International Journal of Foundations of Computer Science

IMPROVED BOUNDS ON THE NUMBER OF AUTOMATA ACCEPTING FINITE LANGUAGES*

MICHAEL DOMARATZKI
School of Computing, Queen's University, Kingston, Ontario
K7L 3N6, Canada. Email: domaratz@cs.queensu.ca

Received (received date)
Revised (revised date)
Communicated by Editor's name

ABSTRACT

We improve the known bounds on the number of pairwise non-isomorphic minimal deterministic finite automata (DFAs) on n states which accept finite languages. The lower bound constructions are iterative approaches which yield recurrence relations.

1. Introduction

The study of enumeration of automata has a long history, including classic papers by Harrison [4], Radke [7], Robinson [8], Liskovets [6] and many papers by Korshunov in Russian, including the survey [5]. For a current list of references see Domaratzki *et al.* [3].

However, little work has been done on enumerating automata which accept finite languages. In this paper, we extend the work of Domaratzki *et al.* [3] by improving the upper and lower bounds on the number of finite languages accepted by deterministic finite automata (DFAs) with n states.

The author [1] has recently presented an extension to the Genocchi numbers which gives an upper bound on the number of finite languages accepted by DFAs with n states. This upper bound is not perfect, however, since it is a labeled enumeration, whereas what we seek is an unlabeled enumeration of the automata.

In the upper bound presented here, we revise the constructions of the previous work, and achieve a better upper bound. Unlike the previous work, however, the new construction does not parallel any studied sequences of numbers.

2. Definitions

We review the basic definitions of automata theory. For additional background, or any unexplained notions, please consult Yu [9]. Let Σ be a finite alphabet of

*A preliminary version of this paper appeared at DLT 2002 [2].

symbols, called *letters*. Then Σ^* is the set of all *words* over Σ, which are finite sequences from Σ. The empty word is denoted by ϵ. The length of a word $w = a_1 a_2 \cdots a_n \in \Sigma^*$ (where $a_i \in \Sigma$ for all $1 \le i \le n$), denoted $|w|$, is n. Note that ϵ is the unique word of length 0. A language is any subset of Σ^*. We will focus in this paper on finite languages, that is, those whose cardinality is finite.

A deterministic finite automaton (DFA) is a 5-tuple $M = (Q, \Sigma, \delta, q_0, F)$ where Q is a set of states, Σ is an alphabet, δ is a transition function $\delta : Q \times \Sigma \to Q$. The start state is $q_0 \in Q$ and the set of final states is $F \subseteq Q$. We extend δ to a function from $Q \times \Sigma^*$ to Q in the usual way: for all $q \in Q$, we set $\delta(q, \epsilon) = q$ and for all $q \in Q$, $w \in \Sigma^*$ and $a \in \Sigma$, we set

$$\delta(q, wa) = \delta(\delta(q, w), a).$$

A DFA is said to be *complete* if $\delta(q, a)$ is defined for all pairs $(q, a) \in Q \times \Sigma$. In what follows, we will assume that all DFAs are complete.

A word $w \in \Sigma^*$ is accepted by a DFA $M = (Q, \Sigma, \delta, q_0, F)$ if $\delta(q_0, w) \in F$. Define the language accepted by a DFA M by

$$L(M) = \{w \in \Sigma^* \ : \ \delta(q_0, w) \in F\}.$$

A DFA $M = (Q, \Sigma, \delta, q_0, F)$ is minimal for a language L if $L(M) = L$ and for all DFAs $M' = (Q', \Sigma, \delta', q'_0, F')$ with $L(M') = L$, we have $|Q| \le |Q'|$.

Let $[[P]]$ denote the boolean value of the expression P, i.e. $[[P]] \in \{0, 1\}$ and $[[P]] = 1$ iff P is true. We use the symbol $\triangle$ to denote symmetric difference of sets. That is, for all sets S_1, S_2, we define $S_1 \triangle S_2 = (S_1 \setminus S_2) \cup (S_2 \setminus S_1)$. For the alphabet $\Sigma = \{a, b\}$, we define the morphism $\bar{\cdot} : \Sigma \to \Sigma$ by the rules $\overline{a} = b$ and $\overline{b} = a$.

3. Auxiliary Results

We first state and prove two auxiliary lemmas which will be useful for demonstrating that DFAs are minimal:

Lemma 1 *Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA such that $L(M)$ is finite. Let $q_1, q_2 \in Q$ be such that*

(i) state q_1 is reachable from the initial state;
(ii) from state q_2, a final state is reachable;
(iii) there exists a word $w \in \Sigma^+$ such that $\delta(q_1, w) = q_2$.

Then there exists another word $x \in \Sigma^*$ such that $\delta(q_1, x) \in F$ and $\delta(q_2, x) \notin F$ or vice-versa.

Proof. Assume that no such word x exists. In particular,

$$\delta(q_1, x) \in F \iff \delta(q_2, x) \in F$$

for all word $x \in \Sigma^*$ (i.e., the states are equivalent). Then let $u \in \Sigma^*$ be a word such that $\delta(q_0, u) = q_1$ and $z \in \Sigma^*$ be a word such that $\delta(q_2, z) \in F$. Then note that $\delta(q_0, uwz) \in F$. Now consider $\delta(q_1, wz) \in F$. Thus, $\delta(q_2, wz) \in F$ also, by assumption, and so $\delta(q_0, uwwz) \in F$. Assume that $\delta(q_0, uw^i z) \in F$ for some $i > 1$.

Then as $\delta(q_0, u) = q_1$, and thus $\delta(q_1, w^i z) \in F$, this implies $\delta(q_2, w^i z) \in F$, by assumption. Thus we may conclude that $\delta(q_0, uww^{i+1}z) \in F$ and so $L(M)$ contains uw^+z. As $L(M)$ is finite, this is a contradiction, as $w \neq \epsilon$. □

The following fact is easily proven:

Fact 1 *Let $M = (Q, \Sigma, \delta, q_0, F)$ be a DFA such that $L(M)$ is finite. Let $q_d \in Q$ be a reachable state such that $\delta(q_d, a) = q_d$ for all $a \in \Sigma$. Then for all $q \in Q \setminus \{q_d\}$, if from q a final state is reachable, there exists a word w such that $\delta(q, w) \in F$ and $\delta(q_d, w) \notin F$.*

4. A Lower Bound

For all $k \geq 1$, let Σ_k be an alphabet of size k, given by $\Sigma_k = \{a_1, a_2, \ldots, a_k\}$.

Definition 1 *Let $k \geq 1$. Let $f_k(n)$ denote the number of non-isomorphic minimal DFAs on n states over a k-letter alphabet which accept finite languages.*

The best known lower bound on $f_k(n)$ is $f_k(n) \geq 2^{n-2}((n-1)!)^{k-1}$, given by Domaratzki *et al.* [3].

We define the function $s_k(n)$ for all $k \geq 2$ as follows:

$$\begin{aligned} s_k(2) &= 1; \\ s_k(n) &= 2((n-1)^k - (n-2)^k) s_k(n-1). \end{aligned}$$

Our goal will be to establish the following lower bound:

Theorem 1 *For all $k \geq 2$ and $n \geq 2$, we have $f_k(n) \geq s_k(n)$.*

To do this, we will describe a set of DFAs $S_{n,k}$ with $|S_{n,k}| = s_k(n)$. Each DFA in $S_{n,k}$ will have state set $\{1, 2, \ldots, n-1, d\}$ ($d \notin \{1, 2, \ldots, n-1\}$ will represent our "dead state") and will accept a different finite language. We let Q_n denote the state set $Q_n = \{1, 2, \ldots, n-1, d\}$.

We define $S_{n,k}$ recursively as follows, beginning with $S_{2,k}$:

$$S_{2,k} = \{M_{2,k} = (Q_2, \Sigma_k, \delta, 1, \{1\})\}$$

with δ defined by $\delta(1, a) = \delta(d, a) = d$ for all $a \in \Sigma_k$. Thus, $M_{2,k}$ accepts the language $\{\epsilon\}$. Now, let $n \geq 2$ and given the set $S_{n,k}$, define the set $S_{n+1,k}$ as follows: For each $M = (Q_n, \Sigma_k, \delta, n-1, F) \in S_{n,k}$ add a new state n. For each possible k-tuple $(i_1, i_2, \ldots, i_k) \in Q_n^k$ (subject to the condition that $i_j = n-1$ for at least one j with $1 \leq j \leq k$), we create two DFAs $M_1 = (Q_{n+1}, \Sigma_k, \delta', n, F_1)$, $M_2 = (Q_{n+1}, \Sigma_k, \delta', n, F_2) \in S_{n+1,k}$. The transition function δ' is inherited from M along with our added transitions for state n:

$$\begin{aligned} \delta'(n, a_j) &= i_j; \\ \delta'(i, a) &= \delta(i, a) \quad \forall i \in Q_n, a \in \Sigma_k. \end{aligned}$$

Finally, $F_1 = F$ and $F_2 = F \cup \{n\}$. Note that by the inheritance of final states, for each $M \in S_{n,k}$, the state $1 \in F$, while $d \notin F$.

We may first state two basic lemmas about DFAs in $S_{n,k}$ for $n \geq 2$.

Lemma 2 *Let $M = (Q_n, \Sigma_k, \delta, n-1, F) \in S_{n,k}$. Then for each state $i \in Q_n \setminus \{d\}$, there exists a word w_i such that $|w_i| = n-1-i$ and $\delta(n-1, w_i) = i$. For $i = d$, there exists a word w_d such that $|w_d| = n-1$ and $\delta(n-1, w_d) = d$.*

Proof. The proof is by induction on n. The base case $n = 2$ is easily verified. Now let the result hold for $n-1$. We prove it for $S_{n,k}$. Let $M \in S_{n,k}$ be arbitrary. By construction, let $M' \in S_{n-1,k}$ be the DFA from whom M inherited its structure on the states in Q_{n-1}. Thus, by induction, for any state $i \in Q_{n-1}$, there is some word w_i such that $|w_i| = n-2-i$ and $\delta'(n-2, w_i) = i$ in M'. Now, assume that $\delta(n-1, a) = n-2$ for some $a \in \Sigma_k$ (we insisted that one of the $a \in \Sigma_k$ satisfied $\delta(n-1, a) = n-2$). Since δ is inherited from δ' we see that $\delta(n-1, aw_i) = i$ for all states $i \in Q_{n-1}$ and $|aw_i| = n-2-i+1 = n-1-i$. Clearly, $\delta(n-1, \epsilon) = n-1$, so the lemma is proven. □

Lemma 3 *Let $M = (Q_n, \Sigma_k, \delta, n-1, F) \in S_{n,k}$. Then for each state $i \in Q_n \setminus \{d\}$, there exists a word x_i such that $|x_i| = i-1$ and $\delta(i, x_i) = 1$.*

Proof. The proof is again by induction on n. The base case $n = 2$ is easily verified. Now, let the result hold for $n-1$. We prove it for $S_{n,k}$. Let $M \in S_{n,k}$ be arbitrary and let $M' \in S_{n-1,k}$ be the DFA from whom M inherited its structure on states Q_{n-1}. Clearly, by Lemma 2, there is some word w_1 of length $n-2$ such that $\delta(n-1, w_1) = 1$. Thus, this word satisfies the conditions we require of x_{n-1}. For $s \neq n-1$, by induction, there is some word x'_s of length $s-1$ such that $\delta'(s, x'_s) = 1$ in M'. But by the inheritance of transitions from M', the same word x'_s will satisfy $\delta(s, x'_s) = 1$ in M. Thus, the lemma is proven. □

These tools will prove sufficient for establishing the following lemma:

Lemma 4 *Let $M_1 \neq M_2 \in S_{n,k}$. Then $L(M_1) \neq L(M_2)$.*

Proof. The result is by induction on n. For $n = 2$, $|S_2| = 1$ so the result is trivially true. Let $M_1 = (Q_n, \Sigma_k, \delta_1, n-1, F_1)$ and $M_2 = (Q_n, \Sigma_k, \delta_2, n-1, F_2)$. Further, let $M'_1, M'_2 \in S_{n-1,k}$ be the DFA from which M_1 and M_2 inherited their structure, respectively. Let $M'_1 = (Q_{n-1}, \Sigma_k, \delta'_1, n-2, F'_1)$ and $M'_2 = (Q_{n-1}, \Sigma_k, \delta'_2, n-2, F'_2)$. There are three cases:

(i) There exists $a \in \Sigma_k$ such that $\delta_1(n-1, a) \neq \delta_2(n-1, a)$: Assume without loss of generality that $\delta_1(n-1, a) = i_1$ and $\delta_2(n-1, a) = i_2$ with $i_1 > i_2$. By Lemma 3, there is some word x of length $i_1 - 1$ such that $\delta_1(i_1, x) = 1$. But since $i_1 > i_2$, we must have that $\delta_2(i_2, x) = d$. Thus $ax_1 \in L(M_1) \setminus L(M_2)$.

(ii) $\delta_1(n-1, a) = \delta_2(n-1, a)$ for all $a \in \Sigma_k$ but $M'_1 \neq M'_2$: By induction $L(M'_1) \neq L(M'_2)$. Thus, let $x \in L(M'_1) \triangle L(M'_2)$. Since by definition, there exists $a \in \Sigma_k$ such that $\delta_1(n-1, a) (= \delta_2(n-1, a)) = n-2$, we must have that one of $ax \in L(M_1) \triangle L(M_2)$.

(iii) $\delta_1(n-1, a) = \delta_2(n-1, a)$ for all $a \in \Sigma_k$ and $M'_1 = M'_2$: Since $M_1 \neq M_2$ we must have $n-1 \in F_1 \triangle F_2$. Thus, the word $\epsilon \in L(M_1) \triangle L(M_2)$.

This establishes the result. □

Thus, all M in $S_{n,k}$ are pairwise non-isomorphic. It remains to be shown that each M is minimal. This is established by the following lemma:

Lemma 5 *Let $M = (Q_n, \Sigma_k, \delta, n-1, F) \in S_{n,k}$. Then M accepts a word of length $n-2$, but no word of length $n-1$ or greater.*

Proof. The fact that M accepts a word of length $n-2$ follows directly from Lemma 2, on taking $i = 1$.

Now, we prove by induction that the longest word accepted by $M \in S_{n,k}$ is $n-2$. For $n = 2$ and all k, the result holds by observation. Assume that it is true for all integers less than n. Let k be arbitrary. Assume that there exists a DFA $M = (Q_n, \Sigma_k, \delta, n-1, F) \in S_{n,k}$ such that there is a word $x \in L(M)$ with $|x| > n-2$. Let $x = cx'$ for some $c \in \Sigma_k$. Let $M' = (Q_{n-1}, \Sigma_k, \delta', n-2, F')$ be the DFA from which M inherited its structure. Then there are two cases:

(i) $\delta(n-1, c) = n-2$: Then note that $x' \in L(M')$ where $M' \in S_{n-1,k}$. So by induction $|x'| \le n-3$. Thus, $|x| = 1 + |x'| \le n-2$, contradicting our choice of x.

(ii) $\delta(n-1, c) \ne n-2$: Then let $\delta(n-1, c) = i$ for some state $i \in Q_{n-2}$. If $i = d$ then clearly $\delta(n-1, cx') \notin F$. Thus M does not accept $x = cx'$, contrary to our assumption. Thus $i \in Q_{n-2} \setminus \{d\}$. By applying Lemma 2 to M', there exists a word w_i such that $|w_i| = (n-2) - i = n-i-2$ and $\delta(n-2, w_i) = i$. Thus $w_i x'$ is accepted by M'. However $|w_i x'| = |w_i| + |x'| > n-i-2+n-3 = 2n-i-5$. As $i \le n-3$, $2n-i-5 \ge n-2$. Thus $|w_i x'| > n-2$. But by induction, the longest word accepted by M' is of length $n-3$. This is a contradiction.

This proves the lemma. □

Thus, we note that M is necessarily minimal, since if it were not, then a DFA M' with less than n states would accept the same language. But then M' accepts a word of length $n-2$ with less than n states. This is impossible if M' accepts a finite language.

We now established all of Theorem 1 except for computing the size of $S_{n,k}$. We can establish the following recurrence: Let $s_k(n) = |S_{n,k}|$. We already know that $s_k(2) = 1$ (we may also set $s_k(1) = 1$ by letting $S_{1,k}$ contain only the DFA whose language accepted is the empty set). Now, for each choice of a DFA M' in $S_{n-1,k}$, we have $(n-1)^k - (n-2)^k$ different choices for the transitions $\delta(n-1, a)$ for all $a \in \Sigma_k$ and subject to our constraints, since there are $(n-1)^k$ ways of setting the k transitions in the set Q_{n-1}, and $(n-2)^k$ ways of setting the k transitions in the set $Q_{n-1} \setminus \{n-2\}$. Further, we may assign either $n-1 \in F$ or not, giving us two more choices. This gives the recurrence:

$$s_k(n) = 2((n-1)^k - (n-2)^k) s_k(n-1) \quad \forall n \ge 3.$$

It is an easy inductive proof to show that $s_2(n) = (2n-3)!/(n-2)!$. Table 1 compares our lower bound to the old lower bound and the actual value for $k = 2$.

n	1	2	3	4	5	6	7
$2^{n-2}(n-1)!$	1	1	4	24	192	1920	23040
$(2n-3)!/(n-2)!$	1	1	6	60	840	15120	332640
$f_2(n)$	1	1	6	60	900	18480	487560

Figure 1: Our lower bound, compared with the actual counts, for $k = 2$. The values of $f_2(n)$ are from Domaratzki *et al.*

5. An Improvement

We may improve the results of Section 4 for the case $k = 2$. We set $\Sigma = \{a, b\}$. For each $n \geq 3$, we will define a set $T(n)$ of minimal, n-state DFAs. To do this, we define $T(n, i)$ for all $0 \leq i \leq \lfloor \frac{n-3}{2} \rfloor$, and set $T(n) = \cup_{i=0}^{\lfloor (n-3)/2 \rfloor} T(n, i)$.

The definition of the sets $T(n, i)$ is as follows: The first set $T(n, 0) = S_{n,2}$ for all $n \geq 3$ (we set $T(2, 0) = S_{2,2}$ as needed). From these sets, we will define each of the sets $T(n, i)$.

For $i < \lfloor \frac{n-3}{2} \rfloor$, define $T(n, i)$ recursively as follows:

(a) for each $M \in T(n-1, i)$, we define $2((n-1)^2 - (n-2)^2) = 4n - 6$ different DFAs in $T(n, i)$, as in the definition of $S_{n,2}$. Thus, let

$$M = (Q_{n-1}, \{a, b\}, \delta, n-2, F).$$

Then for each possible pair of states $(j, k) \in Q_{n-1}^2$ subject to the condition that at least one of $j, k = n - 2$, we create the DFAs

$$\begin{aligned} M' &= (Q_n, \{a, b\}, \delta', n-1, F') \\ M'' &= (Q_n, \{a, b\}, \delta', n-1, F'') \end{aligned}$$

where

$$\delta'(n-1, a) = j, \quad \delta'(n-1, b) = k,$$

and

$$\delta'(\ell, c) = \delta(\ell, c) \quad \forall c \in \{a, b\}, \ell \in Q_{n-1}.$$

Finally, $F' = F$, and $F'' = F \cup \{n-1\}$.

(b) for each $M \in T(n-2, i-1)$, we define $2(2n-5)(n-4)$ DFAs in $T(n, i)$. Let

$$M = (Q_{n-2}, \{a, b\}, \delta, n-3, F) \in T(n-2, i-1)$$

For any such M, define $D(M)$ to be the following set

$$D(M) = \{(\delta(q, a), \delta(q, b), [[q \in F]]) \ : \ q \in Q_{n-2}\}$$

Note that $D(M) \subseteq Q_{n-3}^2 \times \{0, 1\}$. Let D_n denote the set $D_n = Q_{n-3}^2 \times \{0, 1\}$. Now, $|D(M)| = n - 2$, but $|D_n| = 2(n-3)^2$. Thus, for all of the $2(n-3)^2 - (n-2) = (2n-5)(n-4)$ elements $(j, k, B) \in D_n - D(M)$, we define two new DFAs

$$\begin{aligned} M' &= (Q_n, \{a, b\}, \delta', n-1, F'), \\ M'' &= (Q_n, \{a, b\}, \delta', n-1, F''). \end{aligned}$$

where

$$\begin{aligned} \delta'(n-1, a) &= n-2, & \delta'(n-1, b) &= n-3, \\ \delta'(n-2, a) &= j, & \delta'(n-2, b) &= k, \end{aligned}$$

and

$$\delta'(\ell, c) = \delta(\ell, c) \quad \forall c \in \{a, b\}, \ell \in Q_{n-2}.$$

Finally, if $B = 1$ then set $F' = F \cup \{n-2\}$ and $F'' = F \cup \{n-2, n-1\}$. If $B = 0$, then set $F' = F$, $F'' = F \cup \{n-1\}$.

For $i = \lfloor \frac{n-3}{2} \rfloor$, we have two cases:

(i) If n is odd, we define $T(n,i)$ using the construction of type (b) only.
(ii) If n is even, $T(n,i)$ will be defined as usual, that is, with type (a) and (b).

The reason for this difference in definitions is since, for n odd, there is no set $T(n-1,\lfloor\frac{n-3}{2}\rfloor)$ from which to define automata in $T(n,\lfloor\frac{n-3}{2}\rfloor)$ by adding one state. Thus, we construct automata in this set using type (b) constructions only. For n even, both type (a) and (b) constructions are possible. We will continue to use the terms "type (a)" and "type (b)" to refer to the two different constructions used.

Note the following fact:

Fact 2 *For all DFAs $M \in T(n,i)$ for some $n \geq 3$ and $0 \leq i \leq \lfloor (n-3)/2 \rfloor$, $(d,d,0),(d,d,1) \in D(M)$. Thus, we never have $(d,d,B) \in D_n \setminus D(M)$ for any $B \in \{0,1\}$.*

To see this, note that in $M \in S_{2,2} = T(2,0)$, $D(M) = \{(d,d,0),(d,d,1)\}$. Further, note that if M is constructed from M', $D(M) \supset D(M')$.

We introduce the index i for our own use. We will see that the index i is irrelevant in the final analysis of the size of $T(n) = \cup_{i=0}^{\lfloor (n-3)/2 \rfloor} T(n,i)$. However, they will prove very useful in proving the lemmas which will establish our bound to be correct.

Lemma 6 *Let $M \in T(n,i)$ be arbitrary. Then for each $j \in Q_n$, there exists a word $w_j \in \{a,b\}^*$ such that $\delta(n-1,w_j) = j$. Further, if $j \neq n-1$, we may choose $w_j \neq \epsilon$.*

Proof. The proof is by induction on n. For $n \leq 4$, the result holds by Lemma 2, as $T(n,i) = T(n,0) = S_{n,2}$. Let the result hold for all integers less than n. Let $M = (Q_n,\{a,b\},\delta,n-1,F) \in T(n,i)$ and suppose that $M' = (Q_{n-1},\{a,b\},\delta',n-2,F') \in T(n-1,i)$ be the DFA from which M inherited its structure. Then by induction, for all states $j \in Q_{n-1}$ there is a word w_j such that $\delta'(n-2,w_j) = j$ in M'. But now by definition of our type (a) construction, there exists $c \in \{a,b\}$ such that $\delta(n-1,c) = n-2$. Thus, we simply let our word be cw_j for each state $j \in Q_{n-1}$. Further, $\delta(n-1,\epsilon) = n-1$.

For the case where a type (b) construction is used, suppose $M = (Q_n,\{a,b\},\delta,n-1,F) \in T(n,i)$ and that $M' = (Q_{n-2},\{a,b\},\delta',n-3,F') \in T(n-2,i-1)$ is the DFA from which M inherited its structure. Then again, for each $j \in Q_{n-2}$, there is a word w_j such that $\delta'(n-3,w_j) = j$. Since $\delta(n-1,b) = n-3$, we set our word to be bw_j for each $j \in Q_{n-2}$. The result now follows since $\delta(n-1,a) = n-2$ and $\delta(n-1,\epsilon) = n-1$. □

Lemma 7 *Let $M \in T(n,i)$ be arbitrary. Then for each $j \in Q_n \setminus \{d\}$ there exists a word $x_j \in \{a,b\}^*$ such that $\delta(j,x_j) = 1$.*

Proof. The proof is again by induction on n. For $n \leq 4$, the result holds by Lemma 3. Let the result hold for all integers less than n. If $M = (Q_n,\{a,b\},\delta,n-1,F) \in T(n,i)$ and suppose that $M' = (Q_{n-1},\{a,b\},\delta',n-2,F') \in T(n-1,i)$ is the DFA from which M inherited its structure, by induction, for each state $j \in Q_{n-1} \setminus \{d\}$, there is some word x_j such that $\delta'(j,x_j) = 1$ in M'. This word will also work for M, by our definition of M. By Lemma 6, there is some word w_1 such that $\delta(n-1,w_1) = 1$. This word will also work for our choice of x_{n-1}.

If $M = (Q_n, \{a,b\}, \delta, n-1, F) \in T(n,i)$, and $M' = (Q_{n-2}, \{a,b\}, \delta', n-3, F') \in T(n-2, i-1)$ is the DFA from which M inherited its structure, by induction, for each state $j \in Q_{n-2} \setminus \{d\}$, there is a word x_j such that $\delta'(j, x_j) = 1$. Again, this word will also work for M. By Lemma 6, there is some word w_1 such that $\delta(n-1, w_1) = 1$. This word will also work for our choice of x_{n-1}. Finally, let $(k_a, k_b) = (\delta(n-2, a), \delta(n-2, b))$. By Fact 2, $(k_a, k_b) \neq (d, d)$. Thus, let $c \in \{a, b\}$ such that $\delta(n-2, c) = k_c \neq d$. Then, let $x_2 = cx_{k_c}$. Clearly $\delta(n-2, x_2) = 1$. This completes the proof. □

We now prove bounds on the length of words accepted by DFAs in $T(n,i)$. We first introduce the following notation: For any DFA M which accepts a finite language, let $m(M) = \max\{|w| \ : \ w \in L(M)\}$.

Lemma 8 *Let $n \geq 3$ and $0 \leq i \leq \lfloor \frac{n-3}{2} \rfloor$. Let $M = (Q_n, \{a,b\}, \delta, n-1, F) \in T(n,i)$ be constructed from $M' = (Q_{n-1}, \{a,b\}, \delta', n-2, F') \in T(n-1, i)$. Then $m(M) = m(M') + 1$. The same result also holds if $M \in T(n,i)$ is constructed from $M' = (Q_{n-2}, \{a,b\}, \delta', n-3, F') \in T(n-2, i-1)$.*

Proof. The proof is essentially by observation of the type (a) and (b) constructions. Suppose $M = (Q_n, \{a,b\}, \delta, n-1, F) \in T(n,i)$ is constructed from $M' = (Q_{n-1}, \{a,b\}, \delta', n-2, F') \in T(n-1,i)$. Let $w \in L(M')$ be such that $|w| = m(M')$. Further, let $c \in \{a,b\}$ be such that $\delta(n-1, c) = n-2$. Then by definition, $\delta(n-1, cw) \in F$ and so $cw \in L(M)$. Thus $m(M) \geq m(M') + 1$.

Suppose there is a word x such that $|x| > m(M') + 1$ and $x \in L(M)$. Let $x = cx'$ for some $c \in \{a,b\}$. Suppose $\delta(n-1, c) = n-2$. Then $\delta(n-2, x') \in F$ and by the type (a) construction, $x' \in L(M')$. Thus $m(M') < |x'| \leq m(M')$. This is a contradiction.

Now suppose $\delta(n-1, c) = j \neq n-2$ Then $\delta(j, x') \in F$. By Lemma 6 there is some word y such that $\delta(n-2, y) = j$. Thus $yx' \in L(M')$. So $|yx'| = |y| + |x'| \leq m(M')$. But now $|x'| > m(M')$ and $|y| \geq 0$ implies $|y| + |x'| > m(M')$. This again is a contradiction. Thus $m(M) = m(M') + 1$.

We now treat case (b) constructions, where $M = (Q_n, \{a,b\}, \delta, n-1, F) \in T(n,i)$ is constructed from $M' = (Q_{n-2}, \{a,b\}, \delta', n-3, F') \in T(n-2, i-1)$. Let $w \in L(M')$ be such that $|w| = m(M')$. Now, $\delta(n-1, b) = n-3$. Thus $\delta(n-1, bw) \in F$ and so $bw \in L(M)$. Thus $m(M) \geq m(M') + 1$.

Suppose there is a word x such that $|x| > m(M') + 1$ and $x \in L(M)$. Then either $x = ax'$ or $x = bx'$. If $x = bx'$ then $\delta'(n-3, x') \in F$ and so $|x'| \leq m(M')$, by our type (b) construction. Contradiction.

Now, if $x = ax'$ then $\delta(n-2, x') \in F$. Let $x' = cx''$ for some $c \in \{a,b\}$. Then let $\delta(n-2, c) = j$. Note that by our type (b) construction (as $(\delta(n-2,a), \delta(n-2,b)) \in Q_{n-3}^2$ and $n-3 \notin Q_{n-3}$), $j \neq n-3$. By Lemma 6, there exists a word $y \neq \epsilon$ such that $\delta(n-3, y) = j$. Thus $yx'' \in L(M')$ and thus $|x''| + |y| \leq m(M')$. But also $|x''| > m(M') - 1$ and $|y| \geq 1$. Thus $|x''| + |y| > m(M')$. Contradiction. □

Theorem 2 *For all $n \geq 3$ and $0 \leq i \leq \lfloor (n-3)/2 \rfloor$, and for all $M \in T(n,i)$,*

$$m(M) = n - i - 2.$$

Proof. The proof is by induction on i. For $i = 0$, $T(n, 0) = S_{n,2}$. Thus, the result follows by Lemma 5. Assume the result hold for all integers less than i. Note that $\min\{n : T(n, i) \text{ is defined. }\} = 2i+3$. Thus, we first consider $T(2i+3, i)$. Let $M \in T(2i + 3, i)$. For this DFA, we must have that M was constructed from $M' \in T(2i+1, i-1)$ by a type (b) construction. By induction $m(M') = (2i + 1) - (i - 1) - 2 = i$. By Lemma 8, $m(M) = i + 1 = (2i + 3) - i - 2$. Thus, the result holds for $n = 2i + 3$.

There are now two possibilities for each $M \in T(n, i)$ for arbitrary $n > 2i+3$.

(i) $M \in T(n, i)$ was constructed from $M' \in T(n - 2, i - 1)$. By induction (on i), $m(M') = n-2-(i-1)-2 = n-i-3$. Thus by Lemma 8, $m(M) = n-i-2$.

(ii) $M \in T(n, i)$ was constructed from $M' \in T(n - 1, i)$. By induction (this time on n), $m(M) = n-1-i-2 = n-i-3$. Thus by Lemma 8, $m(M) = n-i-2$.

This completes the proof. □

There are two important corollaries of Theorem 2:

Corollary 1 *Let $n \geq 3$ be fixed. For all $0 \leq i_1 < i_2 \leq \lfloor (n-3)/2 \rfloor$, $T(n, i_1) \cap T(n, i_2) = \emptyset$.*

Proof. Let $M_1 \in T(n, i_1)$ and $M_2 \in T(n, i_2)$. Then $L(M_1)$ contains a word of length $n-i_1-2$. This word cannot be accepted by M_2 if $i_1 < i_2$, since the maximum length word M_2 can accept is $n - i_2 - 2 < n - i_1 - 2$. Thus $L(M_1) \neq L(M_2)$ and $T(n, i_1) \cap T(n, i_2) = \emptyset$. □

Corollary 2 *Let $n \geq 3$, $0 \leq i \leq \lfloor (n-3)/2 \rfloor$. Let $M \in T(n, i)$ be arbitrary. Then for all $j \in Q_n \setminus \{n-1\}$, the following holds:*

$$\max\{|y| \ : \delta(j, y) \in F\} < n - i - 2$$

Proof. Let $M \in T(n, i)$. Suppose the result does not hold. Let $j \in Q_n \setminus \{n-1\}$ and y_j be a word such that $|y_j| \geq n - i - 2$ and $\delta(j, y_j) \in F$. Then by Lemma 6, there exists a word $x_j \neq \epsilon$ such that $\delta(n - 1, x_j) = j$. Thus $x_j y_j \in L(M)$. But $|x_j y_j| \geq n - i - 1$. This contradicts Theorem 2. □

We now establish the minimality of each DFA.

Theorem 3 *Let $n \geq 1$ and $0 \leq i \leq \lfloor (n-3)/2 \rfloor$. Then for all $M \in T(n, i)$, M is minimal.*

Proof. The proof is by induction on i. For $i = 0$, $T(n, 0) = S_{n,2}$. Thus, the result holds. Assume that the result holds for all integers less than i. Now, consider i and $n = 2i + 3$. We first argue that all the DFAs in $T(2i + 3, i)$ are minimal. Note that for each $M = (Q_{2i+3}, \{a, b\}, \delta, 2i + 2, F) \in T(2i + 3, i)$, M must have been constructed from a DFA $M' = (Q_{2i+1}, \{a, b\}, \delta', 2i, F') \in T(2i + 1, i - 1)$. By induction, M' is minimal. Now, we argue that M is minimal. The state $2i+2$ cannot be equivalent to any other state in Q_{2i+1} or to state $2i + 1$ by Lemma 1, Fact 1 and Lemma 6. Now, consider the state $2i + 1$. Let $(k_1, k_2) = (\delta(2i + 1, a), \delta(2i + 1, b))$. There are two cases here:

(i) for all states $j \in Q_{2i+1}$, $(\delta(j, a), \delta(j, b)) \neq (k_1, k_2)$. Thus, for every state j, let $c_j \in \{a, b\}$ be the letter (dependent on j) such that $\delta(j, c_j) \neq \delta(2i + 1, c_j)$. We claim the state $\ell_1 = \delta(j, c_j)$ is not equivalent to $\ell_2 = \delta(2i+1, c_j)$. Clearly,

$\ell_1, \ell_2 \in Q_{2i+1}$ are not equivalent in M' by induction. But now the construction of M only involves adding states $2i+1$,$2i+2$, which cannot be reached by the states in Q_{2i+1}, and the transitions and finality of states $2i+1$ and $2i+2$. But clearly after these additions ℓ_1, ℓ_2 are still not equivalent, as we have not changed any properties of them, or any state reachable from them.
Thus, as ℓ_1 and ℓ_2 are not equivalent, $2i+1$ and j are also not equivalent.

(ii) there exists some state $j \in Q_{n-2}$ such that $\delta(j, a) = k_1$ and $\delta(j, b) = k_2$. Then $(k_1, k_2, [[j \in F']]) \in D(M')$. Thus, by our construction algorithm, in order for M to have been added to $T(2i+3, i)$, we must have $j \in F \iff 2i+1 \notin F$. Thus, the word ϵ witnesses the fact that j and $2i+1$ are distinct.

We now prove, by induction on n, that for all $n \geq 2i+3$, each DFA $M \in T(n, i)$ is minimal. Let the result hold for all integers less than n. There are two cases:

(a) the DFA M was constructed from a DFA $M' \in T(n-2, i-1)$ using type (b) construction. In this case, by induction (on i), the DFA M' is minimal. We repeat the argument (i) above for arguing $M \in T(2i+3, i)$ is minimal, but instead arguing on $M \in T(n, i)$ instead of $M \in T(2i+3, i)$. The cases (i) and (ii) above hold.

(b) the DFA M was constructed from a DFA $M' \in T(n-1, i)$ using type (a) construction. In this case, the induction is on n, which implies M' is minimal. By use of Lemma 1, Fact 1 and Lemma 6, the added initial state $n-1$ will not be equivalent to each of the states in Q_{n-1} of M'.

Thus, for all $n \geq 2i+3$, each DFA $M \in T(n, i)$ is minimal. Thus, by induction on i, each DFA is minimal for any choice of $T(n, i)$. □

Our final goal is to prove that each $M \in T(n, i)$ is unique:

Theorem 4 *Let $n \geq 3$ and $0 \leq i \leq \lfloor (n-3)/2 \rfloor$. Then for all $M_1 \neq M_2 \in T(n, i)$, $L(M_1) \neq L(M_2)$.*

The proof of Theorem 4 is by induction on n. To prove this, we examine the following three cases. We prove each of the three cases simultaneously, but as separate lemmas:

(i) M_1, M_2 both arrived at from elements of $T(n-1, i)$ through type (a) construction.

(ii) M_1 arrived at through an element of $T(n-1, i)$ through a type (a) construction, M_2 arrived at through an element of $T(n-2, i-1)$ through a type (b) construction.

(iii) M_1, M_2 both arrived at from elements of $T(n-2, i-1)$ through type (b) constructions.

Of these, case (i) is proven through the exact same proof idea of Lemma 4, so we omit it.

We will first consider case (ii). This case will be where the introduction of the index i into $T(n, i)$ will yield us the most benefit. We prove the following lemma, which involves removing states. By that we mean deleting a state s, and all transitions leading out s. Then we also delete all states (potentially none) that are unreachable after removing s. Consider then the following argument: If we take a DFA $M \in T(n,i)$ arrived at through a type (a) construction, and remove state $n-1$, then we obviously arrive at a DFA $M' \in T(n-1,i)$. Similarly, if take a DFA $M \in T(n,i)$ arrived at through a type (b) construction and remove the states $n-1$ and $n-2$, then we will arrive at a DFA $M' \in T(n-2,i-1)$. However, what can we say about the DFA we get when we delete the states $n-1$ and $n-3$ from M? The following lemma says that we do in fact get back a DFA we have already enumerated.

This lemma gives us a reason why we fix the transitions out of state $n-1$ in a type (b) construction: $\delta(n-1,a) = n-2$ and $\delta(n-1,b) = n-3$. This is to avoid double counting. We now state our lemma, which our index i allows us to formalize easily:

Lemma 9 *Let $n \geq 5$ and $1 \leq i \leq \lfloor (n-3)/2 \rfloor$ and $M = (Q_n, \{a,b\}, \delta, n-1, F) \in T(n,i)$. Suppose M is constructed from a DFA $M' \in T(n-2,i-1)$. Then the DFA defined by deleting the states $n-1$ and $n-3$ and setting the initial state to $n-2$ is isomorphic to an element of $T(m,j)$ for some $m \leq n-2$ and $j \leq i-1$.*

Proof. Fix an arbitrary $n \geq 5$. The proof is then by induction on i. For $i = 1$, note that $M' \in T(n-2,0)$. Thus, each state in $j \in Q_{n-2} \setminus \{1,d\}$ was added one at a time through a type (a) construction to yield the DFA M' in $T(n-2,0)$.

Let $\delta(n-2,a) = i_1$ and $\delta(n-2,b) = i_2$. Let $i_0 = \max\{i_1,i_2\}$ and let $c \in \{a,b\}$ be such that $\delta(n-2,c) = i_0$. Thus, there was a point where i_0 was added as the initial state as a type (a) construction. Then we let $M'' \in T(i_0+2,0)$ be the DFA which agrees with M' on the states of Q_{i_0+1} and which satisfies $\delta(i_0+1,c) = i_0$ and $\delta(i_0+1,\overline{c}) = \min\{i_1,i_2\}$. It is clear that $M'' \in T(i_0+2,0)$ and M'' is isomorphic to the DFA obtained from M by deleting $n-1$ and $n-3$. Clearly, $i_0+2 \leq n-2$, so the result holds for $i = 1$.

Now, suppose the result is true for all integers less than i. Let $M \in T(n,i)$. Let $\delta(n-2,a) = i_1$, $\delta(n-2,b) = i_2$ and $i_0 = \max\{i_1,i_2\}$. There are three cases to consider when we remove the states $n-1$ and $n-3$ from M:

(i) The state i_0 was added during a type (a) construction: Then after removing $n-1$ and $n-3$, and all states which are then unreachable, we arrive at a DFA which is isomorphic to an element of $T(i_0+2,j)$ for some $j \leq i-1$. Thus, we are done.

(ii) The state i_0 was the added initial state during a type (b) construction: This case is identical to case (i); we are again left with a DFA isomorphic to an element of $T(i_0+2,j)$ for some $j \leq i-1$.

(iii) The state i_0 was the non-initial added state during a type (b) construction. Then, adding state i_0 resulted in a DFA $M' \in T(i_0+2,j)$ for some $j \leq i-1$. Note that in this case, the state i_0-1 will necessarily be unreachable after removing states $n-1$ and $n-3$. But now $j < i$ implies that by induction, there is some DFA $M'' \in T(m,\ell)$ which is isomorphic to the DFA obtained from removing states i_0+1 and i_0-1 from M'. Then M'' is also isomorphic to the DFA obtained from removing states $n-1$ and $n-3$ from M.

Thus, by induction, the result holds for all i. The lemma is proven. □

With Lemma 9 in hand, we may now prove case (ii) of Theorem 4:

Lemma 10 *Let $n \geq 5$ and $0 \leq i \leq \lfloor (n-3)/2 \rfloor$. Let $M_1, M_2 \in T(n,i)$ such that M_1 is arrived at through a type (a) construction from $M_1' \in T(n-1,i)$ and M_2 is arrived at through a type (b) construction from $M_2' \in T(n-2,i-1)$. Then $L(M_1) \neq L(M_2)$.*

Proof. Let $M_1 = (Q_n, \{a,b\}, \delta_1, n-1, F_1)$, $M_2 = (Q_n, \{a,b,\}, \delta_2, n-1, F_2)$, $M_1' = (Q_{n-1}, \{a,b\}, \delta_1', n-2, F_1')$, and $M_2' = (Q_{n-2}, \{a,b\}, \delta_2', n-3, F_2')$. We have two cases to consider:

(i) Suppose $\delta_1(n-1,b) = n-2$. By Theorem 3, the DFAs M_1' and M_2' are minimal. Since they have a different number of states, they must accept distinct languages. Namely, let $x \in L(M_1') \triangle L(M_2')$. Then note that $bx \in L(M_1) \triangle L(M_2)$.

(ii) Suppose $\delta_1(n-1,a) = n-2$. By Lemma 9, let $M_2'' \in T(m,j)$ be the DFA we get by deleting states $n-1$ and $n-3$ from M_2. Then $m \leq n-2$ and $j \leq i-1$. Since M_2'' has at most $n-2$ states and M_1' has $n-1$ states, by Theorem 3 again, the DFAs M_1' and M_2'' must accept different languages. Let $x \in L(M_1') \triangle L(M_2'')$. Then note that $ax \in L(M_1) \triangle L(M_2)$.

As one of these two cases must hold, this completes the proof. □

Finally, we prove case (iii) of Theorem 4:

Lemma 11 *Let $n \geq 5$ and $1 \leq i \leq \lfloor (n-3)/2 \rfloor$. Let $M_1 \neq M_2 \in T(n,i)$ such that both M_1, M_2 are constructed from $M_1', M_2' \in T(n-2, i-1)$ using a type (b) construction. Then $L(M_1) \neq L(M_2)$.*

Proof. The proof is by induction on n. The base case when $n = 5$ is easily verified.

(i) Suppose $M_1' \neq M_2'$. Then either by induction on n or by cases (i) or (ii) of Theorem 4, $L(M_1') \neq L(M_2')$. Let $x \in L(M_1') \triangle L(M_2')$. Then clearly $bx \in L(M_1) \triangle L(M_2)$.

(ii) Suppose $M_1' = M_2'$, but $(\delta_1(n-2,a), \delta_1(n-2,b)) \neq (\delta_2(n-2,a), \delta_2(n-2,b))$. By an application of Lemma 9, let M_1'' and M_2'' be the DFA we get by deleting states $n-1$ and $n-3$ from M_1 and M_2, respectively. There are three simple subcases:

  (a) $M_1'', M_2'' \in T(m,j)$ for the same $m \leq n-2$ and $j \leq i-1$. As the transitions leading out of the initial state (isomorphic to $n-2$) in M_1'' and M_2'' are different, $M_1'' \neq M_2''$. Thus again by induction n or by Theorem 4 cases (i) and (ii), as $m \leq n-2$, there is some word $x \in L(M_1'') \triangle L(M_2'')$. Then as the initial state in M_1'' and M_2'' is the same as state $n-2$ in M_1 and M_2, and $\delta_1(n-1,a) = \delta_2(n-1,a) = n-2$, then $ax \in L(M_1) \triangle L(M_2)$.

  (b) $M_1'' \in T(m_1, j_1)$,$M_2'' \in T(m_2, j_2)$ for $m_1 \neq m_2$. Since these DFAs are minimal and have different numbers of states, then there is some word $x \in L(M_1'') \triangle L(M_2'')$. Again, $ax \in L(M_1) \triangle L(M_2)$.

  (c) Finally, suppose that $M_1'' \in T(m, j_1)$, $M_2'' \in T(m, j_2)$ for $j_1 \neq j_2$. Then by Theorem 2, there exists $x \in L(M_1'') \triangle L(M_2'')$, namely x of length $m - \min\{j_1, j_2\} - 2$.

(iii) Suppose $M_1' = M_2'$, and $(\delta_1(n-2,a), \delta_1(n-2,b)) = (\delta_2(n-2,a), \delta_2(n-2,b))$. Then since $M_1 \neq M_2$, we must have that $([[n-1 \in F_1]], [[n-2 \in F_1]]) \neq$

$([[n-1 \in F_2]], [[n-2 \in F_2]])$. Thus, we must have one of $\epsilon, a \in L(M_1) \triangle L(M_2)$. This completes the proof. □

Combining the results of this section, we have established the following theorem:

Theorem 5 *Each DFA $M \in T(n)$ is minimal and accepts a distinct language.*

5.1. A Recurrence for the Size of $T(n)$

As we mentioned, the index i will bear no part in our estimation on the size of $T(n)$. Let $t(n) = |T(n)|$. We state the following recurrence:

Theorem 6 *The following recurrence holds for $t(n)$:*

$$\begin{aligned} t(n) &= 2(2n-3)t(n-1) + 2(2n-5)(n-4)t(n-2) \quad \forall n \geq 5 \\ t(n) &= (2n-3)!/(n-2)! \quad \textit{otherwise.} \end{aligned}$$

Proof. Let $t(n,i) = |T(n,i)|$. If n is odd, $\lfloor (n-3)/2 \rfloor = (n-3)/2$. Then we have that

$$t(n) = \sum_{i=0}^{(n-3)/2} t(n,i).$$

By the algorithm for constructing $T(n,i)$, we have that

$$\begin{aligned} t(n,0) &= 2(2n-3)t(n-1,0), \\ t(n,i) &= 2(2n-3)t(n-1,i) + 2(2n-5)(n-4)t(n-2,i-1), \\ t(n,(n-3)/2) &= 2(2n-5)(n-4)t(n-2,(n-5)/2), \end{aligned}$$

the middle equation being valid for all i with $1 \leq i \leq (n-3)/2 - 1 = (n-5)/2$. Thus, we write

$$\begin{aligned} t(n) &= \sum_{i=0}^{(n-5)/2} 2(2n-3)t(n-1,i) + \sum_{i=1}^{(n-3)/2} 2(2n-5)(n-4)t(n-2,i-1) \\ &= 2(2n-3)\sum_{i=0}^{(n-5)/2} t(n-1,i) + 2(2n-5)(n-4)\sum_{i=0}^{(n-5)/2} t(n-2,i). \end{aligned}$$

As $n-1$ is even, $\lfloor ((n-1)-3)/2 \rfloor = (n-5)/2$. Thus, $\sum_{i=0}^{(n-5)/2} t(n-1,i) = t(n-1)$. Similarly, $\lfloor ((n-2)-3)/2 \rfloor = (n-5)/2$ and so $\sum_{i=0}^{(n-5)/2} t(n-2,i) = t(n-2)$. Thus, we have

$$t(n) = 2(2n-3)t(n-1) + 2(2n-5)(n-4)t(n-2).$$

Consider now n even, which is very similar. In this case, $\lfloor (n-3)/2 \rfloor = (n-4)/2$. Thus,

$$t(n) = \sum_{i=0}^{(n-4)/2} t(n,i).$$

From the description of type (a) and (b) constructions, we have

$$\begin{aligned} t(n,0) &= 2(2n-3)t(n-1,0), \\ t(n,i) &= 2(2n-3)t(n-1,i) + 2(2n-5)(n-4)t(n-2,i-1), \end{aligned}$$

the last equation being valid for all i with $1 \leq i \leq (n-4)/2$. Thus, we write

$$\begin{aligned} t(n) &= \sum_{i=0}^{(n-4)/2} 2(2n-3)t(n-1,i) + \sum_{i=1}^{(n-4)/2} 2(2n-5)(n-4)t(n-2,i-1) \\ &= 2(2n-3)\sum_{i=0}^{(n-4)/2} t(n-1,i) + 2(2n-5)(n-4)\sum_{i=0}^{(n-6)/2} t(n-2,i). \end{aligned}$$

As $n-1$ is odd, $\lfloor((n-1)-3)/2\rfloor = (n-4)/2$. Thus, $\sum_{i=0}^{(n-4)/2} t(n-1,i) = t(n-1)$. Similarly, as $n-2$ is even, $\lfloor((n-2)-3)/2\rfloor = (n-6)/2$. Thus, $\sum_{i=0}^{(n-6)/2} t(n-2,i) = t(n-2)$. Thus, we may conclude that

$$t(n) = 2(2n-3)t(n-1) + 2(2n-5)(n-4)t(n-2).$$

This completes the proof. □

We demonstrate the effect of $t(n)$ compared to the lower bound in Figure 2.

n	1	2	3	4	5	6	7
$(2n-3)!/(n-2)!$	1	1	6	60	840	15120	332640
$t(n)$	1	1	6	60	900	17880	441960
$f_2(n)$	1	1	6	60	900	18480	487560

Figure 2: Our lower bounds for an alphabet of size 2, compared with the actual counts.

We now show that $c_1^{n-2} \leq t(n)/s_2(n) \leq c_2^{n-2}$ for constants $c_1 \simeq 1.066946710$ and $c_2 \simeq 1.207106781$.

Let $\{f_n\}_{n\geq 3}$ be the infinite sequence given by the linear recurrence

$$\begin{aligned} f_n &= 1 \quad \forall n < 5; \\ f_n &= f_{n-1} + \frac{1}{14} f_{n-2} \quad \forall n \geq 5. \end{aligned}$$

Then we can give a lower bound to the growth of $t(n)$ compared to $s_2(n) = (2n-3)!/(n-2)!$ as follows:

Lemma 12 *For all $n \geq 1$,*

$$t(n) \geq f_n \frac{(2n-3)!}{(n-2)!}.$$

Proof. The proof is by induction on n. For $n < 5$, the result holds since $f_n = 1$ and this agrees with our definition that $t(n) = (2n-3)!/(n-2)!$ for $n < 5$.

Let $n \geq 5$. Then by induction

$$\begin{aligned} t(n-1) &\geq f_{n-1}\frac{(2n-5)!}{(n-3)!}; \\ t(n-2) &\geq f_{n-2}\frac{(2n-7)!}{(n-4)!}. \end{aligned}$$

Then

$$\begin{aligned}
t(n) &= 2(2n-3)t(n-1) + 2(2n-5)(n-4)t(n-2) \\
&\geq 2(2n-3)f_{n-1}\frac{(2n-5)!}{(n-3)!} + 2(2n-5)(n-4)f_{n-2}\frac{(2n-7)!}{(n-4)!} \\
&= f_{n-1}\frac{(2n-3)!}{(n-2)!} + \frac{(n-4)}{2(2n-3)}f_{n-2}\frac{(2n-3)!}{(n-2)!} \\
&= (f_{n-1} + \frac{(n-4)}{2(2n-3)}f_{n-2})\frac{(2n-3)!}{(n-2)!}.
\end{aligned}$$

Thus, it suffices to show that $\frac{(n-4)}{2(2n-3)} \geq 1/14$ for all $n \geq 5$. But this is easily verified. □

As f_n obeys such a simple linear recurrence, we can estimate its growth as

$$f_n = \frac{\alpha_1^{n-2} - \alpha_2^{n-2}}{\alpha_1 - \alpha_2}$$

where $\alpha_1 = 1/2 + 3\sqrt{7}/14$ and $\alpha_2 = 1/2 - 3\sqrt{7}/14$ are the solutions to the quadratic equation $x^2 - x - 1/14 = 0$. Thus, we may conclude that $f_n \sim \alpha_1^{n-2}$, and

$$t(n) \in \Omega\left(\alpha_1^{n-2}\frac{(2n-3)!}{(n-2)!}\right).$$

Note that $\alpha_1 \simeq 1.06694$.

We may also give a similar upper bound on $t(n)$ compared to $(2n-3)!/(n-2)!$. Let $\{g_n\}_{n\geq 3}$ be given by the linear recurrence

$$\begin{aligned}
g_n &= 1 \quad \forall n < 5; \\
g_n &= g_{n-1} + \frac{1}{4}g_{n-2} \quad \forall n \geq 5.
\end{aligned}$$

Lemma 13 *For all $n \geq 1$,*

$$t(n) \leq g_n\frac{(2n-3)!}{(n-2)!}.$$

Proof. The proof is identical to the proof of Lemma 12, except for noting that $\frac{(n-4)}{2(2n-3)} \leq 1/4$ for $n \geq 5$. □

Solving for g_n explicitly, we get

$$g_n = \frac{\beta_1^{n-2} - \beta_2^{n-2}}{\beta_1 - \beta_2}$$

where $\beta_1 = (1 + \sqrt{2})/2$ and $\beta_2 = (1 - \sqrt{2})/2$ are the solutions to the equation $x^2 - x - 1/4 = 0$. Thus, $g_n \sim \beta_1^{n-2}$. Further,

$$t(n) \in O\left(\beta_1^{n-2}\frac{(2n-3)!}{(n-2)!}\right).$$

Note in this case $\beta_1 \simeq 1.2071068$.

6. An Upper Bound

In this section, we present our upper bound argument. We first review the work of the author in [1]. Consider the following iterative construction of a DFA accepting a finite language over the alphabet $\{a, b\}$: Let $M = (Q, \{a, b\}, \delta, q_0, F)$ be our DFA, with states $Q = \{q_0, q_1, \ldots, q_{n-1}\}$. We may assume that may assume that if $\delta(q_i, c) = q_j$ is a transition, then $i < j$ or $i = j = n - 1$ [3]. We would like to ensure that each DFA is initially connected, that is, that for each state q_i, there is some word w_i such that $\delta(q_0, w_i) = q_i$. To this end, we ensure that each state q_i has an edge entering it. Since the states are ordered, this must mean that $\delta(q_j, a) = q_i$ or $\delta(q_j, b) = q_i$ for some $j < i$. Since our automata are deterministic, we may assign $\delta(q_j, a)$ and $\delta(q_j, b)$ only once. Thus, an upper bound is to simply count the number of ways we may do this.

For state q_1, there are at most two transitions that may enter it: $\delta(q_0, a)$ and $\delta(q_0, b)$. Thus, we can make the assignment in $\sum_{i_1=1}^{2} \binom{2}{i_1}$ ways. In general, consider having done this for the previous $j - 1$ states, with k transitions remaining unassigned: then we have $\sum_{i_j=1}^{k} \binom{k}{i_j}$ ways of assigning at least one of the unassigned transitions to enter state q_j. This suggests defining the function $A_n(i, k)$ by

$$A_n(j, k) = \sum_{i=1}^{k} \binom{k}{i} A_n(j + 1, k - i + 2).$$

Thus, $A_n(j, k)$ denotes the number of ways of defining the transitions of states $q_j, q_{j+1}, \ldots, q_{n-1}$ knowing that we have k unassigned transitions from states q_0 through q_{j-1}. We must only determine our terminating condition. Since all unassigned transitions after adding states $q_0, q_1, \ldots, q_{n-2}$ must necessarily enter q_{n-1}, we get that $A_n(n, k) = 1$ for all k. Under these conditions, it is proven [1] that $A_n(2, 2)$ is equivalent to the ordinary Genocchi numbers.

However, this argument does not yield a tight upper bound. Consider Figure 3: the DFAs M_1 and M_2, which are both identical after state 3, will both be counted by our upper bound. The main idea of our new upper bound is to alleviate this

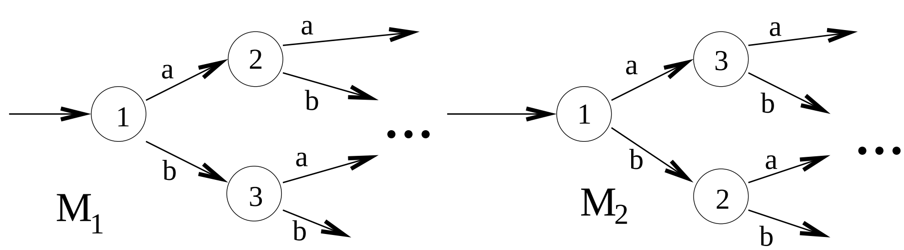

Figure 3: Problematic DFAs M_1 and M_2.

problem.

6.1. An Improved Upper Bound

The upper bound from the last section only considered adding single states at a time. However, we may also consider adding pairs of states. This way we can control the double counting exhibited in the last section. In this section, let $M = (\{1, 2, 3, \dots, n\}, \Sigma, \delta, 1, F)$.

Let c denote the size of our alphabet. Thus, we define a new function, call it $B_n^{(c)}(j, k)$, which will count the number of ways of defining the transitions of an n state DFA, having already defined states $1, \dots, j-1$ and having k extra transitions to assign to states $j, \dots, n$. Further, the underlying graph will guaranteed to be initially connected and acyclic, and we will add two states at a time.

We let $\binom{k}{k_1 \; k_2}$ for $k_1 + k_2 \leq k$ denote the multinomial coefficient given by $k!/(k_1!k_2!(k-k_1-k_2)!)$. The function $B_n^{(c)}(j, k)$ is defined for $n \geq 4$ by

$$\begin{aligned} B_n^{(c)}(j,k) &= \sum_{k_1=1}^{k} \sum_{k_2=0}^{k-k_1} \sum_{i=1}^{c} \binom{k}{k_1 \; k_2} \binom{c}{i} B_n^{(c)}(j+2, k-k_1-k_2-i+2c) \\ &+ \frac{1}{2} \sum_{k_1=1}^{k-1} \sum_{k_2=1}^{k-k_1} \binom{k}{k_1 \; k_2} B_n^{(c)}(j+2, k-k_1-k_2+2c); \\ B_n^{(c)}(n,k) &= 1 \quad \forall k \geq 0; \\ B_n^{(c)}(n-1,k) &= 2^k - 1 \quad \forall k \geq 0. \end{aligned}$$

Recall the definition of $f_c(n)$ from Definition 1. Then we have the following theorem:

Theorem 7 *Let $c, n \geq 2$ and $B_n^{(c)}$ be defined as above. Then $f_c(n) \leq 2^{n-2} B_n^{(c)}(2, c)$.*

Proof. Consider constructing an DFA M with states $1, 2, 3, \dots, n$. Assume we have added states $1, 2, \dots, j-1$ and k transitions from those states remain unassigned. We now add states j and $j+1$. There are two cases: There is a transition from j to $j+1$ or not.

Consider the first case. This is depicted in Figure 4. As we can see, we must

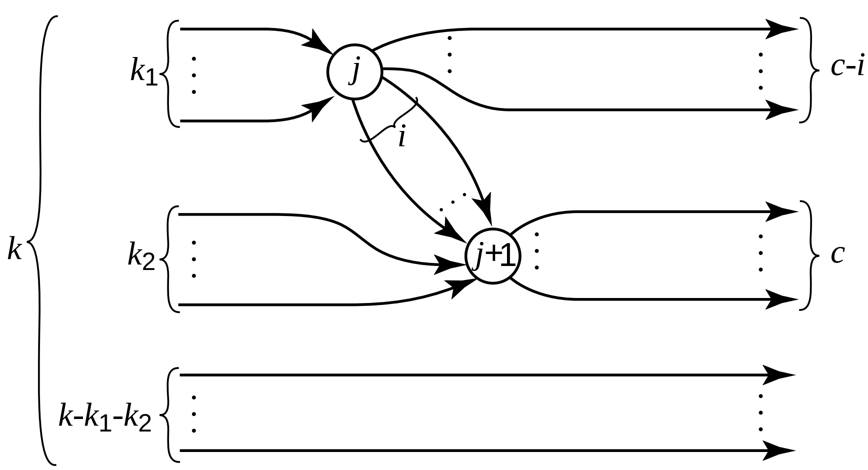

Figure 4: Case 1: adding two states which have a transition between them.

insist that $k_1 \neq 0$, since at least one transition must enter state j. Further, we insist that $i \neq 0$ so we are certain to be in case 1. Now note that there are $k - k_1 - k_2 - i + 2c$ transitions unassigned as we prepare to add state $j+2$. Summing over all possible choices of subsets of size k_1 and k_2 and subsets of size i, this gives the first term in the expression for $B_n^{(c)}(j,k)$:

$$\sum_{k_1=1}^{k} \sum_{k_2=0}^{k-k_1} \sum_{i=1}^{c} \binom{k}{k_1 \quad k_2} \binom{c}{i} B_n^{(c)}(j+2, k-k_1-k_2-i+2c).$$

If we are not in case 1, then certainly we are in case 2, depicted in Figure 5. In

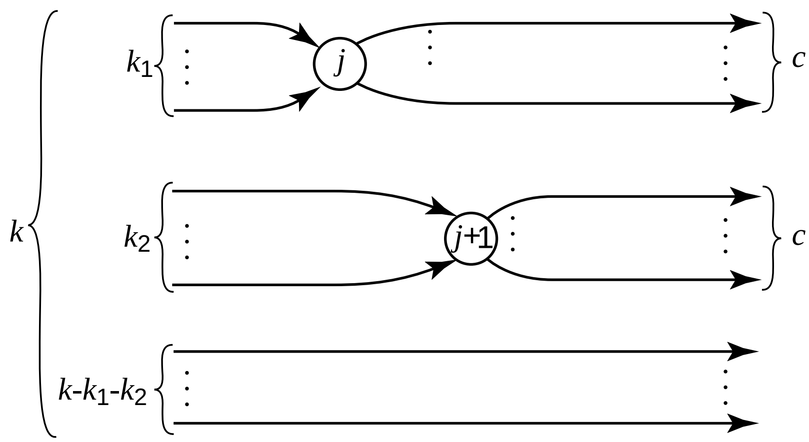

Figure 5: Case 2: adding two states which have no transitions between them.

this case, we insist that $k_1 \neq 0$ since there must be at least one transition entering j. Further, we insist that $k_2 \neq 0$ (and thus $k_1 \neq k$) because we must also ensure that at least one transition enters $j+1$. Now, clearly the number of transitions which are unassigned after $j+1$ is $k - k_1 - k_2 + 2c$. Summing over all possibilities gives

$$\sum_{k_1=1}^{k-1} \sum_{k_2=1}^{k-k_1} \binom{k}{k_1 \quad k_2} B_n^{(c)}(j+2, k-k_1-k_2+2c).$$

However, we now get to divide the entire expression by 2, to eliminate the double counting associated with renaming the states j and $j+1$. This gives our recurrence expression for $B_n^{(c)}(j,k)$.

The terminating conditions are as follows: $B_n^{(c)}(n,k) = 1$ since any unassigned transitions after adding states $1, 2, ..., n-1$ must enter state n. Further $B_n^{(c)}(n-1,k) = 2^k - 1$, since there are 2^k ways of distributing k transitions between states n and $n-1$. However, we must have at least one transition entering state $n-1$, which eliminates one of these 2^k possibilities.

Thus, our upper bound is $2^{n-2} B_n^{(c)}(2,c)$. The factor of 2^{n-2} comes from the possible assignments of the final states, subject to the fact that state n is not a final state, while state $n-1$ is final [3]. □

We may compare this upper bound for alphabets of size $c = 2$ to our old upper bounds [1], the best known lower bounds from Section 5 and the actual values [3] in Figure 6.

n	1	2	3	4	5	6	7
$f_2(n)$	1	1	6	60	900	18480	487560
$2^{n-2}B_n^{(2)}(2,2)$	1	1	6	64	1120	26432	889216
$2^{n-2}A_n(2,2)$	1	1	6	68	1240	33168	1223264

Figure 6: Summary of upper bounds.

Acknowledgments

I must thank Kai Salomaa and Galina Jirásková, who both read this paper and made many suggestions and corrections, and the anonymous referees, who made several useful comments. I am grateful to Alexander Okhotin for producing figures 4 and 5.

References

1. M. Domaratzki A generalization of the Genocchi numbers with applications to enumeration of finite automata. *Submitted (Preliminary version available at http://www.cs.queensu.ca/TechReports/Reports/2001-449.ps)* (2001).
2. M. Domaratzki, "Improved bounds on the number of automata accepting finite languages," in *DLT 2002: Developments in Language Theory, Sixth International Conference*, eds. M. Ito and M. Toyama (2002) pp. 159–181.
3. M. Domaratzki, D. Kisman and J. Shallit, "On the number of distinct languages accepted by finite automata with n states," *J. Automata, Languages and Combinatorics 7*, 4 (2002) 469–486.
4. M. Harrison, "A census of finite automata," *Can. J. Math 17* (1965), 100–113.
5. A. Korshunov, "Enumeration of finite automata" (In Russian), *Problemy Kibernetiki 34* (1978), 5–82.
6. V. Liskovets, "The number of connected initial automata," *Cybernetics 5* (1969), 259–262.
7. C. Radke, "Enumeration of strongly connected sequential machines," *Information and Control 8* (1965), 377–389.
8. R. Robinson, "Counting strongly connected finite automata," in *Graph Theory with Applications to Algorithms and Computer Science*, eds. Y. Alavi, G. Chartrand, D. Lick, and C. Wall, (Wiley, 1985) pp. 671–685.
9. S. Yu, "Regular languages," in *Handbook of Formal Languages, Vol. I*, eds. G. Rozenberg and A. Salomaa, (Springer-Verlag, 1997) pp. 41–110.