Journal of Automata, Languages and Combinatorics **22** (2017) 1–3, 123 – 143


# THE COMPLEXITY OF LANGUAGES RESULTING FROM THE CONCATENATION OPERATION

Galina Jirásková[(A,B)] Alexander Szabari[(C,D)] Juraj Šebej[(C,D)]

[(A)] *Mathematical Institute, Slovak Academy of Sciences*
*Grešákova 6, 040 01 Košice, Slovakia*
jiraskov@saske.sk

[(C)] *Institute of Computer Science, Faculty of Science, P.J. Šafárik University*
*Jesenná 5, 040 01 Košice, Slovakia*
alexander.szabari@gmail.com juraj.sebej@gmail.com

ABSTRACT

We prove that for all $m, n$, and $\alpha$ with $1 \le \alpha \le f(m,n)$, where $f(m,n)$ is the state complexity of the concatenation operation, there exist a minimal $m$-state deterministic finite automaton $A$ and a minimal $n$-state deterministic finite automaton $B$, both defined over an alphabet $\Sigma$ with $|\Sigma| \le 2n+4$, such that the minimal deterministic finite automaton for the language $L(A)L(B)$ has exactly $\alpha$ states. This improves a similar result in the literature that uses an exponential alphabet.

*Keywords:* regular languages, deterministic finite automata, concatenation, state complexity, the magic number problem

## 1. Introduction

In 2000, Iwama et al. [3] stated the question of whether there always exists a minimal nondeterministic finite automaton (NFA) of $n$ states whose equivalent minimal deterministic finite automaton (DFA) has $\alpha$ states for all integers $n$ and $\alpha$ satisfying $n \leqslant \alpha \leqslant 2^n$. The question was also considered by Iwama et al. [4], and answered positively in [7] for a ternary alphabet. However, in the unary case, the existence of holes, so called "magic numbers" was proved by Geffert [1]. The binary case is still open.

The same problem on subregular language families was studied by Holzer et al. [2]. It turned out that the existence of non-trivial magic numbers is rare, and that the ranges of possible complexities are usually contiguous. One interesting exception was obtained by Čevorová [14]. She studied the star operation on unary languages, and proved that there are two linear segments of magic numbers in the range from 1 to

[(B)] Research supported by grant VEGA 2/0084/15 and grant APVV-15-0091.

[(D)] Research supported by grant VEGA 1/0142/15 and grant APVV-15-0091.

$(n-1)^2+1$, that is, of values that cannot be met by the state complexity of the star of a unary language accepted by a minimal $n$-state DFA.

A similar problem for the reversal, star, and concatenation operation was studied in [6, 8], where it was shown that for all three operations the whole range of possible complexities up to known upper bounds can be produced using an exponential alphabet. The result for reversal and star was improved in [9, 15] by showing that a linear alphabet is enough to produce the whole range of complexities.

In this paper we complement these results, and show that a linear alphabet can also be used for the concatenation operation. We prove that for all $m, n$, and $\alpha$ with $1 \leq \alpha \leq f(m,n)$, where $f(m,n)$ is the state complexity of the concatenation operation, there exist a minimal $m$-state DFA $A$ and a minimal $n$-state DFA $B$, both defined over an alphabet $\Sigma$ with $|\Sigma| \leq 2n+4$, such that the minimal DFA for the language $L(A)L(B)$ has exactly $\alpha$ states.

To get complexities from 1 to $m+n-1$, we use a known result from [6]. We deal with the value $m+n$ separately, and use a binary alphabet here. To get complexities larger than $m+n$, we describe three constructions. Using these constructions, we are able to get $m$-state and $(n+1)$-state DFAs $A_i, B_i$ for $i = 1,2,3$ from $m$-state and $n$-state DFAs $A$ and $B$, by adding a new state to $B$, and by adding transitions on two new symbols. Moreover, if the state complexity of the concatenation of $L(A)$ and $L(B)$ is $\alpha$, then the state complexity of the concatenation of $L(A_i)$ and $L(B_i)$, $i = 1,2,3$, is $2\alpha, 2\alpha-1$, and $\alpha+1$, respectively. As a result, we get a contiguous range of complexities from $m+n+1$ up to the known upper bound for a linear alphabet.

The paper is organized as follows. The next section contains some definitions and preliminary results. In Section 3, we recall known results concerning the state complexity of concatenation. In Section 4, we prove that the range of possible complexities for the languages resulting from the concatenation operation is contiguous from 1 up to the known upper bound, and we show that a linear alphabet is enough for this. Section 5 contains some concluding remarks.

## 2. Preliminaries

In this section we give some basic definitions and preliminary results. For details, the reader may refer to [13, 16].

Let $\Sigma$ be a finite alphabet of symbols. Then $\Sigma^*$ denotes the set of strings over $\Sigma$ including the empty string $\varepsilon$. The length of a string $w$ is denoted by $|w|$, and the number of occurrences of a symbol $a$ in a string $w$ is denoted by $\#_a(w)$. A language is any subset of $\Sigma^*$. The concatenation of languages $K$ and $L$ is the language $KL = \{ uv \mid u \in K \text{ and } v \in L \}$. The cardinality of a finite set $A$ is denoted by $|A|$, and its power-set by $2^A$.

A *nondeterministic finite automaton* (NFA) is a quintuple $A = (Q, \Sigma, \cdot, I, F)$, where $Q$ is a finite set of states, $\Sigma$ is a finite alphabet, $\cdot : Q \times \Sigma \to 2^Q$ is the transition function which is extended to the domain $2^Q \times \Sigma^*$ in the natural way, $I \subseteq Q$ is the set of initial states, and $F \subseteq Q$ is the set of final states. The *language accepted by $A$* is the set $L(A) = \{ w \in \Sigma^* \mid I \cdot w \cap F \neq \emptyset \}$.

For a symbol $a$, we say that $(p, a, q)$ is a transition in NFA $A$ if $q \in p \cdot a$, and for a string $w$, we write $p \xrightarrow{w} q$ if $q \in p \cdot w$. We say that $(p, a, q)$ is an in-transition going to the state $q$ on symbol $a$.

An NFA $A$ is *deterministic* (DFA) (and complete) if $|I| = 1$ and $|q \cdot a| = 1$ for each $q$ in $Q$ and each $a$ in $\Sigma$. In such a case, we write $q \cdot a = q'$ instead of $q \cdot a = \{q'\}$.

Two automata are *equivalent* if they accept the same language. A DFA $A$ is *minimal* if each DFA that is equivalent to $A$ has at least as many states as the DFA $A$. It is well known that every regular language has a unique minimal DFA, up to the naming of its states.

The *state complexity* of a regular language $L$, $\mathrm{sc}(L)$, is the number of states in the minimal DFA accepting the language $L$. The state complexity of a binary regular operation $\circ$ is defined as a function $f(m, n)$ given by

$$f(m, n) = \max\{\, \mathrm{sc}(K \circ L) \mid K, L \subseteq \Sigma^*, \mathrm{sc}(K) = m, \mathrm{sc}(L) = n \,\}.$$

Every nondeterministic finite automaton $A = (Q, \Sigma, \cdot, I, F)$ can be converted to an equivalent deterministic automaton $A' = (2^Q, \Sigma, \cdot', I, F')$, where $R \cdot' a = R \cdot a$ and $F' = \{\, R \in 2^Q \mid R \cap F \neq \emptyset \,\}$ [12]. The DFA $A'$ is called the *subset automaton* of the NFA $A$. The subset automaton may not be minimal since some of its states may be unreachable or equivalent to other states.

In the following proposition, we provide a sufficient condition for a nondeterministic finite automaton which guarantees that the corresponding subset automaton does not have equivalent states.

**Proposition 1.** *Let $N = (Q, \Sigma, \cdot, I, F)$ be an NFA. Assume that for each state $q$ in $Q$, there is a string $w_q$ in $\Sigma^*$ which is accepted by $N$ from and only from the state $q$, that is, we have $p \cdot w_q \cap F \neq \emptyset$ if and only if $p = q$. Then the subset automaton of $N$ does not have equivalent states.*

*Proof.* Let $S$ and $T$ be two distinct subsets of the subset automaton. Then, without loss of generality, there is a state $q$ with $q \in S \setminus T$. Then the string $w_q$ is accepted by the subset automaton from the subset $S$, but it is rejected from $T$. $\square$

We say that a transition $(p, a, q)$ of an NFA $N$ is a *unique in-transition* (going to the state $q$ on symbol $a$) if there is no state $r$ with $r \neq p$ such that $(r, a, q)$ is a transition in the NFA $N$. To describe a string $w_q$ which is accepted by an NFA from and only from a state $q$, we almost always use the next observation.

**Proposition 2.** *Let a string $w_q$ be accepted by an NFA $N$ from and only from a state $q$. If $(p, a, q)$ is the unique in-transition going to the state $q$ by $a$, then the string $aw_q$ is accepted by $N$ from and only from the state $p$.*

*Proof.* The string $aw_q$ is accepted by the NFA $N$ from the state $p$ since $(p, a, q)$ is a transition in $N$ and $w_q$ is accepted from $q$. Moreover, the string $aw_q$ is accepted only from the state $p$ since $w_q$ is accepted only from $q$ and $p$ is the only state which goes to $q$ on $a$. $\square$

In what follows, we often need to show how the family of all the reachable subsets in a subset automaton looks. To do this, the following observation is useful.

**Proposition 3.** *Let $D$ be the subset automaton of an NFA $N = (Q, \Sigma, \cdot, I, F)$. Let $\mathcal{R}$ be a family of subsets of $Q$ such that*

*(1) each subset in $\mathcal{R}$ is reachable in $D$,*

*(2) $I \in \mathcal{R}$, and*

*(3) for each $S$ in $\mathcal{R}$ and each symbol $\sigma$ in $\Sigma$, the set $S \cdot \sigma$ is in $\mathcal{R}$.*

*Then $\mathcal{R}$ is the family of all the reachable subsets of DFA $D$, that is, no subset outside $\mathcal{R}$ is reachable in $D$.*

*Proof.* Each set in $\mathcal{R}$ is reachable in $D$ by (1). Let $S$ be a reachable subset of $D$. Then there is a string $w$ in $\Sigma^*$ such that $S = I \cdot w$. We prove the proposition by induction on $|w|$. If $|w| = 0$, then $w = \varepsilon$ and $S = I \cdot \varepsilon = I$, which is in $\mathcal{R}$ by (2). Now let $w = v\sigma$ for a string $v$ and a symbol $\sigma$. By the induction hypothesis, the set $S' = I \cdot v$ is in $\mathcal{R}$. Then $S = S' \cdot \sigma$, so $S$ is in $\mathcal{R}$ by (3). □

## 3. State Complexity of Concatenation

Consider a minimal $m$-state DFA $A$ and a minimal $n$-state DFA $B$. Without loss of generality, we assume that the state set of $A$ is $\{q_0, q_1, \ldots, q_{m-1}\}$ with the initial state $q_0$, and the state set of $B$ is $\{0, 1, \ldots, n-1\}$ with the initial state 0. Moreover, in both $A$ and $B$, let us denote the transition function by $\cdot$. This is not confusing since the state sets of $A$ and $B$ are disjoint. First, let us recall the construction of an NFA for the language $L(A)L(B)$.

**Construction of NFA for concatenation:**

( DFA $A$ and DFA $B \to$ NFA $N$ for $L(A)L(B)$ )

Let $A = (\{q_0, q_1, \ldots, q_{m-1}\}, \Sigma, \cdot, q_0, F_A)$ and $B = (\{0, 1, \ldots, n-1\}, \Sigma, \cdot, 0, F_B)$ be DFAs. Construct NFA $N$ from DFAs $A$ and $B$ as follows:

(a) for each symbol $a$ and each state $q_i$ with $q_i \cdot a \in F_A$, add transition $(q_i, a, 0)$;

(b) the set of initial states of $N$ is $\{q_0\}$ if $q_0 \notin F_A$, and it is $\{q_0, 0\}$ otherwise;

(c) the set of final state of $N$ is $F_B$.

Let $N$ be the NFA for $L(A)L(B)$ constructed as above. In the subset automaton of the NFA $N$, each reachable subset is of the form $\{q_i\} \cup S$, where $S \subseteq \{0, 1, \ldots, n-1\}$ since $A$ is deterministic and complete. Moreover, if $q_i$ is a final state of $A$, then $0 \in S$ since $N$ has the transition $(q, a, 0)$ whenever a state $q$ of $A$ goes to a final state $q_i$ on a symbol $a$. It follows that no subset containing a final state of $A$ and not containing the state 0 can be reachable.

Thus if $|F_A| = k$, then the subset automaton of $N$ has at most $m2^n - k2^{n-1}$ reachable subsets. This number is maximal if $k = 1$. The upper bound $m2^n - 2^{n-1}$ is known to be tight already in the binary case if $m \geq 1$ and $n \geq 2$ [5, 11, 17]. If $m \geq 1$

and $n = 1$, then $L = \emptyset$ or $L = \Sigma^*$, and the tight upper bound in this case is $m$ [17]. Hence we get the following result.

**Proposition 4 [5, 11, 17].** *Let $m, n \geq 1$ and $f(m,n)$ be the state complexity of the concatenation operation on languages over an alphabet of size at least two defined by $f(m,n) = \max\{\, \mathrm{sc}(KL) \mid K, L \subseteq \Sigma^*, |\Sigma| \geq 2, \mathrm{sc}(K) = m, \mathrm{sc}(L) = n \,\}$. Then we have*

$$f(m,n) = \begin{cases} m, & \text{if } n = 1; \\ m2^n - 2^{n-1}, & \text{if } n \geq 2. \end{cases}$$

## 4. The Range of Possible Complexities

The aim of this section is to show that the whole range of complexities from 1 up to the upper bound $f(m,n)$ for the concatenation operation can be produced using an alphabet that grows linearly with $n$.

As shown in [6], the complexities from 1 to $m + n - 1$ can be produced using a binary alphabet. For the sake of completeness, we give a simplified proof here.

**Lemma 5 [6, Lemma 5].** *Let $m, n \geq 1$. For each $\alpha$ with $1 \leq \alpha \leq m+n-1$, there exist binary languages $K$ and $L$ such that $\mathrm{sc}(K) = m$, $\mathrm{sc}(L) = n$, and $\mathrm{sc}(KL) = \alpha$.*

*Proof.* (1) First, let $1 \leq \alpha \leq m - 1$. Then $m \geq 2$. If $m = n = 2$ and $\alpha = 1$, we set

$$K = (b + a(a+b))^*,$$
$$L = (a + b(a+b))^*.$$

Then $KL = \{a, b\}^*$, so $\mathrm{sc}(KL) = 1$. Otherwise, we define unary languages $K$ and $L$ as follows:

$$K = \begin{cases} (a^2)^*, & \text{if } m = 2; \\ \{\, a^k \mid k \bmod m \in \{\alpha - 1, \alpha\} \,\}, & \text{if } m \geq 3; \end{cases}$$

$$L = \begin{cases} a^*, & \text{if } n = 1; \\ \{\, a^k \mid k \bmod n \neq n - 1 \,\}, & \text{if } n \geq 2. \end{cases}$$

Then $\mathrm{sc}(K) = m, \mathrm{sc}(L) = n$, and $KL = \{\, a^k \mid k \geq \alpha - 1 \,\}$, so $\mathrm{sc}(KL) = \alpha$.

(2) Now, let $m \leq \alpha \leq m + n - 1$. Then there is an integer $\beta$ with $0 \leq \beta \leq n - 1$ such that $\alpha = m + \beta$. We set

$$K = \{\, a^k \mid k \geq m - 1 \,\},$$
$$L = \{\, a^k \mid k \bmod n = \beta \,\}.$$

Then $\mathrm{sc}(K) = m, \mathrm{sc}(L) = n$, and $KL = \{\, a^k \mid k \geq m + \beta - 1 \,\}$. Therefore we have $\mathrm{sc}(KL) = m + \beta = \alpha$. □

The next lemma shows that the complexity $m+n$ can be produced with a binary alphabet provided that $m, n \geq 2$.

**Lemma 6.** *Let $m \geq 2, n \geq 2$. There exist binary regular languages $K$ and $L$ with* $\mathrm{sc}(K) = m$ *and* $\mathrm{sc}(L) = n$ *such that* $\mathrm{sc}(KL) = m + n$.

*Proof.* Let $K$ and $L$ be the languages accepted by minimal DFAs $A$ and $B$ shown in Figure 1, where for each $i$ in $\{0, 1, \ldots, m-1\}$ and $j$ in $\{0, 1, \ldots, n-1\}$, we have

$q_i \cdot a = q_{i+1}$ if $i \neq m-1$, $q_{m-1} \cdot a = q_{m-1}$ and $q_i \cdot b = q_{m-1}$;

$j \cdot a = j+1$ if $j \neq n-1$, $(n-1) \cdot a = 0$, and $j \cdot b = n-1$.

Construct an NFA $N$ for $KL$ from DFAs $A$ and $B$ by adding the transitions $(q_{m-2}, a, 0)$, $(q_{m-1}, a, 0)$, and $(q_i, b, 0)$ for each $i$; the initial state of $N$ is $q_0$, and the set of final states is $\{n-1\}$. In the subset automaton of $N$, we have

$$\{q_0\} \xrightarrow{a^i} \{q_i\} \text{ for } i = 1, 2, \ldots, m-2,$$

$$\{q_{m-2}\} \xrightarrow{a} \{q_{m-1}, 0\} \xrightarrow{a^j} \{q_{m-1}, 0, 1, \ldots, j\} \text{ for } j = 1, 2, \ldots, n-1,$$

$$\{q_{m-1}, 0, 1, \ldots, n-1\} \xrightarrow{a} \{q_{m-1}, 0, 1, \ldots, n-1\},$$

$$\{q_{m-1}, 0\} \xrightarrow{b} \{q_{m-1}, 0, n-1\} \xrightarrow{a} \{q_{m-1}, 0, 1\}.$$

It follows that the subset automaton has $m+n$ reachable subsets. Let $\mathcal{R}$ be the family of these $m+n$ reachable subsets. The initial subset $\{q_0\}$ is in $\mathcal{R}$. Let $S \in \mathcal{R}$. The transitions given above imply that $S \cdot a \in \mathcal{R}$. Next, we have $S \cdot b = \{q_{m-1}, 0\}$ if $q_{m-1} \notin S$ and $S \cdot b = \{q_{m-1}, 0, n-1\}$ if $q_{m-1} \in S$. Hence $S \cdot b \in \mathcal{R}$. By Proposition 3, no other set is reachable.

To prove distinguishability, let $\{q_i\} \cup S$ and $\{q_j\} \cup T$ be two distinct reachable subsets. Since NFA $N$ accepts the string $a^{n-1-t}$ from and only from the state $t$ $(0 \leq t \leq n-1)$, the two subsets are distinguishable if $S \neq T$. If $S = T$, then we must have $S = T = \emptyset$ and, without loss of generality, we may assume that $i < j \leq m-2$.

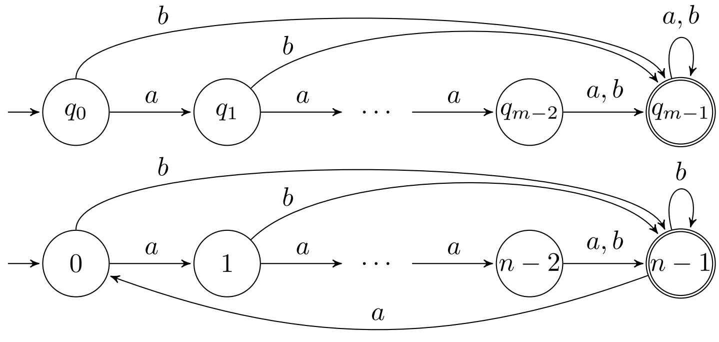


Figure 1: The minimal DFAs $A$ and $B$ with $\mathrm{sc}(L(A)L(B)) = m + n$.

After reading the string $a^{m-1-j}$, we get

$$\{q_i\} \xrightarrow{a^{m-1-j}} \{q_{m-1-(j-i)}\} \text{ and } \{q_j\} \xrightarrow{a^{m-1-j}} \{q_{m-1}, 0\}.$$

The resulting subsets are distinguishable since they differ in a state of $B$. This proves distinguishability and concludes the proof. □

Now, our aim is to prove that all the complexities from $m + n + 1$ up to the upper bound $f(m, n)$ can be produced for a linear alphabet in the case of $m, n \geq 2$.

To prove this, consider minimal DFAs $A = (\{q_0, q_1, \ldots, q_{m-1}\}, \Sigma, \cdot, q_0, \{q_{m-1}\})$, and $B = (\{0, 1, \ldots, n-1\}, \Sigma, \cdot, 0, \{1\})$, where $\{a, b, c, a_2, a_3, \ldots, a_{n-1}\} \subseteq \Sigma$. Notice that the initial state of $A$ is $q_0$ and the sole final state is $q_{m-1}$, while the initial state of $B$ is 0 and its sole final state is 1. Construct an NFA $N$ for $L(A)L(B)$ as described in Section 3. Let $D$ be the subset automaton of $N$, and $\mathcal{R}$ the family of all the reachable subsets in $D$. We assume that $A, B, N, D$, and $\mathcal{R}$ satisfy the following four conditions.

(1) The transitions on symbols $a, b, c$ in states in $\{q_0, q_1, \ldots, q_{m-1}\} \cup \{0, 1\}$ are defined as in Figure 2, that is,

$$\begin{aligned}
&q_i \cdot a = q_{i+1} \text{ if } 0 \leq i \leq m-2 \text{ and } q_{m-1} \cdot a = q_{m-1}, \\
&q_i \cdot b = q_{m-2} \text{ if } 0 \leq i \leq m-3, q_{m-2} \cdot b = q_{m-2}, \text{ and } q_{m-1} \cdot b = q_{m-1}, \\
&q_i \cdot c = q_{m-2} \text{ if } 0 \leq i \leq m-3, q_{m-2} \cdot c = q_{m-1}, \text{ and } q_{m-1} \cdot c = q_{m-2},
\end{aligned}$$

$$\begin{aligned}
&0 \cdot a = 0 \text{ and } 1 \cdot a = 1, \\
&0 \cdot b = 1 \text{ and } 1 \cdot b = 1, \\
&0 \cdot c = 1 \text{ and } 1 \cdot c = 0.
\end{aligned}$$

(2) If $(q_i, \sigma, q_0)$ is a transition in $A$ for some $\sigma$ in $\Sigma$, then $i = 0$ or $i = m-1$.

(3) In the subset automaton $D$, each set in $\mathcal{R} \setminus \{\{q_0\}\}$ is reachable from $\{q_1, 0\}$ if $m = 2$ and from $\{q_1\}$ if $m \geq 3$.

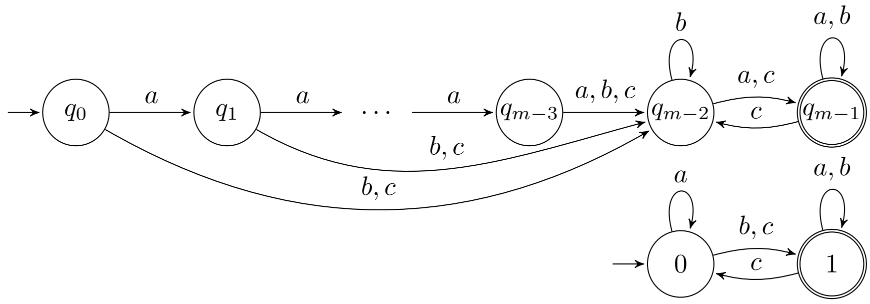


Figure 2: Transitions on $a, b, c$ for states in $\{q_0, q_1, \ldots, q_{m-1}\} \cup \{0, 1\}$.

(4) For each state $q$ of $N$, there exists a string $w_q$ in $\Sigma^*$ accepted by $N$ from and only from the state $q$. Moreover, we have

$$\begin{aligned} w_1 &= \varepsilon, \\ w_0 &= c, \\ w_j &= a_j \text{ for } j = 2, 3, \ldots, n-1, \\ w_{q_{m-1}} &= bc, \\ w_{q_{m-2}} &= cbc, \text{ and} \\ w_{q_i} &= a^{m-2-i}cbc \text{ for } i = 0, 1, \ldots, m-3. \end{aligned}$$

**Proposition 7.** *Let $m, n \geq 2$. Let $A$, $B$, $N$, $D$, and $\mathcal{R}$ satisfy (1)–(4). Then*

*(a) The sets $\{q_{m-1}, 0\}$, $\{q_{m-1}, 0, 1\}$, $\{q_{m-2}, 0, 1\}$ are in $\mathcal{R}$. If $m \geq 3$, then $\{q_1\}$ is in $\mathcal{R}$ as well.*

*(b) The initial subset $\{q_0\}$ of the subset automaton $D$ cannot be reached from any other reachable subset of $D$.*

*(c) The subset automaton $D$ does not have equivalent states, so* $\mathrm{sc}(L(A)L(B)) = |\mathcal{R}|$.

*Proof.* (a) By condition (1), the transitions on $a, b, c$ are as depicted in Figure 2. The NFA $N$ for $L(A)L(B)$, restricted to states $q_i$ with $0 \leq i \leq m-1$, 0, and 1, and to symbols $a, b, c$, is shown in Figure 3. If $m = 2$, then $m-1 = 1, m-2 = 0$, and in the subset automaton $D$ of the NFA $N$, we have $\{q_0\} \xrightarrow{a} \{q_1, 0\} \xrightarrow{b} \{q_1, 0, 1\} \xrightarrow{c} \{q_0, 0, 1\}$. If $m \geq 3$, then $\{q_0\} \xrightarrow{a} \{q_1\} \xrightarrow{a^{m-2}} \{q_{m-1}, 0\} \xrightarrow{b} \{q_{m-1}, 0, 1\} \xrightarrow{c} \{q_{m-2}, 0, 1\}$.

(b) Assume, for contradiction, that there is a set $R$ in $\mathcal{R} \setminus \{\{q_0\}\}$ and a symbol $\sigma$ such that $R \cdot \sigma = \{q_0\}$. Then $R = \{q_i\} \cup S$, where $q_i$ is a state of $A$ and $S$ is a set of states of $B$. The set $S$ must be empty because otherwise $R \cdot \sigma$ would contain a state

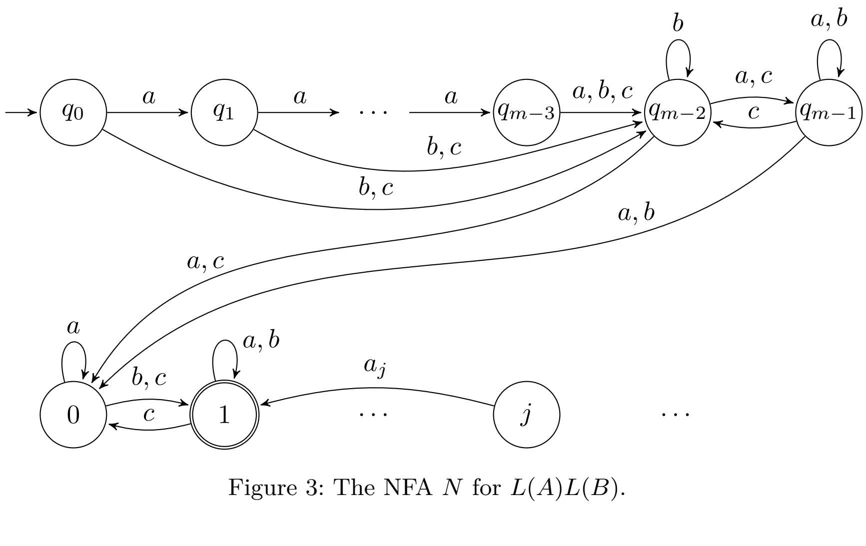


Figure 3: The NFA $N$ for $L(A)L(B)$.

of $B$. By condition (2), we must have $i = 0$ or $i = m-1$. However, if $i = m-1$ then $S$ must contain the state 0, thus $S$ cannot be empty. Hence $R = \{q_0\}$, a contradiction.

(c) By (4), the NFA $N$ satisfies the condition in Proposition 1. Therefore the subset automaton $D$ of $N$ does not have equivalent states, and $\mathrm{sc}(L(A)L(B)) = |\mathcal{R}|$. $\square$

Now let $A, B, N, D, \mathcal{R}$ satisfy conditions (1)–(4). Our goal is to construct a minimal $m$-state DFA $A_i$ and a minimal $(n+1)$-state DFA $B_i$, for $i = 1, 2, 3$, over the alphabet $\Sigma \cup \{a_n, b_n\}$ in such a way that $A_i$ and $B_i$, the NFA $N_i$ for $L(A_i)L(B_i)$, the subset automaton $D_i$ of $N_i$ and the family $\mathcal{R}_i$ of reachable states of $D_i$ satisfy conditions (1)–(4). Moreover, if $|\mathcal{R}| = \alpha$, then we want to have $|\mathcal{R}_1| = 2\alpha$, $|\mathcal{R}_2| = 2\alpha - 1$, and $|\mathcal{R}_3| = \alpha + 1$.

We construct minimal DFAs $A_i$ and $B_i$ $(i = 1, 2, 3)$ from automata $A$ and $B$ by adding a new state $n$ to the DFA $B$, and by adding transitions on two new symbols $a_n$ and $b_n$. The transitions on $a_n$ are the same in all three constructions, and they guarantee that the string $a_n$ is accepted by $N_i$ only from the state $n$. The transitions on $b_n$ are used to reach the set $\{q_0, n\}$ in $D_1$, the set $\{q_1, n\}$ in $D_2$ and the set $\{q_{m-1}, 0, n\}$ in $D_3$. We have to be careful with condition (4), especially in the third construction.

**Construction 1.** $(\alpha \to 2\alpha)$
Construct DFAs $A_1$ and $B_1$ from DFAs $A$ and $B$ as follows:

(1) add a new state $n$ to DFA $B$ going to itself on each old symbol $\sigma$ in $\Sigma$;

(2) add the transitions on two new symbols $a_n$ and $b_n$ as shown in Table 1 in column C1, that is,

$q_i \cdot a_n = q_{m-1}$ $(0 \le i \le m-1)$, $j \cdot a_n = 0$ $(0 \le j \le n-1)$, and $n \cdot a_n = 1$;
$q_i \cdot b_n = q_{m-1}$ $(0 \le i \le m-2)$, $q_{m-1} \cdot b_n = q_0$, and $j \cdot b_n = n$ $(1 \le j \le n)$.

**Construction 2.** $(\alpha \to 2\alpha - 1)$
Construct DFAs $A_2$ and $B_2$ from DFAs $A$ and $B$ as follows:

(1) add a new state $n$ to DFA $B$ going to itself on each old symbol $\sigma$ in $\Sigma$;

(2) add the transitions on two new symbols $a_n$ and $b_n$ as shown in Table 1 in column C2, that is, the transitions on $a_n$ are the same as in Construction 1, and

$q_i \cdot b_n = q_{m-1}$ $(0 \le i \le m-2)$, $q_{m-1} \cdot b_n = q_1$, and $j \cdot b_n = n$ $(1 \le j \le n)$.

**Construction 3.** $(\alpha \to \alpha + 1)$
Construct DFAs $A_3$ and $B_3$ from DFAs $A$ and $B$ as follows:

(1) add a new state $n$ to DFA $B$ with $n \cdot c = 0$ and $n \cdot \sigma = 0 \cdot \sigma$ if $\sigma \in \Sigma \setminus \{c\}$;

(2) add the transitions on two new symbols $a_n$ and $b_n$ as shown in Table 1 in column C3, that is, the transitions on $a_n$ are the same as in Construction 1, and

$q_i \cdot b_n = q_{m-1}$ $(0 \le i \le m-1)$ and $j \cdot b_n = n$ $(1 \le j \le n)$.

Now we prove that all three constructions preserve conditions (1)–(4).

**Lemma 8.** *Let $A, B, N, D, \mathcal{R}$ satisfy conditions (1)–(4). Let $A_i, B_i$ for $i = 1, 2, 3$ be the DFAs resulting from Constructions 1, 2, 3, respectively. Let $N_i$ be an NFA for $L(A_i)L(B_i)$ constructed as described in Section 3, $D_i$ be the corresponding subset*

Table 1: New transitions; $i \in \{0, 1, \ldots, m-1\}$, $j \in \{0, 1, \ldots, n-1\}$.

| | C1 | C2 | C3 |
|---|---|---|---|
| $\sigma \in \Sigma$ | $n \to n$ | $n \to n$ | $n \xrightarrow{c} 0$ |
| | | | $n \xrightarrow{\sigma} 0 \cdot \sigma$ if $\sigma \neq c$ |
| $a_n$ | $q_i \to q_{m-1}$ | $q_i \to q_{m-1}$ | $q_i \to q_{m-1}$ |
| | $n \to 1$ | $n \to 1$ | $n \to 1$ |
| | $j \to 0$ | $j \to 0$ | $j \to 0$ |
| $b_n$ | $q_{m-1} \to q_0$ | $q_{m-1} \to q_1$ | |
| | $q_i \to q_{m-1}$ if $i \neq m-1$ | $q_i \to q_{m-1}$ if $i \neq m-1$ | $q_i \to q_{m-1}$ |
| | $n \to n$ | $n \to n$ | $n \to n$ |
| | $j \to n$ | $j \to n$ | $j \to n$ |

*automaton, and $\mathcal{R}_i$ be the family of all the reachable subsets in DFA $D_i$. Then all these automata satisfy conditions (1)–(4). Moreover, if $|\mathcal{R}| = \alpha$, then $|\mathcal{R}_1| = 2\alpha$, $|\mathcal{R}_2| = 2\alpha - 1$, and $|\mathcal{R}_3| = \alpha + 1$. Next, if $A$ and $B$ are minimal, then $A_i$ and $B_i$, $i = 1, 2, 3$, are minimal as well.*

*Proof.* Since we do not change transitions on symbols in $\Sigma$ on states of $A$ and $B$, condition (1) is satisfied in each $A_i$ and $B_i$. Since the only new transition to $q_0$ is $(q_{m-1}, b_n, q_0)$ in Construction 1, condition (2) is satisfied in each $A_i$.

In each $N_i$, the string $a_n$ is accepted only from the state $n$. Moreover, in $B_1$ and $B_2$, the state $n$ goes to itself on each symbol in $\Sigma$. In $B_3$, we have $n \cdot c = 0$ and $n \cdot b = 0 \cdot b = 1$. It follows that the transitions $(0, c, 1)$ and $(q_{m-1}, b, 0)$, which were unique in-transitions in $N$, remain unique in each $N_i$. It follows that each $N_i$ satisfies condition (4). All the unique in-transitions are illustrated in Figure 4.

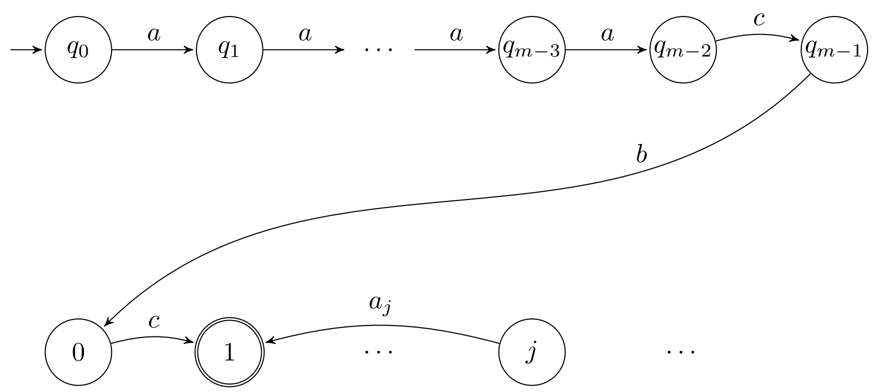


Figure 4: Unique in-transitions in the NFAs $N_i$.

Now consider the subset automata $D_1, D_2, D_3$. Since we did not change transitions on symbols in $\Sigma$ for states in $A$ and $B$, and the initial subset of $D$ and each $D_i$ is $\{q_0\}$,

we have $\mathcal{R} \subseteq \mathcal{R}_i$ for $i = 1, 2, 3$. Let us show that
$\mathcal{R}_1 = \mathcal{R} \cup \{ S \cup \{n\} \mid S \in \mathcal{R} \}$,
$\mathcal{R}_2 = \mathcal{R} \cup \{ S \cup \{n\} \mid S \in \mathcal{R} \text{ and } S \neq \{q_0\} \}$,
$\mathcal{R}_3 = \mathcal{R} \cup \{ \{q_{m-1}, 0, n\} \}$.
If $S$ is in $\mathcal{R}$ then $S$ is reachable in $D$, so $S$ can be reached from the initial subset $\{q_0\}$ by a string $u_S$ over $\Sigma$. If $S \neq \{q_0\}$, then, by (3), $S$ is reached from $\{q_1, 0\}$ if $m = 2$ and from $\{q_1\}$ if $m \geq 3$ by a string $v_S$.

In the subset automaton $D_1$ of $N_1$, we have

$$\begin{array}{ll} \{q_0\} \xrightarrow{a} \{q_1, 0\} \xrightarrow{b_n} \{q_0, n\} \xrightarrow{u_S} S \cup \{n\} & \text{if } m = 2, \\ \{q_0\} \xrightarrow{a} \{q_1\} \xrightarrow{a^{m-2}} \{q_{m-1}, 0\} \xrightarrow{b_n} \{q_0, n\} \xrightarrow{u_S} S \cup \{n\} & \text{if } m \geq 3. \end{array}$$

Thus $\mathcal{R} \cup \{ S \cup \{n\} \mid S \in \mathcal{R} \} \subseteq \mathcal{R}_1$, and every new set $S \cup \{n\}$ can be reached from $\{q_1, 0\}$ or from $\{q_1\}$, respectively. Let us show that no other set is reachable in $D_1$. For each set $S$ in $\mathcal{R}$ and each $\sigma$ in $\Sigma$, we have
$S \cdot \sigma \in \mathcal{R}$,
$S \cdot a_n = \{q_{m-1}, 0\}$,
$S \cdot b_n \in \{\{q_0, n\}, \{q_{m-1}, 0\}, \{q_{m-1}, 0, n\}\}$,
$(S \cup \{n\}) \cdot \sigma = S \cdot \sigma \cup \{n\}$,
$(S \cup \{n\}) \cdot a_n = \{q_{m-1}, 0, 1\}$, and
$(S \cup \{n\}) \cdot b_n \in \{\{q_0, n\}, \{q_{m-1}, 0, n\}\}$.
Using Proposition 7(a), we get that all the resulting sets are in $\mathcal{R} \cup \{ S \cup \{n\} \mid S \in \mathcal{R} \}$. By Proposition 3, we have $\mathcal{R}_1 = \mathcal{R} \cup \{ S \cup \{n\} \mid S \in \mathcal{R} \}$. Moreover, $\mathcal{R}_1$ satisfies condition (3).

Next, in the subset automaton $D_2$ of $N_2$, we have

$$\begin{array}{ll} \{q_0\} \xrightarrow{a} \{q_1, 0\} \xrightarrow{b_n} \{q_1, 0, n\} \xrightarrow{v_S} S \cup \{n\} & \text{if } m = 2 \text{ and } S \neq \{q_0\}, \\ \{q_0\} \xrightarrow{a} \{q_1\} \xrightarrow{a^{m-2}} \{q_{m-1}, 0\} \xrightarrow{b_n} \{q_1, n\} \xrightarrow{v_S} S \cup \{n\} & \text{if } m \geq 3 \text{ and } S \neq \{q_0\}. \end{array}$$

So, every new set $S \cup \{n\}$ is reached from $\{q_1, 0\}$ or from $\{q_1\}$. The transitions on each $\sigma$ in $\Sigma$ and on $a_n$ are the same as in Construction 1, and for each $S$ in $\mathcal{R}$, we have
$S \cdot b_n \in \{\{q_1, 0\}, \{q_1, 0, n\}\}$ if $m = 2$, and
$S \cdot b_n \in \{\{q_1, n\}, \{q_{m-1}, 0\}, \{q_{m-1}, 0, n\}\}$ if $m \geq 3$,
$(S \cup \{n\}) \cdot b_n = \{q_1, 0, n\}$ if $m = 2$, and
$(S \cup \{n\}) \cdot b_n \in \{\{q_1, n\}, \{q_{m-1}, 0, n\}\}$ if $m \geq 3$.
All the resulting sets are in $\mathcal{R} \cup \{ S \cup \{n\} \mid S \in \mathcal{R} \text{ and } S \neq \{q_0\} \}$. By Proposition 3, we have $\mathcal{R}_2 = \mathcal{R} \cup \{ S \cup \{n\} \mid S \in \mathcal{R} \text{ and } S \neq \{q_0\} \}$. Moreover, $\mathcal{R}_2$ satisfies condition (3).

Finally, in the subset automaton $D_3$ of $N_3$, we have

$$\begin{array}{ll} \{q_0\} \xrightarrow{a} \{q_1, 0\} \xrightarrow{b_n} \{q_1, 0, n\} & \text{if } m = 2, \\ \{q_0\} \xrightarrow{a} \{q_1\} \xrightarrow{a^{m-2}} \{q_{m-1}, 0\} \xrightarrow{b_n} \{q_{m-1}, 0, n\} & \text{if } m \geq 3. \end{array}$$

Thus the new set $\{q_{m-1}, 0, n\}$ is reached from $\{q_1, 0\}$ or from $\{q_1\}$. The transitions on $a_n$ are the same as above, and for each $S$ in $\mathcal{R}$ and each $\sigma$ in $\Sigma$, we have $S \cdot b_n \in \{\{q_{m-1}, 0\}, \{q_{m-1}, 0, n\}\}$. Next, for the new set $\{q_{m-1}, 0, n\}$, we have
$\{q_{m-1}, 0, n\} \cdot c = \{q_{m-2}, 0, 1\}$,
$\{q_{m-1}, 0, n\} \cdot \sigma = \{q_{m-1}, 0\} \cdot \sigma$ if $\sigma \in \Sigma$ and $\sigma \neq c$;
$\{q_{m-1}, 0, n\} \cdot b_n = \{q_{m-1}, 0, n\}$.
All the resulting subsets are in $\mathcal{R} \cup \big\{\{q_{m-1}, 0, n\}\big\}$. By Proposition 3, we have $\mathcal{R}_3 = \big\{\mathcal{R} \cup \{q_{m-1}, 0, n\}\big\}$, and again, $\mathcal{R}_3$ satisfies condition (3).

Finally, let $A$ and $B$ be minimal. Then each $A_i$ is minimal since we do not change the transitions on symbols in $\Sigma$ in states of $A$. Each $B_i$ is minimal since $B$ is minimal, the state $n$ is reachable in $B_i$, and the string $a_n$ is accepted by $B_i$ only from the state $n$. Our proof is complete. $\square$

Recall that $f(m, n) = m2^n - 2^{n-1}$ is the state complexity of concatenation if $n \geq 2$. Our aim is to show that each value in the range from $m + n + 1$ to $f(m, n)$ may be attained by the state complexity of concatenation of $m$-state and $n$-state DFA languages provided that $m \geq 2$. We are going to show this by induction. The next lemma proves the base case for this induction.

**Lemma 9.** *Let $m \geq 2$ and $n = 2$. For each $\alpha$ with $m + 3 \leq \alpha \leq f(m, 2) = 4m - 2$, there exist a minimal $m$-state DFA $A$ and a minimal 2-state DFA $B$, both defined over an alphabet $\Sigma$ with $|\Sigma| \leq 7$, such that $\mathrm{sc}(L(A)L(B)) = \alpha$. Moreover, automata $A$ and $B$, the corresponding NFA $N$ for $L(A)L(B)$, the subset automaton $D$ of $N$, and the set $\mathcal{R}$ of reachable states of $D$ satisfy conditions (1)–(4) on page 129.*

*Proof.* First, let $\alpha = m + 3$. Define a minimal $m$-state DFA

$$A_0 = (\{q_0, q_1, \ldots, q_{m-1}\}, \{a, b, c\}, \cdot, q_0, \{q_{m-1}\}),$$

where for each $i$ in $\{0, 1, \ldots, m - 1\}$,
$q_i \cdot a = q_{i+1}$ if $i \neq m - 1$ and $q_{m-1} \cdot a = q_{m-1}$,
$q_i \cdot b = q_{m-2}$ if $i \neq m - 1$ and $q_{m-1} \cdot b = q_{m-1}$, and
$q_i \cdot c = q_{m-2}$ if $i \neq m - 2$ and $q_{m-2} \cdot c = q_{m-1}$.
Define a minimal two-state DFA $B_0 = (\{0, 1\}, \{a, b, c\}, \cdot, 0, \{1\})$ where
$0 \cdot a = 0$ and $1 \cdot a = 1$,
$0 \cdot b = 1$ and $1 \cdot b = 1$,
$0 \cdot c = 1$ and $1 \cdot c = 0$.
Hence the transitions on $a, b, c$ are defined as in condition (1). Condition (2) is satisfied as well. Construct NFA $N_0$ for $L(A_0)L(B_0)$ as shown in Figure 5.

Notice that the transitions $(q_i, a, q_{i+1})$ $(0 \leq i \leq m-3)$, $(q_{m-2}, c, q_{m-1})$, $(q_{m-1}, b, 0)$, and $(0, c, 1)$ are unique in-transitions. The state 1 is a sole final state of $N_0$. It follows that $N_0$ satisfies condition (4). In the subset automaton $D_0$ of $N_0$, we have
$\{q_0\} \xrightarrow{a^i} \{q_i\}$ for $i = 0, 1, \ldots, m - 2$,
$\{q_{m-2}\} \xrightarrow{a} \{q_{m-1}, 0\} \xrightarrow{b} \{q_{m-1}, 0, 1\} \xrightarrow{c} \{q_{m-2}, 0, 1\} \xrightarrow{b} \{q_{m-2}, 1\}$.
Thus the subset automaton has $m + 3$ reachable subsets. Notice that each of these $m + 3$ subsets goes to some of them by each symbol in $\{a, b, c\}$. By Proposition 3, no

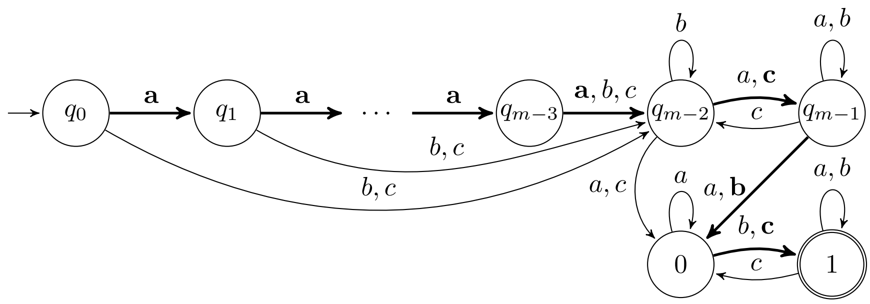


Figure 5: The NFA $N_0$ for $L(A_0)L(B_0)$ in the base case.

other set is reachable, so the complexity of $L(A_0)L(B_0)$ is $m+3$. Moreover, condition (3) is satisfied as well.

Now we consider the values $\alpha = i(m-2)+6$ for $i = 1,2,3,4$. We construct appropriate DFAs from automata $A_0$ and $B_0$ by adding transitions on new symbols. Thus we do not change the transitions on symbols $a,b,c$, and therefore the conditions (1) and (4) are always satisfied. Moreover, for each new symbol, the new transition is defined in such a way that condition (2) is satisfied as well. Finally, notice that $\{q_{m-1},0\}$ is reachable from $\{q_1\}$ by $a^{m-2}$ in the subset automaton $D_0$ if $m \geq 3$. In what follows, we always reach new subsets in the corresponding subset automata for concatenation from the subset $\{q_{m-1},0\}$. Hence condition (3) is always satisfied.

First let $i=1$, so $\alpha = (m-2)+6 = m+4$. Construct DFAs $A_{1,0}$ and $B_{1,0}$ from DFAs $A_0$ and $B_0$, respectively, by adding the transitions on a new symbol $d$ defined by

$$q_i \cdot d = q_{m-2} \ (0 \leq i \leq m-1);$$
$$0 \cdot d = 0 \text{ and } 1 \cdot d = 1.$$

Construct the NFA $N_{1,0}$ for $L(A_{1,0})L(B_{1,0})$. In the subset automaton $D_{1,0}$, all the sets that were reachable in the subset automaton $D_0$ are reachable as well, since the transitions on the old symbols $a,b,c$ are the same. In $D_{1,0}$ we have

$$\{q_{m-1},0\} \xrightarrow{d} \{q_{m-2},0\}.$$

No other set is reachable, so the complexity of $L(A_{1,0})L(B_{1,0})$ is $m+4$. Notice that all the possible subsets containing states $q_{m-1}$ and $q_{m-2}$ are reachable in $D_{1,0}$.

Let $\alpha = 2(m-2)+6$. Construct DFAs $A_{2,0}, B_{2,0}$ from DFAs $A_{1,0}, B_{1,0}$ by adding the transitions on a new symbol $e_0$ defined by

$$q_i \cdot e_0 = q_{m-1} \ (0 \leq i \leq m-2),$$
$$q_{m-1} \cdot e_0 = q_0;$$
$$0 \cdot e_0 = 0,$$
$$1 \cdot e_0 = 0.$$

Construct the NFA $N_{2,0}$ for $L(A_{2,0})L(B_{2,0})$. In the subset automaton $D_{2,0}$, all the sets that were reachable in the subset automaton $D_{1,0}$ are reachable as well, since the transitions on the old symbols $a, b, c, d$ are the same. In $D_{2,0}$ we have

$$\{q_{m-1}, 0\} \xrightarrow{e_0} \{q_0, 0\} \xrightarrow{a^i} \{q_i, 0\} \text{ for } i = 1, 2, \ldots, m-3.$$

No other new set is reachable since each set $\{q_i, 0\}$ goes either to a set $\{q_j, 0\}$ or to a set containing $q_{m-2}$ or $q_{m-1}$ by each symbol in $\{a, b, c, d, e_0\}$; and moreover, by $e_0$, each set goes either to $\{q_0, 0\}$ or to a set containing $q_{m-1}$. Therefore the resulting complexity of the concatenation $L(A_{2,0})L(B_{2,0})$ is $2(m-2)+6$.

In a similar way, we construct DFAs $A_{3,0}, B_{3,0}$ from $A_{2,0}, B_{2,0}$ by adding transitions on a new symbol $e_{01}$ defined by

$$\begin{aligned}
&q_i \cdot e_{01} = q_{m-1} \ (0 \le i \le m-2),\\
&q_{m-1} \cdot e_{01} = q_0;\\
&0 \cdot e_{01} = 0,\\
&1 \cdot e_{01} = 1.
\end{aligned}$$

This results in the reachability of $m-2$ new subsets $\{q_i, 0, 1\}$ in the subset automaton of $N_{3,0}$. Since no other new set is reachable, the complexity of $L(A_{3,0})L(B_{3,0})$ is $3(m-2)+6$.

Finally, construct DFAs $A_{4,0}, B_{4,0}$ from $A_{3,0}, B_{3,0}$ by adding the transitions on a new symbol $e_1$ defined by

$$\begin{aligned}
&q_i \cdot e_1 = q_{m-1} \ (0 \le i \le m-2),\\
&q_{m-1} \cdot e_1 = q_0;\\
&0 \cdot e_1 = 1,\\
&1 \cdot e_1 = 1.
\end{aligned}$$

This results in the reachability of subsets $\{q_i, 1\}$ in the subset automaton of $N_{4,0}$, and the complexity of $L(A_{4,0})L(B_{4,0})$ is $4(m-2)+6$.

Up to now we have defined appropriate minimal DFAs $A_{i,0}$ and $B_{i,0}$ for the values $\alpha = i(m-2)+6$ for $i = 1, 2, 3, 4$. Now let us consider an intermediate value $\alpha = i(m-2)+6+j$ where $1 \le i \le 3$ and $1 \le j \le m-3$. Construct DFAs $A_{i,j}$ and $B_{i,j}$ from DFAs $A_{i,0}$ and $B_{i,0}$ by adding the transitions on a new symbol $f_1$ defined by

$$\begin{aligned}
&q_i \cdot f_1 = q_{m-1} \ (0 \le i \le m-2),\\
&q_{m-1} \cdot f_1 = q_{m-2-j};\\
&0 \cdot f_1 = 1,\\
&1 \cdot f_1 = 1.
\end{aligned}$$

This results in the reachability of the following $j$ new subsets in the subset automaton of $N_{i,j}$:

$$\{q_{m-1}, 0\} \xrightarrow{f_1} \{q_{m-2-j}, 1\} \xrightarrow{a} \{q_{m-2-j+1}, 1\} \xrightarrow{a} \cdots \xrightarrow{a} \{q_{m-3}, 1\}.$$

Recall that the subset automaton of $N_{i,0}$ has $i(m-2)+6$ reachable states, and since $i \leq 3$, the subsets $\{q_i, 1\}$ are unreachable in the subset automaton of $N_{i,0}$. Hence the resulting complexity of of $L(A_{i,j})L(B_{i,j})$ is $i(m-2)+6+j$ as desired. Moreover, all the automata satisfy conditions (1)–(4). This concludes our proof. $\square$

Now we are ready to prove the main lemma. Recall that the state complexity of concatenation is $f(m,n) = (m-1)2^n + 2^{n-1}$ if $n \geq 2$. Moreover, notice that we have

$$f(m, n+1) = (m-1)2^{n+1} + 2^n = 2((m-1)2^n + 2^{n-1}) = 2f(m,n).$$

**Lemma 10.** *Let $m \geq 2$ and $n \geq 2$. For each $\alpha$ with $m+n+1 \leq \alpha \leq f(m,n)$, there exist a minimal $m$-state DFA $A$ and a minimal $n$-state DFA $B$, both defined over an alphabet $\Sigma$ with $|\Sigma| \leq 2n+4$, such that* $\mathrm{sc}(L(A)L(B)) = \alpha$.

*Proof.* We prove the claim by induction on $n$. Moreover, in the induction hypothesis, we assume that DFAs $A$ and $B$, the corresponding NFA $N$ for $L(A)L(B)$ constructed as in Section 3, the subset automaton $D$ of $N$, and the set $\mathcal{R}$ of reachable states of $D$ satisfy conditions (1)–(4) on page 129.

The basis, in which we have $m \geq 2$, $n = 2$, and $m+3 \leq \alpha \leq f(m,2) = 4m-2$, is proved in Lemma 9.

Let $m \geq 2$, $n \geq 2$, and assume that for each $\beta$ with $m+n+1 \leq \beta \leq f(m,n)$, there exist a minimal $m$-state DFA $A$ and a minimal $n$-state DFA $B$, both defined over an alphabet $\Sigma$ with $|\Sigma| \leq 2n+4$, such that $\mathrm{sc}(L(A)L(B)) = \beta$. Moreover, assume that DFAs $A$ and $B$, the NFA $N$ for $L(A)L(B)$, the subset automaton $D$ of $N$, and the set of reachable states $\mathcal{R}$ of $D$ satisfy conditions (1)–(4) on page 129. Let us show that the claim holds for $n+1$. To this aim let $\alpha$ be an integer with $m+(n+1)+1 \leq \alpha \leq f(m,n+1)$.

First, let $2m+2n+2 \leq \alpha \leq f(m,n+1)$ and $\alpha$ be even. Let $\beta = \alpha/2$. Then $m+n+1 \leq \beta \leq f(m,n)$, and by the induction hypothesis, there exist a minimal $m$-state DFA $A$ and a minimal $n$-state DFA $B$, both defined over an alphabet $\Sigma$ with $|\Sigma| \leq 2n+4$, such that $\mathrm{sc}(L(A)L(B)) = \beta$. Moreover, conditions (1)–(4) are satisfied for $A, B, N, D, \mathcal{R}$. We use Construction 1, in which we add a new state to DFA $B$ and the transitions on two new symbols to get a minimal $m$-state $A_1$ and a minimal $(n+1)$-state DFA $B_1$. By Lemma 8, all conditions (1)–(4) are satisfied for $A_1, B_1, N_1, D_1$, and $R_1$. It follows that $\mathrm{sc}(L(A_1)L(B_1)) = 2\beta = \alpha$.

Now, let $2m+2n+1 \leq \alpha \leq f(m,n+1)-1$ and $\alpha$ be odd. In this case we set $\beta = (\alpha+1)/2$. Then $m+n+1 \leq \beta \leq f(m,n)$, and we use the induction hypothesis and our Construction 2 to get automata $A_2$ and $B_2$ over $\Sigma \cup \{a_n, b_n\}$ satisfying (1)–(4) such that $\mathrm{sc}(L(A_2)L(B_2)) = 2\beta - 1 = \alpha$.

Finally, if $m+(n+1)+1 \leq \alpha \leq 2m+2n$, we set $\beta = \alpha - 1$. Then we have $m+n+1 \leq \beta \leq 2m+2n-1$. Let us show that $2m+2n-1 \leq f(m,n)$ whenever $m \geq 3$ and $n \geq 2$, or if $m \geq 2$ and $n \geq 3$.

If $m \geq 3$ and $n \geq 2$, then $m-1 \geq 2$, $2^{n-2} \geq 1$, $2^{n-1} \geq n$, and therefore

$$f(m,n) = m2^n - 2^{n-1} = 2m2^{n-2} + (m-1)2^{n-1} \geq 2m + 2n - 1.$$

If $m = 2$ and $n \geq 3$, then $f(m,n) = 2^{n+1} - 2^{n-1} \geq 3 \cdot 2^{n-1} \geq 3n \geq 2n+3$. Thus if $m \geq 3$ and $n \geq 2$ or $m \geq 2$ and $n \geq 3$, then we use the induction hypothesis and

Construction 3 to get appropriate automata $A_3$ and $B_3$ satisfying (1)–(4) such that $\mathrm{sc}(L(A_3)L(B_3)) = \beta + 1 = \alpha$.

It remains to solve the case of $m = 2$ and $n = 3$. The base case, in which we have $m = 2$, $n = 2$, and $\alpha \in \{5, 6\}$, and our three constructions provide appropriate automata for $m = 2, n = 3$, and $\alpha \in \{6, 7, 9, 10, 11, 12\}$. Thus we need to get automata for $m = 2, n = 3$, and $\alpha = 8$. Such DFAs are shown in Figure 6. Notice that (1)–(4) are satisfied here. This concludes our proof. □

The case of $m = 1$ and $n \geq 2$ is slightly different, although the main idea is the same. Recall that $f(m, n) = m2^n - 2^{n-1}$ is the state complexity of the concatenation operation if $n \geq 2$. Next, recall that the case of $1 \leq \alpha \leq m + n - 1$ is covered by Lemma 5.

**Lemma 11.** *Let $m = 1$ and $n \geq 2$. For each $\alpha$ with $n + 1 \leq \alpha \leq f(1, n) = 2^{n-1}$, there exist a minimal 1-state DFA $A$, and a minimal $n$-state DFA $B$, both defined over an alphabet $\Sigma$ with $|\Sigma| \leq 2n - 3$, such that $\mathrm{sc}(L(A)L(B)) = \alpha$.*

*Proof.* First notice that for $n = 2$, we have $n + 1 = 3$ and $f(1, n) = 2$. This means that no value of $\alpha$ satisfies the condition in the lemma. Therefore, the lemma holds true for $n = 2$. In what follows we assume that $n \geq 3$. Let $\{a, b, c\} \subseteq \Sigma$.

Let $A$ be a 1-state DFA accepting $\Sigma^*$. Let $B = (\{0, 1, \dots, n-1\}, \Sigma, \cdot, 0, \{1\})$ be a minimal $n$-state DFA with the initial state 0 and the sole final state 1. The NFA $N$ for $\Sigma^* B$ can be constructed from $B$ by adding a loop in the initial state 0 on each input symbol in $\Sigma$. Let $D$ be the subset automaton of $N$, and $\mathcal{R}$ be the family of all the reachable subsets in $D$. We prove the lemma by induction on $n$ again, where we assume that the following conditions hold for $B, N, D$, and $\mathcal{R}$:

(1') In the DFA $B$, the transitions on $a, b, c$ in states $0, 1, 2$ are as in Figure 7.

(2') In the DFA $B$, we have $0 \cdot \sigma \neq 0$ for each $\sigma \in \Sigma$.

(3') In the subset automaton $D$, each subset in $\mathcal{R} \setminus \{\{0\}\}$ is reached from $\{0, 1\}$.

(4') The NFA $N$ satisfies the condition in Proposition 1, that is, for each state $j$ of $N$, there exists a string $w_j$ in $\Sigma^*$ which is accepted by $N$ from and only from

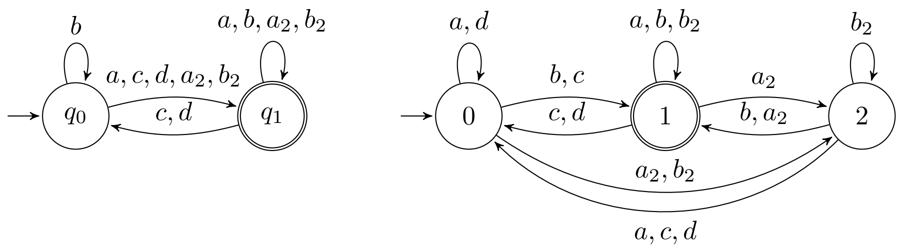


Figure 6: The minimal 2-state DFA $A$ and the minimal 3-state DFA $B$ with $\mathrm{sc}(L(A)L(B)) = 8$.

Table 2: The three constructions in the case of $m = 1$; $j \in \{0, 1, \dots, n-1\}$.

| | C1 | C2 | C3 |
|---|---|---|---|
| $\sigma \in \Sigma$ | $n \to n$ | $n \to n$ | $n \to 0$ |
| $a_n$ | $n \to 1$<br>$j \to n$ | $n \to 2$<br>$0 \to 1$<br>$1 \to n$<br>$j \to 0$ if $j \geq 2$ | $n \to 1$<br>$j \to 2$ |
| $b_n$ | $-$ | $-$ | $n \to n$<br>$j \to n$ |
| $w_n$ | $a_n$ | $a_n c$ | $a_n$ |

the state $j$. Moreover, we have

$$w_0 = a,$$
$$w_1 = \varepsilon,$$
$$w_2 = c.$$

For the induction step, we again describe three constructions: We construct $(n+1)$-state DFAs $B_1, B_2, B_3$ from DFA $B$ by adding a new state $n$, and by adding transitions on new symbol $a_n, b_n$, as shown in Table 2 in columns C1, C2, and C3, respectively. Next we show that for $i = 1, 2, 3$, the DFA $B_i$, the NFA $N_i$ for $\Sigma^* L(B_i)$, the subset automaton $D_i$ of $N_i$, and the family of all the reachable subsets in $D_i$ satisfy conditions (1')–(4'). Moreover, if $|\mathcal{R}| = \alpha$, then $|\mathcal{R}_1| = 2\alpha$, $|\mathcal{R}_2| = 2\alpha - 1$, $|\mathcal{R}_3| = \alpha + 1$.

Since we do not change the transitions on $a, b, c$ in states $0, 1, 2$ in any of our three constructions, each $B_i$ satisfies condition (1'). Condition (2') is satisfied for all symbols in $\Sigma$. Next, we have $0 \cdot a_n \neq 0$ and $0 \cdot b_n \neq 0$ in each of the three constructions. Hence each $B_i$ satisfies condition (2').

Since we do not change the transitions on symbols in $\Sigma$ in the states of $B$, we have $\mathcal{R} \subseteq \mathcal{R}_i$. Notice that the subsets $\{0,1\}, \{0,2\}$, and $\{0,1,2\}$ are in $\mathcal{R}$ since $\{0\} \xrightarrow{a} \{0,1\} \xrightarrow{b} \{0,1,2\}$ and $\{0\} \xrightarrow{c} \{0,2\}$. Next, each $S$ in $\mathcal{R}$ is reached in $D$ from $\{0\}$ by a string $u_S$ over $\Sigma$, and by condition (3'), each $S$ in $\mathcal{R} \setminus \{\{0\}\}$ is reached in $D$ from $\{0,1\}$ by a string $v_S$ over $\Sigma$.

Consider Construction 1 given by column C1. In $D_1$ we have

$$\{0\} \xrightarrow{a} \{0,1\} \xrightarrow{a_n} \{0,n\} \xrightarrow{u_S} S \cup \{n\}$$

for every $S \in \mathcal{R}$. Thus $\mathcal{R} \cup \{\, S \cup \{n\} \mid S \in \mathcal{R} \,\} \subseteq \mathcal{R}_1$. Let us show that no other set is reachable in $D_1$. For each $S$ in $\mathcal{R}$ and each $\sigma$ in $\Sigma$, we have $S \cdot \sigma \in \mathcal{R}$ and $S \cdot a_n = \{0, n\}$. Next $(S \cup \{n\}) \cdot \sigma = S \cdot \sigma \cup \{n\}$ and $(S \cup \{n\}) \cdot a_n = \{0, 1, n\}$, where $\{0,1\} \in \mathcal{R}$. Thus we have exactly $2|\mathcal{R}|$ reachable subsets in $D_1$. Next, each new subset $S \cup \{n\}$ is reachable from $\{0,1\}$. Therefore $D_1$ satisfies (3').

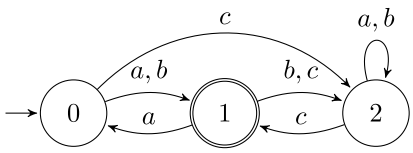


Figure 7: The base case if $m = 1$: The minimal 3-state DFA $B$ with $sc(\Sigma^* L(B)) = 4$.

Consider Construction 2 given by column C2. In $D_2$ we have

$$\{0\} \xrightarrow{a_n} \{0,1\} \xrightarrow{a_n} \{0,1,n\} \xrightarrow{v_S} S \cup \{n\}$$

where $S \neq \{0\}$. Thus $\mathcal{R} \cup \{\, S \cup \{n\} \mid S \in \mathcal{R} \setminus \{0\} \,\} \subseteq \mathcal{R}_2$, and each new subset $S \cup \{n\}$ is reachable from $\{0,1\}$. Let us show that no other set is reachable in $D_2$. For each $S$ in $\mathcal{R}$ and each $\sigma$ in $\Sigma$, we have $S \cdot \sigma \in \mathcal{R}$, $S \cdot a_n \in \{\{0,1\}, \{0,1,n\}\}$. Next $(S \cup \{n\}) \cdot \sigma = S \cdot \sigma \cup \{n\}$ where $S \cdot \sigma \neq \{0\}$ since $0 \cdot \sigma \neq 0$. Finally, $(S \cup \{n\}) \cdot a_n \in \{\{0,1,2\}, \{0,1,2,n\}\}$, where $\{0,1,2\}$ is in $\mathcal{R}$. Thus we have exactly $2|\mathcal{R}| - 1$ reachable subsets in $D_2$. Moreover, $D_2$ satisfies (3').

Consider Construction 3 given by column C3. In $D_3$ we have

$$\{0\} \xrightarrow{a} \{0,1\} \xrightarrow{b_n} \{0,n\}.$$

Thus $\mathcal{R} \cup \{\{0,n\}\} \subseteq \mathcal{R}_3$, and the new subset $\{0,n\}$ is reachable from $\{0,1\}$. Let us show that no other set is reachable in $D_3$. For each $S$ in $\mathcal{R}$ and each $\sigma$ in $\Sigma$, we have $S \cdot \sigma \in \mathcal{R}$, $S \cdot a_n = \{0,2\}$, and $S \cdot b_n = \{0,n\}$. Next $\{0,n\} \cdot \sigma = 0 \cdot \sigma$, $\{0,n\} \cdot a_n = \{0,1,2\}$ and $\{0,1,2\}$ is in $\mathcal{R}$, $\{0,n\} \cdot b_n = \{0,n\}$. Thus we have exactly $|\mathcal{R}| + 1$ reachable subsets in $D_3$, and $D_3$ satisfies condition (3').

Finally we show that each $N_i$ satisfies condition (4). For each state $j$ of $N_i$, except for the state $n$, we have the same string $w_j$ as in $N$. In $N_1$ and $N_3$, the state $n$ is the only state which goes to state 1 on $a_n$. Moreover, the state $n$ goes to itself on each symbol in $\Sigma$ in $N_1$ and the state $n$ goes to state 0 on each symbol in $\Sigma$ in $N_3$. It follows that in $N_1$ and $N_3$ we have $w_n = a_n$ and condition (4) is satisfied. In $N_2$, the state $n$ is the only state which goes to state 2 on $a_n$, and the state 2 is the only state which goes to state 1 on $c$. Therefore $w_n = a_n c$. The state $n$ goes to itself on each symbol in $\Sigma$. It follows that (4) is satisfied for $N_2$ as well.

Next, if $B$ is minimal then $B_i$, $i = 1,2,3$, is minimal as well since in $B_i$ the transitions on symbols in $\Sigma$ in states in $\{0,1,\ldots,n-1\}$ are the same as in $B$, and the string $w_n$ is accepted by $B_i$ only from the state $n$.

Now we are ready to prove the lemma by induction on $n$. The basis, in which we have $n = 3$ and $n + 1 = f(1,3) = 4$, holds true since the 3-state DFA $B$ shown in Figure 7 satisfies (1')–(4'). Moreover, the family of reachable subsets in the subset automaton $D$ of the NFA $N$ for $\Sigma^* L(B)$ is $\mathcal{R} = \{\{0\}, \{0,1\}, \{0,2\}, \{0,1,2\}\}$.

Now let $n \geq 3$ and assume that for each $\beta$ with $n + 1 \leq \beta \leq 2^{n-1}$, there exists a minimal $n$-state DFA $B$ such that $\mathrm{sc}(\Sigma^* L(B)) = \beta$. Moreover, assume that $B, N, D$, and $\mathcal{R}$ satisfy (1')–(4'). Let us show that the claim holds for $n + 1$. To this aim let $n + 2 \leq \alpha \leq 2^n$.

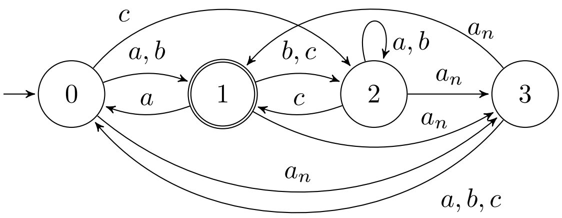


Figure 8: The minimal 4-state DFA $B$ with $\mathrm{sc}(\Sigma^* L(B)) = 6$.

First, let $2n+2 \leq \alpha \leq 2^n$ and $\alpha$ is even. Set $\beta = \alpha/2$. Then $n+1 \leq \beta \leq 2^{n-1}$, and we use the induction hypothesis and Construction 1 to get a minimal DFA $B_1$ satisfying (1')–(4') such that $\mathrm{sc}(\Sigma^* L(B_1)) = 2\beta = \alpha$.

Now, let $2n+1 \leq \alpha \leq 2^n - 1$ and $\alpha$ is odd. Set $\beta = (\alpha+1)/2$. Then we have $n+1 \leq \beta \leq 2^{n-1}$, and we use the induction hypothesis and Construction 2 to get a minimal DFA $B_2$ satisfying (1')–(4') such that $\mathrm{sc}(\Sigma^* L(B_2)) = 2\beta - 1 = \alpha$.

Finally, let $n+2 \leq \alpha \leq 2n$. Set $\beta = \alpha - 1$. Then $n+1 \leq \beta \leq 2n-1$. We have $2n-1 \leq 2^{n-1}$ whenever $n \geq 4$. In such a case, we can use the induction hypothesis and Construction 3 to get a minimal DFA $B_3$ satisfying (1')–(4') such that $\mathrm{sc}(\Sigma^* L(B_3)) = \beta + 1 = \alpha$.

To finish the proof, we need to solve the case of $n = 4$. Using the base case with $n = 3$ and $\alpha = 4$, and our three constructions, we get the values of $\alpha$ in $\{5, 7, 8\}$ for $n = 4$. It remains to get $\alpha = 6$. The DFA in Figure 8 shows such a 4-state DFA. Notice that (1')–(4') are satisfied. Our proof is complete. □

The next theorem summarizes our results, and shows that the whole range of complexities for the concatenation operation can be produced using an alphabet which grows linearly with $n$.

**Theorem 12.** *Let $m, n \geq 1$. Let $f(m,n)$ be the state complexity of the concatenation operation given by*

$$f(m,n) = \begin{cases} m, & \text{if } n = 1; \\ m2^n - 2^{n-1}, & \text{if } n \geq 2. \end{cases}$$

*For each $\alpha$ with $1 \leq \alpha \leq f(m,n)$, there exist regular languages $K$ and $L$ defined over an alphabet $\Sigma$ with $|\Sigma| \leq 2n+4$ such that $\mathrm{sc}(K) = m$, $\mathrm{sc}(L) = n$, and $\mathrm{sc}(KL) = \alpha$.*

*Proof.* Table 3 shows the lemmata dealing with particular cases. □

## 5. Conclusions

We investigated the state complexity of languages resulting from the concatenation operation. We proved that for all $m, n, \alpha$ with $m, n \geq 1$ and $1 \leq \alpha \leq f(m,n)$, where $f(m,n)$ is the state complexity of the concatenation operation, there exist regular

Table 3: The map of cases.

| | | | |
|---|---|---|---|
| $m \geq 1,$ | $n \geq 1$ | $1 \leq \alpha \leq m+n-1$ | Lemma 5 |
| $m = 1,$ | $n \geq 2$ | $m+n \leq \alpha \leq f(1,n)$ | Lemma 11 |
| $m \geq 2,$ | $n \geq 2$ | $\alpha = m+n$ | Lemma 6 |
| $m \geq 2,$ | $n \geq 2$ | $m+n+1 \leq \alpha \leq f(m,n)$ | Lemma 10 |

languages $K$ and $L$ defined over an alphabet of size at most $2n+4$ such that the minimal DFA for $K$ has $m$ states, the minimal DFA for $L$ has $n$ states, and the minimal DFA for $KL$ has $\alpha$ states. This improves the result from [6], where an alphabet of size growing exponentially with $n$ is used to produce the whole range of complexities for the concatenation operation. Our result complements similar results from [9, 15], where a linear alphabet is used to get the whole range of complexities for the reversal and Kleene closure operations.

A similar problem for the square operation, defined as $L^2 = LL$, remains open even for an exponential alphabet.

## References


[1] V. Geffert, Magic numbers in the state hierarchy of finite automata. *Inform. Comput.* **205** (2007), 1652–1670.

[2] M. Holzer, S. Jakobi, M. Kutrib, The magic number problem for subregular language families. *Int. J. Found. Comput. Sci.* **23** (2012), 115–131.

[3] K. Iwama, Y. Kambayashi, 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.

[4] K. Iwama, A. Matsuura, M. Paterson, A family of NFAs which need $2^n - \alpha$ deterministic states. *Theoret. Comput. Sci.* **301** (2003), 451–462.

[5] J. Jirásek, G. Jirásková, A. Szabari, State complexity of concatenation and complementation. *Int. J. Found. Comput. Sci.* **16** (2005), 511–529.

[6] G. Jirásková, Concatenation of regular languages and descriptional complexity. *Theory Comput. Syst.* **49** (2011), 306–318.

[7] G. Jirásková, Magic numbers and ternary alphabet. *Int. J. Found. Comput. Sci.* **22** (2011), 331–344.

[8] G. Jirásková, The ranges of state complexities for complement, star, and reversal of regular languages. *Int. J. Found. Comput. Sci.* **25** (2014), 101–124.

[9] G. Jirásková, M. Palmovský, J. Šebej, Kleene closure on regular and prefix-free languages. In: *Implementation and Application of Automata (CIAA 2014)*. LNCS 8587, Springer, 2014, 226–237.

[10] H. Jürgensen, R. Reis (eds.), *Descriptional Complexity of Formal Systems (DCFS 2013)*. LNCS 8031, Springer, 2013.

[11] A. Maslov, Estimates of the number of states of finite automata. *Soviet Math. Doklady* **11** (1970), 1373–1375.

[12] M. O. Rabin, D. S. Scott, Finite automata and their decision problems. *IBM J. Research Develop.* **3** (1959), 114–125.

[13] M. Sipser, *Introduction to the Theory of Computation*. PWS Publishing Company, 1997.

[14] K. Čevorová, Kleene star on unary regular languages. In: [10], 2013, 277–288.

[15] J. Šebej, Reversal on regular languages and descriptional complexity. In: [10], 2013, 265–276.

[16] S. Yu, Regular languages. In: *Handbook of Formal Languages*. Vol.1, Springer, 1997, 41–110.

[17] S. Yu, Q. Zhuang, K. Salomaa, The state complexities of some basic operations on regular languages. *Theoret. Comput. Sci.* **125** (1994), 315–328.

(Received: November 20, 2016; revised: May 1, 2017)