exists an NFA M such that the following are true:*
   (a) *For all $i \ge 2n \ln n$, M accepts a^i.*
   (b) *For all $i \in \{\alpha n, n\}$, M rejects a^i.*
   (c) *We have no comment on the behavior of M for any other a^i's.*
   (d) *M has $\le 2n^{1/2} \ln n + \widetilde{O}(1)$ states.*
2. *There exists an NFA on $\le 2n^{1/2} \ln n + \widetilde{O}(1)$ states that accepts $\mathrm{MN}(\alpha n, n)$.*

Proof. Let $c = \lceil n^{1/2} \rceil + 1$ and $e = n + 1 \pmod{c}$. Note that $e \le c$.
1) Let M' be $\mathrm{LOOP}(c, c+1, e)$. By the proof of Theorem 3.6 we have:

1. For all $i \ge n + 2\lceil n^{1/2} \rceil$, M' accepts a^i. Note that for all $i \ge 2n \ln n$, M' accepts a^i.
2. M' rejects a^n.
3. We have no comment on the behavior of M' on other a^i.

4. M' has $\leq n^{1/2} + O(1)$ states.

Case 1: M' rejects $a^{\alpha n}$. Then take M to be M'.
Case 2: M' accepts $a^{\alpha n}$. We use the very acceptance of $a^{\alpha n}$ to find an NFA M that satisfies the theorem.
Claim 1: There exists a unique x, y, such that $cx + (c+1)y + e = \alpha n$. Both x, y are $\leq c - 1 \leq n^{1/2}$.
Proof of Claim: Since M' accepts $a^{\alpha n}$ there is at least one such x, y such that:

$$\text{Eq 1 } cx + (c+1)y + e = \alpha n$$

$x \leq c - 1$ since otherwise Eq 1 implies:

$$\alpha n = cx + (c+1)y + e \geq c^2 + (c+1)y + e \geq c^2 = n.$$

$y \leq c - 1$ by a similar argument.
Assume that x', y' also works.

$$\text{Eq 2 } cx' + (c+1)y' + e = \alpha n$$

By the same reasoning that $x \leq c - 1$, we have $x' \leq c - 1$, so $|x - x'| \leq c - 1$.
Subtract the second equation from the first to obtain:

$$c(x - x') + (c+1)(y - y') = 0$$

By Lemma 5.1, $x = x'$ and $y = y'$.
End of Proof of Claim 1
Let p be the least prime that does not divide yc (hence does not divide y or c). Since $y \leq c^{1/2}$ and $c = \lceil n^{1/2} \rceil + 1$, $yc \leq n + n^{1/2} + O(1)$. By Lemma 3.5.3, $p \leq (1+o(1)) \ln(n+n^{1/2}+O(1)) \leq 2\ln(n)+O(1)$. Let M be LOOP$(c, p(c+1), e)$. Note that M has $\leq 2n^{1/2} \ln n + \widetilde{O}(1)$ states. We need to show that (1) for all $i \geq n \ln n$, M accepts a^i, and (2) M rejects $a^{\alpha n}$ and a^n. Note that

$$cp(c+1) - c - cp - p + e = c^2p + cp - c - cp - p + e = c^2p - c - p + e \leq 2n \ln n - 1$$

By Lemma 3.3

- For all $i \geq 2n \ln n$, M accepts a^i.
- For all C, D, M rejects a^i where $i = c^2p - c - p + e - Cc - D(p(c+1))$. (We will not be using this.)

Claim 2: M rejects $a^{\alpha n}$.
Proof of Claim 2:

Assume, by way of contradiction, that M accepts αn. Then there exists x', y' such that

$$\text{Eq 1 } \alpha n = cx' + (c+1)y' + e$$

Recall that from Claim 1 there exists unique x, y such that

$$\text{Eq 2 } \alpha n = cx + (c+1)y + e.$$

Hence $x = x'$ and $y = y'p$. This contradicts that p does not divide y.
End of Proof of Claim 2
Claim 3: M rejects a^n.
Proof of Claim 3:

Assume, by way of contradiction, that M accepts n. Then there exists x', y' such that

$$n = cx' + p(c+1)y' + e$$

Then a^n is accepted by LOOP$(c, c+1, e)$, which is a contradiction.
End of Proof of Claim 3

2) This proof is similar to that of Theorem 4.2.2.

□

6 Small NFA's for MN(A) where A has large gaps

Theorem 6.1. *Let $A \subseteq \mathsf{N}$ with maximum element n'. Let $n' < n$. Then there is an NFA for* MN$(A \cup \{n\})$ *of size $n^{1/2} + n' + \widetilde{O}(1)$.*

Proof. By Theorem 3.6 there exists an NFA M' for MN$(n-n')$ of size $(n-n')^{1/2} + \widetilde{O}(1) \le n^{1/2} + \widetilde{O}(1)$. We form M as follows:

1. Add states $0, 1, \ldots, n'$ where state n' is the start state of M'.

2. 0 is the start state of M.

3. For $0 \le i \le n' - 1$ we have transitions $\delta(i, a) = i + 1$.

4. For all $0 \le i \le n'$, make i an accept state iff $i \in A$.

Clearly M'' accepts $A \cup \{n\}$ and has $n^{1/2} + n' + \widetilde{O}(1)$ states. □

Corollary 6.2.

1. *For all n, for all $0 < \delta < 1$, there is an NFA for $MN(n^{\delta}, n)$ of size $n^{1/2} + n^{\delta} + \widetilde{O}(1)$.*

2. *For all n, for all $\beta \in \mathsf{R}^{+}$, there is an NFA for $MN((\log n)^{\beta}, n)$ of size $n^{1/2} + \widetilde{O}(1)$.*

7 Every NFA for $MN(n)$ has $\geq n^{1/2}$ States

This section is due to Jeff Shallit who shared it with us.

Chroback [2] proved the following.

Theorem 7.1. *Let L be a cofinite unary regular language. If there is an NFA for L with n states then there is an NFA for L of the following form:*

- *There is a sequence of $\leq n^2$ states from the start state to a state we will call X. Note that there is no nondeterminism involved yet.*
- *From X there are e-transitions to $X_1, \ldots, X_m$. (This is nondeterministic.)*
- *Each X_i is part of a cycle C_i. All of the C_i are disjoint.*

Theorem 7.2.

1. *Let L be a cofinite unary language where the shortest string that is not in L is of length n. Then any NFA for L requires $n^{1/2}$ states*

2. *Any NFA for $MN(n)$ has $n^{1/2}$ states (this follows from part 1).*

Proof. Assume there was an NFA with $< n^{1/2}$ states for L. Then by Theorem 7.1 there would be an NFA for L with a path from the start state to a state X of length $< n$ and then from X a branch to many cycles. Let X_i and cycle's C_i as described in Theorem 7.1.

Run a^n through the NFA and try out all paths. For each i there will be a point in C_i that you end up at. Let n_i be the length of C_i. For every i there is a state on C_i that rejects. Hence the strings $a^{n+Kn_1n_2\cdots n_m}$ are all rejected. This is an infinite number of strings. This is a contradiction. $\square$

8 Empirical Results

We have written a program that, given n, tries to find the smallest NFA for MN(n). We first set $c = \lceil\sqrt{n}\rceil$, $d = c + 1$, and e such that LOOP(c, d, e) (1) rejects a^n and (2) for all $i \geq n + 1$, accepts a^i. We then looked at sets of prime powers (these work as well as primes) so that the usual M_{p^b} machines will accept all a^i such that $i \leq n - 1$. We took the smallest NFA among all of these choices. We ran this program for $1 \leq n \leq 10^{27}$. Here is what we discovered:

1. The smallest NFA for MN(n) was around $n^{1/2} + (\ln n)^g$ where g had the following values:

   (a) For $1 \leq n \leq 10^6$ g decreases from 2 to around 1.55.

   (b) For $10^6 \leq n \leq 10^{27}$ g fluctuates around 1.55 but slowly increases.

The log-term is actually the inverse of Landau's function. As such, it is known that g is bounded by 2.

We also wrote a second program that tries to find smaller NFAs than the first program. How? Note that, in the first program, we found a set of M_{p^b} machines to accept all a^i such that $i \leq n-1$. However, LOOP(c, d, e) already accepts some of those strings. Our second program finds a set of M_{p^b} machines that accepts all a^i that LOOP(c, d, e) did not accept. We ran this program for $1 \leq n \leq 1700$ (this program took much longer to run then the first one). Here is what we discovered:

1. For slightly more than half of the n, we found a smaller NFA this way.

2. The most common improvement was 1. Then 2. ... Then 6. There were no improvements bigger than 6. There was a slight tendency of getting bigger improvements for bigger n.

We conjecture that, for all L, there exists n, so that the second program will produce an NFA that has at least L states fewer than the first program.

9 Open Questions

We conjecture that every cofinite unary language has a small NFA; however, this is hard to state rigorously.

The NFA for MN(n) is optimal up to $\widetilde{O}(1)$ terms. We would like to know if the other NFA's we have presented are optimal up to $\widetilde{O}(1)$ terms.

10 Acknowledgments

We would like to thank Jeff Shallit for his slides [3] that got us started on this subject, his sharing the proof of Theorem 7.2 with us, and for his encouragement to pursue this work.

References

[1] Christian Axler. On the sum of the first n prime numbers, 2014. https://arxiv.org/pdf/1409.1777.pdf.

[2] M. Chrobak. Finite automata and unary languages. *TCS*, 47:149–158, 1986. Erratum for paper is at https://dl.acm.org/citation.cfm?id=860232.

[3] J. Shallit. The Frobenius problem and its generalization. Slides: https://cs.uwaterloo.ca/~shallit/Talks/frob14.pdf.