Theoretical Computer Science 125 (1994) 315–328
Elsevier


# The state complexities of some basic operations on regular languages*

Sheng Yu and Qingyu Zhuang

*Department of Computer Science, University of Western Ontario, London, Ont., Canada N6A 5B7*

Kai Salomaa**

*Department of Mathematics, University of Turku, SF-20500 Turku, Finland*

Communicated by G. Rozenberg
Received February 1992

*Abstract*

Yu, S., Q. Zhuang and K. Salomaa, The state complexities of some basic operations on regular languages, Theoretical Computer Science 125 (1994) 315–328.

We consider the state complexities of some basic operations on regular languages. We show that the number of states that is sufficient and necessary in the worst case for a deterministic finite automaton (DFA) to accept the catenation of an $m$-state DFA language and an $n$-state DFA language is exactly $m2^n - 2^{n-1}$, for $m, n \geqslant 1$. The result of $2^{n-1} + 2^{n-2}$ states is obtained for the star of an $n$-state DFA language, $n > 1$. State complexities for other basic operations and for regular languages over a one-letter alphabet are also studied.

## 1. Introduction

Motivated by the recently renewed interest in regular languages [4, 7, 8], we consider the following problems in quantifying the basic operations on DFAs. Let $m, n$ be nonnegative integers and $A$ and $B$ be two arbitrary DFAs of $m$ states and $n$ states, respectively. (1) What is the *exact* number of states that is sufficient and necessary in the worst case for a DFA to accept the catenation of $L(A)$ and $L(B)$? (2) What is the exact number of states that is sufficient and necessary in the worst case for a DFA to

*Correspondence to*: S. Yu, Department of Computer Science, University of Western Ontario, London, Ontario, Canada N6A 5B7.

* This research is supported by the Natural Sciences and Engineering Research Council of Canada grants OGP0041630.

** This work, in part, has been done during Dr. K. Salomaa's stay at the Department of Computer Science, University of Western Ontario supported by the NSERC International Fellowship.


*SSDI* 0304-3975(92)00011-D

accept $(L(B))^*$? (3) The same question for other operations. It seems that these fundamental questions should have been answered long ago. Indeed, it has been shown in [5] that $2^n$ is the tight upper bound on the number of states necessary for a DFA to accept the reversal of an $n$-state DFA language. Also in [6], it has been shown that $2^n$ is the tight upper bound on the number of states necessary for a DFA to accept an $n$-state NFA language. However, the same question (exact bound) for catenation and star operations on regular languages remains open. In [8], it is shown that for any $n>0$ there exists a 2-state DFA language and an $n$-state DFA language such that any DFA accepting the catenation of the two languages needs at least $2^{n-1}$ states. In [8], it is also shown that for any integer $n>0$ there exists an $n$-state DFA $A$ such that any DFA accepting $(L(A))^*$ needs at least $2^{n-1}$ states. In this paper, we improve the above results and obtain exact bounds. We show that $m2^n-2^{n-1}$ is the optimal upper bound for catenation for any $m, n \geqslant 1$. We also show that the answer to the same question for star operation is exactly $2^{n-1}+2^{n-2}$. In our proofs, we use very small alphabets. However, for regular languages over a one-letter alphabet, we show that $(n-1)^2+1$ is the tight upper bound for star operation and $mn$ for catenation. Other operations such as left quotient and right quotient, reversal, as well as union, intersection, etc. are also considered.

A deterministic finite automaton (DFA) is denoted by a quintuple $(Q, \Sigma, \delta, q_0, F)$ where $Q$ is the finite set of states, $\Sigma$ is the finite alphabet, $\delta: Q \times \Sigma \to Q$ is the transition function, $q_0 \in Q$ is the start state, and $F \subseteq Q$ is the set of final states. In this paper, *all the DFAs are assumed to be complete DFAs*. By a complete DFA we mean that there is a transition defined for each letter of the alphabet from each state. For any $x \in \Sigma^*$, we use $\#(x)$ to denote the length of $x$ and $\#_a(x)$ for some $a \in \Sigma$ to denote the number of appearances of $a$ in $x$. The empty word is denoted by $\varepsilon$. The transition function $\delta$ of a DFA is extended to $\hat{\delta}: Q \times \Sigma^* \to Q$ by setting $\hat{\delta}(q, \varepsilon)=q$ and $\hat{\delta}(q, ax)=\hat{\delta}(\delta(q, a), x)$ for $q \in Q$, $a \in \Sigma$, and $x \in \Sigma^*$. In the following, we simply use $\delta$ to denote $\hat{\delta}$ if there is no confusion. A nondeterministic finite automaton (NFA) is also denoted by a quintuple $(Q, \Sigma, \eta, q_0, F)$ where $\eta \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times Q$ is a transition relation rather than a function, and $Q$, $\Sigma$, $q_0$, and $F$ are defined similarly as in a DFA. For a set $s$, we use $|s|$ to denote the cardinality of $s$. For background knowledge in automata theory, the reader may refer to [3, 9].

## 2. State complexity of catenation of two regular languages

In this section, we first give a general example which shows that for any $m \geqslant 1$ and $n>1$ there exist an $m$-state DFA $A$ and an $n$-state DFA $B$ such that any DFA accepting $L(A)L(B)$ needs at least $m2^n-2^{n-1}$ states. Then we show that for any pair of complete $m$-state DFA $A$ and $n$-state DFA $B$ defined on the same alphabet $\Sigma$, there exists a DFA with at most $m2^n-2^{n-1}$ states which accepts $L(A)L(B)$. In the case of $n=1$ and $m \geqslant 1$, we show that $m$ states are sufficient and necessary in the worst case for a DFA to accept $L(A)L(B)$.

**Theorem 2.1.** *For any integers $m \geqslant 1$ and $n \geqslant 2$, there exist a DFA $A$ of $m$ states and a DFA $B$ of $n$ states such that any DFA accepting $L(A)L(B)$ needs at least $m2^n - 2^{n-1}$ states.*

**Proof.** We first consider the cases when $m=1$ and $n \geqslant 2$. Let $\Sigma = \{a, b\}$. Since $m=1$, $A$ is a one-state DFA accepting $\Sigma^*$. Choose $B=(P, \Sigma, \delta_B, p_0, F_B)$ (Fig. 1) where $P=\{p_0, \ldots, p_{n-1}\}$, $F_B=\{p_{n-1}\}$, and $\delta_B(p_0, a)=p_0$, $\delta_B(p_0, b)=p_1$, $\delta_B(p_i, a)=p_{i+1}$, $1 \leqslant i \leqslant n-2$, $\delta_B(p_{n-1}, a)=p_1$, $\delta_B(p_i, b)=p_i$, $1 \leqslant i \leqslant n-1$. It is easy to see that

$$L(A)\,L(B)=\{w \in \Sigma^* \mid w=ubv,\ \#_a(v) \equiv n-2 \bmod (n-1)\}.$$

Let $(i_1, \ldots, i_{n-1}) \in \{0, 1\}^{n-1}$ and denote

$$w(i_1, \ldots, i_{n-1}) = b^{i_1} a b^{i_2} \cdots a b^{i_{n-1}}.$$

Then, for every $j \in \{0, \ldots, n-2\}$, $w(i_1, \ldots, i_{n-1})\, a^j \in L(A)\,L(B)$ iff $i_{j+1}=1$. Thus, a DFA accepting $L(A)L(B)$ needs at least $2^{n-1}$ states.

Now we consider the cases when $m \geqslant 2$ and $n \geqslant 2$.

Let $\Sigma=\{a, b, c\}$. Define $A=(Q, \Sigma, \delta_A, q_0, F_A)$ where $Q=\{q_0, \ldots, q_{m-1}\}$; $F_A=\{q_{m-1}\}$, for each $i$, $0 \leqslant i \leqslant m-1$,

$$\delta_A(q_i, X)=\begin{cases} q_j,\ j=(i+1) \bmod m, & \text{if } X=a,\\ q_0 & \text{if } X=b,\\ q_i & \text{if } X=c. \end{cases}$$

Define $B=(P, \Sigma, \delta_B, p_0, F_B)$ where $P=\{p_0, \ldots, p_{n-1}\}$, $F_B=\{p_{n-1}\}$, and for each $i$, $0 \leqslant i \leqslant n-1$,

$$\delta_B(p_i, X)=\begin{cases} p_j,\ j=(i+1) \bmod n, & \text{if } X=b,\\ p_i & \text{if } X=a,\\ p_1 & \text{if } X=c. \end{cases}$$

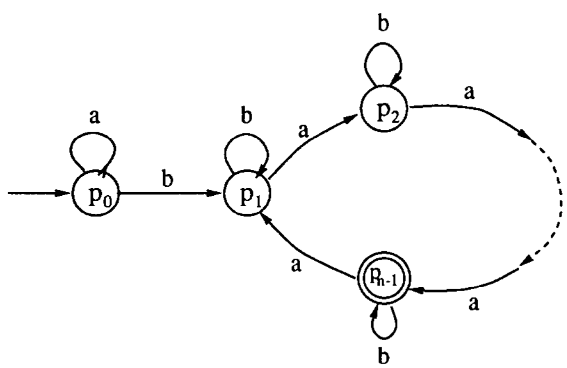


Fig. 1. DFA $B$.

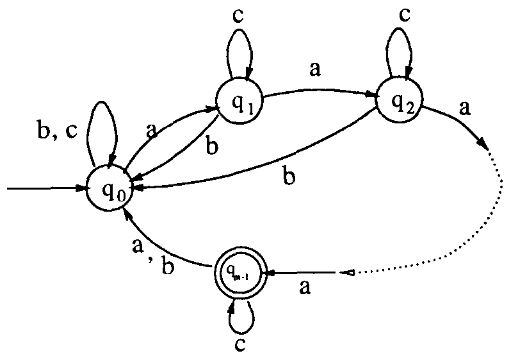


Fig. 2. DFA $A$.

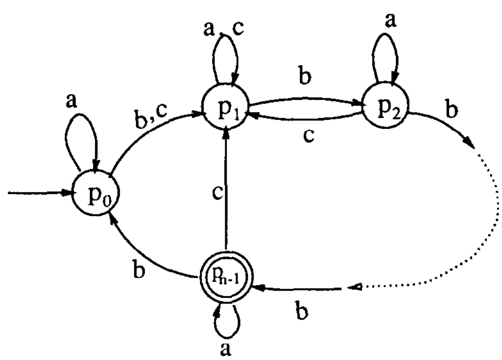


Fig. 3. DFA $B$.

The DFA $A$ and $B$ are shown in Figs. 2 and 3, respectively. The reader can verify that

$$L(A)=\{xy \mid x\in(\Sigma^*\{b\})^*,\ y\in\{a,c\}^* \text{ and } \#_a(y)=m-1 \bmod m\},$$

and

$$L(B)\cap\{a,b\}^*=\{x\in\{a,b\}^* \mid \#_b(x)=n-1 \bmod n\}.$$

Now we consider the catenation of $L(A)$ and $L(B)$, i.e. $L(A)L(B)$.

**Fact 2.2.** *For* $m>1$, $L(A)\cap\Sigma^*\{b\}=\emptyset$.

For each $x\in\{a,b\}^*$, we define

$$S(x)=\{i \mid x=uv \text{ such that } u\in L(A) \text{ and } i=\#_b(v) \bmod n\}.$$

Consider $x, y \in \{a, b\}^*$ such that $S(x) \neq S(y)$. Let $k \in S(x) - S(y)$ (or $S(y) - S(x)$). Then it is clear that $xb^{n-1-k} \in L(A)L(B)$ but $yb^{n-1-k} \notin L(A)L(B)$. So $x$ and $y$ are in different equivalence classes of the right-invariant relation induced by $L(A)L(B)$ [3].

For each $x \in \{a, b\}^*$, define $T(x) = \max\{\#(z) \mid x = yz \text{ and } z \in a^*\}$. Consider $u, v \in \{a, b\}^*$ such that $S(u) = S(v)$ and $T(u) > T(v) \bmod m$. Let $i = T(u) \bmod m$ and $w = ca^{m-1-i}b^{n-1}$. Then clearly $uw \in L(A)L(B)$ but $vw \notin L(A)L(B)$. Notice that there does not exist a word $w \in \Sigma^*$ such that $0 \notin S(w)$ and $T(w) = m-1$, since the fact that $T(w) = m-1$ guarantees that $0 \in S(w)$.

For each subset $s = \{i_1, \ldots, i_t\}$ of $\{0, \ldots, n-1\}$, where $i_1 > \cdots > i_t$, and an integer $j \in \{0, \ldots, m-1\}$ except the case when both $0 \notin s$ and $j = m-1$ are true, there exists a word

$$x = a^{m-1}b^{i_1-i_2}a^{m-1}b^{i_2-i_3}a^{m-1} \ldots a^{m-1}b^{i_t+n}a^j$$

such that $S(x) = s$ and $T(x) = j$. Thus, there are at least $m2^n - 2^{n-1}$ distinct equivalence classes. □

The next theorem gives an upper bound which coincides exactly with the above lower bound. Therefore, the bound is tight.

**Theorem 2.3.** *Let $A$ and $B$ be two complete DFAs defined on the same alphabet, where $A$ has $m$ states and $B$ has $n$ states, and let $A$ have $k$ final states, $0 < k < m$. Then there exists a $(m2^n - k2^{n-1})$-state DFA which accepts $L(A)L(B)$.*

**Proof.** Let $A = (Q, \Sigma, \delta_A, q_0, F_A)$ and $B = (P, \Sigma, \delta_B, p_0, F_B)$. Construct $C = (R, \Sigma, \delta_C, r_0, F_C)$ such that

$R = Q \times 2^P - F_A \times 2^{P - \{p_0\}}$ where $2^X$ denotes the power set of $X$,

$r_0 = \langle q_0, \emptyset \rangle$ if $q_0 \notin F_A$, $r_0 = \langle q_0, \{p_0\} \rangle$ otherwise,

$F_C = \{\langle q, T \rangle \in R \mid T \cap F_B \neq \emptyset\}$;

$\delta_C(\langle q, T \rangle, a) = \langle q', T' \rangle$, for $a \in \Sigma$, where $q' = \delta_A(q, a)$ and

$T' = \delta_B(T, a) \cup \{p_0\}$ if $q' \in F_A$, $T' = \delta_B(T, a)$ otherwise.

Intuitively, $R$ is a set of pairs such that the first component of each pair is a state in $Q$ and the second component is a subset of $P$. $R$ does not contain those pairs whose first component is a final state of $A$ and whose second component does not contain the initial state of $B$. Clearly, $C$ has $m2^n - k2^{n-1}$ states. The reader can easily verify that $L(C) = L(A)L(B)$. □

We still need to consider the cases when $m \geqslant 1$ and $n = 1$. We have the following result.

**Theorem 2.4.** *The number of states that is sufficient and necessary in the worst case for a DFA to accept the catenation of an m-state DFA language and a 1-state DFA language is m.*

**Proof.** Let $\Sigma$ be an alphabet and $a \in \Sigma$. Clearly, for any integer $m > 0$, the language $L = \{w \in \Sigma^* \mid \#_a(w) \equiv m-1 \bmod m\}$ is accepted by an $m$-state DFA. Note that $\Sigma^*$ is accepted by a one-state DFA. It is easy to see that any DFA accepting $L\Sigma^* = \{w \in \Sigma^* \mid \#_a(w) \geqslant m-1\}$ needs at least $m$ states. So we have proved the necessary condition.

Let $A$ and $B$ be an $m$-state DFA and a 1-state DFA, respectively. Since $B$ is a complete DFA, $L(B)$ is either $\emptyset$ or $\Sigma^*$. We need to consider only the case $L(B) = \Sigma^*$. Let $A = (Q, \Sigma, \delta_A, q_0, F_A)$. Define $C = (Q, \Sigma, \delta_C, q_0, F_A)$, where for any $X \in \Sigma$ and $q \in Q$,

$$\delta_C(q, X) = \begin{cases} \delta_A(q, X) & \text{if } q \notin F_A, \\ q & \text{if } q \in F_A. \end{cases}$$

The automaton $C$ is exactly as $A$ except that the final states are made to be sink states: when the computation has reached some final state $q$, it remains there. Now it is clear that $L(C) = L(A)\Sigma^*$. $\square$

## 3. State complexity of star operation on regular languages

In [8], an example is given to show that any DFA accepting the star of an $n$-state DFA language needs at least $2^{n-1}$ states in some cases for $n > 0$. Here we improve that result and show that $2^{n-1} + 2^{n-2}$ is necessary in the worst case for a DFA to accept the star of an $n$-state DFA language for each $n > 1$. We use a very different technique and use a two-letter alphabet. However, we give the *sufficient* condition first.

**Theorem 3.1.** *For any n-state DFA $A = (Q, \Sigma, \delta, q_0, F)$ such that $|F - \{q_0\}| = k \geqslant 1$ and $n > 1$, there exists a DFA of at most $2^{n-1} + 2^{n-k-1}$ states that accepts $(L(A))^*$.*

**Proof.** Let $A = (Q, \Sigma, \delta, q_0, F)$ and $L = L(A)$. Denote $F - \{q_0\}$ by $F_0$. Then $|F_0| = k \geqslant 1$. We construct a DFA $A' = (Q', \Sigma, \delta', q_0', F')$ where

$q_0' \notin Q$ is a new start state,

$Q' = \{q_0'\} \cup \{P \mid P \subseteq (Q - F_0) \text{ and } P \neq \emptyset\}$

$\cup \{R \mid R \subseteq Q \text{ and } q_0 \in R \text{ and } R \cap F_0 \neq \emptyset\}$,

$\delta'(q_0', a) = \{\delta(q_0, a)\}$ for any $a \in \Sigma$, and $\delta'(R, a) = \delta(R, a)$ for $R \subseteq Q$ and

$a \in \Sigma$ if $\delta(R, a) \cap F_0 = \emptyset$, $\delta'(R, a) = \delta(R, a) \cup \{q_0\}$ otherwise,

$F' = \{q_0'\} \cup \{R \mid R \subseteq Q \text{ and } R \cap F \neq \emptyset\}$.

The reader can verify that $L(A')=L^*$. Now we consider the number of states in $Q'$. Note that in the second term of the union for $Q'$, there are $2^{n-k}-1$ states. In the third term, there are $(2^k-1)2^{n-k-1}$ states. So $|Q'|=2^{n-1}+2^{n-k-1}$. □

Note that if $q_0$ is the only final state of $A$, $(L(A))^*=L(A)$.

**Corollary 3.2.** *For any n-state DFA A, $n>1$, there exists a DFA $A'$ of at most $2^{n-1}+2^{n-2}$ states such that $L(A')=(L(A))^*$.*

**Proof.** Let $k$ be defined as in the proof above. If $k=0$, then $A'$ simply needs $n$ states. If $k\geqslant 1$, then the claim is clearly true by Theorem 3.1. □

**Theorem 3.3.** *For any integer $n\geqslant 2$, there exists a DFA A of n states such that any DFA accepting $(L(A))^*$ needs at least $2^{n-1}+2^{n-2}$ states.*

**Proof.** For $n=2$, it is clear that $L=\{w\in\{a,b\}^* \mid \#_a(w) \text{ is odd}\}$ is accepted by a two-state DFA, and $L^*=\{\varepsilon\}\cup\{w\in\{a,b\}^* \mid \#_a(w)\geqslant 1\}$ cannot be accepted by a DFA with less than 3 states.

For $n>2$, we give the following construction: $A_n=(Q_n, \Sigma, \delta_n, 0, \{n-1\})$ where $Q_n=\{0,\ldots,n-1\}$, $\Sigma=\{a,b\}$, $\delta(i,a)=(i+1) \bmod n$ for each $0\leqslant i<n$, $\delta(i,b)=(i+1) \bmod n$ for each $1\leqslant i<n$ and $\delta(0,b)=0$. $A_n$ is shown in Fig. 4.

We construct the DFA $A'_n=(Q'_n, \Sigma, \delta'_n, q'_0, F'_n)$ from $A_n$ exactly as described in the proof of the previous theorem. We need to show that (I) every state 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 state set. (Note that each state is a subset of $Q_n$ except $q'_0$.)

Consider all $q$ such that $q\in Q'$ and $|q|=1$. We have $\{0\}=\delta'_n(q'_0,b)$ and $\{i\}=\delta'_n(i-1,a)$ for each $0<i<n-1$.

Assume that all $q$ such that $|q|<k$ are reachable. Consider $q$ where $|q|=k$. Let $q=\{i_1,i_2,\ldots,i_k\}$ such that $0\leqslant i_1<i_2<\cdots<i_k<n-1$ if $n-1\notin q$, $i_1=n-1$ and $0=i_2<\cdots<i_k<n-1$ otherwise. There are four cases:

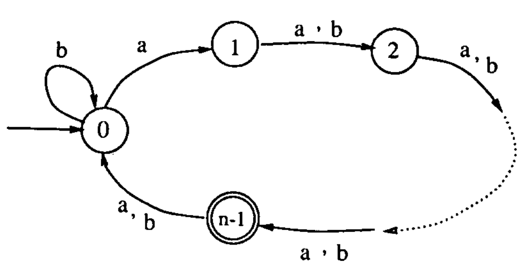


Fig. 4. DFA $A_n$.

(i) $i_1=n-1$ and $i_2=0$. Then $q=\delta_n'(\{n-2, i_3-1, \ldots, i_k-1\}, a)$ where the latter state contains $k-1$ states.

(ii) $i_1=0$ and $i_2=1$. Then $q=\delta_n'(q', a)$ where $q'=\{n-1, 0, i_3-1, \ldots, i_k-1\}$ which is considered in case (i).

(iii) $i_1=0$ and $i_2=1+t$ for $t>0$. Then $q=\delta_n'(q', b^t)$ where $q'=\{0, 1, i_3-t, \ldots, i_k-t\}$. The latter state is considered in case (ii).

(iv) $i_1=t>0$. Then $q=\delta_n'(q', a^t)$ where $q'=\{0, i_2-t, \ldots, i_k-t\}$ is considered in either case (ii) or case (iii).

To prove (II), let $i\in p-q$ for some $p, q\in Q_n'$ and $p\neq q$. Then $\delta_n'(p, a^{n-1-i})\in F_n'$ but $\delta_n'(q, a^{n-1-i})\notin F_n'$. □

Note that a DFA accepting the star of a 1-state DFA language may need up to two states. For example, $\emptyset$ is accepted by a 1-state DFA and any DFA accepting $\emptyset^*=\{\varepsilon\}$ has at least two states.

## 4. Left and right quotient, reversal and other operations

**Theorem 4.1.** *For any integer $n>0$, $2^n-1$ states are sufficient and necessary in the worst case for a DFA to accept the left quotient of an n-state DFA language R by an arbitrary language L ($L\backslash R$).*

**Proof.** We show that $2^n-1$ states are sufficient in the following. Let $M=(Q, \Sigma, \delta, s, F)$ be a complete DFA of $n$ states and $R=L(M)$. For each $q\in Q$, denote by $L(M_q)$ the set $\{w\in\Sigma^* \mid \delta(s, w)=q\}$. As above we construct an NFA $M'$ with multiple initial states to accept $L\backslash R$ as follows. $M'$ is the same as $M$ except that the initial state $s$ of $M$ is replaced by the set of initial states $S=\{q \mid L(M_q)\cap L\neq\emptyset\}$. By using the standard subset construction, the reader can easily verify that there exists a DFA of no more than $2^n-1$ states that is equivalent to $M'$. (Note that $\emptyset$ is not a state of $M'$.)

Now we show that $2^n-1$ states are necessary in the worst case. For any integer $n>0$, let $M=(Q, \Sigma, \delta, 0, F)$ be an $n$-state DFA shown in Fig. 4, where $Q=\{0, \ldots, n-1\}$ and $F=\{n-1\}$, and $R=L(M)$. Let $L=\Sigma^*$. We construct an NFA with multiple initial states $N=(Q, \Sigma, \delta, S, F)$ where $S=Q$. Clearly, $L(N)=L\backslash R$. Let the DFA $N'$ be $(Q', \Sigma, \delta', s', F')$ such that $Q'=2^Q-\{\emptyset\}$, $\delta'(X, a)=\{q\in Q \mid \exists p\in X$ such that $\delta(p, a)=q\}$ for each $X\in Q'$ and $a\in\Sigma$, $s'=S$, and $F'=\{X\in Q' \mid n-1\in X\}$. It is easy to see that $N'$ is equivalent to $N$. It remains to prove that $N'$ is minimal, i.e. (1) each state of $N'$ is reachable from the initial state $s'$ and (2) each state defines a distinct class of the right-invariant relation of the regular language $L(N')=L\backslash R$. For (1), the reader can verify that each state $X\in Q'$ can be reached from $s'$ on the string $x_0x_{n-1}\ldots x_1$ where, for each $0\leqslant j\leqslant n-1$, $x_j=a$ if $j\in X$ and $x_j=b$, otherwise. For (2), consider two arbitrary states $X, Y\in Q'$ and $X\neq Y$. Let $i\in X-Y$ (or $Y-X$). Then it is clear that $\delta'(X, a^{n-1-i})\in F'$ but $\delta'(Y, a^{n-1-i})\notin F'$(or vice versa). □

In the first part of the above proof, in order to make the construction effective, one needs to impose some restrictions, e.g., context-freeness, on the language $L$.

For a DFA to accept the right quotient of an $n$-state DFA language $R$ by an arbitrary language $L$, $n$ states are sufficient and necessary in the worst case. Let $A=(Q, \Sigma, \delta, s, F)$ be the $n$-state DFA accepting $R$. Then $R/L$ is accepted by a DFA which is exactly the same as $A$ except that the final state set is the set of all states $q \in Q$ such that there exists a word $w \in L$ such that $\delta(q, w) \in F$. The necessity can be shown by letting $L=\{\varepsilon\}$.

It is clear that any DFA accepting the reversal of an $n$-state DFA language does not need more than $2^n$ states. But can this upper bound be reached? In [1], a result on alternating finite automata (Theorem 5.3) implies a positive answer to the above question in the case where $n$ is in the form $2^k$ for some integer $k \geqslant 0$. Leiss has solved this problem in [5] for all $n>0$. A modification of Leiss's solution is shown in Fig. 5.

**Theorem 4.2.** *In the worst case, $2^n$ states are both sufficient and necessary for a DFA to accept the reversal of an n-state DFA language.*

The next theorem is obvious.

**Theorem 4.3.** *In the worst case, $m \cdot n$ states are both sufficient and necessary for a DFA to accept the intersection (union) of an m-state DFA language and an n-state DFA language.*

**Proof.** For intersection, let $L_1=\{x \in \{a, b\}^* \mid \#_a(x)=0 \bmod m\}$ and $L_2=\{y \in \{a, b\}^* \mid \#_b(x)=0 \bmod n\}$. For union, use $\overline{L_1}$ and $\overline{L_2}$. $\square$

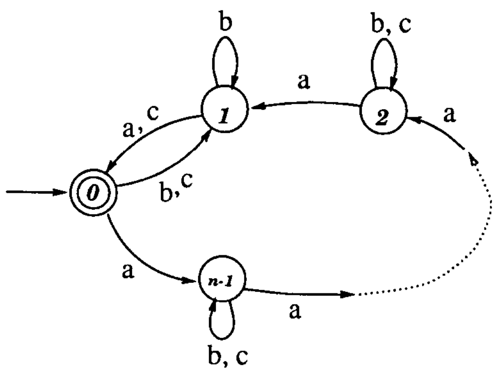


Fig. 5. DFA $B_n$.

## 5. One-letter regular languages

For regular languages over a one-letter alphabet, the results above do not hold in general. For example, it is obvious that a regular language over a one-letter alphabet has the same state complexity as its reversal, while in the two-letter alphabet case, the complexity can be much higher. In the following, we show that the optimal upper bound for the number of states which is needed for a DFA to accept the star of an $n$-state DFA language over a one-letter alphabet is $(n-1)^2+1$, and this upper bound can be reached for any $n>1$. For the catenation of an $m$-state DFA language and an $n$-state DFA language, the optimal upper bound is $mn$ in general, and we show that this bound can be reached for any $m, n \geqslant 1$ such that $(m, n)=1$ ($m$ and $n$ are relatively prime). Again we assume that all the DFAs are complete. Therefore, there is one and exactly one loop in the transition diagram of each DFA over a one-letter alphabet.

The following lemma is essential to the next two results. Although its proof uses only elementary number theory, for the sake of completeness we prove one case as an example.

**Lemma 5.1.** *Let $m$, $n>0$ be two arbitrary integers such that $(m, n)=1$ ($m$ and $n$ are relatively prime).*

(i) *The largest integer that cannot be presented as $cm+dn$ for any integers $c, d>0$ is $mn$.*

(ii) *The largest integer that cannot be presented as $cm+dn$ for any integers $c>0$ and $d \geqslant 0$ is $(m-1)n$.*

(iii) *The largest integer that cannot be presented as $cm+dn$ for any integers $c, d \geqslant 0$ is $mn-(m+n)$.*

**Proof.** Let us consider (ii) only. (i) and (iii) can be proved similarly. It suffices to show that $(m-1)n$ cannot be presented as $cm+dn$ for any integers $c>0$ and $d \geqslant 0$, but $(m-1)n+i$ can be presented for any integer $i$, $1 \leqslant i \leqslant m$.

Assume that $(m-1)n = cm+dn$ for some $c>0$ and $d \geqslant 0$. Then

$$n \mid (m-1)n \Rightarrow n \mid (cm+dn) \Rightarrow n \mid cm \Rightarrow n \mid c.$$

Since $c<n$, this is a contradiction.

Define a mapping $f: \{1, \ldots, m\} \rightarrow \{1, \ldots, m\}$ as follows. For $k \in \{1, \ldots, m\}$, let $f(k)$ be the integer $i \in \{1, \ldots, m\}$ such that $kn \equiv i \pmod{m}$, i.e. $kn = i + j_i m$ where $0 \leqslant j_i < n$. Note that $f$ is bijective since $(m, n)=1$. Thus, every $i \in \{1, \ldots, m\}$ can be written in the form $f^{-1}(i)n - j_i m$ where $0 \leqslant j_i < n$. Then

$$(m-1)n+i = (m-1)n + f^{-1}(i)n - j_i m = (n-j_i)m + (f^{-1}(i)-1)n$$

where $n-j_i>0$ and $f^{-1}(i)-1 \geqslant 0$. $\square$

**Fact 5.2.** *Let $R \subseteq \Sigma^*$ be a regular language. If there exists an integer $n$ such that*

$$\max\{\#(w) \mid w \in \Sigma^* \text{ and } w \notin R\} = n,$$

*then any DFA accepting R needs at least $n+2$ states. In particular, if $\Sigma$ is a singleton, the minimal DFA accepting R uses exactly $n+2$ states.*

**Theorem 5.3.** *The number of states which is sufficient and necessary in the worst case for a DFA to accept the star of an n-state DFA language, $n>1$, over a one-letter alphabet is* $(n-1)^2+1$.

**Proof.** For $n=2$, the necessity is shown by a 2-state DFA which accepts $(aa)^*$. For each $n>2$, the necessary condition can be shown by the DFA $A=(\{0,\ldots,n-1\},\{a\},\delta,0,\{n-1\})$ where $\delta(i,a)=i+1 \bmod n$ for each $i$, $0\leqslant i\leqslant n-1$. The star of $L(A)$ is the language $\{a^i \mid i=c(n-1)+dn$, for some integers $c>0$ and $d\geqslant 0$, or $i=0\}$. By (ii) of Lemma 5.1, the largest $i$ such that $a^i\notin(L(A))^*$ is $(n-2)n$. So the minimal DFA that accepts $(L(A))^*$ has $(n-2)n+2$, i.e. $(n-1)^2+1$, states.

The proof for showing that $(n-1)^2+1$ states are sufficient is more interesting. Let $A=(Q,\{a\},\delta,s,F)$ be an arbitrary $n$-state DFA, $n>1$ and $R=L(A)$. If $s$ is the only final state of $A$, then $R^*=R$. So we assume that there is at least one final state $f$ such that $f\neq s$. Clearly, $R^*$ (excluding $\varepsilon$ if $s\notin F$) is accepted by the NFA $A'=(Q,\{a\},\delta',s,F)$ where $\delta'=\delta\cup\{(q,\varepsilon,s)\mid q\in F\}$. For any $X\subseteq Q$, denote by $closure(X)$ the set $X\cup\{q\in Q\mid (p,\varepsilon,q)\in\delta'$ for some $p\in X\}$. Now we follow the subset construction approach to build a DFA $B=(P,\{a\},\eta,\{s\},F_P)$ from $A'$ to accept $R^*$ such that $P\subseteq 2^Q$, $\eta(X,a)=closure(\{q\in Q\mid$ there exists $p\in X$ such that $(p,a,q)\in\delta'\})$, and $F_P=\{X\in P\mid X\cap F\neq\emptyset$ or $X=\{s\}\}$. Let $f$ be the first final state from $s$ in $A$ and let $a^t$ be the shortest word such that $\delta(s,a^t)=f$. Then $\eta(\{s\},a^t)=\{s,f\}$. Denote by $p_{k_i}$ the state $\eta(\{s\},a^{it})$ in $P$, $i\geqslant 0$, which is a subset of $Q$.

We claim that $p_{k_i}\supseteq p_{k_{i-1}}$ for all $i\geqslant 1$. It is true for $i=1$ because $\eta(\{s\},a^t)=\{s,f\}$, and also true for $i>1$ since

$$\begin{aligned}p_{k_i}&=\eta(\{s\},a^{it})=\eta(\{s,f\},a^{(i-1)t})=\eta(\{s\},a^{(i-1)t})\cup\eta(\{f\},a^{(i-1)t})\\&=p_{k_{i-1}}\cup\eta(\{f\},a^{(i-1)t}).\end{aligned}$$

Then one of the following must be true:

(1) $p_{k_i}=p_{k_{i-1}}$ for some $i\leqslant n-1$,

(2) $p_{k_{n-1}}=Q$.

This is because if (1) is false, $p_{k_{n-1}}$ contains at least $n$ states and, therefore, (2) is true. Note that if (2) is true, then $\eta(p_{k_{n-1}},a)=p_{k_{n-1}}$. In any of the cases, the number of states of $B$ is no more than $t(n-1)+1$ which is at most $(n-1)^2+1$. □

**Theorem 5.4.** *Let $m,n$ be two arbitrary positive integers such that $(m,n)=1$. Then there exist an m-state DFA language $R_1$ and an n-state DFA language $R_2$, over a one-letter alphabet, such that any DFA accepting $R_1R_2$ needs at least mn states.*

**Proof.** Let $R_1=a^{m-1}(a^m)^*$ and $R_2=a^{n-1}(a^n)^*$. Obviously, $R_1$ and $R_2$ can be accepted by an $m$-state DFA and an $n$-state DFA, respectively. Then $R_1R_2=\{a^i\mid i=$

$(m-1)+(n-1)+cm+dn$ for some integers $c,d\geqslant 0\}$. By Lemma 5.1 (iii), the largest $i$ such that $a^i\notin R_1R_2$ is $mn-2$. So the minimal DFA that accepts $R_1R_2$ has $mn$ states. $\square$

**Theorem 5.5.** *For any integers $m,n\geqslant 1$, let $A$ and $B$ be an $m$-state DFA and an $n$-state DFA, respectively, over a one-letter alphabet. Then there exists a DFA of at most $mn$ states that accepts $L(A)L(B)$.*

**Proof.** The cases when $m=1$ or $n=1$ are trivial. We assume that $m,n\geqslant 2$ in the following. Let $A=(Q_A,\{a\},\delta_A,s_A,F_A)$ and $B=(Q_B,\{a\},\delta_B,s_B,F_B)$. By a variation of the subset construction, we know that $L(A)L(B)$ is accepted by the DFA $C=(Q_C,\{a\},\delta_C,s_C,F_C)$ where

$$Q_C=\{\langle q,P\rangle \mid q\in Q_A \text{ and } P\subseteq Q_B\},$$

$$s_C=\langle s_A,\emptyset\rangle \text{ if } s_A\notin F_A \text{ and } s_C=\langle s_A,\{s_B\}\rangle \text{ if } s_A\in F_A,$$

$$\delta_C(\langle q,P\rangle,a)=\langle q',P'\rangle \text{ where } q'=\delta_A(q,a) \text{ and } P'=\delta_B(P,a)\cup\{s_B\} \text{ if}$$
$$q'\in F_A,\ P'=\delta_B(P,a) \text{ otherwise;}$$

$$F_C=\{\langle q,P\rangle \mid P\cap F_B\neq\emptyset\}.$$

Now we show that at most $mn$ states of $Q_C$ are reachable from $s_C$.

First we assume that in $A$ there is a final state $f$ in the loop of the transition diagram of $A$. Then $\delta_A(s_A,a^t)=f$ and $\delta_A(f,a^l)=f$ for some nonnegative integers $t<m$ and $l\leqslant m$. Let $j_1,\ldots,j_r, 0<j_1<\cdots<j_r<l$, be all the integers such that $\delta_A(f,a^{j_i})\in F_A$ for each $1\leqslant i\leqslant r$. Denote

$$P_0=\{s_B\},$$

$$P_1=\{\delta_B(s_B,a^l),\delta_B(s_B,a^{l-j_1}),\ldots,\delta_B(s_B,a^{l-j_r})\},$$

and for $i\geqslant 2$ we define

$$P_i=\delta_B(P_{i-1},a^l).$$

Let $\delta_C(s_C,a^t)=\langle f,S\rangle$. Denote $S_0=S-\{s_B\}$ and $S_i=\delta_B(S_{i-1},a^l)$ for each $i\geqslant 1$. Then we have the following state transition sequence of $C$:

(1) $s_C \vdash_C^t \langle f,P_0\cup S_0\rangle$

(2) $\vdash_C^l \langle f,P_0\cup P_1\cup S_1\rangle$

(3) $\cdots\cdots\cdots$

(4) $\vdash_C^l \langle f,P_0\cup P_1\cup\cdots\cup P_{n-1}\cup S_{n-1}\rangle$

(5) $\vdash_C^l \langle f,P_0\cup P_1\cup\cdots\cup P_n\cup S_n\rangle.$

Here $p\vdash_C^k q$ stands for $\delta_C(p,a^k)=q$. Denote $P_0\cup\cdots\cup P_i$ by $\mathscr{P}_i$, $i\geqslant 0$. Let $i$ be the smallest integer such that $\mathscr{P}_{i-1}=\mathscr{P}_i$. It is clear that $i\leqslant n$ since $B$ has $n$ states. If $i=n$, then $\mathscr{P}_{n-1} = Q_B$ and

$$\langle f, \mathscr{P}_{n-1} \cup S_{n-1} \rangle = \langle f, \mathscr{P}_n \cup S_n \rangle = \langle f, Q_B \rangle.$$

Therefore, $C$ needs at most $m + l(n-1) \leqslant m + m(n-1) = mn$ states. If $i < n$, consider the set $S'_{i-1} = S_{i-1} - \mathscr{P}_{i-1}$. Note that every state in $S'_{i-1}$ is in the loop of the transition diagram of $B$. If for each element $r$ of $S'_{i-1}$, there exists $j$, $0 \leqslant j \leqslant n-i$, such that $\delta_B(r, a^{jl}) \in \mathscr{P}_{i-1}$ (i.e. $\mathscr{P}_{n-1}$), then the proof is concluded as above. Otherwise, there is an element $r_0$ of $S'_{i-1}$ and a transition sequence

$$r_0 \vdash_B^l r_1 \vdash_B^l \cdots \vdash_B^l r_{n-i}$$

such that, for some $j, k \leqslant n-i$ and $j < k$, $r_j = r_k$. (There are at most $n-i$ states not in $\mathscr{P}_{i-1}$.) Then it is easy to verify that $S_{i-1+j} = S_{i-1+k}$. Therefore, $\langle f, \mathscr{P}_{i-1+j} \cup S_{i-1+j} \rangle = \langle f, \mathscr{P}_{i-1+k} \cup S_{i-1+k} \rangle$. Thus, the number of states that are reachable from $s_C$ is at most $t + 1 + l(n-1) \leqslant (m-1) + 1 + m(n-1) = mn$.

Finally, we consider the case when no final states of $A$ are in the loop. Let $Q_A = \{0, \ldots, m-1\}$ where $s_A = 0$ and $\delta_A(0, a^i) = i$ for $0 \leqslant i \leqslant m-1$. We can assume that $m-2$ is a final state and $m-1$ loops to itself. Otherwise, $L(A)$ can be accepted by a complete DFA with less than $m$ states. Consider the following $m+n-1$ transition steps of $C$

$$s_C \vdash_C^{m-2} \langle m-2, T \rangle \vdash_C \langle m-1, T_0 \rangle \vdash_C \langle m-1, T_1 \rangle \vdash_C \cdots \vdash_C \langle m-1, T_n \rangle.$$

Let the state $\delta_B(s_B, a^{i+1})$ be $t_i$, for each $i \geqslant 0$. Note that $s_B \in T$ and $t_i$ is in $T_i$. It is clear that there exist $j, k$ such that $0 \leqslant j < k \leqslant n$ and $t_j = t_k$. Then it is not difficult to see that $\langle m-1, T_j \rangle = \langle m-1, T_k \rangle$. Therefore, at most $m+n$ states are necessary for $C$. ($m + n < mn$ for $m, n \geqslant 2$.) $\square$

## 6. Open problems

For the problems on catenations, we have considered the three-letter alphabet case and the one-letter alphabet case. We do not know whether the results in the three-letter alphabet case hold if the size of the alphabet is two.

## References

[1] A.K. Chandra, D.C. Kozen and L.J. Stockmeyer, Alternation, *J. ACM* **28** (1981) 114–133.

[2] A. Fellah, H. Jürgensen and S. Yu, Constructions for alternating finite automata, *Internat. J. Comput. Math.* **35** (1990) 117–132.

[3] J.E. Hopcroft and J.D. Ullman, *Introduction to Automata Theory, Languages, and Computation* (Addison-Wesley, Reading, MA, 1979).

[4] T. Jiang and B. Ravikumar, Minimal NFA problems are hard, in: *Proc. 18th ICALP*, Lecture Notes in Computer Science, Vol. 510 (Springer, Berlin, 1991) 629–640.

[5] E. Leiss, Succinct representation of regular languages by boolean automata, *Theoret. Comput. Sci.* **13** (1981) 323–330.

[6] A.R. Meyer and M.J. Fischer, Economy of description by automata, grammars, and formal systems, *FOCS* **12** (1971) 188–191.
[7] B. Ravikumar, Some applications of a technique of Sakoda and Sipser, *SIGACT News* **21** (4) (1990) 73–77.
[8] B. Ravikumar and O.H. Ibarra, Relating the type of ambiguity of finite automata to the succinctness of their representation, *SIAM J. Comput.* **18** (6) (1989) 1263–1282.
[9] A. Salomaa, *Theory of Automata* (Pergamon, Oxford, 1969).