Small NFA's for Cofinite Unary Languages

William Gasarch [*] Erik Metz [†] Eric Shen [‡]
Zan Xu [§] Sam Zbarsky [¶]

Abstract

Let $n \in \mathbb{N}$ and let $\text{MN}(n) = \{a^y : y \neq n\}$. It is easy to show that any DFA for $\text{MN}(n)$ requires $n+2$ states. What about an NFA? It is a folklore theorem that there is an NFA for $\text{MN}(n)$ with $O(\sqrt{n})$ states.

We consider a generalization of this problem. Let A be a set and let $\text{MN}(A) = \{a^y : y \notin A\}$. ($MN$ stands for *Missing Number*.) We show that for many finite sets A the NFA for $\text{MN}(A)$ is much smaller than the DFA for $\text{MN}(A)$. In the process we get more refined bounds on the size of an NFA for $\text{MN}(n)$.

1 Introduction

Consider the language

$$\text{MN}(n) = \{a^y : y \neq n\}.$$

(MN stands for *Missing Number*.)

It is easy to show that (1) there is a DFA for $\text{MN}(n)$ with $n+2$ state, and (2) any DFA for $\text{MN}(n)$ has at least $n+2$ states. What about an NFA for $\text{MN}(n)$? What about other cofinite unary sets?

[*]University of Maryland. gasarch@cs.umd.edu
[†]University of Maryland. emetz1618@gmail.com
[‡]Montgomery Blair High School. eric.shen2000@gmail.com
[§]Montgomery Blair High School. zaxu@mbhs.edu
[¶]Princeton University, zbarskysam@gmail.com

Def 1.1 If $A \subseteq \mathbb{N}$ then

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

We will only use this definition when A is finite.

Def 1.2 If L is any language then *a small* NFA *for* L is an NFA of size $\ll$ the number of states for an optimal DFA for L.

In Section 3 we show (1) An NFA for MN(1000) of size 59, (2) An NFA for MN$(999, 1000, 1001, 1002, 1003)$ of size 316 (and one of size 190, and finally one of size 99). The proofs contains many of the ideas for later results. In Section 4 and 5 we give small NFA's for partial sets. That means that there will be many strings that we do not care if the NFA accepts them or not. In Section 6, 7, and 6.2 we use the results of Sections 4 and 5 to show there are small NFA's for many sets of the form MN(A). In Section 8 we show that any NFA for MN$(n, \ldots, n+k-1)$ requires $\max\{k, \sqrt{n})\}$ states. In Section 9 we discuss open problems and present some empirical results.

We will often want to ignore polylog terms, so we use the following notation.

Notation 1.3 If f and g are functions then $f(n) = \tilde{O}(g(n))$ means that there exists i such that $f(n) = O(g(n)\log^i(n))$.

2 The LOOP(c, d, e) NFA

We will need a certain NFA in all of our constructions:

Def 2.1 Let $c, d, e \in \mathbb{N}$ such that $1 \leq c \leq d-1$. We describe LOOP(c, d, e) for when $e < d$ and consider the $e \geq d$ case later.

1. states $0, 1, \ldots, d-1$ with 0 as the start state and e as the only accept state,
2. for all $0 \leq i \leq d-1$ with $i \neq c$, $\delta(i, a) = i + 1 \pmod{d}$,
3. $\delta(c-1, a) = \{0, c\}$ (so the NFA is a cycle with a chord that is a shortcut back to the start state).

Note that to get to the accept state, we need to go through several (possibly 0) cycles, each of length c or d, and then e more steps, so LOOP(c, d, e) accepts

$$L = \{a^y : (\exists C, D \in \mathbb{N})[y = cC + dD + e]\}$$

and has d states.

If $e > d$ then the NFA will have a string of states of length e from the start state to the only accept state. That accept state will then have the d-loop and the c-shortcut. This NFA recognizes L and has $d + e$ states.

Lemma 2.2 *Let c, d, e be such that $c < d$ are relatively prime. There exists an* NFA *M such that:*

1. *M accepts all elements of $\{a^y : y \geq cd - c - d + e + 1\}$.*
2. *M rejects $a^{cd-c-d+e}$.*
3. *M rejects strings of the form $a^{cd-Cc-Dd+e}$ where $C, D \geq 1$.*
4. *We have no comment on any other strings.*
5. *If $e < d$ then M has d states.*
6. *If $e \geq d$ then M has $d + e$ states.*

Proof: Let $M =$ LOOP(c, d, e).

The following is well known. See Alfonsin [1], page 31, for several proofs and references.

1. All natural numbers $\geq cd - c - d + 1$ are in $\{xc + yd : x, y \in \mathbb{N}\}$.
2. $cd - c - d \notin \{xc + yd : x, y \in \mathbb{N}\}$.

Hence for all c, d relatively prime and $e \in \mathbb{N}$:

1. All natural numbers $\geq cd - c - d + e + 1$ are in $\{xc + yd + e : x, y \in \mathbb{N}\}$. Hence $M =$ LOOP(c, d, e) accepts all elements of $\{a^y : y \geq c+d+e+1\}$.
2. $cd - c - d + e \notin \{xc + yd + e : x, y \in \mathbb{N}\}$. Hence $M =$ LOOP(c, d, e) rejects $a^{cd-c-d+e}$.

Assume, by way of contradiction, that M accepts some string of the form $a^{cd-Cc-Dd+e}$ where $C, D \geq 1$. Then

$$cd - cC - Dd + e \in \{xc + yd + e : x, y \in \mathbb{N}\}.$$

so

$$cd - cC - Dd + e = xc + yd + e$$

Add $c(C-1) + d(D-1)$ to both sides to get

$$cd - c - d + e = (x + C - 1)c + (y + D - 1)d + e.$$

Since $C, D \geq 1$, $x + C - 1, y + D - 1 \geq 0$. Therefore $cd - c - d + e \in \{xc + yd + e : x, \in \mathbb{N}\}$ which is false. ∎

3 Examples of Small NFAs for MN(A)

3.1 A Small NFA for MN(1000)

Theorem 3.1 *There exists an* NFA *for* MN(1000) *with 59 States.*

Proof: Let M' be the NFA from Lemma 2.2 with $c = 32$, $d = 33$, and $d = 9$. By that lemma (1) M accepts all elements in $\{a^y : y \geq 32 \times 33 - 33 - 32 + 9 + 1 = 1001\}$, (2) M rejects a^{1000}, and (3) M has 33 states.

Note that M rejects some a^y with $y \leq 999$ that we want to accept.

For $q \in \{4, 5, 7, 9\}$ create a DFA M_q that accepts $\{a^y : y \not\equiv 1000 \pmod{q}\}$. Note that M_q has q states.

Our final NFA M has a start state and ϵ transitions to M', M_4, M_5, M_7, M_9. The number of states is $1 + 33 + 4 + 5 + 7 + 9 = 59$.

Let a^y be a string rejected by M. By the nature of M', $y \leq 999$. By the definition of M_4, M_5, M_7, M_9 we have

$$\begin{aligned} y &\equiv 1000 \pmod{4} \\ y &\equiv 1000 \pmod{5} \\ y &\equiv 1000 \pmod{7} \\ y &\equiv 1000 \pmod{9} \end{aligned}$$

Since $4, 5, 7, 9$ are relatively prime and $4 \times 5 \times 7 \times 9 = 1260 > 1000$, $y = 1000$.

Therefore M accepts MN(1000). ∎

3.2 A Small NFA for MN$(999, 1000, 1001, 1002, 1003)$

Theorem 3.2

1. *There exists an* NFA *for* MN$(999, 1000, 1001, 1002, 1003)$ *with 316 states.*
2. *There exists an* NFA *for* MN$(999, 1000, 1001, 1002, 1003)$ *with 190 states.*
3. *There exists an* NFA *for* MN$(999, 1000, 1001, 1002, 1003)$ *with 99 states.*

Proof:

All three proofs begin the same way.

Let M be the NFA from Lemma 2.2 with $c = 34$, $d = 35$ $e = 18$. By that lemma (1) M accepts all elements in $\{a^y : y \geq 35 \times 34 - 35 - 34 + 18 + 1 = 1139\}$, (2) M rejects a^{1138}, and (3) M has 35 states. One can also show that M rejects $\{a^{999}, a^{1000}, a^{1001}, a^{1002}, a^{1003}\}$.

Note that M' rejects some a^y with $y \leq 1138$ that we want to accept. We show three ways to deal with this.

1) Let $Q = \{2, 3, 5, 11, 13, 17, 19, 23, 29, 31, 37, 41, 49\}$. Note that

$$\prod_{q \in Q} q = 2,129,751,844,690,470 > 1,916,984,564,383,744 = (1139)^5$$

and

$$\sum_{q \in Q} q = 280.$$

For each $q \in Q$ create a DFA M_q that accepts

$$\{a^y : y \not\equiv 999, 1000, 1001, 1002, 1003 \pmod{q}\}.$$

Note that M_q has q states.

Our final NFA M has a start state and ϵ transitions to M' and to all M_q. Hence the number of states in M is $1 + 35 + 280 = 316$.

Let a^y be a string rejected by M. By the nature of M', $y \leq 1138$. By the definition of M_q we have the following.

(For reasons of space we abbreviate $\pmod p$ by (p).)

$y \equiv 999(2)$ or	$y \equiv 1000(2)$ or	$y \equiv 1001(2)$ or	$y \equiv 1002(2)$ or	$y \equiv 1003(2)$
$y \equiv 999(3)$ or	$y \equiv 1000(3)$ or	$y \equiv 1001(3)$ or	$y \equiv 1002(3)$ or	$y \equiv 1003(3)$
$y \equiv 999(5)$ or	$y \equiv 1000(5)$ or	$y \equiv 1001(5)$ or	$y \equiv 1002(5)$ or	$y \equiv 1003(5)$
$y \equiv 999(11)$ or	$y \equiv 1000(11)$ or	$y \equiv 1001(11)$ or	$y \equiv 1002(11)$ or	$y \equiv 1003(11)$
$y \equiv 999(13)$ or	$y \equiv 1000(13)$ or	$y \equiv 1001(13)$ or	$y \equiv 1002(13)$ or	$y \equiv 1003(13)$
$y \equiv 999(17)$ or	$y \equiv 1000(17)$ or	$y \equiv 1001(17)$ or	$y \equiv 1002(17)$ or	$y \equiv 1003(17)$
$y \equiv 999(19)$ or	$y \equiv 1000(19)$ or	$y \equiv 1001(19)$ or	$y \equiv 1002(19)$ or	$y \equiv 1003(19)$
$y \equiv 999(23)$ or	$y \equiv 1000(23)$ or	$y \equiv 1001(23)$ or	$y \equiv 1002(23)$ or	$y \equiv 1003(23)$
$y \equiv 999(29)$ or	$y \equiv 1000(29)$ or	$y \equiv 1001(29)$ or	$y \equiv 1002(29)$ or	$y \equiv 1003(29)$
$y \equiv 999(31)$ or	$y \equiv 1000(31)$ or	$y \equiv 1001(31)$ or	$y \equiv 1002(31)$ or	$y \equiv 1003(31)$
$y \equiv 999(37)$ or	$y \equiv 1000(37)$ or	$y \equiv 1001(37)$ or	$y \equiv 1002(37)$ or	$y \equiv 1003(37)$
$y \equiv 999(41)$ or	$y \equiv 1000(41)$ or	$y \equiv 1001(41)$ or	$y \equiv 1002(41)$ or	$y \equiv 1003(41)$
$y \equiv 999(49)$ or	$y \equiv 1000(49)$ or	$y \equiv 1001(49)$ or	$y \equiv 1002(49)$ or	$y \equiv 1003(49)$

Let

$$Q_{999} = \{q \in Q : y \equiv 999 \pmod q\}$$

$$Q_{1000} = \{q \in Q : y \equiv 1000 \pmod q\}$$

$$Q_{1001} = \{q \in Q : y \equiv 1001 \pmod q\}$$

$$Q_{1002} = \{q \in Q : y \equiv 1002 \pmod q\}$$

$$Q_{1003} = \{q \in Q : y \equiv 1003 \pmod q\}$$

Every $q \in Q$ is in at least one of Q_{999}, Q_{1000}, Q_{1001}, Q_{1002}, Q_{1003}. Let

$$R_{999} = \prod_{q \in Q_{999}} q$$

$$R_{1000} = \prod_{q \in Q_{1000}} q$$

$$R_{1001} = \prod_{q \in Q_{1001}} q$$

$$R_{1002} = \prod_{q \in Q_{1002}} q$$

$$R_{1003} = \prod_{q \in Q_{1003}} q$$

Since every $q \in Q$ is in some R we have

$$R_{999}R_{1000}R_{1001}R_{1002}R_{1003} \geq \prod_{q \in Q} q > (1139)^5$$

Hence there is an $x \in \{999, 1000, 1001, 1002, 1003\}$ such that $R_x \geq 1139$. Since $y \leq 1138$, $y \equiv x$ mod each of the primes in Q_x, and $R_x \geq 1139$, $y = x$. Hence $y \in \{999, 1000, 1001, 1002, 1003\}$.

2) Let $Q = \{1 \times 7, 2 \times 7, 3 \times 7, 5 \times 7, 11 \times 7\}$. Note that the lcm of Q is $7 \times 2 \times 3 \times 5 \times 11 = 2310 > 1139$ and $\sum_{q \in Q} q = 154$.

For each $q \in Q$ create a DFA M_q that accepts

$$\{a^y : y \not\equiv 999, 1000, 1001, 1002, 1003 \pmod{q}\}.$$

Note that M_q has q states.

Our final NFA M has a start state and ϵ transitions to M' and to all M_q. Hence the number of states in M is $1 + 35 + 154 = 190$.

Let a^y be a string rejected by M. By the nature of M', $y \leq 1138$. By the definition of M_q we have

$y \equiv 999(7)$ or	$y \equiv 1000(7)$ or	$y \equiv 1001(7)$ or	$y \equiv 1002(7)$ or	$y \equiv 1003(7)$
$y \equiv 999(14)$ or	$y \equiv 1000(14)$ or	$y \equiv 1001(14)$ or	$y \equiv 1002(14)$ or	$y \equiv 1003(14)$
$y \equiv 999(21)$ or	$y \equiv 1000(21)$ or	$y \equiv 1001(21)$ or	$y \equiv 1002(21)$ or	$y \equiv 1003(21)$
$y \equiv 999(35)$ or	$y \equiv 1000(35)$ or	$y \equiv 1001(35)$ or	$y \equiv 1002(35)$ or	$y \equiv 1003(35)$

Let $x_1 \in \{999, 1000, 1001, 1002, 1003\}$ be such that $y \equiv x \pmod{7}$. We claim that $y \equiv x \pmod{14}$ (and mod 21 and mod 35). Since $y \equiv x_1 \pmod{7}$, there exists m_1 such that

$$y = x_1 + 7m_1$$

Let x_2 be such that $y \equiv x_2 \pmod{14}$. Then there exists m_2 such that

$$y = x_2 + 14m_2$$

Subtract these two equations to obtain

$$0 = (x_1 - x_2) + 7m_1 - 14m_2$$

$$x_1 \equiv x_2 \pmod 7.$$

Since no two elements of $\{999, 1000, 1001, 1002, 1003\}$ are equivalent mod 7 (which is why we picked 7) we have $x_1 = x_2$.
Similarly $y \equiv x$ mod 21 and 35.
Since $y \equiv x \pmod 7$, $y \equiv x \pmod{14}$, $y \equiv x \pmod{21}$ and $y \equiv x \pmod{35}$, we have $y \equiv x \pmod{lcm(7, 14, 21, 35)}$. Since the lcm is ≥ 1139 we have $y = x$.

3) Use M and also M_q where $q \in \{9, 13, 19, 23\}$. We omit the proof that this works. The number of states is $1+35+9+13+19+23 = 99$. We found this NFA by a computer search, so there is no general theorem based on it. ∎

Several of the methods in the above constructions can be generalized. We call the technique used in the proof of Theorem 3.2.1 (Theorem 3.2.2, Theorem 3.2.3) by the name "T1" (T2, T3).

1. T1 is generalized in Lemma 4.1. T2 is generalized in Lemma 5.7. T3 was found by brute force and hence cannot be generalized.

2. Let A be a finite set. If we want to get a small NFA for MN(A) should we use T1 or T2?

   - If A does not have that many elements in it then use T1. In particular if there are ℓ elements then you will need a set of primes whose product is $\geq N^\ell$ for some N. If ℓ is small this set of primes will have a small sum.
   - If there is small prime p such that no two elements of A are congruent mod p then use T2. You will need a set of primes such that the product of p with those primes is $\geq N$; however, the part of the NFA will have uses this will have size p times the sum of the primes.

- Exercise: Work out the small NFA's for $\{999, 1003\}$ using both T1 and T2. T1 will do better than T2 here.

3. For large n the number of states in the LOOP part will dominate the number of states in the primes-part.

4 A Small NFA for a Particular Partial Cofinite Set

We use the following is the $d = c + 1$ case of Lemma 2.2:

Lemma 4.1 *Let $c, e \in \mathbb{N}$. There exists an* NFA *M such that the following all hold:*

1. *M accepts all elements of $\{a^y : y \geq c^2 - c + e\}$.*
2. *M rejects $a^{c^2-c+e-1}$.*
3. *M rejects strings of the form $a^{c(c+1)-Cc-D(c+1)+e}$ where $C, D \geq 1$.*
4. *M may or may not accept any other string.*
5. *If $e < c + 1$ then M has $c + 1$ states.*
6. *If $e \geq c + 1$ then M has $c + e + 1$ states.*

Lemma 4.2 *Let $k, n \in \mathbb{N}$.*

$$\frac{(k+1) + \sqrt{4n + (k+1)^2 + 4k + 4}}{2} \leq \sqrt{n} + k + 2$$

Proof:

$$\frac{(k+1) + \sqrt{4n + (k+1)^2 + 4k + 4}}{2} \leq \frac{(k+1) + \sqrt{4n} + \sqrt{(k+3)^2}}{2}$$

$$\leq \frac{2\sqrt{n} + 2k + 4}{2} \leq \sqrt{n} + k + 2$$

∎

Lemma 4.3 *Let $k, n \in \mathbb{N}$. There exists c, e such that the following hold:*

1. $n = c^2 - kc - k + e$ *OR* $n = c^2 - (k+1)c - (k+1) + e$.
2. $c \leq \sqrt{n} + k + 2$.
3. $0 \leq e \leq k + c$.

Proof:

Consider the two sets of intervals (of naturals) indexed by c.

$$I_c^1 = [c^2-(k+1)c-(k+1)+0, c^2-(k+1)c+(c-1)] = [c^2-(k+1)c-(k+1), c^2-kc-1]$$

$$I_c^2 = [c^2 - kc - k + 0, c^2 - kc - k + (c-1)] = [c^2 - kc - k, c^2 - (k-1)c - k - 1]$$

Since the left end point of I_c^2 is $\leq$ the right endpoint of I_c^1, the union of all of these intervals is

$$\bigcup_{c=1}^{\infty} [c^2 - (k+1)c - (k+1), c^2 - (k-1)c - k - 1]$$

This covers all of $\mathbb{N}$ since

$$(c+1)^2 - (k+1)(c+1) - (k+1) = c^2 + 2c + 1 - (k+1)(c+1) - (k+1) =$$

$$c^2 - (k-1)c - k$$

and the prior interval ended at $c^2 - (k-1)c - k - 1$.

So, given n one of the following holds:

Case 1: There is a c such that $n \in [c^2 - (k+1)c - (k+1), c^2 - kc - 1]$. Clearly n is of the form $n = c^2 - (k+1)c - (k+1) + e$ with $e \leq c + k$. Hence

$$c^2 - (k+1)c + (e - k - n - 1) = 0$$

We use the quadratic formula. We ignore the case of $-\sqrt{b^2 - 4ac}$ since the upper bound in that case is better than the one we get in the $+\sqrt{b^2 - 4ac}$ case.

$$c = \frac{(k+1) + \sqrt{(k+1)^2 + 4(n-e+k+1)}}{2}$$

By Lemma 4.2 $c \leq \sqrt{n} + k + 2$.

Case 2: There is a c such that $n \in [c^2 - kc - k, c^2 - (k-1)c - k - 1]$. Clearly n is of the form $c^2 - kc - k + e$ where $e \leq c - 1$.

$$n = c^2 - kc - k + e$$

$$c^2 - kc + (e - n - k) = 0$$

We use the quadratic formula. We ignore the case of $-\sqrt{b^2 - 4ac}$ since the upper bound in that case is better than the one we get in the $+\sqrt{b^2 - 4ac}$ case.

$$c = \frac{k + \sqrt{k^2 + 4(n+k-e)}}{2}$$

By Lemma 4.2 $c \leq \sqrt{n} + k + 2$. ∎

Lemma 4.4 *Let $k \geq 2$. For all n there is an* NFA *M such that*

1. *M accepts a^y for all $y \geq n + k\sqrt{n} + k^2 + 3k$.*
2. *M does not accept $a^n, \ldots, a^{n+k-1}$.*
3. *We do not care what M does in the cases not specified above.*
4. *If $k = O(1)$ then M is of size $\sqrt{n} + O(1)$.*
5. *If $k = O(n^\delta)$ then M is of size $\leq 2n^{\max\{1/2,\delta\}} + O(1)$.*

Proof:

Let c be as in Lemma 4.3. Note that $c \leq \sqrt{n} + k + 2$. We will use $\text{LOOP}(c, c+1, e)$ for some $e \leq c + k$. Hence:

- If $k = O(1)$ then $k \leq c$ and $\text{LOOP}(c, c+1, e)$ is of size $c+1 \leq \sqrt{n} + O(1)$.

- If $k = O(n^\delta)$ then either (1) $\delta \leq 1/2$ and LOOP$(c, c+1, e)$ is of size $c+1 = n^{1/2} + n^\delta \leq 2n^{1/2} + O(1)$ or (2) $\delta > 1/2$ and LOOP$(c, c+1, e)$ is of size $\leq n^{1/2} + n^\delta \leq 2n^\delta + O(1)$. Hence the NFA is of size $\leq 2n^{\max\{1/2,\delta\}} + O(1)$.

Case 1: $n = c^2 - kc - k + e$. Let M be the NFA from Lemma 4.1 with c, e.
Then M accepts $\{a^y : y \geq c^2 - c - 1 + e\}$.
Note that

$$c^2 - c - 1 + e = (c^2 - kc - k + e) + kc + k - c - 1 = n + kc + k - c - 1$$

$$\leq n + kc + k \leq n + k(\sqrt{n} + k + 2) + k \leq n + k\sqrt{n} + k^2 + 3k.$$

Hence M accepts $\{a^y : y \geq n + k\sqrt{n} + k^2 + 3k\}$.
Since, for all $C, D \in \mathbb{N}^{\geq 1}$ M does not accept $a^{c(c+1)-Cc-D(c+1)+e}$, for $0 \leq i \leq k-1$, using $C = i+1$ and $D = k-i$ we get that for all $0 \leq i \leq k-1$, M does not accept a^{n+i}.
Case 2: $n = c^2 - (k+1)c - (k+1) + e$. Let M be the NFA from Lemma 4.1 with c, e.
Then M accepts $\{a^y : y \geq c^2 - c - 1 + e\}$.
Note that

$$c^2 - c - 1 + e = (c^2 - (k+1)c - (k+1) + e) + (k+1)c + (k+1) - c - 1 = n + kc + k$$

As in Case 1 this is $\leq n + k\sqrt{n} + k^2 + 3k$.
Hence M accepts $\{a^y : y \geq n + k\sqrt{n} + k^2 + 3k\}$.
Since for all $C, D \in \mathbb{N}^{\geq 1}$ M does not accept $a^{c(c+1)-Cc-D(c+1)+e}$, for $0 \leq i \leq k-1$, let $C = i+1$ and $D = k+1-i$ we get that for all $0 \leq i \leq k-1$, M does not accept a^{n+i}. ∎

5 Small NFAs for Partial Finite Sets

5.1 A Small NFA for a Large Partial Finite Set

Def 5.1 Let $N \in \mathbb{N}$. A set Q of naturals is *N-cool* if

1. Q is a set of relatively prime numbers.
2. $\prod_{q \in Q} q \geq N$.

Let $f(N)$ be the min sum over all N-cool sets Q.

Lemma 5.2 $f(N) \leq \ln^2 N$.

Proof:
A function related to f has been well studied:

Def 5.3 Let $S \in \mathbb{N}$. $L(S)$ is the maximum least common multiple of a partition of S. L is known as *Landau's function*.

Massias et al. [4] shows that, for $S \geq 2$, $L(S) \geq e^{\sqrt{S \ln S}}$. So, if $e^{\sqrt{S \ln S}} \geq N$, we can find a set Q of relatively prime numbers such that the product over Q is at least N and the sum over Q is at most S. Hence, for $N \geq 2$, $f(N) \leq \ln^2 N$ (better bounds are possible but will not help us). ∎

Lemma 5.4 *Let $\ell < N$. Let $A \subseteq \{1, \ldots, N\}$ of size ℓ. There is an* NFA *M such that*

1. *If M rejects a^y then either $y \in A$ or $y \geq N$.*
2. *We do not care what M does in the cases not specified above.*
3. *M is of size $O(\ell^2 \log^2 N)$.*
4. *If $N = n^\Delta$ and $\Delta = O(1)$ then M is of size $\tilde{O}(\ell^2)$. (This follows from the previous part.)*

Proof:
By Lemma 5.2 there exists an N^ℓ-cool set Q such that $\sum_{q \in Q} q \leq \ln^2 N^\ell = \ell^2 \ln^2 N$.

For each $q \in Q$ let M_q be the DFA that accepts the following set:

$$\{a^y : (\forall x \in A)[y \not\equiv x \pmod q]\}.$$

(It is possible for some of the DFA's to not accept any strings. We discuss this point after the proof.)

Let M be the NFA that has an ϵ transition to each M_q.

Let y be such that a^y is rejected by M and $y \leq N$. Then

$$(\forall q \in Q)(\exists x \in A)[y \equiv x \pmod{q}].$$

We call this *The condition.*

For every $x \in A$ let

$$Q_x = \{q \in Q : y \equiv x \pmod{q}\}$$

$$R_x = \prod_{q \in Q_x} q$$

By condition 2 every $q \in Q$ is a factor of at least one R_x. Hence

$$\prod_{x \in Q} R_x \geq \prod_{q \in Q} q \geq N^\ell$$

Since the left hand side is the product of ℓ numbers one of them is $\geq N$. Therefore there exists x such that $R_x \geq N$. Hence there exists a set Q_x such that

1. $(\forall q \in Q_x)[y \equiv x \pmod{p}]$
2. $\prod_{q \in Q_x} q \geq N$.

Hence $y = x$. Therefore the only strings rejected are in $\{a^y : y \in A\}$.

The size of the DFA is clearly

$$1 + \sum_{q \in Q} q \leq \ell^2 \log^2 N + O(1) = O(\ell^2 \log^2 N).$$

∎

Note 5.5 In the proof of Lemma 5.4, for every $q \in Q$, we have a DFA that accepts a^y iff

$$(\exists x \in A)[y \not\equiv x \pmod{q}].$$

If $x, x+1 \in A$ and $q = 2$ then this DFA will reject every string. There are many similar scenarios where the DFA rejects every string. We could

leave out all such DFA's associated to $q < k$; however, we have not been able to use this to obtain a smaller DFA. In the current proof having all of the DFA's makes the analysis easier, though if we removed the ones that reject everything we could still make essentially the same proof work.

5.2 A Small NFA for a Small Partial Finite Set

Lemma 5.4 is not useful when $\ell \geq \sqrt{n}$. We will prove a lemma that will be helpful when A is a large contiguous set. We first need some known facts from number theory:

Lemma 5.6 *Let $m \in \mathbb{N}$. Let $q_1, \ldots, q_m$ be the first m primes.*

1. $\prod_{i=1}^{m} q_i = e^{(1+o(1))m \log m}$.
2. $\sum_{i=1}^{m} q_i \sim \frac{m^2}{2} \log m$.

Proof:
1) From Rosser and Shoenfeld [5], equations 3.12 through 3.15, one can easily derive the result.
2) This result is well known. Axler [2] has references and better bounds.
∎

Lemma 5.7 *Let $A \subseteq \{1, \ldots, N\}$. Let p be such that for all $x_1, x_2 \in A$, $x_1 \not\equiv x_2 \pmod{p}$. Let $m \geq 2$. There is an* NFA *M such that*

1. *If M rejects a^y then either $y \in A$ or $y \geq N$.*
2. *We do not care what M does in the cases not specified above.*
3. *M is of size $O(p \log^3(N))$.*

Proof:

Let q_i be the ith prime. Let m be the least number such that $q_1 \cdots q_m \geq N$ (we only need $\geq \frac{N}{p}$). Since $\Pi_{i=1}^{m} q_i \geq m^m$ we have $m \leq \log N$. Since $\sum_{i=1}^{m} q_i \leq m^2 \log m$ we have $\sum_{i=1}^{m} q_i \leq \log^3 N$.

For $i = 0$ to m let M_{pq_i} be the DFA that accepts the following set:

$$\{a^y : (\forall x \in A)[y \not\equiv x \pmod{pq_i}]\}.$$

Let M be the NFA that has ϵ transitions to each M_{pq_i}. Note that the number of states in M is

$$p + \sum_{i=1}^{m} pq_i = O(p(\log(N/p))^3).$$

Let y be such that a^y is rejected by M. Since a^y is rejected by M_p there is an $x \in A$ such that $y \equiv x \pmod{p}$.

Claim: For all $1 \leq i \leq m$, $y \equiv x \pmod{pq_i}$.

Proof of Claim: Let $1 \leq i \leq m$. Since M rejected a^y there is an $x' \in A$ such that $y \equiv x' \pmod{pq_i}$. Hence $y \equiv x' \pmod{p}$. Since $y \equiv x \pmod{p}$ we have $x \equiv x' \pmod{p}$. Since no two elements of A are congruent mod p, $x = x'$.

End of Proof of Claim

We now have that, for all $i = 1$ to m, $y \equiv x \pmod{pq_i}$. Hence $y \equiv x \pmod{pq_1 q_2 \cdots q_m}$. Therefore either $y = x$ or $y \geq pq_1 \cdots q_m = N$. ∎

We now combine Lemmas 5.4 and 5.7.

Lemma 5.8 *Let $0 < \epsilon \leq \delta < 1$ and $\Delta = O(1)$. (We assume that $n^\epsilon, n^\delta, n^\Delta$ are integers. Modifications for when it is not are left to the reader.) Let $\delta' > \delta$ (we think of $\delta' - \delta$ as being very small). Let $A \subseteq \{n, n+1, \ldots, n+n^\delta\}$ of size n^ϵ. There is an* NFA *M such that*

1. *If M rejects a^y then either $y \in A$ or $y \geq n^\Delta$.*
2. *We do not care what M does in the cases not specified above.*
3. *M is of size $\tilde{O}(n^{\min\{2\epsilon, \delta'\}})$.*

Proof: By Lemma 5.4 there is an NFA satisfying (1) and (2) with $O(n^{2\epsilon} \log^2 n^\Delta) = O(\Delta n^{2\epsilon} \log^2 n) = \tilde{O}(n^{2\epsilon})$ states.

Let p be a prime between n^δ and $2n^\delta$. Note that if $x_1, x_2 \in A$ then $x_1 \not\equiv x_2 \pmod{p}$. Let $N = n^\Delta$. Let A be as above, though view it as a subset of $\{1, \ldots, n^\Delta\}$. Apply Lemma 5.7 with $p = \Theta(n^\delta)$ as above, and $N = n^\Delta$, to obtain the desired NFA of size

$$O(p \log^3 N) = \tilde{O}(n^\delta)$$

Take the smaller of the two NFAs to obtain the result. ∎

6 Small NFA for MN(A)

6.1 Small NFA for Subsets of $\{n, \ldots, n+k-1\}$

Theorem 6.1 *Let $n, k, \ell \in \mathbb{N}$. Let $A \subseteq \{n, \ldots, n+k-1\}$ of size ℓ.*

1. *If $k = O(1)$ (and hence $\ell = O(1)$) then there is an* NFA *for* MN(A) *of size $\sqrt{n} + \tilde{O}(1)$.*
2. *If $k = O(n^\delta)$ and $\ell = O(1)$ then there is an* NFA *for* MN(A) *of size $2n^{\max\{1/2,\delta\}} + \tilde{O}(1)$.*
3. *If $k = O(n^\delta)$ and $\ell = O(n^\epsilon)$ then there is an* NFA *for* MN(A) *of size $2n^{\max\{1/2,\delta\}} + \tilde{O}(n^{\min\{2\epsilon,\delta'\}})$ where δ' is any number that is $> \delta$.*

Proof: Let M_1 be the NFA from Lemma 4.4 such that

1. M accepts a^y for all $y \geq n + k\sqrt{n} + k^2 + 3k$.
2. M does not accept $a^n, \ldots, a^{n+k-1}$.
3. We do not care what M does in the cases not specified above.
4. If $k = O(1)$ then M is of size $\sqrt{n} + O(1)$
5. If $k = O(n^\delta)$ then M is of size $\leq 2n^{\max\{1/2,\delta\}} + O(1)$.

Let M_2 be the NFA from Lemma 5.8, with Δ such that $N = n + k\sqrt{n} + k^2 + 3k \leq O(n^\Delta)$. Note that for $k \ll n$ (which are the only cases below) $\Delta \leq 2$ suffices.

1. If M_2 rejects a^y then either $y \in A$ or $y \geq n + k\sqrt{n} + k^2 + 3k$.
2. We do not care what M does in the cases not specified above.
3. M_2 is of size $\tilde{O}(n^{\min\{2\epsilon,\delta'\}})$ where $k = O(n^\delta)$, and $\delta' > \delta$. Note that the following hold:
   (a) If $\ell = O(1)$ then M_2 is of size $\tilde{O}(1)$.
   (b) If $\ell = O(n^\epsilon)$ then M_2 is of size $\tilde{O}(n^{\min\{2\epsilon,\delta'\}})$.

Let M be the NFA that has a start state that has an ϵ transition into both M_1 and M_2. The result follows from the bounds on M_1, M_2 given above. ∎

6.2 A Small NFA for $\mathrm{MN}(n, kn)$

If $A = \{n, 2n\}$ then Theorem 6.1 yields an NFA of size $O(n)$ for $\mathrm{MN}(A)$. This is not a small NFA. Is there a small NFA for $\mathrm{MN}(A)$? Yes:

Theorem 6.2 *Let $k \in \mathbb{Q}$ with $0 < k < 1$. Let n be such that $kn \in \mathbb{N}$. There is an* NFA *for $\mathrm{MN}(kn, n)$ with $2\sqrt{n}\log n + \tilde{O}(1)$ states.*

Proof: We first construct an NFA M with the following three properties: (1) there exists $1 < \alpha < \log kn + O(1)$ such M accepts $\{a^y \mid y \geq \alpha n\}$, (2) M does not accept a^{kn}, (3) M does not accept a^n, (4) M has $2\sqrt{n}\log n + O(1)$ states.

Let c, e be such that $c(c-1) < n+1 \leq (c+1)c$ and $n = c(c-1) + e - 1$. Note that $c, c+1$ are $\leq \sqrt{n} + O(1)$ and $e \leq 2\sqrt{n} + O(1)$.

By Lemma 2.2 with $c, c-1, e$ there exists an NFA M such that (1) M accepts all elements of $\{a^y \mid y \geq n+1\}$, and (2) does not accept a^n. We will use $\mathrm{LOOP}(c, c+1, e)$ in the first case directly, in the second case indirectly.

Case 1: $\mathrm{LOOP}(c, c+1, e)$ does not accept a^{kn}. Let $M = \mathrm{LOOP}(c, c+1, e)$.

M does not accept a^{kn} by the premise.

As noted above M does not accept a^n and it does accept $\{a^y \mid y \geq n+1\}$. Formally we take $\alpha = 2$.

M has $c + 1 + e \leq 3\sqrt{n} + O(1) \leq 2\sqrt{n}\log n + O(1)$ states.

Case 2: a^{kn} is accepted by $\mathrm{LOOP}(c, c+1, e)$. Then there exists $C, D \in \mathbb{N}$ such that

$$kn = C(c) + D(c+1) + e$$

$$kn - e = C(c) + D(c+1)$$

Note that a is relatively prime to $c+1$, $c(c+1) > kn - e$, and $C, D \geq 0$. We show that the choice of C and D is unique: If $C(c) + D(c+1) = C'(c) + D'(c+1)$, then $(C - C')(c) = (D' - D)(c+1)$. Since c and $c+1$ are coprime, we have $z(c+1) = (C - C')$ and $z(c) = (D' - D)$ for some integer z. Note $C < c+1$ and $D < c$, so z positive implies C' negative, and z negative implies D' negative. Hence C, D are unique.

Let p be a small prime that does not divide D or c (one can show that $p = \log(Dc) + O(1) \leq \log(kn) + O(1)$). Let $M = \mathrm{LOOP}(c, p(c+1), e)$.

M accepts a^y such that $y \geq c(c+1)p - c - p(c+1) + 1 + e = (c-1)(c+1)p - c + 1 + e$. Note that

$$n < (c-1)(c+1)p - c + 1 + e < (n + \sqrt{n}) \log kn + \sqrt{n} + O(1)$$

Hence there is a constant $1 < \alpha < \log kn + O(1)$ such that M accepts $\{a^y \mid y \geq \alpha n\}$. M may accept more strings. We need to show that M does not accept a^n or a^{kn}.

If M accepts a^{kn} then

$$kn = C''(c) + D''p(c+1) + e = C''c + pD''(c+1) + e$$

Since C and D are unique we have $C = C''$ and $D = pD''$. But then p divides D which is impossible.

If M accepts a^n then

$$n = C'(c) + D'p(c+1) + c = C'c + pD'(c+1) + e$$

hence a^n is accepted by $\text{LOOP}(c, c+1, e)$ which is impossible.

M has

$$p(c+1) \leq \sqrt{n} \log kn + O(1) \leq 2\sqrt{n} \log n + O(1)$$

states.

By Case 1 and Case 2 there is an NFA satisfying the four conditions above.

By Lemma 5.4 with $A = \{kn, n\}$, $N = \alpha n$, $\ell = 2$ there is an NFA M' such that (1) if M' rejects a^y then $y \in \{kn, n\}$ or $y \geq \alpha n$, (2) M' is of size $O(\log^2 \alpha n) = O(\log^2(n \log kn)) = \tilde{O}(1)$.

Our final NFA has a start state that has ϵ transitions to both M and M'. This clearly accepts $\text{MN}(kn, n)$ and has $2\sqrt{n} \log kn + \tilde{O}(1)$ states. ∎

Theorem 6.3 *Let $k \in \mathsf{Q}$ with $1 < k$. Let n be such that $kn \in \mathbb{N}$. There is an* NFA *for* $\text{MN}(n, kn)$ *with* $2\sqrt{kn} \log n + \tilde{O}(1)$ *states.*

Proof: Let $m = kn$. Note that $n = \frac{m}{k} \in \mathbb{N}$. Let $k' = \frac{1}{k}$. Note that $0 < k' < 1$. Apply Theorem 6.2 to $(k'm, m)$ to get the result. ∎

7 A Small NFA for $\mathrm{MN}(A \cup B)$ where $A \ll B$

If $A = \{n, n^2\}$ then Theorem 6.1 yields an NFA of size $O(n^2)$ for $\mathrm{MN}(A)$. This is not a small NFA. Is there a small NFA for $\mathrm{MN}(A)$? Yes. In fact, we state a theorem from which a small NFA for $\mathrm{MN}(A)$ is a corollary.

Theorem 7.1 *Let A_1 be a subset of $\mathbb{N}$. Let ℓ_1 be the largest element of A_1 and let $\ell_1 < \ell_2$. There is an* NFA *for $\mathrm{MN}(A_1 \cup \{\ell_2\})$ of size*

$$\leq 2\ell_2 + \sqrt{\ell_1} + O(\log^2(\ell_1)).$$

Proof:

By Theorem 6.1 there is an NFA M' for $\mathrm{MN}(\ell_1 - \ell_2)$ of size

$$\sqrt{\ell_2 - \ell_1} + O(\log^2(\ell_2 - \ell_1)) \leq \sqrt{\ell_2} + O(\log^2(\ell_2)).$$

Let s be its start state. Take this NFA and (a) make s a non-start state and call it r_{ℓ_1}, (b) add a (rejecting) start state r_0 and a sequence of $\ell_1 - 1$ reject states $r_1, r_2, \ldots, r_{\ell_1 - 1}$, (c) add transitions $\delta(r_i, a) = r_{i+1}$.

We call the new NFA M_1'. Note that it accepts $\mathrm{MN}(0, 1, 2, \ldots, \ell_1, \ell_2)$ and has $\ell_1 + \sqrt{\ell_2 - \ell_1} + O(\log^2(\ell_2))$ states.

Let M_2 be the DFA that accepts $\{a^y : y \leq \ell_1 \wedge y \notin A_1\}$. Note that M_2 has $\ell_1 + O(1)$ states.

Our final NFA M has a start state with a transition to M_1' and M_2. It clearly accepts $\mathrm{MN}(A_1 \cup A_2)$ and is of size

$$\leq 2\ell_1 + \sqrt{\ell_2} + O(\log^2(\ell_2))$$

∎

Corollary 7.2 *Let f be a function such that $\lim_{n\to\infty} \frac{f(n)}{n} = \infty$ (so $n \ll f(n)$). Let n be large. Let A be a set with largest element $f(n)$ and second largest element a^n. Then* $\mathrm{MN}(A)$ *has a small* NFA.

Proof: The smallest DFA for $\mathrm{MN}(A)$ has $f(n) + O(1)$ states. By Theorem 7.1 there is an NFA M for $\mathrm{MN}(A)$ of size

$$2n + \sqrt{f(n)} + O(\log^2(f(n)))$$

Since $n \ll f(n)$, $\sqrt{f(n)} \ll f(n)$, $\log^2 f(n) \ll f(n)$, M is a small NFA.

∎

8 $\mathrm{MN}(n, \ldots, n+k-1)$ requires $\geq \max\{k, \sqrt{n}\}$ States

Chrobak [3] proved the following.

Theorem 8.1 *Let L be a co-finite 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 ϵ 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.*

Jeffery Shallit proved the following and emailed it to us.

Theorem 8.2

1. *Let L be a cofinite unary language where the longest string that is not in L is of length n. Then any* NFA *for L requires $\geq \sqrt{n}$ states.*
2. *If A has least element n then any* NFA *for $\mathrm{MN}(A)$ has $\geq \sqrt{n}$ states. (This follows from part 1.)*

Proof:

Assume there was an NFA with $< \sqrt{n}$ states for L. Then by Theorem 8.1 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 8.1.

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. ∎

Theorem 8.3

1. *Let L be a unary language such that there exists k, n with $a^n, a^{n+1}, \ldots, a^{n+k-1} \notin L$ and $a^{n+k} \in L$. Then any* NFA *for L has $\geq k$ states.*

2. *Any* NFA *for* $\mathrm{MN}(n, \ldots, n+k-1)$ *has* $\geq k$ *states. (This follows from part 1.)*

Proof:

Let q_0 be the start state of an NFA that accepts L. Since $a^{n+k} \in L$ there is an accepting path of (not necessarily distinct) states

$$q_0, q_1, q_2, \ldots, q_n, q_{n+1}, \ldots, q_{n+k-1}, q_{n+k}$$

such that $\delta(q_0, a^i) = q_i$ and q_{n+k} is an accept state. Note that $q_n, \ldots, q_{n+k-1}$ are all reject states. We show they are all different. If there exists $n \leq i < j \leq n+k-1$ such that $q_i = q_j$ then the computation path

$$q_0, q_1, q_2, \ldots, q_n, \ldots, q_i, q_{j+1}, \ldots, q_{n+k-1}, q_{n+k}$$

accepts $a^{n+k-(j-i)}$. Since $1 \leq j-i \leq k-1$, $n+1 \leq n+k-(j-i) \leq n+k-1$. Hence $a^{n+k-(j-i)} \notin L$, which contradicts it being accepted. ∎

Theorem 8.4 *Any* NFA *for* $\mathrm{MN}(n, \ldots, n+k-1)$ *has* $\geq \max\{k, \sqrt{n}\}$ *states.*

Proof: This follows from Theorems 8.3 and 8.2. ∎

9 Open Problems

1) Open Problem I:

Let $0 < \epsilon < \delta < 1$ where $\delta > \frac{1}{2}$ and $\delta = 2\epsilon$. $A \subseteq \{n, \ldots, n + O(n^\delta)\}$ of size $O(n^\epsilon)$.

1. By Theorem 6.1 there exists an NFA for $\mathrm{MN}(A)$ with $\leq 2n^{\max\{1/2,\delta\}} + \tilde{O}(n^{\min\{2\epsilon,\delta\}}) = O(n^\delta)$ states.

2. By Theorem 8.3 any NFA for $\mathrm{MN}(A)$ has $\geq \Omega(n^{1/2})$ states.

Since $1/2 < \delta$ there is a gap between he upper and lower bounds on the number of states for the minimal NFA for $\mathrm{MN}(A)$. We would like to narrow or close the gap between upper and lower bounds. We tend to think the upper bounds are close to optimal.

II) Open Problem II:

Consider MN(1000). We know there is an NFA of size 59. By Theorem 8.3 any NFA is of size ≥ 33. It would be of interest to narrow the gap in this and other concrete cases.

10 Acknowledgment

We thank Jeffrey Shallit because (1) his slides [6] on the Frobenius problem, which included the folklore theorem that MN(n) has an NFA with $O(\sqrt{n})$ states, inspired our paper, (2) he emailed us a proof of Theorem 8.2, and (3) he proofread an early draft of the paper.

References

[1] J. R. Alfonsin. *The diophantine Frobenius problem.* Oxford University Press, Oxford, 2006.

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

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

[4] J.-P. Massias, J.-L. Nicolas, and G. Robin. Effective bounds for the maximal order of an element in the symmetric group. *Mathematics of Computation*, 53:665–678, 1989. `http://math.univ-lyon1.fr/~nicolas/gdenMathComp.pdf`.

[5] J. B. Rosser and L. Schoenfeld. Approximate formulas for some functions of prime numbers. *Illinois Journal of Mathematics*, 6:64–94, 1962.

[6] J. Shallit. The Frobenius problem and its generalizations. `https://cs.uwaterloo.ca/~shallit/Talks/frob6.pdf`.