

DETERMINISTIC BLOW-UPS OF MINIMAL NFA'S*

GALINA JIRÁSKOVÁ¹

Abstract. The paper treats the question whether there always exists a minimal nondeterministic finite automaton of n states whose equivalent minimal deterministic finite automaton has α states for any integers n and α with $n \leq \alpha \leq 2^n$. Partial answers to this question were given by Iwama, Kambayashi, and Takaki (2000) and by Iwama, Matsuura, and Paterson (2003). In the present paper, the question is completely solved by presenting appropriate automata for all values of n and α . However, in order to give an explicit construction of the automata, we increase the input alphabet to exponential sizes. Then we prove that $2n$ letters would be sufficient but we describe the related automata only implicitly. In the last section, we investigate the above question for automata over binary and unary alphabets.

Mathematics Subject Classification. 68Q45, 68Q19.

INTRODUCTION

Finite automata are one of the simplest computational models. Many of their properties have been extensively studied. Nevertheless, some important problems concerning finite automata are still open. For instance, we recall the question of how many states are sufficient and necessary for two-way deterministic finite automata to simulate two-way nondeterministic finite automata. The problem is related to the famous open question whether or not DLOGSPACE equals to NLOGSPACE [3, 19].

In the last few years, a renewed interest for researchers in automata theory can be observed; for a discussion we refer to [9, 22]. Some aspects in this field are now intensively and deeply investigated. One such aspect is descriptonal

Keywords and phrases. Regular languages, deterministic finite automata, nondeterministic finite automata, state complexity.

* Research supported by the VEGA grant No. 2/3164/23.

¹ Mathematical Institute, Slovak Academy of Sciences, Grešákova 6, 040 01 Košice, Slovakia;
jiraskov@saske.sk

© EDP Sciences 2006

complexity which studies the costs of the description of languages by different formal systems. This paper treats the question of which kind of relations between the sizes of minimal nondeterministic finite automata and minimal deterministic finite automata are possible.

The subset construction [18] assures that 2^n states are sufficient for a DFA to simulate any NFA of n states. Several examples of worst-case blow-ups are known for a long time [15–17]. Another famous example with an exponential blow-up is the language $\{x1y \mid x \in \{0, 1\}^*, y \in \{0, 1\}^{n-1}\}$ which is accepted by an $(n + 1)$ -state NFA but cannot be accepted by any DFA with less than 2^n states. On the other hand, the DFA which counts the number of 1s modulo n needs n states and equivalent NFAs need the same number of states. Thus for some languages, the introduction of nondeterminism can help to reduce exponentially the number of states needed for describing the languages, while for other languages the nondeterminism is useless in such a reduction.

Iwama, Kambayashi, and Takaki [11] stated the question whether there always exists a minimal NFA of n states whose equivalent minimal DFA has α states for any integers n and α satisfying that $n \leq \alpha \leq 2^n$. The question was also considered by Iwama, Matsuura, and Paterson [12]. In these two papers, the authors succeeded in finding minimal n -state NFAs for some special values of α (for about $3n$ values of α between 2^{n-1} and 2^n).

In the present paper, the above question is completely solved by presenting appropriate n -state NFAs for all values of α between n and 2^n . However, in contrast to the previous papers, which discuss automata over a binary alphabet, in order to give an explicit construction of the NFAs, we increase the input alphabet to exponential sizes. Then we prove that $2n$ letters would be sufficient but we can describe the related NFAs only implicitly. In the case of a binary alphabet, we do not provide a complete solution but we complement the solutions given in [11] and [12] by presenting minimal n -state NFAs for further $n^2/2$ values of α . In the unary case, the deterministic blow-ups cannot be close to 2^n since it is known [4] that any unary n -state NFA can be simulated by an $O(F(n))$ -state DFA, where $F(n) \approx e\sqrt{n \cdot \ln(n)}$.

Note that in [11] and [12] the values of α were greater than 2^{n-1} , and so it was not necessary to give explicit proofs for the minimality of NFAs. This is not a case in this paper. Therefore, we introduce a technique based on simple communication complexity theoretical results to state sufficient conditions for the minimality of NFAs. For a specific class of NFAs, we can besides guarantee that the DFAs obtained by the common subset construction are minimal as well. Thus in contrast to [11, 12], we remove the need for proofs of inequivalence of states.

The paper consists of five sections, including this introduction. The next section contains basic definitions, notations, and preliminary results. We present main results in Section 3. Section 4 deals with automata over binary and unary alphabets. The last section contains some concluding remarks and open problems.

1. PRELIMINARIES

In this section, we give basic definitions, notations, and preliminary results. For further details, we refer to [20, 21].

Let Σ be an alphabet and Σ^* the set of all strings over Σ including the empty string ε . A language over Σ is a subset of Σ^* . The cardinality of a finite set A is denoted by $|A|$ and its power-set by 2^A .

A *deterministic finite automaton* (DFA) is a 5-tuple $M = (Q, \Sigma, \delta, q_0, F)$, where Q is a finite set of states, Σ is a finite input alphabet, $\delta : Q \times \Sigma \rightarrow Q$ is the transition function, $q_0 \in Q$ is the initial state, and $F \subseteq Q$ is the set of accepting states. The transition function δ can be extended to a function from $Q \times \Sigma^*$ to Q by defining $\delta(q, \varepsilon) = q$ and $\delta(q, wa) = \delta(\delta(q, w), a)$ for each $w \in \Sigma^*$ and $a \in \Sigma$. A string $w \in \Sigma^*$ is accepted by the DFA M if $\delta(q_0, w)$ is an accepting state of M . If $p = \delta(q, w)$, we say that the state p is reachable from the state q after reading the string w . A state p is said to be *reachable* if it is reachable from the initial state q_0 after reading a string $w \in \Sigma^*$.

A *nondeterministic finite automaton* (NFA) is a 5-tuple $M = (Q, \Sigma, \delta, q_0, F)$, where $Q, \Sigma, q_0,$ and F are defined as for a DFA, and $\delta : Q \times \Sigma \rightarrow 2^Q$ is the transition function which can be extended to the domain $Q \times \Sigma^*$ by defining $\delta(q, \varepsilon) = \{q\}$ and $\delta(q, wa) = \bigcup_{p \in \delta(q, w)} \delta(p, a)$, for each $w \in \Sigma^*$ and $a \in \Sigma$. Furthermore, it can be extended to the domain $2^Q \times \Sigma^*$ by defining $\delta(P, w) = \bigcup_{p \in P} \delta(p, w)$. A string $w \in \Sigma^*$ is accepted by the NFA M if $\delta(q_0, w) \cap F \neq \emptyset$.

The *language accepted by* a finite automaton M , denoted $L(M)$, is the set of all strings accepted by M . Two automata are said to be *equivalent* if they accept the same language. A DFA (NFA) M is called *minimal* if all DFAs (NFAs, resp.) that are equivalent to M have at least as many states as M . It is known [18] that a DFA $M = (Q, \Sigma, \delta, q_0, F)$ is minimal if (i) all its states are reachable from the initial state q_0 and (ii) no two of its states are equivalent (two states p and q are said to be *equivalent* if for all $w \in \Sigma^*$, $\delta(p, w) \in F$ iff $\delta(q, w) \in F$).

Any NFA $M = (Q, \Sigma, \delta, q_0, F)$ can be converted to an equivalent DFA $M' = (2^Q, \Sigma, \delta', q'_0, F')$ using an algorithm known as the “subset construction” [18] in the following way. Every state of the DFA M' is a subset of the state set Q . The initial state of M' is $\{q_0\}$. A state $P \subseteq Q$ is an accepting state of M' if it includes at least one accepting state of M . The transition function δ' is defined using the transition function δ as follows: $\delta'(P, a) = \bigcup_{p \in P} \delta(p, a)$. Note that the DFA M' need not be minimal since some of its states may be unreachable or equivalent.

By a well-known result, for any regular language, there is a unique minimal DFA, up to isomorphism, but the same result does not hold for NFAs. To prove the minimality of NFAs we will use a fooling-set technique known from communication complexity theory [8], cf. also [1, 2, 6]. Although lower bounds based on fooling sets may be exponentially smaller than the size of minimal NFAs [10, 13], in this paper, the technique does help us to prove the minimality of all considered NFAs.

After defining a fooling set, we give the lemma from [1] describing the fooling-set lower-bound technique. For the sake of completeness, we recall its proof here. Then, we give an example.

Definition 1.1. A set of pairs of strings $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ is said to be a *fooling set* for a language L if for any i and j in $\{1, 2, \dots, n\}$,

- (1) the string $x_i y_i$ is in the language L ;
- (2) if $i \neq j$, then at least one of the strings $x_i y_j$ and $x_j y_i$ is not in L .

Lemma 1.2 (Birget [1]). (Lower-bound argument for the state complexity of NFAs.) *Let a set of pairs $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ be a fooling set for a regular language L . Then any NFA accepting the language L needs at least n states.*

Proof. Let $M = (Q, \Sigma, \delta, q_0, F)$ be any NFA accepting the language L . Since $x_i y_i \in L$, there is a state p_i in Q such that $p_i \in \delta(q_0, x_i)$ and $\delta(p_i, y_i) \cap F \neq \emptyset$ (i.e., p_i is a state on an accepting computation on $x_i y_i$ that is reached after reading x_i). Assume that a fixed choice of p_i has been made for any i in $\{1, 2, \dots, n\}$. We prove that $p_i \neq p_j$ for $i \neq j$. Suppose by contradiction that $p_i = p_j$ for some $i \neq j$. Then the NFA M accepts both strings $x_i y_j$ and $x_j y_i$ which contradicts the assumption that the set $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ is a fooling set for the language L . Hence the NFA M has at least n states. \square

Example 1.3. Let n be arbitrary but fixed positive integer and let $a \in \Sigma$. Consider the regular language

$$L = \{w \in \Sigma^* \mid \text{the number of } a\text{'s in } w \text{ is a multiple of } n\}.$$

The set of pairs $\{(a, a^{n-1}), (a^2, a^{n-2}), \dots, (a^n, \varepsilon)\}$ is a fooling set for the language L because for any i and j in $\{1, 2, \dots, n\}$,

- (1) $a^i a^{n-i} = a^n$ and a^n is in L ;
- (2) if $1 \leq i < j \leq n$, then $a^i a^{n-j} = a^{n-(j-i)}$ and $a^{n-(j-i)}$ is not in L since $0 < n - (j - i) < n$.

Thus by Lemma 1.2, any NFA accepting the language L needs at least n states. Since the language L is accepted by an n -state DFA, the minimal DFA and any minimal NFA for the language L have n states. \square

The following lemma describes a specific class of minimal n -state NFAs, for which the DFAs obtained by the common subset construction do not contain equivalent states. The lemma removes the need for proofs of inequivalence of states in the following sections.

Lemma 1.4. *Let $a \in \Sigma$. Let $M = (\{q_1, q_2, \dots, q_n\}, \Sigma, \delta, q_1, \{q_n\})$ be an n -state NFA such that $\delta(q_i, a) = \{q_{i+1}\}$, for $i = 1, 2, \dots, n - 1$, and $\delta(q_n, a) = \emptyset$ (i.e., the transitions on reading a look like in Fig. 1, the other transitions may be arbitrary). Then*

- (i) M is a minimal NFA for the language it accepts;
- (ii) no two different states of the DFA M' obtained from the NFA M by the subset construction are equivalent.

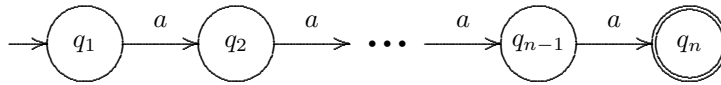


FIGURE 1. Transitions on reading a of the NFA M .

Proof. Let $M = (Q_M, \Sigma, \delta, q_1, F_M)$, where $Q_M = \{q_1, q_2, \dots, q_n\}$ and $F_M = \{q_n\}$, be an n -state NFA satisfying the conditions in the lemma.

To prove (i) consider the set of pairs $\{(\varepsilon, a^{n-1}), (a, a^{n-2}), \dots, (a^{n-1}, \varepsilon)\}$. It is a fooling set for the language $L(M)$ since for any i and j in $\{1, 2, \dots, n\}$,

- (1) $a^{i-1}a^{n-i} = a^{n-1}$ and a^{n-1} is in $L(M)$;
- (2) if $1 \leq i < j \leq n$, then $a^{i-1}a^{n-j} = a^{n-1-(j-i)}$ and

$a^{n-1-(j-i)}$ is not in $L(M)$ since $n - 1 - (j - i) < n - 1$.

By Lemma 1.2, any NFA for the language $L(M)$ needs at least n states, so the NFA M is a minimal NFA for the language $L(M)$.

To prove (ii) note that for any $i = 1, 2, \dots, n$, the string a^{n-i} is accepted by the NFA M starting in state q_i , but it is not accepted by M starting in any other state. Now, let P and R be two different states of the DFA M' , *i.e.*, P and R are subsets of the state set Q_M . Without loss of generality, there is a state q_i in Q_M such that $q_i \in P$ and $q_i \notin R$. Then, the string a^{n-i} is accepted by the DFA M' starting in state P since it is accepted by the NFA M starting in state q_i which is in P . On the other hand, the string a^{n-i} is not accepted by the DFA M' starting in state R since it is not accepted by the NFA M starting in any state different from q_i . Hence no two different states of the DFA M' are equivalent and the proof is complete. \square

2. MAIN RESULTS

In this section, we show that all relations between the sizes of minimal NFAs and equivalent minimal DFAs are possible. For any integers n and α with $n \leq \alpha \leq 2^n$, we describe a minimal n -state NFA whose equivalent minimal DFA has α states. However, in order to present an explicit construction of the NFAs, we increase the input alphabet to exponential sizes. In the second part of this section, we prove that $2n$ letters would be sufficient, but we can describe the related NFAs only implicitly.

In the first theorem, we give minimal binary NFAs for some special values of α . We use these binary automata in the second theorem to prove the main result of the paper.

Theorem 2.1. *For any integers n and k such that $1 \leq k \leq n$, there exists a minimal binary NFA of n states whose equivalent minimal DFA has $2^k + n - k$ states.*

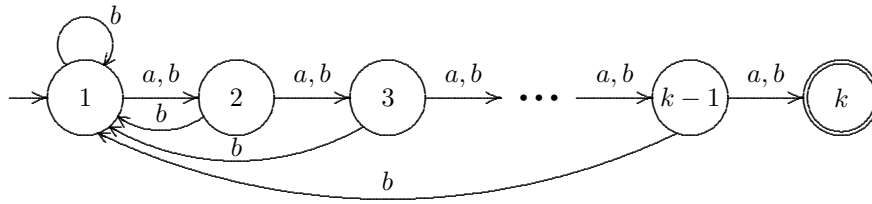


FIGURE 2. The nondeterministic finite automaton A_k .

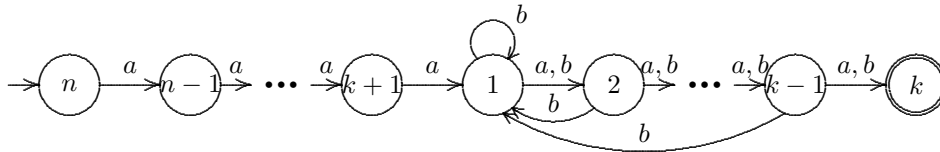


FIGURE 3. The nondeterministic finite automaton $B_{n,k}$.

Proof. Let $\Sigma = \{a, b\}$. Let n and k be arbitrary but fixed integers with $1 \leq k \leq n$. We first define a k -state NFA A_k . Then, by adding $n - k$ new states, we construct an n -state NFA $B_{n,k}$ satisfying the theorem.

Define a k -state NFA $A_k = (Q_A, \Sigma, \delta_A, 1, F_A)$, where $Q_A = \{1, 2, \dots, k\}$, $F_A = \{k\}$, and for any $i \in Q_A$ and any $X \in \Sigma$,

$$\delta_A(i, X) = \begin{cases} \{i + 1\}, & \text{if } i < k \text{ and } X = a, \\ \{1, i + 1\}, & \text{if } i < k \text{ and } X = b, \\ \emptyset, & \text{if } i = k, \end{cases}$$

i.e., the NFA A_k moves from state i to state $i + 1$ on reading a , and it can move from state i either to the initial state 1 or to state $i + 1$ on reading b , except for the accepting state k , see Figure 2.

Now, construct an n -state NFA $B_{n,k}$ from the NFA A_k by adding new states $k + 1, k + 2, \dots, n$, and new transitions on reading symbol a for these states.

Define an n -state NFA $B = B_{n,k} = (Q_B, \Sigma, \delta_B, n, F_B)$, where $Q_B = \{1, 2, \dots, n\}$, $F_B = \{k\}$, and for any $i \in Q_B$ and any $X \in \Sigma$,

$$\delta_B(i, X) = \begin{cases} \delta_A(i, X), & \text{if } i \leq k, \\ \{1\}, & \text{if } i = k + 1 \text{ and } X = a, \\ \{i - 1\}, & \text{if } i > k + 1 \text{ and } X = a, \\ \emptyset, & \text{if } i \geq k + 1 \text{ and } X = b. \end{cases}$$

The NFA $B_{n,k}$ is shown in Figure 3.

Let $B' = (2^{Q_B}, \Sigma, \delta', \{n\}, F')$ be the DFA obtained from the NFA B by the subset construction. By Lemma 1.4, the NFA B is minimal and no two different states of the DFA B' are equivalent. Thus, to prove the theorem we only need to show that the DFA B' has $2^k + n - k$ reachable states.

Let \mathcal{R} be the following system of sets

$$\mathcal{R} = \{\{i\} \mid k + 1 \leq i \leq n\} \cup \{S \mid S \subseteq \{1, 2, \dots, k\}\}.$$

We will show that any set in \mathcal{R} is a reachable state of the DFA B' and no other states are reachable in B' . Since the system \mathcal{R} contains $2^k + n - k$ sets, the theorem then follows immediately.

The singletons $\{n\}, \{n-1\}, \dots, \{k+1\}$ are reachable since $\{n-i\} = \delta'(\{n\}, a^i)$ for $i = 0, 1, \dots, n - k - 1$.

Next, we show that any subset of $\{1, 2, \dots, k\}$ is a reachable state in the DFA B' . We prove this by induction on the size of sets. The empty set and the singletons $\{1\}, \{2\}, \dots, \{k\}$ are reachable because we have

$$\emptyset = \delta'(\{n\}, a^n) \text{ and } \{i\} = \delta'(\{k+1\}, a^i) \text{ for } i = 1, 2, \dots, k.$$

Assume that for any integer m with $2 \leq m \leq k$, any subset of $\{1, 2, \dots, k\}$ of size $m - 1$ is reachable in the DFA B' . Let $\{i_1, i_2, \dots, i_m\}$, where $1 \leq i_1 < i_2 < \dots < i_m \leq k$, be a subset of size m . Then we have

$$\{i_1, i_2, \dots, i_m\} = \delta'(\{i_2 - i_1, i_3 - i_1, \dots, i_m - i_1\}, ba^{i_1-1}),$$

where the latter set of size $m - 1$ is reachable by induction (note that $1 \leq i_j - i_1 < k$ for $j = 2, 3, \dots, m$). It follows that the set $\{i_1, i_2, \dots, i_m\}$ is reachable. Thus, we have shown that any set of \mathcal{R} is a reachable state of the DFA B' .

To prove that no other set is reachable it is sufficient to show that for any set S in \mathcal{R} and any symbol X in Σ , the set $\delta'(S, X)$ is in \mathcal{R} . If S is a singleton $\{i\}$, where $k + 1 \leq i \leq n$, then the set $\delta'(S, a)$ is a singleton set and the set $\delta'(S, b)$ is the empty set, so both are in \mathcal{R} . If S is a subset of $\{1, 2, \dots, k\}$, then the sets $\delta'(S, a)$ and $\delta'(S, b)$ are also subsets of $\{1, 2, \dots, k\}$, and so are in \mathcal{R} . It follows that no other set is reachable in the DFA B' .

Hence the DFA B' has $2^k + n - k$ reachable states and the theorem follows. \square

In the next theorem, we prove the main result of the paper by presenting minimal n -state NFAs which need α deterministic states, for all values of n and α .

Theorem 2.2. *For any integers n and α such that $1 \leq n \leq \alpha \leq 2^n$, there exists a minimal NFA of n states whose equivalent minimal DFA has α states.*

Proof. Let n and α be arbitrary but fixed integers with $1 \leq n \leq \alpha \leq 2^n$.

If $\alpha = n$, then a minimal NFA for the language

$$\{w \mid \text{the number of } a\text{'s in } w \text{ is a multiple of } n\}$$

can be taken to prove the theorem, see Example 1.3.

If α can be expressed as $2^k + n - k$ for some integer k such that $1 \leq k \leq n$, then the NFA $B_{n,k}$ described in the proof of Theorem 2.1 satisfies the theorem.

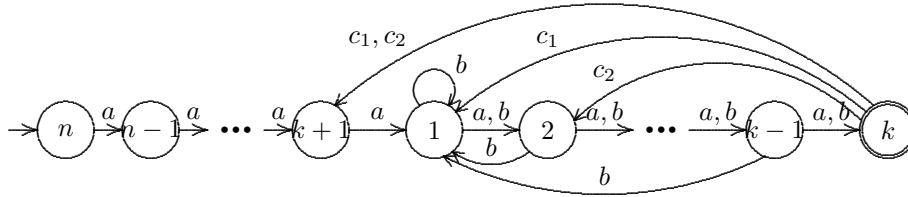


FIGURE 4. The nondeterministic finite automaton $C_{n,k,2}$.

Otherwise, α can be expressed as

$$\alpha = 2^k + n - k + m$$

for some integers k and m such that $1 \leq k < n$ and $1 \leq m \leq 2^k - 2$. In this case, to define an appropriate NFA we will use new input symbols. So consider an $(m + 2)$ -letter input alphabet

$$\Sigma = \{a, b\} \cup \{c_i \mid i = 1, 2, \dots, m\}.$$

Let $S_1, S_2, \dots, S_{2^k-2}$ be all the subsets of $\{1, 2, \dots, k+1\}$ containing the state $k+1$, except for the singleton $\{k+1\}$ and the set $\{1, 2, \dots, k+1\}$, somehow ordered. Let $B_{n,k} = (Q_B, \Sigma, \delta_B, n, F_B)$ be the NFA described in the proof of Theorem 2.1. We construct an n -state NFA $C_{n,k,m}$ from the NFA $B_{n,k}$ by adding a new transition from state k to the set S_j on reading symbol c_j for any $j = 1, 2, \dots, m$.

Formally, $C = C_{n,k,m} = (Q_C, \Sigma, \delta_C, n, F_C)$, where $Q_C = \{1, 2, \dots, n\}$, $F_C = \{k\}$, and for any $i \in Q_C$ and any $X \in \Sigma$,

$$\delta_C(i, X) = \begin{cases} \delta_B(i, X), & \text{if } i \neq k \text{ and } X \in \{a, b\}, \\ S_j, & \text{if } i = k \text{ and } X = c_j \text{ (} 1 \leq j \leq m \text{)}, \\ \emptyset, & \text{otherwise.} \end{cases}$$

The NFA $C_{n,k,2}$ is shown in Figure 4; here, $S_1 = \{1, k+1\}$ and $S_2 = \{2, k+1\}$.

Let $C' = (2^{Q_C}, \Sigma, \delta', \{n\}, F')$ be the DFA obtained from the NFA C by the subset construction. By Lemma 1.4, the NFA C is minimal and no two different states of the DFA C' are equivalent. Thus to prove the theorem we only need to show that the DFA C' has $2^k + n - k + m$ reachable states.

Let \mathcal{R} be the following systems of sets

$$\mathcal{R} = \{\{i\} \mid k+1 \leq i \leq n\} \cup \{S \mid S \subseteq \{1, 2, \dots, k\}\} \cup \{S_j \mid 1 \leq j \leq m\}.$$

We will show that any set in \mathcal{R} is a reachable state of the DFA C' and no other state is reachable. Since the system \mathcal{R} contains $2^k + n - k + m$ sets, the theorem then follows.

Clearly, the sinletons $\{n\}, \{n-1\}, \dots, \{k+1\}$, and all subsets of $\{1, 2, \dots, k\}$ are reachable in the DFA C' since they are reachable in the DFA B' obtained from

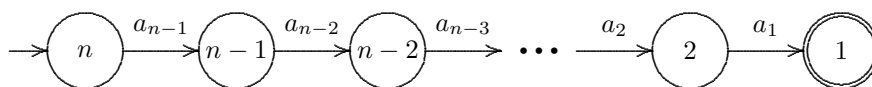


FIGURE 5. Transitions on reading a_1, a_2, \dots, a_{n-1} of the NFA M .

the NFA $B_{n,k}$ by the subset construction, see the proof of Theorem 2.1, and the transitions on a and b in the NFA C are the same as in the NFA $B_{n,k}$. Moreover, the sets S_1, S_2, \dots, S_m are reachable in the DFA C' because $S_j = \delta'(\{k\}, c_j)$ for $j = 1, 2, \dots, m$.

To show that no other set is reachable note that

- if S is a singleton $\{i\}$, where $k + 1 \leq i \leq n$, then the set $\delta'(S, a)$ is a singleton set, the set $\delta'(S, b)$ and the sets $\delta'(S, c_j), j = 1, 2, \dots, m$, are empty;
- if S is a subset of $\{1, 2, \dots, k + 1\}$, then $\delta'(S, a)$ and $\delta'(S, b)$ are some subsets of $\{1, 2, \dots, k\}$, and for all $j = 1, 2, \dots, m$, the set $\delta'(S, c_j)$ is empty if $k \notin S$, and $\delta'(S, c_j) = S_j$ if $k \in S$.

Thus, for any set S in \mathcal{R} and any symbol X in Σ , the set $\delta'(S, X)$ is again in \mathcal{R} . It follows that no other set is reachable in the DFA C' .

So the DFA C' has $2^k + n - k + m$ reachable states and the proof is complete. \square

To prove the above theorem we use automata defined over alphabets, the size of which grows exponentially with n . In the second part of this section, we show that the input alphabet size can be decreased to $2n$. First, we prove the following lemma describing another class of minimal NFAs, for which the DFAs obtained by the subset construction do not contain equivalent states.

Lemma 2.3. *Let $n \geq 2$ and let $a_j \in \Sigma$ for $j = 1, 2, \dots, n - 1$. Let $M = (Q_M, \Sigma, \delta, n, F_M)$ be an n -state NFA such that $Q_M = \{1, 2, \dots, n\}$, $F_M = \{1\}$, and for any $i \in Q_M$ and any $j = 1, 2, \dots, n - 1$, $\delta(i, a_j) = \{i - 1\}$ if $i = j + 1$ and $\delta(i, a_j) = \emptyset$ if $i \neq j + 1$, see Figure 5. Then*

- (i) M is a minimal NFA for the language it accepts;
- (ii) no two different states of the DFA M' obtained from the NFA M by the subset construction are equivalent.

Proof. The lemma can be proved in the similar way as Lemma 1.4. The set of pairs $\{(\varepsilon, a_{n-1}a_{n-2} \dots a_1), (a_{n-1}, a_{n-2}a_{n-3} \dots a_1), \dots, (a_{n-1}a_{n-2} \dots a_1, \varepsilon)\}$ is a fooling set for the language $L(M)$ of size n . By Lemma 1.2, any NFA for the language $L(M)$ needs at least n states. Next, the string $a_{i-1}a_{i-2} \dots a_1$ ($1 \leq i < n$) is accepted by the NFA M starting in state i but it is not accepted by M starting in any other state. This immediately implies that no two different states of the DFA M' are equivalent which completes the proof. \square

The next theorem shows that an arbitrary blow-up between NFAs and DFAs is possible likewise for alphabets of size $2n$. Since the cases $n = 1$ and $\alpha = n$ are trivial, we assume $\alpha > n \geq 2$ in the theorem.



FIGURE 6. The NFAs $D_{2,3}$ and $D_{2,4}$.

Theorem 2.4. *Let $n \geq 2$ and let $\Sigma_n = \{a_i, b_i \mid 1 \leq i < n\}$ be a $(2n - 2)$ -letter alphabet. For any integer α such that $n < \alpha \leq 2^n$, there exists a minimal n -state NFA $D_{n,\alpha}$ with the input alphabet Σ_n , whose equivalent minimal DFA has α states.*

Proof. We prove this theorem by induction on n . Denote by Q_n the set $\{1, 2, \dots, n\}$.

For $n = 2$, let $D_{2,3}$ and $D_{2,4}$ be the NFAs shown in Figure 6.

By Lemma 2.3, the NFAs $D_{2,3}$ and $D_{2,4}$ are minimal and the corresponding DFAs $D'_{2,3}$ and $D'_{2,4}$ obtained by the subset construction do not contain equivalent states. Since the DFA $D'_{2,3}$ has 3 reachable states $(\emptyset, \{1\}, \{2\})$ and the DFA $D'_{2,4}$ has 4 reachable states $(\emptyset, \{1\}, \{2\}, \{1, 2\})$, the theorem holds for $n = 2$.

Let $n \geq 2$ and assume that for any α with $n < \alpha \leq 2^n$, there exists a minimal n -state NFA $D_{n,\alpha}$ over the alphabet Σ_n satisfying the conditions in Lemma 2.3 and requiring α deterministic states. Using this assumption we will show that for any β with $n + 1 < \beta \leq 2^{n+1}$, there is a minimal $(n + 1)$ -state NFA $D_{n+1,\beta}$ over the alphabet Σ_{n+1} which satisfies the conditions in Lemma 2.3 and needs β deterministic states.

First, let $n + 1 < \beta \leq 2n$. Define NFA $D_{n+1,\beta} = (Q_{n+1}, \Sigma_{n+1}, \delta', n + 1, \{1\})$, where $\delta'(i, a_{i-1}) = \{i - 1\}$ for $i = 2, 3, \dots, n$, $\delta'(1, b_j) = \{1, j + 1\}$ for $j = 1, 2, \dots, \beta - n - 2$, and $\delta'(i, X) = \emptyset$ otherwise. By Lemma 2.3, the NFA $D_{n+1,\beta}$ is minimal and no two different states of the DFA $D'_{n+1,\beta}$ obtained by the subset construction are equivalent. Since the DFA $D'_{n+1,\beta}$ has exactly β reachable states

$$\emptyset, \{1\}, \{2\}, \dots, \{n + 1\}, \{1, 2\}, \{1, 3\}, \dots, \{1, \beta - n - 1\},$$

we are ready in this case.

Now, let $2n < \beta \leq 2^{n+1}$ and let β is even. Set $\alpha = \beta/2$. Then $n < \alpha \leq 2^n$, and by induction, there is a minimal n -state NFA $D_{n,\alpha} = (Q_n, \Sigma_n, \delta, n, \{1\})$ satisfying the conditions in Lemma 2.3 and requiring α deterministic states. We construct an $(n + 1)$ -state NFA $D_{n+1,\beta} = (Q_{n+1}, \Sigma_{n+1}, \delta', n + 1, \{1\})$ from the NFA $D_{n,\alpha}$ by adding a new initial state $n + 1$, a new transition from $n + 1$ to n on reading a new symbol a_n , and new transitions from i to $\{i, n + 1\}$ on reading a new symbol b_n for $i = 1, 2, \dots, n$, see Figure 7. Formally, we define δ' so that for any $i \in Q_{n+1}$ and any $X \in \Sigma_{n+1}$,

$$\delta'(i, X) = \begin{cases} \delta(i, X), & \text{if } i \in Q_n \text{ and } X \in \Sigma_n, \\ \{n\}, & \text{if } i = n + 1 \text{ and } X = a_n, \\ \{i, n + 1\}, & \text{if } i \in Q_n \text{ and } X = b_n, \\ \emptyset, & \text{otherwise.} \end{cases}$$

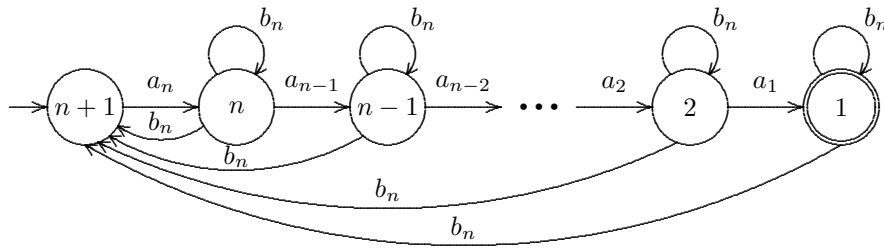


FIGURE 7. Transitions on a_1, \dots, a_n, b_n of the NFA $D_{n+1, \beta}, \beta$ even.

The NFA $D_{n+1, \beta}$ satisfies the conditions in Lemma 2.3, so we only need to show that the DFA $D'_{n+1, \beta}$ obtained by the subset construction has β reachable states. Note that if a set P is reachable in the DFA $D'_{n, \alpha}$, then the sets P and $P \cup \{n+1\}$ are reachable in the DFA $D'_{n+1, \beta}$ since $\delta'(\{n+1\}, a_n) = \{n\}$ and $\delta'(P, b_n) = P \cup \{n+1\}$. Thus we have $2\alpha = \beta$ reachable sets in the DFA $D'_{n+1, \beta}$. To see that no other set is reachable note that if S is one of these β reachable sets, then for any $X \in \Sigma_{n+1}$, the set $\delta(S, X)$ is also one of these β sets. Thus the DFA $D'_{n+1, \beta}$ has β reachable states.

If $2n < \beta \leq 2^{n+1}$ and β is odd, we set $\alpha = (\beta + 1)/2$. Then $n < \alpha \leq 2^n$, and so there is an n -state NFA $D_{n, \alpha}$ satisfying the induction hypothesis. We construct an $(n + 1)$ -state NFA $D_{n+1, \beta}$ from the NFA $D_{n, \alpha}$ by adding a new initial state $n + 1$, a new transition from $n + 1$ to n on reading a_n , and new transitions from i to $\{i, n + 1\}$ on reading b_n for $i = 1, 2, \dots, n - 1$, *i.e.*, we construct this automaton similarly as the NFA for an even β above, but we do not add the transition from n to $\{n, n + 1\}$ on reading b_n . The NFA $D_{n+1, \beta}$ satisfies the conditions in Lemma 2.3. Next, if a set P is reachable in the DFA $D'_{n, \alpha}$, then the sets P and $P \cup \{n + 1\}$ are reachable in the DFA $D'_{n+1, \beta}$ except for the set $\{n, n + 1\}$. Since no other set is reachable, the DFA $D'_{n+1, \beta}$ has $2\alpha - 1 = \beta$ reachable states. This completes the proof. \square

3. AUTOMATA OVER BINARY AND UNARY ALPHABETS

In this section, we investigate the question whether there exists a minimal n -state NFA whose equivalent minimal DFA has α states, where $n \leq \alpha \leq 2^n$, for binary and unary alphabets.

Iwama *et al.* in [11, 12] discussed finite automata over a binary alphabet and succeeded in finding minimal binary n -state NFAs for about $3n$ values of α between 2^{n-1} and 2^n , namely for $\alpha = 2^n - 2^k$ or $\alpha = 2^n - 2^k - 1$, where $0 \leq k \leq n/2 - 2$ [11], and for $\alpha = 2^n - m$, where $n \geq 7$, $5 \leq m \leq 2n - 2$, and some coprimality condition holds [12].

In Theorem 2.1, we showed that appropriate minimal binary n -state NFAs exist for further n values of α (for $\alpha = 2^k + n - k$, where $1 \leq k \leq n$). The next theorem

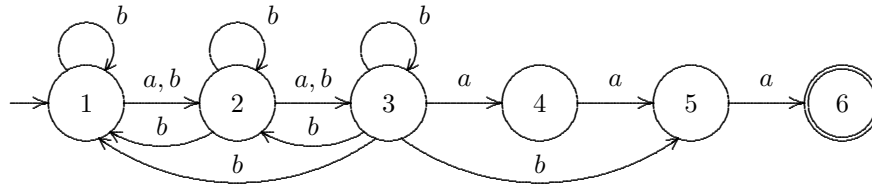


FIGURE 8. The nondeterministic finite automaton $M_{6,2,2}$.

describes minimal binary n -state NFAs requiring α deterministic states for about $n^2/2$ values of α between n and n^2 .

Theorem 3.1. *For any integers n and α such that $1 \leq n \leq \alpha \leq 1 + n(n + 1)/2$, there exists a minimal binary NFA of n states whose equivalent minimal DFA has α states.*

Proof. Let $\Sigma = \{a, b\}$. Let n and α be arbitrary but fixed integers with $1 \leq n \leq \alpha \leq 1 + n(n + 1)/2$.

Since the minimal DFA and any minimal NFA for the language $\{w \in \Sigma^* \mid \text{the number of } a\text{'s in } w \text{ is a multiple of } n\}$ have n states, see Example 1.3, the theorem holds for $\alpha = n$.

In the case of $\alpha = n + 1$, we can consider a minimal n -state NFA accepting all strings of length $n - 1$, whose equivalent minimal DFA has $n + 1$ states.

Now, let $n + 2 \leq \alpha \leq 1 + n(n + 1)/2$. Since $1 + n(n + 1)/2 = 1 + n + (n - 1) + (n - 2) + \dots + 2 + 1$, we can express α as

$$\alpha = 1 + n + (n - 1) + \dots + (n - k) + m$$

for some integers k and m such that $0 \leq k < n$ and $1 \leq m \leq n - (k + 1)$.

Define an n -state NFA $M = M_{n,k,m} = (Q_M, \Sigma, \delta_M, 1, F_M)$, where $Q_M = \{1, 2, \dots, n\}$, $F_M = \{n\}$, and for any $i \in Q_M$ and any $X \in \Sigma$,

$$\delta_M(i, X) = \begin{cases} \{i + 1\}, & \text{if } i < n \text{ and } X = a, \\ \{1, 2, \dots, i + 1\}, & \text{if } i \leq k \text{ and } X = b, \\ \{1, 2, \dots, k + 1\} \cup \{n - m + 1\}, & \text{if } i = k + 1 \text{ and } X = b \\ \emptyset, & \text{otherwise;} \end{cases}$$

note that $k + 2 \leq n - m + 1 \leq n$. The NFA $M_{6,2,2}$ is depicted in Figure 8.

Let $M' = (2^{Q_M}, \Sigma, \delta', \{1\}, F')$ be the DFA obtained from the NFA M by the subset construction. By Lemma 1.4, the NFA M is minimal and no two different states of the DFA M' are equivalent. So we only need to show that the DFA M' has $1 + n + (n - 1) + \dots + (n - k) + m$ reachable states.

Clearly, the empty set and all the singleton sets are reachable in the DFA M' . The following sets are also reachable

$$\begin{aligned} A_{1,1} &= \{1, 2\}, A_{1,2} = \{2, 3\}, \dots, A_{1,n-1} = \{n - 1, n\}, \\ A_{2,1} &= \{1, 2, 3\}, A_{2,2} = \{2, 3, 4\}, \dots, A_{2,n-2} = \{n - 2, n - 1, n\}, \dots, \\ A_{k,1} &= \{1, 2, \dots, k + 1\}, \dots, A_{k,n-k} = \{n - k, n - k + 1, \dots, n\}, \end{aligned}$$

because we have $A_{i,j} = \delta'(\{i\}, ba^{j-1})$ for $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, n - i$.
 Next, there are m more reachable sets

$$\begin{aligned} S_1 &= \{1, 2, \dots, k + 1\} \cup \{n - m + 1\}, \\ S_2 &= \{2, 3, \dots, k + 2\} \cup \{n - m + 2\}, \\ &\vdots \\ S_m &= \{m, m + 1, \dots, k + m\} \cup \{n\}, \end{aligned}$$

because we have

$$S_j = \delta'(\{k + 1\}, ba^{j-1}) \text{ for } j = 1, 2, \dots, m.$$

It follows that the DFA M' has at least $1 + n + (n - 1) + \dots + (n - k) + m$ reachable states.

To see that no other set is reachable in the DFA M' note that

- for any $i = 1, 2, \dots, k$ and any $j = 1, 2, \dots, n - i$, we have
 - $\delta'(A_{i,j}, a) = A_{i,j+1}$ if $j < n - i$,
 - $\delta'(A_{i,j}, a) = A_{i-1,j+1}$ if $i > 1$ and $j = n - i$,
 - $\delta'(A_{1,n-1}, a) = \{n\}$;
- next, we have
 - $\delta'(\{n\}, a) = \emptyset$ and $\delta'(\{i\}, a) = \{i + 1\}$ for $i = 1, 2, \dots, n - 1$,
 - $\delta'(S_j, a) = S_{j+1}$ for $j = 1, 2, \dots, m - 1$, and
 - $\delta'(S_m, a) = \{m + 1, m + 2, \dots, m + k + 1\} = A_{k,m+1}$;
- finally, for any subset S of Q_M , we have
 - $\delta'(S, b) = \emptyset$ if $S \subseteq \{k + 2, \dots, n\}$,
 - $\delta'(S, b) = S_1$ if $k + 1 \in S$, and
 - $\delta'(S, b) = \{1, 2, \dots, i + 1\}$ if $i \leq k, i \in S$ and $S \cap \{i + 1, i + 2, \dots, k + 1\} = \emptyset$.

Hence the DFA M' has $1 + n + (n - 1) + (n - 2) + \dots + (n - k) + m$ reachable states and the proof is complete. □

The above theorem shows that there are $\Omega(n^2)$ values of α for which there is a minimal binary n -state NFA requiring α deterministic states. We are not able to prove that appropriate NFAs exist for all integers α with $n \leq \alpha \leq 2^n$ but it seems to be true for us. We have computationally verified the conjecture for $n \leq 12$, *i.e.*, for all n and α such that $n \leq 12$ and $n \leq \alpha \leq 2^n$, we have found a minimal binary n -state NFA whose equivalent minimal DFA has α states.

The last part of the paper is devoted to a few remarks on unary automata. For the state complexity of unary automata, a crucial role is played by the function $F(n) = \max\{\text{lcm}(x_1, \dots, x_k) \mid x_1 + \dots + x_k = n\}$. It is known that $F(n) \in e^{\Theta(\sqrt{n \ln n})}$ and that $O(F(n))$ states suffice to simulate any unary n -state NFA by a DFA [4, 5]. This means that deterministic blow-ups of minimal unary n -state NFAs cannot reach any value between $e^{O(\sqrt{n \ln n})}$ and 2^n . On the other hand, such blow-ups are not bounded by any polynomial because there are n -state unary NFAs which need at least $F(n - 1)$ deterministic states [4, 7, 14]. For $n \leq 10$, we have computationally verified that for any α with $n \leq \alpha \leq (n - 1)^2 + 2$, there is a minimal unary n -state NFA whose equivalent minimal DFA has α states.

4. CONCLUSIONS

In this paper, we investigated the relations between the sizes of minimal NFAs and corresponding minimal DFAs. We showed that for all integers n and α with $1 \leq n \leq \alpha \leq 2^n$, there exists a minimal NFA of n states whose equivalent minimal DFA has α states. However, in order to present an explicit construction of the NFAs, we increased the input alphabet to exponential sizes. Then we proved that $2n$ letters would be sufficient but we described the related NFAs only implicitly. In the case of a binary alphabet, we complemented the known solutions by presenting appropriate NFAs for further $n^2/2$ values of α . It remains open whether such an arbitrary blow-up between NFAs and DFAs holds likewise for alphabets of fixed size.

Acknowledgements. I would like to thank Juraj Hromkovič for his comments concerning this work. I am also grateful to Jozef Jirásek for his help with the computational verification of some conjectures.

REFERENCES

- [1] J.C. Birget, Intersection and union of regular languages and state complexity. *Inform. Process. Lett.* **43** (1992) 185–190.
- [2] J.C. Birget, Partial orders on words, minimal elements of regular languages, and state complexity. *Theoret. Comput. Sci.* **119** (1993) 267–291.
- [3] P. Berman and A. Lingas, *On the complexity of regular languages in terms of finite automata*. Technical Report 304, Polish Academy of Sciences (1977).
- [4] M. Chrobak, Finite automata and unary languages. *Theoret. Comput. Sci.* **47** (1986) 149–158.
- [5] M. Chrobak, Errata to: “Finite automata and unary languages” [*Theoret. Comput. Sci.* **47** (1986) 149–158]. *Theoret. Comput. Sci.* **302** (2003) 497–498.
- [6] I. Glaister and J. Shallit, A lower bound technique for the size of nondeterministic finite automata. *Inform. Process. Lett.* **59** (1996) 75–77.
- [7] M. Holzer and M. Kutrib, Nondeterministic descriptive complexity of regular languages. *Internat. J. Found. Comput. Sci.* **14** (2003) 1087–1102.
- [8] J. Hromkovič, *Communication Complexity and Parallel Computing*. Springer-Verlag, Berlin, Heidelberg (1997).
- [9] J. Hromkovič, Descriptive complexity of finite automata: concepts and open problems. *J. Autom. Lang. Comb.* **7** (2002) 519–531.
- [10] J. Hromkovič, S. Seibert, J. Karhumäki, H. Klauck and G. Schnitger, Communication complexity method for measuring nondeterminism in finite automata. *Inform. Comput.* **172** (2002) 202–217.
- [11] K. Iwama, Y. Kambayashi and K. Takaki, Tight bounds on the number of states of DFAs that are equivalent to n -state NFAs. *Theoret. Comput. Sci.* **237** (2000) 485–494.
- [12] K. Iwama, A. Matsuura and M. Paterson, A family of NFAs which need $2^n - \alpha$ deterministic states. *Theoret. Comput. Sci.* **301** (2003) 451–462.
- [13] G. Jirásková, Note on minimal automata and uniform communication protocols, in *Grammars and Automata for String Processing: From Mathematics and Computer Science to Biology, and Back*, edited by C. Martin-Vide, V. Mitrana, Taylor and Francis, London (2003) 163–170.

- [14] G. Jirásková, State complexity of some operations on regular languages, in *Proc. 5th Workshop Descriptive Complexity of Formal Systems*, edited by E. Csuhaaj-Varjú, C. Kintala, D. Wotschke, Gy. Vaszil, MTA SZTAKI, Budapest (2003) 114–125.
- [15] O.B. Lupanov, A comparison of two types of finite automata. *Problemy Kibernetiki* **9** (1963) 321–326 (in Russian).
- [16] A.R. Meyer and M.J. Fischer, Economy of description by automata, grammars and formal systems, in *Proc. 12th Annual Symposium on Switching and Automata Theory* (1971) 188–191.
- [17] F.R. Moore, On the bounds for state-set size in the proofs of equivalence between deterministic, nondeterministic, and two-way finite automata. *IEEE Trans. Comput.* **20** (1971) 1211–1214.
- [18] M. Rabin and D. Scott, Finite automata and their decision problems. *IBM Res. Develop.* **3** (1959) 114–129.
- [19] M. Sipser, Lower bounds on the size of sweeping automata. *J. Comput. System Sci.* **21** (1980) 195–202.
- [20] M. Sipser, *Introduction to the Theory of Computation*. PWS Publishing Company, Boston (1997).
- [21] S. Yu, Chapter 2: Regular languages, in *Handbook of Formal Languages - Vol. I*, edited by G. Rozenberg, A. Salomaa, Springer-Verlag, Berlin, New York (1997) 41–110.
- [22] S. Yu, A renaissance of automata theory? *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* **72** (2000) 270–272.

Communicated by J. Hromkovič.

Received March 8, 2002. Accepted February 2005.