International Journal of Foundations of Computer Science

State Complexity of Two Combined Operations: Catenation-Union and Catenation-Intersection*

Bo Cui, Yuan Gao, Lila Kari, and Sheng Yu

Department of Computer Science,
The University of Western Ontario,
London, Ontario, Canada N6A 5B7
{*bcui2, ygao72, lila, syu*} *@csd.uwo.ca*

Received (Day Month Year)
Accepted (Day Month Year)
Communicated by (xxxxxxxxxx)

In this paper, we study the state complexities of two particular combinations of operations: catenation combined with union and catenation combined with intersection. We show that the state complexity of the former combined operation is considerably less than the mathematical composition of the state complexities of catenation and union, while the state complexity of the latter one is equal to the mathematical composition of the state complexities of catenation and intersection.

1. Introduction

State complexity is a type of descriptional complexity for regular languages based on the deterministic finite automaton (DFA) model [22]. The state complexity of an operation on regular languages is the number of states that are necessary and sufficient in the worst case for the minimal, complete DFA that accepts the resulting language of the operation [8]. Many results on the state complexities of individual operations have been obtained, e.g. union, intersection, catenation, star, etc [1, 2, 3, 4, 9, 11, 12, 15, 16, 18, 20, 22].

However, in practice, the operation to be performed is often a combination of several individual operations in a certain order, rather than only one individual operation. The research on state complexity of combined operations started in 2005. Up to now, a number of papers on this topic have been published [4, 5, 6, 7, 13, 14, 17, 19]. It has been shown that the state complexity of a combined operation is not simply a mathematical composition of the state complexities of its component operations. It appears that the state complexity of a combined operation in general is more

*This work is supported by Natural Science and Engineering Council of Canada Discovery Grant R2824A01, Canada Research Chair Award, and Natural Science and Engineering Council of Canada Discovery Grant 41630. All correspondence should be directed to Yuan Gao at ygao72@csd.uwo.ca.

difficult to obtain than that of an individual operation, especially the tight lower bound of the operation. This is because the resulting languages of the worst case of one operation may not be among the worst case input languages of the subsequent operation.

The study on state complexity of individual operations has already greatly relied on computer software to test and verify the results. One could say that, without the use of computer software, there would be no results on the state complexity of combined operations.

Although there is only a limited number of individual operations, the number of combined operations is unlimited. It is impossible to study the state complexity of all the combined operations. However, we consider that, besides the study of estimation and approximation of state complexity of general combined operations [6, 7], establishing the exact state complexity of some commonly used and basic combined operations is helpful to reveal the mutual influence between the component operations. For example, the state complexities of union and intersection on regular languages are known to be the same [15, 20]. However, the state complexities of $(L_1 \cup L_2)^*$ and $(L_1 \cap L_2)^*$ have been proved to be different [19].

In this paper, we study the state complexities of catenation combined with union, i.e., $(L(A)(L(B) \cup L(C)))$, and catenation combined with intersection, i.e., $(L(A)(L(B) \cap L(C)))$, for DFAs A, B and C of sizes $m, n, p \geq 1$, respectively. Both of them are basic combined operations and are commonly used in practice. For $L(A)(L(B) \cup L(C))$, we show that its state complexity is $(m-1)(2^{n+p} - 2^n - 2^p + 2) + 2^{n+p-2}$, for $m, n, p \geq 1$ (except the situations when $m \geq 2$ and $n = p = 1$), which is much smaller than $m2^{np} - 2^{np-1}$, the mathematical composition of the state complexities of union and catenation [15, 20]. On the other hand, for $L(A)(L(B) \cap L(C))$, we show that the mathematical composition of the individual state complexities of this combined operation is $m2^{np} - 2^{np-1}$, i.e., exactly equal to the state complexity of the operation (also except the cases when $m \geq 2$ and $n = p = 1$). Note that the individual state complexity of union and that of intersection are exactly the same. However, when they combined with catenation, the resulting state complexities are so different.

In the next section, we introduce the basic definitions and notation used in the paper. Then we prove our results on catenation combined with union and catenation combined with intersection in Sections 3 and 4, respectively. We conclude the paper in Section 5.

2. Preliminaries

A *non-deterministic finite automaton* (NFA) is a quintuple $A = (Q, \Sigma, \delta, s, F)$, where Q is a finite set of states, $s \in Q$ is the start state, and $F \subseteq Q$ is the set of final states, and $\delta : Q \times \Sigma \to 2^Q$ is the transition function. If $|\delta(q, a)| \leq 1$ for any $q \in Q$ and $a \in \Sigma$, then this automaton is called a *deterministic finite automaton* (DFA). A DFA is said to be complete if $|\delta(q, a)| = 1$ for all $q \in Q$ and $a \in \Sigma$. All the DFAs we mention in this paper are assumed to be complete. We extend δ to $Q \times \Sigma^* \to Q$ in the usual way. Then the word $w \in \Sigma^*$ is accepted by the automaton if $\delta(s, w) \cap F \neq \emptyset$. Two states in a finite automaton A are said to be *equivalent* if and only if for every word $w \in \Sigma^*$, if A is started in either state with w as input, it either accepts in both cases or rejects in both cases. It is well-known that a language which is accepted by an NFA can be accepted by a DFA, and such a language is said to be *regular*. The language accepted by a DFA A is denoted by $L(A)$. The reader may refer to [10, 21] for more details about regular languages and finite automata.

The *state complexity* of a regular language L, denoted by $sc(L)$, is the number of states of the minimal complete DFA that accepts L. The state complexity of a class S of regular languages, denoted by $sc(S)$, is the supremum among all $sc(L)$, $L \in S$. The state complexity of an operation on regular languages is the state complexity of the resulting languages from the operation as a function of the state complexity of the operand languages. For example, we say that the state complexity of the intersection of an m-state DFA language and an n-state DFA language is exactly mn. This implies that the largest number of states of all the minimal complete DFAs that accept the intersection of an m-state DFA language and an n-state DFA language is mn, and such languages exist. Thus, in a certain sense, the state complexity of an operation is a worst-case complexity.

3. Catenation combined with union

In this section, we consider the state complexity of $L(A)(L(B) \cup L(C))$ for three DFAs A, B, C of sizes $m, n, p \geq 1$, respectively. We first obtain the following upper bound $(m-k)(2^{n+p} - 2^n - 2^p + 2) + k2^{n+p-2}$ (Theorem 1), and then show that this bound is tight for $m, n, p \geq 1$, except the situations when $m \geq 2$ and $n = p = 1$ (Theorems 2 and 4).

Theorem 1. *For integers $m, n, p \geq 1$, let A, B and C be three DFAs with m, n and p states, respectively, where A has k final states. Then there exists a DFA of at most $(m-k)(2^{n+p} - 2^n - 2^p + 2) + k2^{n+p-2}$ states that accepts $L(A)(L(B) \cup L(C))$.*

Proof. Let $A = (Q_1, \Sigma, \delta_1, s_1, F_1)$ where $|F_1| = k$, $B = (Q_2, \Sigma, \delta_2, s_2, F_2)$, and $C = (Q_3, \Sigma, \delta_3, s_3, F_3)$. We construct $D = (Q, \Sigma, \delta, s, F)$ such that

$$
\begin{aligned}
Q = \{&\langle q_1, q_2, q_3\rangle \mid q_1 \in Q_1 - F_1, q_2 \in 2^{Q_2} - \{\emptyset\}, q_3 \in 2^{Q_3} - \{\emptyset\}\} \\
&\cup \{\langle q_1, \emptyset, \emptyset\rangle \mid q_1 \in Q_1 - F_1\} \\
&\cup \{\langle q_1, \{s_2\} \cup q_2, \{s_3\} \cup q_3\rangle \mid q_1 \in F_1, q_2 \in 2^{Q_2 - \{s_2\}}, q_3 \in 2^{Q_3 - \{s_3\}}\},
\end{aligned}
$$

$$s = \langle s_1, \emptyset, \emptyset\rangle \text{ if } s_1 \notin F_1, s = \langle s_1, \{s_2\}, \{s_3\}\rangle \text{ otherwise,}$$

$$F = \{\langle q_1, q_2, q_3\rangle \in Q \mid q_2 \cap F_2 \neq \emptyset \text{ or } q_3 \cap F_3 \neq \emptyset\},$$

$$
\begin{aligned}
\delta(\langle q_1, q_2, q_3\rangle, a) = \langle q_1', q_2', q_3'\rangle,\ &\text{for } a \in \Sigma, \text{ where } q_1' = \delta_1(q_1, a) \text{ and,} \\
&\text{for } i \in \{2, 3\},\ q_i' = S_i \cup \{s_i\} \text{ if } q_1' \in F_1,\ q_i' = S_i \text{ otherwise,} \\
&\text{where } S_i = \cup_{r \in q_i}\{\delta_i(r, a)\}.
\end{aligned}
$$

Intuitively, Q is a set of triples such that the first component of each triple is a state in Q_1 and the second and the third components are subsets of Q_2 and Q_3, respectively.

We notice that if the first component of a state is a non-final state of Q_1, the other two component are either both the empty set or both nonempty sets. This is because the two components always change from the empty set to a non-empty set at the same time. This is the reason to have the first and second terms of Q.

Also, we notice that if the first component of a state of D is a final state of A, then the second component and the third component of the state must contain the initial state of B and C, respectively. This is described by the third term of Q.

Clearly, the size of Q is $(m-k)(2^{n+p}-2^n-2^p+2)+k2^{n+p-2}$. Moreover, one can easily verify that $L(D)=L(A)(L(B)\cup L(C))$. $\square$

In the following, we consider the conditions under which this bound is tight. We know that a complete DFA of size 1 only accepts either $\emptyset$ or Σ^*. Thus, when $n=p=1$, $L(A)(L(B)\cup L(C))=L(A)\Sigma^*$ if either $L(B)=\Sigma^*$ or $L(C)=\Sigma^*$, and $L(A)(L(B)\cup L(C))=\emptyset$ otherwise. Therefore, in such cases, the state complexity of $L(A)(L(B)\cup L(C))$ is m as shown in [20].

Now, we consider the case when $n=1$ and $p\geq 2$. Since $L(B)\cup L(C)=L(C)$ when $L(B)=\emptyset$, it is clear that the state complexity of $L(A)(L(B)\cup L(C))$ is equal to that of $L(A)L(C)$, $m2^p-k2^{p-1}$ given in [20], which coincides with the upper bound obtained in Theorem 1. The situation is analogous to the case when $n\geq 2$ and $p=1$.

Next, we consider the case when $m=1$ and $n,p\geq 2$.

Theorem 2. *Let A be a DFA of size 1 over a four-letter alphabet. Then for any integers $n,p\geq 2$, there exist DFAs B and C with n and p states, respectively, defined over the same alphabet such that any DFA accepting $L(A)(L(B)\cup L(C))$ needs at least 2^{n+p-2} states.*

Proof. We use a four-letter alphabet $\Sigma=\{a,b,c,d\}$, and let A be the DFA accepting Σ^*.

Let $B=(Q_2,\Sigma,\delta_2,0,\{n-1\})$, as shown in Figure 1, where $Q_2=\{0,1,\ldots,n-1\}$, and the transitions are given as

- $\delta_2(i,a)=i+1 \bmod n$, for $i\in\{0,\ldots,n-1\}$,
- $\delta_2(i,x)=i$ for $i\in Q_2$, where $x\in\{b,d\}$,
- $\delta_2(0,c)=0$, $\delta_2(i,c)=i+1 \bmod n$, for $i\in\{1,\ldots,n-1\}$.

Let $C=(Q_3,\Sigma,\delta_3,0,\{p-1\})$be a DFA, as shown in Figure 1, where $Q_3=\{0,1,\ldots,p-1\}$, and the transitions are given as

- $\delta_3(i,x)=i$ for $i\in Q_3$, where $x\in\{a,c\}$,
- $\delta_3(i,b)=i+1 \bmod p$, for $i\in\{0,\ldots,p-1\}$,
- $\delta_3(0,d)=0$, $\delta_3(i,d)=i+1 \bmod p$, for $i\in\{1,\ldots,p-1\}$.

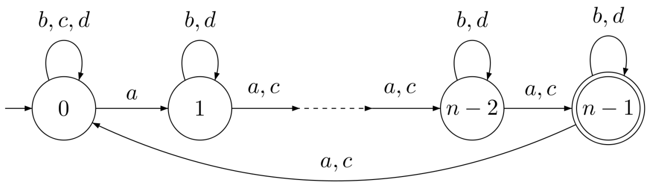

Fig. 1. The DFA B showing that the upper bound in Theorem 1 is reachable when $m = 1$ and $n, p \geq 2$

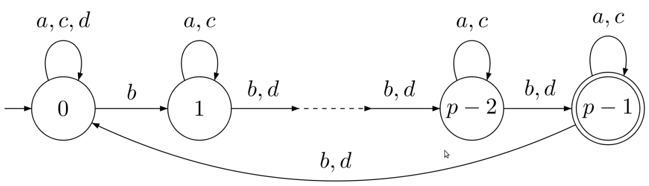

Fig. 2. The DFA C showing that the upper bound in Theorem 1 is reachable when $m = 1$ and $n, p \geq 2$

Let $D = (Q, \{a, b, c, d\}, \delta, \langle 0, \{0\}, \{0\}\rangle, F)$ be the DFA for accepting the language $L(A)(L(B) \cup L(C))$ constructed from those DFAs exactly as described in the proof of Theorem 1, where

$$Q = \{\langle 0, \{0\} \cup q_2, \{0\} \cup q_3\rangle \mid q_2 \in 2^{Q_2 - \{0\}}, q_3 \in 2^{Q_3 - \{0\}}\},$$
$$F = \{\langle q_1, q_2, q_3\rangle \in Q \mid n - 1 \in q_2 \text{ or } p - 1 \in q_3\}.$$

We omit the definition of the transitions.

Then we prove that the size of Q is minimal by showing that (I) any state in Q can be reached from the initial state, and (II) no two different states in Q are equivalent.

For (I), we first show that all the states $\langle 0, q_2, q_3\rangle$ such that $q_3 = \{0\}$ are reachable by induction on the size of q_2.

The basis clearly holds, since the initial state is the only state whose second component is of size 1.

In the induction steps, we assume that all states $\langle 0, q_2, \{0\}\rangle$ such that $|q_2| < k$ are reachable. Then we consider the states $\langle 0, q_2, \{0\}\rangle$ where $|q_2| = k$. Let $q_2 = \{0, j_2, \ldots, j_k\}$ such that $0 < j_2 < j_3 < \ldots < j_k \leq n - 1$. Note that the states such that $j_2 = 1$ can be reached as follows

$$\langle 0, \{0, 1, j_3, \ldots, j_k\}, \{0\}\rangle = \delta(\langle 0, \{0, j_3 - 1, \ldots, j_k - 1\}, \{0\}\rangle, a),$$

where $\{0, j_3 - 1, \ldots, j_k - 1\}$ is of size $k - 1$. Then the states such that $j_2 > 1$ can be reached from these states as follows

$$\langle 0, \{0, j_2, \ldots, j_k\}, \{0\}\rangle = \delta(\langle 0, \{0, 1, j_3 - t, \ldots, j_k - t\}, \{0\}\rangle, c^t), \text{ where } t = j_2 - 1.$$

After this induction, all the states such that the third component is $\{0\}$ have been reached. Then it is clear that, from each of these states $\langle 0, q_2, \{0\}\rangle$, all the states in Q such that the second component is q_2 and the size of their third component is larger than 1 can be reached by using the same induction steps but using the transitions on letters b and d.

Next, we show that any two distinct states $\langle 0, q_2, q_3\rangle$ and $\langle 0, q_2', q_3'\rangle$ in Q are not equivalent. We only consider the situations where $q_2 \neq q_2'$, since the other case can be shown analogously. Without loss of generality, there exists a state r such that $r \in q_2$ and $r \notin q_2'$. It is clear that $r \neq 0$. Let $w = d^{p-1}c^{n-1-r}$. Then $\delta(\langle 0, q_2, q_3\rangle, w) \in F$ but $\delta(\langle 0, q_2', q_3'\rangle, w) \notin F$. □

Then we consider the more general case when $m, n, p \geq 2$.

Example 3. *We use a five-letter alphabet* $\Sigma = \{a, b, c, d, e\}$ *in the following three DFAs, which are modified from the two DFAs in the proof of Theorem 1 in* [20].

Let $A = (Q_1, \Sigma, \delta_1, 0, \{m-1\})$ *be a DFA, where* $Q_1 = \{0, \ldots, m-1\}$ *and, for each state* $i \in Q_1$, $\delta_1(i, a) = j$, $j = (i+1)$ *mod* m, $\delta_1(i, x) = 0$, *if* $x \in \{b, d\}$, *and* $\delta_1(i, x) = i$, *if* $x \in \{c, e\}$.

Let $B = (Q_2, \Sigma, \delta_2, 0, \{n-1\})$ *be a DFA, where* $Q_2 = \{0, \ldots, n-1\}$ *and, for each state* $i \in Q_2$, $\delta_2(i, b) = j$, $j = (i+1)$ *mod* m, $\delta_2(i, c) = 1$, *and* $\delta_2(i, x) = i$, *if* $x \in \{a, d, e\}$.

Let $C = (Q_3, \Sigma, \delta_3, 0, \{p-1\})$ *be a DFA, where* $Q_3 = \{0, \ldots, p-1\}$ *and, for each state* $i \in Q_3$, $\delta_3(i, d) = j$, $j = (i+1)$ *mod* m, $\delta_3(i, e) = 1$, *and* $\delta_3(i, x) = i$, *if* $x \in \{a, b, c\}$.

Following the construction in the proof of Theorem 1, the DFA D can be constructed from the DFAs in Example 3 for showing that the upper bound is attainable for $m, n, p \geq 2$. We note that, similar to the proof of Theorem 2, DFAs B and C in this example change their states on disjoint letter sets, $\{b, c\}$ and $\{d, e\}$. Thus, by using a proof that is similar to the proof of Theorem 1 in [20], that shows the upper bound for the state complexity of catenation can be reached, we can easily verify that there are at least $(m-1)(2^{n+p} - 2^n - 2^p + 2) + 2^{n+p-2}$ distinct equivalence classes of the right-invariant relation induced by $L(A)(L(B) \cup L(C))$ [10]. Therefore, the upper bound can be attained and the following theorem holds.

Theorem 4. *Given three integers* $m, n, p \geq 2$, *there exist a DFA* A *of* m *states, a DFA* B *of* n *states, and a DFA* C *of* p *states over the same five-letter alphabet such that any DFA accepting* $L(A)(L(B) \cup L(C))$ *needs at least* $(m-1)(2^{n+p} - 2^n - 2^p + 2) + 2^{n+p-2}$ *states.*

A natural question is that, if we reduce the size of the alphabet used in DFAs A, B, C, using a three-letter alphabet, can we attain the upper bound as well? We give a positive answer in the next theorem under the condition $m, n, p \geq 3$.

Theorem 5. *For integers $m, n, p \geq 3$, there exist DFAs A, B and C of m, n, and p states, respectively, defined over a three-letter alphabet, such that any DFA that accepts $L(A)(L(B) \cup L(C))$ has at least $(m-1)(2^{n+p} - 2^n - 2^p + 2) + 2^{n+p-2}$ states.*

Proof. We define the following three automata over the three-letter alphabet $\Sigma = \{a, b, c\}$.

Let $A = (Q_1, \Sigma, \delta_1, 0, \{m-1\})$ be a DFA, where $Q_1 = \{0, 1, \ldots, m-1\}$, and the transitions are given as follows:

- $\delta_1(i, a) = i + 1$ for $i \in \{0, \ldots, m-2\}$, $\delta_1(m-1, a) = 0$;
- $\delta_1(i, e) = i$ for $i \in Q_1$, where $e \in \{b, c\}$.

Let $B = (Q_2, \Sigma, \delta_2, 0, \{n-1\})$ be a DFA, where $Q_2 = \{0, 1, \ldots, n-1\}$, and the transitions are given as follows:

- $\delta_2(i, a) = i$ for $i \in \{0, \ldots, n-3\}$, $\delta_2(n-2, a) = n-1$, $\delta_2(n-1, a) = n-2$;
- $\delta_2(i, b) = i + 1$ for $i \in \{0, \ldots, n-2\}$, $\delta_2(n-1, b) = n-1$;
- $\delta_2(i, c) = i$ for $i \in Q_2$.

Let $C = (Q_3, \Sigma, \delta_3, 0, \{p-1\})$ be a DFA, where $Q_3 = \{0, 1, \ldots, p-1\}$, and the transitions are given as follows:

- $\delta_3(i, a) = i$ for $i \in \{0, \ldots, p-3\}$, $\delta_3(p-2, a) = p-1$, $\delta_3(p-1, a) = p-2$;
- $\delta_3(i, b) = i$ for $i \in Q_3$;
- $\delta_3(i, c) = i + 1$ for $i \in \{0, \ldots, p-2\}$, $\delta_3(p-1, c) = p-1$.

Let $D = (Q, \{a, b, c\}, \delta, \langle 0, \emptyset, \emptyset \rangle, F)$ be the DFA that accepts the language $L(A)(L(B) \cup L(C))$ constructed from those DFAs exactly as described in the proof of Theorem 1, where

$$
\begin{aligned}
Q = {} & \{\langle q_1, q_2, q_3 \rangle \mid q_1 \in Q_1 - \{m-1\}, q_2 \in 2^{Q_2} - \{\emptyset\}, q_3 \in 2^{Q_3} - \{\emptyset\}\} \\
& \cup \{\langle q_1, \emptyset, \emptyset \rangle \mid q_1 \in Q_1 - \{m-1\}\} \\
& \cup \{\langle m-1, \{0\} \cup q_2, \{0\} \cup q_3 \rangle \mid q_2 \in 2^{Q_2 - \{0\}}, q_3 \in 2^{Q_3 - \{0\}}\}, \\
F = {} & \{\langle q_1, q_2, q_3 \rangle \in Q \mid n-1 \in q_2 \text{ or } p-1 \in q_3\}.
\end{aligned}
$$

We omit the definition of transitions.

Then we prove that the size of Q is minimal by showing that (I) any state in Q can be reached from the initial state and (II) no two different states in Q are equivalent.

Now we consider (I). It is clear that states $\langle q_1, \emptyset, \emptyset \rangle$, for $q_1 \in Q_1 - \{m-1\}$, are reachable from the initial state on strings a^{q_1}, and the state $\langle m-1, \{0\}, \{0\} \rangle$ can be reached from $\langle m-2, \emptyset, \emptyset \rangle$ on the letter a.

We first show by induction on the size of the second component that any remaining state in Q such that its third component is $\{0\}$ can be reached from the state $\langle m-1, \{0\}, \{0\} \rangle$. We only use strings over the letters a, b. Thus, the last component remains $\{0\}$.

Basis: for any $i \in \{0, \ldots, m-2\}$, the state $\langle i, \{0\}, \{0\} \rangle$ can be reached from the state $\langle m-1, \{0\}, \{0\} \rangle$ on the string a^{i+1}. Then for any $i \in \{0, \ldots, m-2\}$ and $j \in \{1, \ldots, n\}$,

$$\langle i, \{j\}, \{0\} \rangle = \delta(\langle i, \{0\}, \{0\} \rangle, b^j).$$

Induction step: for $i \in \{0, \ldots, m-1\}$, assume that all states $\langle i, q_2, \{0\} \rangle$ such that $|q_2| < k$ are reachable. Then we consider the states $\langle i, q_2, \{0\} \rangle$ where $|q_2| = k$. Let $q_2 = \{j_1, j_2, \ldots, j_k\}$ such that $0 \leq j_1 < j_2 < \cdots < j_k \leq n-1$.

Note that the states such that $j_1 = 0$ are reachable as follows. If either (i) $j_k \leq n-3$, or (ii) $j_{k-1} = n-2$ and $j_k = n-1$, we have

$$\langle m-1, \{0, j_2, \ldots, j_k\}, \{0\} \rangle = \delta(\langle m-2, \{j_2, \ldots, j_k\}, \{0\} \rangle, a).$$

If $j_k = n-2$, the states $\langle m-1, \{0, j_2, \ldots, j_k\}, \{0\} \rangle$ can be reached from the states $\langle m-2, \{j_2, \ldots, j_{k-1}, n-1\}, \{0\} \rangle$ by reading the letter a. If $j_k = n-1$ and $j_{k-1} \neq n-2$, the states $\langle m-1, \{0, j_2, \ldots, j_k\}, \{0\} \rangle$ can be reached from states $\langle m-2, \{j_2, \ldots, j_{k-1}, n-2\}, \{0\} \rangle$ by reading the letter a. In all the cases, we reach the state from a state such that $|q_2| = k-1$. Similarly, we can easily verify that, by reading the letter a, states $\langle 0, \{0, \ldots, j_k\}, \{0\} \rangle$ can be reached from the states $\langle m-1, \{0, \ldots, j_k\}, \{0\} \rangle$. Note that the state $\langle 0, q', \{0\} \rangle$ is not simply reached from $\langle m-1, q', \{0\} \rangle$ by reading the letter a. We still need to consider the previous cases, and these cases apply to the following states as well. For $i \in \{1, \ldots, m-2\}$, the states $\langle i, \{0, \ldots, j_k\}, \{0\} \rangle$ can be reached from the states $\langle i-1, \{0, \ldots, j_k\}, \{0\} \rangle$.

Next, we show that all states such that $0 \notin q_2$ are reachable. Note that the first component of these states cannot be $m-1$. Thus, for $i \in \{0, \ldots, m-2\}$, we have

$$\langle i, \{j_1, \ldots, j_k\}, \{0\} \rangle = \delta(\langle i, \{0, j_2 - j_1, \ldots, j_k - j_1\}, \{0\} \rangle, b^{j_1}).$$

After the induction step, we can verify that all states in Q such that the third component is $\{0\}$ have been reached.

In the following, we consider the states whose third component is non-empty but not $\{0\}$. Note that if the second component of a state does not contain the states $n-2$ and $n-1$ or contains both of them, this component does not change by reading the letter a. Thus, by using the letter c instead of the letter b in the same induction step, we can show that, for $i \in \{0, \ldots, m-1\}$, the states $\langle i, q_2, q_3 \rangle$ in Q such that $q_2 \cap \{n-2, n-1\} = \emptyset$ or $\{n-2, n-1\} \subseteq q_2$ are reachable from the state $\langle 0, q_2, \{0\} \rangle$. The remaining states to be considered are the states $\langle i, q_2, q_3 \rangle$ such that q_2 contains either $n-2$ or $n-1$ but not both, for $i \in \{0, \ldots, m-1\}$. Assume q_2 contains $n-2$. Then by the same induction with the letters a, c, we can reach the states $\langle i, q_2, q_3 \rangle$ and states $\langle i', q_2', q_3' \rangle$, $i, i' \in \{0, \ldots, m-1\}$, from the state $\langle 0, q_2, \{0\} \rangle$ such that $q_2' = (q_2 \cup \{n-1\}) - \{n-2\}$. Moreover, if we replace q_2' with q_2, the union of these two types of states is exactly all states in Q such that their second component is q_2. It is clear that those states $\langle i', q_2, q_3' \rangle$ are reachable from the state $\langle 0, q_2', \{0\} \rangle$ by following the same induction step with letters a, c. An analogous argument can be applied to the states such that q_2 contains $n-1$ but not $n-2$.

Now all the states in Q are reachable, and next we will show that the states of the DFA D are pairwise inequivalent. Let $\langle i, q_2, q_3\rangle$ and $\langle j, q_2', q_3'\rangle$ be two different states. We consider the following two cases:

(1) $i < j$. Then the string $a^{m-1-i}b^{n-1}c^{p-1}a$ is accepted by the DFA D starting from the state $\langle i, q_2, q_3\rangle$, but it is not accepted starting from the state $\langle j, q_2', q_3'\rangle$.

(2) $i = j$. We only prove for the situation where $q_2 \neq q_2'$, since the proof is analogous when $q_3 \neq q_3'$. Without loss of generality, there exists a state r such that $r \in q_2$ and $r \notin q_2'$.

If $i = j \neq m-1$, we can verify that $c^{p-1}b^{n-r-2}a$ is accepted by D from the state $\langle i, q_2, q_3\rangle$ but not from the state $\langle j, q_2', q_3'\rangle$.

If $i = j = m-1$, it is clear that $r \neq 0$. We consider the following three cases.

(a) $r \in \{1, \ldots, n-3\}$. After reading the letter a, i and j become 0 and we still have $r \in q_2$ and $r \notin q_2'$. Thus, the resulting situation has just been considered.

(b) $r = n-2$. Then the state $\langle i, q_2, q_3\rangle$ reaches a final state on $ac^{p-1}ab$, but the state $\langle j, q_2', q_3'\rangle$ does not on the same string.

(c) $r = n-1$. Then the state $\langle i, q_2, q_3\rangle$ reaches a final state by reading $ac^{p-1}a$, but the state $\langle j, q_2', q_3'\rangle$ does not. □

4. Catenation combined with intersection

In this section, we investigate the state complexity of $L_1(L_2 \cap L_3)$, and show that its upper bound (Theorem 6) coincides with its lower bound (Theorems 7 and 8). The following theorem shows an upper bound for the state complexity of this combined operation.

Theorem 6. *Let L_1, L_2 and L_3 be three regular languages accepted by an m-state, an n-state and a p-state DFA, respectively, for m, n, $p \geq 1$. Then there exists a DFA of at most $m2^{np} - 2^{np-1}$ states that accepts $L_1(L_2 \cap L_3)$. However, when $m \geq 1$, $n = p = 1$, the number of states can be lowered to m.*

Theorem 6 gives a general upper bound of the state complexity of $L_1(L_2 \cap L_3)$ because $m2^{np} - 2^{np-1}$ is the mathematical composition of the state complexities of the individual component operations. Thus, we omit the proof of this upper bound. When $m \geq 1$, $n = p = 1$, $L(A)(L(B) \cap L(C)) = L(A)\Sigma^*$ if both $L(B)$ and $L(C)$ are Σ^*. The resulting language is $\emptyset$ otherwise. Thus, the state complexity of $L(A)(L(B) \cap L(C))$ in this case is the same as that of $L(A)\Sigma^*$: namely, m [20].

When $m \geq 1$, $n = 1$, $p \geq 2$, $L(A)(L(B) \cap L(C)) = \emptyset$, if $L(B) = \emptyset$, and $L(A)L(C)$ if $L(B) = \Sigma^*$. In this case, the state complexity of the combined operation is $m2^p - 2^{p-1}$ which is the same as that of $L(A)L(C)$ [20] and meets the upper bound in Theorem 6. Similarly, when $m \geq 1$, $n \geq 2$, $p = 1$, the state complexity of

$L(A)(L(B) \cap L(C))$ is $m2^n - 2^{n-1}$ which also attains the upper bound in Theorem 6. Next, we show the upper bound $m2^{np} - 2^{np-1}$ is attainable when $m, n, p \geq 2$.

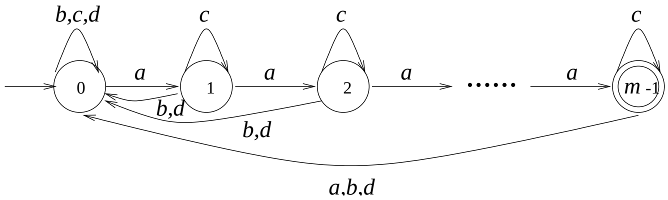

Fig. 3. The DFA A showing that the upper bound in Theorem 6 is attainable when $m \geq 2$ and $n, p \geq 1$

Theorem 7. *Given three integers $m, n, p \geq 2$, there exists a DFA A of m states, a DFA B of n states and a DFA C of p states over the same four-letter alphabet such that any DFA accepting $L(A)(L(B) \cap L(C))$ needs at least $m2^{np} - 2^{np-1}$ states.*

Proof. Let $A = (Q_A, \Sigma, \delta_A, 0, F_A)$ be a DFA, as shown in Figure 3, where $Q_A = \{0, 1, \ldots, m-1\}$, $F_A = \{m-1\}$, $\Sigma = \{a, b, c, d\}$ and the transitions are given as:

- $\delta_A(i, a) = i + 1 \bmod m$, $i = 0, \ldots, m-1$,
- $\delta_A(i, x) = 0$, $i = 0, \ldots, m-1$, where $x \in \{b, d\}$,
- $\delta_A(i, c) = i$, $i = 0, \ldots, m-1$.

Let $B = (Q_B, \Sigma, \delta_B, 0, F_B)$ be a DFA, as shown in Figure 4, where $Q_B = \{0, 1, \ldots, n-1\}$, $F_B = \{n-1\}$ and the transitions are given as:

- $\delta_B(i, x) = i$, $i = 0, \ldots, n-1$, where $x \in \{a, d\}$,
- $\delta_B(i, b) = i + 1 \bmod n$, $i = 0, \ldots, n-1$,
- $\delta_B(i, c) = 1$, $i = 0, \ldots, n-1$.

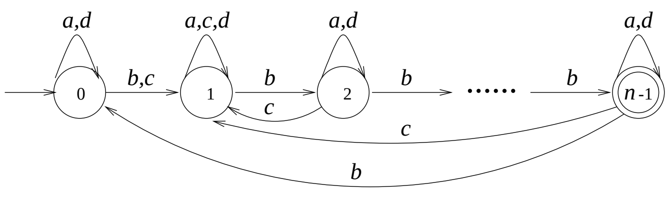

Fig. 4. The DFA B showing that the upper bound in Theorem 6 is attainable when $m \geq 2$ and $n, p \geq 1$

Let $C = (Q_C, \Sigma, \delta_C, 0, F_C)$ be a DFA, as shown in Figure 5, where $Q_C = \{0, 1, \ldots, p-1\}$, $F_C = \{p-1\}$ and the transitions are given as:

- $\delta_C(i, x) = i$, $i = 0, \ldots, p-1$, where $x \in \{a, b\}$,
- $\delta_C(i, c) = 1$, $i = 0, \ldots, p-1$,
- $\delta_C(i, d) = i + 1 \bmod p$, $i = 0, \ldots, p-1$.

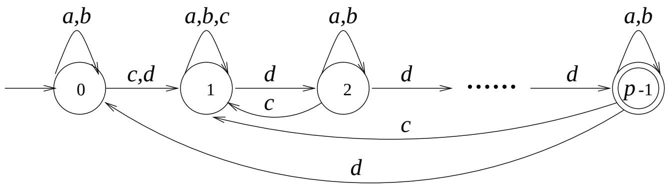

Fig. 5. The DFA C showing that the upper bound in Theorem 6 is attainable when $m \geq 2$ and $n, p \geq 1$

We construct the DFA $D = (Q_D, \Sigma, \delta_D, s_D, F_D\}$, where

$$\begin{aligned} Q_D &= \{\langle u, v\rangle \mid u \in Q_B,\, v \in Q_C\}, \\ s_D &= \langle 0, 0\rangle, \\ F_D &= \{\langle n-1, p-1\rangle\}, \end{aligned}$$

and for each state $\langle u, v\rangle \in Q_D$ and each letter $e \in \Sigma$,

$$\delta_D(\langle u, v\rangle, e) = \langle u', v'\rangle \text{ if } \delta_B(u, e) = u',\, \delta_C(v, e) = v'.$$

Clearly, there are $n \cdot p$ states in D and $L(D) = L(B) \cap L(C)$. Now we construct another DFA $E = (Q_E, \Sigma, \delta_E, s_E, F_E\}$, where

$$\begin{aligned} Q_E &= \{\langle q, R\rangle \mid q \in Q_A - F_A,\, R \subseteq Q_D\} \cup \{\langle m-1, S\rangle \mid s_D \in S,\, S \subseteq Q_D\}, \\ s_E &= \langle 0, \emptyset\rangle, \\ F_E &= \{\langle q, R\rangle \mid R \cap F_D \neq \emptyset,\, \langle q, R\rangle \in Q_E\}, \end{aligned}$$

and for each state $\langle q, R\rangle \in Q_E$ and each letter $e \in \Sigma$,

$$\delta_E(\langle q, R\rangle, e) = \begin{cases} \langle q', R'\rangle \text{ if } \delta_A(q, e) = q' \neq m-1,\, \delta_D(R, e) = R', \\ \langle q', R'\rangle \text{ if } \delta_A(q, e) = q' = m-1,\, R' = \delta_D(R, e) \cup \{s_D\}. \end{cases}$$

It is easy to see that $L(E) = L(A)(L(B) \cap L(C))$. There are $(m-1) \cdot 2^{np}$ states in the first term of the union for Q_E. In the second term, there are $1 \cdot 2^{np-1}$ states. Thus,

$$|Q_E| = (m-1) \cdot 2^{np} + 1 \cdot 2^{np-1} = m2^{np} - 2^{np-1}.$$

In order to show that E is minimal, we need to show that (I) every state in E is reachable from the start state and (II) each state defines a distinct equivalence class.

We prove (I) by induction on the size of the second component of states in Q_E. First, any state $\langle q, \emptyset\rangle$, $0 \leq q \leq m-2$, is reachable from s_E by reading the word a^q.

The we consider all states $\langle q, R\rangle$ such that $|R| = 1$. In this case, let $R = \{\langle x, y\rangle\}$. We have

$$\langle q, \{\langle x, y\rangle\}\rangle = \delta_E(\langle 0, \emptyset\rangle, a^m b^x d^y a^q).$$

Notice that the only state $\langle q, R\rangle$ in Q_E such that $q = m-1$ and $|R| = 1$ is $\langle m-1, \{\langle 0, 0\rangle\}\rangle$ since the fact that $q = m-1$ guarantees $\langle 0, 0\rangle \in R$.

Assume that all states $\langle q, R\rangle$ such that $|R| < k$ are reachable. Consider $\langle q, R\rangle$ where $|R| = k$. Let $R = \{\langle x_i, y_i\rangle \mid 1 \le i \le k\}$ such that $0 \le x_1 \le x_2 \le \ldots \le x_k \le n-1$ if $q \ne m-1$ and $0 = x_1 \le x_2 \le \ldots \le x_k \le n-1$, $y_1 = 0$, otherwise. We have $\langle q, R\rangle = \delta_E(\langle 0, R'\rangle, a^m b^{x_1} d^{y_1} a^q)$, where

$$R' = \{\langle x_j - x_1, (y_j - y_1 + n) \bmod n\rangle \mid 2 \le j \le k\}.$$

The state $\langle 0, R'\rangle$ is attainable from the start state, since $|R'| = k-1$. Thus, $\langle q, R\rangle$ is also reachable.

To prove (II), let $\langle q_1, R_1\rangle$ and $\langle q_2, R_2\rangle$ be two different states in E. We consider the following two cases.

1. $q_1 \ne q_2$. Without loss of generality, we may assume that $q_1 > q_2$. There always exists a string $t = ca^{m-1-q_1} b^{n-1} d^{p-1}$ such that
$$\delta_E(\langle q_1, R_1\rangle, t) \in F_E \text{ and } \delta_E(\langle q_2, R_2\rangle, t) \notin F_E.$$
2. $q_1 = q_2$, $R_1 \ne R_2$. Without loss of generality, we may assume that $|R_1| \ge |R_2|$. Let $\langle x, y\rangle \in R_1 - R_2$. Then
$$\delta_E(\langle q_1, R_1\rangle, b^{n-1-x} d^{p-1-y}) \in F_E,$$
$$\delta_E(\langle q_2, R_2\rangle, b^{n-1-x} d^{p-1-y}) \notin F_E.$$

Thus, the minimal DFA accepting $L(A)(L(B) \cap L(C))$ needs at least $m2^{np} - 2^{np-1}$ states for $m, n, p \ge 2$. □

Now we consider the case when $m = 1$, i.e., $L(A) = \Sigma^*$.

Theorem 8. *Given two integers $n, p \ge 2$, there exists a DFA A of one state, a DFA B of n states and a DFA C of p states over the same five-letter alphabet such that any DFA accepting $L(A)(L(B) \cap L(C))$ needs at least 2^{np-1} states.*

Proof. When $m = 1$, $n \ge 2$, $p \ge 2$, we give the following construction. Let $A = (\{0\}, \Sigma, \delta_A, 0, \{0\})$ be a DFA, where $\Sigma = \{a, b, c, d, e\}$ and $\delta_A(0, t) = 0$ for any letter $t \in \Sigma$. It is clear that $L(A) = \Sigma^*$.

Let $B = (Q_B, \Sigma, \delta_B, 0, F_B)$ be a DFA, where $Q_B = \{0, 1, \ldots, n-1\}$, $F_B = \{n-1\}$ and the transitions are given by

- $\delta_B(i, a) = i + 1 \bmod n$, $i = 0, \ldots, n-1$;
- $\delta_B(i, b) = i$, $i = 0, \ldots, n-1$;
- $\delta_B(0, c) = 1$, $\delta_B(j, c) = j$, $j = 1, \ldots, n-1$;
- $\delta_B(0, d) = 0$, $\delta_B(j, d) = j + 1$, $j = 1, \ldots, n-2$, $\delta_B(n-1, d) = 1$;

- $\delta_B(i, e) = i$, $i = 0, \ldots, n-1$.

Let $C = (Q_C, \Sigma, \delta_C, 0, F_C)$ be a DFA, where $Q_C = \{0, 1, \ldots, p-1\}$, $F_C = \{p-1\}$ and the transitions are given by

- $\delta_C(i, a) = i$, $i = 0, \ldots, p-1$;
- $\delta_C(i, b) = i + 1 \bmod p$, $i = 0, \ldots, p-1$;
- $\delta_C(0, c) = 1$, $\delta_C(j, c) = j$, $j = 1, \ldots, p-1$;
- $\delta_C(i, d) = i$, $i = 0, \ldots, p-1$;
- $\delta_C(0, e) = 0$, $\delta_C(j, e) = j + 1$, $j = 1, \ldots, p-2$, $\delta_C(p-1, e) = 1$.

Construct the DFA $D = (Q_D, \Sigma, \delta_D, \langle 0, 0\rangle, F_D)$ that accepts $L(B) \cap L(C)$ in the same way as the proof of Theorem 7, where

$$Q_D = \{\langle u, v\rangle \mid u \in Q_B,\ v \in Q_C\},$$
$$F_D = \{\langle n-1, p-1\rangle\},$$

and for each state $\langle u, v\rangle \in Q_D$ and each letter $t \in \Sigma$,

$$\delta_D(\langle u, v\rangle, t) = \langle u', v'\rangle \text{ if } \delta_B(u, t) = u',\ \delta_C(v, t) = v'.$$

Now we construct the DFA $E = (Q_E, \Sigma, \delta_E, s_E, F_E)$, where

$$Q_E = \{\langle 0, R\rangle \mid \langle 0, 0\rangle \in R, R \subseteq Q_D\},$$
$$s_E = \langle 0, \{\langle 0, 0\rangle\}\rangle,$$
$$F_E = \{\langle 0, R\rangle \in Q_E \mid R \cap F_D \neq \emptyset\},$$

and for each state $\langle 0, R\rangle \in Q_E$ and each letter $t \in \Sigma$,

$$\delta_E(\langle 0, R\rangle, t) = \langle 0, R'\rangle \text{ where } R' = \delta_D(R, t) \cup \{\langle 0, 0\rangle\}.$$

Note that $\langle 0, 0\rangle \in R$ for every state $\langle 0, R\rangle \in Q_E$, since 0 is the only state in A and it is both initial and final. It is easy to see that $L(E) = L(A)(L(B) \cap L(C))$ and E has $2^{np} - 2^{np-1} = 2^{np-1}$ states in total. Now we show that E is a minimal DFA by (I) every state in E is reachable from the initial state and (II) each state defines a distinct equivalence class.

We again prove (I) by induction on the size of the second component of states in Q_E. First, the only state in $\langle 0, R\rangle \in Q_E$ such that $|R| = 1$ is the initial state, $\langle 0, \{\langle 0, 0\rangle\}\rangle$.

Assume that all states $\langle 0, R\rangle$ such that $|R| \leq k$ are reachable. Consider $\langle 0, R\rangle$ where $|R| = k + 1$. Let $R = \{\langle 0, 0\rangle, \langle x_1, y_1\rangle, \ldots, \langle x_k, y_k\rangle\}$ such that $0 \leq x_1 \leq x_2 \leq \ldots \leq x_k \leq n-1$. We consider the following three cases.

Case 1. $\langle 0, y_1\rangle \in R$, $y_1 \geq 1$. If there exists $\langle 0, y_i\rangle \in R$, $y_i \geq 1$, $1 \leq i \leq k$, then $x_1 = 0$ and $y_1 \geq 1$, since $0 \leq x_1 \leq x_2 \leq \cdots \leq x_k \leq n-1$. For this case, we have

$$\langle 0, R\rangle = \delta_E(\langle 0, R_1\rangle, be^{y_1-1}),$$

where $R_1 = \{\langle 0, 0\rangle\} \cup S_1 \cup T_1$,

$$S_1 = \{\langle x_j, p-1\rangle \mid \langle x_j, 0\rangle \in R, x_j \neq 0\},$$

$$T_1 = \{\langle x_j, (y_j - y_1 + p - 1) \bmod (p-1)\rangle \mid \langle x_j, y_j\rangle \in R, y_j \neq 0, 2 \leq j \leq k\}.$$

Notice that $\langle 0,0\rangle \notin S_1 \cup T_1$ and $S_1 \cap T_1 = \emptyset$. So the state $\langle 0, R\rangle$ is reachable from the initial state, since $|R_1| = k$ and $\langle 0, R_1\rangle$ is reachable.

Case 2. $x_1 \geq 1$, $\langle x_i, 0\rangle \in R$, $1 \leq i \leq k$. It is easy to see that every $x_i \geq 1$ because $x_i \geq x_1$. We have

$$\langle 0, R\rangle = \delta_E(\langle 0, R_2\rangle, ad^{x_i - 1}),$$

where $R_2 = \{\langle 0,0\rangle\} \cup T_2$,

$$T_2 = \{\langle (x_j - x_i + n - 1) \bmod (n-1), y_j\rangle \mid \langle x_j, y_j\rangle \in R, 1 \leq j \leq k, j \neq i\}.$$

There are k elements in R_2. So the state $\langle 0, R\rangle$ is also reachable for this case.

Case 3. $x_1 \geq 1$, $y_i \geq 1$, $1 \leq i \leq k$, because every $x_i \geq x_1 \geq 1$, we have

$$\langle 0, R\rangle = \delta_E(\langle 0, R_3\rangle, cd^{x_1 - 1}e^{y_1 - 1}),$$

where $R_3 = \{\langle 0,0\rangle\} \cup T_3$,

$$T_3 = \{\langle (x_j - x_1 + 1), (y_j - y_1 + p - 1) \bmod (p-1) + 1\rangle \mid \langle x_j, y_j\rangle \in R, 2 \leq j \leq k\}.$$

So every state $\langle 0, R\rangle$ in E is reachable when $|R| = k + 1$.

To prove (II), let $\langle 0, R\rangle$ and $\langle 0, R'\rangle$ be two different states in E. Without loss of generality, we may assume that $|R| \geq |R'|$. So we can always find $\langle x, y\rangle \in R - R'$. Clearly, $\langle x, y\rangle \neq \langle 0, 0\rangle$. So there exists a string $w = a^{n-1-x}b^{p-1-y}$ such that $\delta_E(\langle 0, R\rangle, w) \in F_E$ and $\delta_E(\langle 0, R'\rangle, w) \notin F_E$.

Thus, the minimal DFA that accepts $\Sigma^*(L(B) \cap L(C))$ has at least 2^{np-1} states for $m = 1$, $n, p \geq 2$. $\square$

This lower bound coincides with the upper bound given in Theorem 6. Thus, the bounds are tight for the case when $m = 1$, $n, p \geq 2$.

5. Conclusion

In this paper, we have studied the state complexities of two basic combined operations: catenation combined with union and catenation combined with intersection. We have proved that the state complexity of $L(A)(L(B) \cup L(C))$ is $(m-1)(2^{n+p} - 2^n - 2^p + 2) + 2^{n+p-2}$ for $m, n, p \geq 1$ (except the situations when $m \geq 2$ and $n = p = 1$), which is significantly less than the mathematical composition of state complexities of its component operations, $m2^{np} - 2^{np-1}$. We have also proved that the state complexity of $L(A)(L(B) \cap L(C))$ is $m2^{np} - 2^{np-1}$ for $m, n, p \geq 1$ (except the cases when $m \geq 2$ and $n = p = 1$), which is exactly the mathematical composition of state complexities of its component operations.

Acknowledgement

We wish to express our gratitude to the anonymous referees for their valuable comments on the earlier version of this paper.

References

[1] J. Birget: State-complexity of finite-state devices, state compressibility and incompressibility, *Mathematical Systems Theory*, 26(3) (1993) 237–269

[2] C. Campeanu, K. Culik, K. Salomaa, S. Yu: State complexity of basic operations on finite language, in:*Proceedings of the Fourth International Workshop on Implementing Automata* VIII 1-11, LNCS 2214, 1999, 60–70

[3] M. Domaratzki: State complexity and proportional removals, *Journal of Automata, Languages and Combinatorics*, 7 (2002) 455–468

[4] M. Domaratzki, A. Okhotin: State complexity of power, *Theoretical Computer Science* 410(24-25) (2009) 2377–2392

[5] Z. Ésik, Y. Gao, G. Liu, S. Yu: Estimation of state complexity of combined operations, *Theoretical Computer Science* 410 (35) (2009) 3272–3280.

[6] Y. Gao, K. Salomaa, S. Yu: The state complexity of two combined operations: star of catenation and star of Reversal, *Fundam. Inform.* 83 (1-2) (2008) 75–89

[7] Y. Gao and S. Yu: State complexity approximation, in:*Proceedings of Descriptional Complexity of Formal Systems* (2009) 163–174.

[8] Y. Han and K. Salomaa: State complexity of basic operations on suffix-free regular languages, in:*Theoretical Computer Science* 410 (27-29) (2009) 2537–2548

[9] M. Holzer, M. Kutrib: State complexity of basic operations on nondeterministic finite automata, in:*Proceedings of International Conference on Implementation and Application of Automata 2002* LNCS 2608, 2002, 148–157

[10] J. E. Hopcroft, J. D. Ullman: *Introduction to Automata Theory, Languages, and Computation*, Addison Wesley, 1979

[11] J. Jirásek, G. Jirásková, A. Szabari: State complexity of concatenation and complementation of regular languages, *International Journal of Foundations of Computer Science* 16 (2005) 511–529

[12] G. Jirásková: State complexity of some operations on binary regular languages, *Theoretical Computer Science* 330 (2005) 287–298

[13] G. Jirásková, A. Okhotin: On the state complexity of star of union and star of intersection, *Turku Center for Computer Science TUCS Technical Report* No. 825, 2007

[14] G. Liu, C. Martin-Vide, A. Salomaa, S. Yu: State complexity of basic language operations combined with reversal, *Information and Computation* 206 (2008) 1178–1186

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

[16] G. Pighizzini, J. Shallit: Unary language operations, state complexity and Jacobsthal's function, *International Journal of Foundations of Computer Science* 13 (2002) 145–159

[17] N. Rampersad: The state complexity of L^2 and L^k, *Information Processing Letters*, 98 (2006) 231–234

[18] A. Salomaa, D. Wood, S. Yu: On the state complexity of reversals of regular languages, *Theoretical Computer Science* 320 (2004) 293–313

[19] A. Salomaa, K. Salomaa, S. Yu: State complexity of combined operations, *Theoretical Computer Science* 383 (2007) 140–152

[20] S. Yu, Q. Zhuang, K. Salomaa: The state complexity of some basic operations on regular languages, *Theoretical Computer Science* 125 (1994) 315-328

[21] S. Yu: Regular languages, in: G. Rozenberg, A. Salomaa (Eds.),*Handbook of Formal Languages*, Vol. 1, Springer-Verlag, 1997, 41–110

[22] S. Yu: State complexity of regular languages, *Journal of Automata, Languages and Combinatorics* 6(2) (2001) 221–234