RAIRO-Theor. Inf. Appl. 52 (2018) 153–168
https://doi.org/10.1051/ita/2018011

RAIRO - Theoretical Informatics and Applications
www.rairo-ita.org

THE COMPLEXITY OF CONCATENATION ON DETERMINISTIC AND ALTERNATING FINITE AUTOMATA☆

Michal Hospodár* and Galina Jirásková

Abstract. We study the state complexity of the concatenation operation on regular languages represented by deterministic and alternating finite automata. For deterministic automata, we show that the upper bound $m2^n - k2^{n-1}$ on the state complexity of concatenation can be met by ternary languages, the first of which is accepted by an m-state DFA with k final states, and the second one by an n-state DFA with ℓ final states for arbitrary integers m, n, k, ℓ with $1 \leq k \leq m-1$ and $1 \leq \ell \leq n-1$. In the case of $k \leq m-2$, we are able to provide appropriate binary witnesses. In the case of $k = m-1$ and $\ell \geq 2$, we provide a lower bound which is smaller than the upper bound just by one. We use our binary witnesses for concatenation on deterministic automata to describe binary languages meeting the upper bound $2^m + n + 1$ for the concatenation on alternating finite automata. This solves an open problem stated by Fellah *et al.* [*Int. J. Comput. Math.* **35** (1990) 117–132].

Mathematics Subject Classification. 68Q19, 68Q45.

Accepted November 21, 2018.

1. Introduction

Concatenation is a binary operation on formal languages defined as $KL = \{uv \mid u \in K \text{ and } v \in L\}$. It is known that if a language K is accepted by an m-state deterministic finite automaton (DFA) and L is accepted by an n-state DFA, then the concatenation KL is accepted by a DFA of at most $m2^n - 2^{n-1}$ states [10].

Ternary languages meeting this upper bound were described by Yu *et al.* [14]. Maslov [10] proposed binary witnesses for concatenation, but he did not provide any proof. The tightness of this upper bound in the binary case was proven in [6].

However, if the minimal DFA recognizing the first language has more than one final state, then the upper bound $m2^n - 2^{n-1}$ on the state complexity of concatenation cannot be met; here, the state complexity of a regular language is the number of states in the minimal DFA for the language, and the state complexity of a regular operation is the number of states that are sufficient and necessary in the worst case for a DFA to recognize the language resulting from the operation considered as a function of the state complexities of the operands. Yu *et al.* [14] showed that the state complexity of concatenation is at most $m2^n - k2^{n-1}$, where k is the number of final states in the minimal DFA for the first language. The binary languages meeting this upper bound were described for each k with $1 \leq k \leq m-1$ in Theorem 1 of [5], but there are some errors in the proof

☆Research supported by VEGA grant 2/0084/15 and grant APVV-15-0091. This work was conducted as a part of PhD study of the first author at Comenius University in Bratislava.

Keywords and phrases: Regular languages, finite automata, concatenation, state complexity.

Mathematical Institute, Slovak Academy of Sciences, Grešákova 6, 040 01 Košice, Slovakia.

* Corresponding author: hosmich@gmail.com

of this theorem, and one of our aims is to fix them. We also show that the witnesses from [10, 14] meet the upper bound $m2^n - k2^{n-1}$ if we make the k last states final in the DFA for the first language.

Then we study the complexity of concatenation also in the case where the second automaton has more than one final state. Our motivation comes from the paper by Fellah *et al.* [3], where the authors consider the concatenation operation on languages represented by alternating finite automata (AFA), and get an upper bound $2^m + n + 1$. They also write: *"We conjecture that this number of states is actually necessary in the worst case, but have no proof."*

It is known ([3], Thm 4.1, Cor. 4.2) and ([7], Lem. 1, Lem. 2) that a language L is accepted by an n-state AFA if and only if its reversal L^R is accepted by a 2^n-state DFA with 2^{n-1} states final. Hence to get a lower bound for concatenation on AFAs, we need two languages represented by DFAs with half of states final that are hard for concatenation on DFAs.

We first inspect the witnesses from [5, 10, 14] and show that none of them meets the upper bound $m2^n - k2^{n-1}$ if the second automaton has more then one final state. Then we describe ternary languages meeting this bound for all m, n, k, ℓ, where m and k is the number of states and the number of final states in the minimal DFA for the first language, and n and ℓ is the number of states and the number of final states in the minimal DFA for the second language. Then, in the case of $k \leq m - 2$, that is, if the first automaton has at least two non-final states, we describe appropriate binary languages. Finally, we consider the case of $k = m - 1$ and $\ell \geq 2$ over a binary alphabet. In such a case, the upper bound is $(m + 1)2^{n-1}$, and we provide languages meeting the bound $(m + 1)2^{n-1} - 1$. We strongly conjecture that this lower bound is tight, but have no proof.

We use the binary witnesses for the concatenation on DFAs to define binary languages K and L accepted by an m-state and n-state AFA, respectively, such that the minimal AFA for KL requires $2^m + n + 1$ states. This proves that the upper bound $2^m + n + 1$ from [3] is tight, and solves the open problem stated in Theorem 9.3 of [3].

2. Preliminaries

In this section, we give some basic definitions and preliminary results. For details and all unexplained notions, the reader may refer to [4, 12, 13].

Let Σ be a finite alphabet of symbols. Then Σ^* denotes the set of strings over Σ including the empty string ε. A language is any subset of Σ^*. The concatenation of languages K and L is the language $KL = \{uv \mid u \in K$ and $v \in L\}$. The cardinality of a finite set A is denoted by $|A|$, and its power-set by 2^A. We define an operator $\ominus$ as follows: If $i, j \in \{0, 1, \dots, n-1\}$, then $j \ominus i = (j - i) \bmod n$, and if $S \subseteq \{0, 1, \dots, n-1\}$, then $S \ominus i = \{j \ominus i \mid j \in S\}$.

A *nondeterministic finite automaton* (NFA) is a quintuple $N = (Q, \Sigma, \cdot, I, F)$, where Q is a finite set of states, Σ 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* N is the set $L(N) = \{w \in \Sigma^* \mid I \cdot w \cap F \neq \emptyset\}$. For a symbol a and states p and q, we say that (p, a, q) is a transition in NFA N if $q \in p \cdot a$. For a string w, we write $p \xrightarrow{w} q$ if $q \in p \cdot w$.

An NFA N is *deterministic* (DFA) and complete if $|I| = 1$ and $|q \cdot a| = 1$ for each q in Q and each a in Σ. In such a case, we write $q \cdot a = q'$ instead of $q \cdot a = \{q'\}$. The *state complexity* of a regular language L, $\mathrm{sc}(L)$, is the smallest number of states in any DFA for L.

The reversal L^R of a language L is defined as $L^R = \{w^R \mid w \in L\}$, where w^R is the mirror image of the string w. For every finite automaton $N = (Q, \Sigma, \cdot, I, F)$ we can construct the automaton $N^R = (Q, \Sigma, \cdot^R, F, I)$ where $p \in q \cdot^R a$ iff $q \in p \cdot a$ for every p, q in Q and every a in Σ. Then $L(N^R) = (L(N))^R$.

Every NFA $N = (Q, \Sigma, \cdot, I, F)$ can be converted into an equivalent DFA $D = (2^Q, \Sigma, \cdot', I, F')$ where $F' = \{S \in 2^Q \mid S \cap F \neq \emptyset\}$, and for every set S in 2^Q and every symbol a, we have $S \cdot' a = S \cdot a$ [11]. The DFA D is called the *subset automaton* of the NFA N. The subset automaton may not be minimal since some of its states may be unreachable or equivalent to other states.

We say that a state q of an NFA $N = (Q, \Sigma, \cdot, I, F)$ is *uniquely distinguishable* if there is a string w which is accepted by N from and only from the state q, that is, if we have $p \cdot w \cap F \neq \emptyset$ iff $p = q$.

Proposition 2.1. *If two subsets of states of an* NFA *differ in a uniquely distinguishable state, then the two subsets are distinguishable in the subset automaton .*

Proof. Let S and T be two subsets of states of an NFA N. Let q be a uniquely distinguishable state of N such that, without loss of generality, $q \in S \setminus T$. Then there is a string w which is accepted by N from and only from q. It follows that w is accepted by the subset automaton of N from S and rejected from T. Hence S and T are distinguishable in the subset automaton of N. □

We say that a transition (p, a, q) is a *unique in-transition* in an NFA N if there is no state r with $r \neq p$ such that (r, a, q) is a transition in N. We say that a state q is *uniquely reachable* from a state p if there is a sequence of unique in-transitions (q_{i-1}, a_i, q_i) for $i = 1, 2, \ldots, k$ such that $k \geq 1$, $q_0 = p$, and $q_k = q$.

Proposition 2.2. *Let a uniquely distinguishable state q be uniquely reachable from a state p. Then the state p is uniquely distinguishable.*

Proof. Let a string w be accepted by an NFA N from and only from a state q. If (p, a, q) is a unique in-transition, then the string aw is accepted by N from and only from the state p. Now the claim follows by induction. □

3. Construction of NFA for concatenation

Let K and L be accepted by minimal DFAs A and B, respectively. Without loss of generality, we may 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, we denote the transition function in both A and B by $\cdot$; there is no room for confusion since A and B have distinct state sets. We first recall the construction of an NFA for the concatenation of languages K and L.

Construction 3.1. (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* DFA*s. We construct* NFA $N = (\{q_0, q_1, \ldots, q_{m-1}\} \cup \{0, 1, \ldots, n-1\}, \Sigma, \cdot, I, F_B)$ *from* DFA*s* A *and* B *as follows:*

- *for each a in Σ and each state q_i of A, if $q_i \cdot a \in F_A$, then add the transition $(q_i, a, 0)$;*
- *the set I of initial states of N is $\{q_0\}$ if $q_0 \notin F_A$, and it is $\{q_0, 0\}$ otherwise;*
- *the set of final states of N is F_B.*

Using Construction 3.1, we get an upper bound on the state complexity of concatenation. Notice that the bound depends on the number of final states in the DFA for the first language.

Proposition 3.2 (Concatenation: Upper Bound if $|F_A| = k$). *Let A be an m-state* DFA *with k final states and let B be an n-state* DFA*. Then we have* $\mathrm{sc}(L(A)L(B)) \leq m2^n - k2^{n-1}$.

Proof. Consider DFAs $A = (\{q_0, q_1, \ldots, q_{m-1}\}, \Sigma, \cdot, q_0, F_A)$, where $|F_A| = k$, and $B = (\{0, 1, \ldots, n-1\}, \Sigma, \cdot, 0, F_B)$. Construct an NFA N for $L(A)L(B)$ as in the construction above, and consider the corresponding subset automaton D. Since A is deterministic and complete, each reachable subset in D is of the form $\{q_i\} \cup S$, where $S \subseteq \{0, 1, \ldots, n-1\}$. Moreover, if q_i is a final state of A, then $0 \in S$ since the NFA N has the transition $(q, a, 0)$ whenever a state q of A goes to a final state q_i on a symbol a. If q_0 is final, then D starts in $\{q_0, 0\}$. It follows that no subset containing a final state of A and not containing state 0 is reachable in D. Hence the subset automaton D has at most $m2^n - k2^{n-1}$ reachable subsets. □

Since $m2^n - k2^{n-1}$ is maximal if $k = 1$, we get the following upper bound on the state complexity of concatenation [10, 14].

Corollary 3.3 (Concatenation: Upper Bound). *Let A and B be an m-state and n-state* DFA*, respectively. Then* $\mathrm{sc}(L(A)L(B)) \leq m2^n - 2^{n-1}$. □

4. Ternary and binary witness languages

Motivated by the open problem from [3] concerning the tightness of the upper bound $2^m + n + 1$ for concatenation on alternating automata, we study the state complexity of the concatenation of languages represented by deterministic finite automata that have more than one final state. Let us start with the following observation in which we assume that the state complexity of the second language is one.

Observation 4.1. *Let $m \geq 1$ and $1 \leq k \leq m$. Let A be an m-state* DFA *with k final states and B be a 1-state* DFA*, both over an alphabet Σ. Then $\operatorname{sc}(L(A)L(B)) \leq m - k + 1$, and the bound is tight if $|\Sigma| \geq 1$.*

Proof. If a complete DFA B has one state, then either $L(B) = \emptyset$ or $L(B) = \Sigma^*$. Since $L(A)\,\emptyset = \emptyset$, and hence $\operatorname{sc}(L(A)\,\emptyset) = 1$, we assume that $L(B) = \Sigma^*$. We construct the DFA for $L(A)L(B)$ from A as follows: for every final state p and every a in Σ, we replace the transition (p, a, q) by the transition (p, a, p). The resulting automaton is deterministic and complete, has m states and k final states. All the final states are equivalent since every string is accepted from any of them. Thus we can merge all final states into a single final state. This gives the upper bound $m - k + 1$.

To prove tightness, let us consider the unary deterministic finite automaton $A = (\{0, 1, \ldots, m-1\}, \{a\}, \cdot, 0, \{q \mid m - k \leq q \leq m - 1\})$, where $q \cdot a = q + 1 \bmod m$ for $q = 0, 1, \ldots, m - 1$. For each final state p, we remove all the transitions going from p, and add the transition (p, a, p) to get a DFA for $L(A)\,\Sigma^*$. Then we merge all final states into a single final state. The resulting minimal automaton accepts the language $a^{m-k}a^*$ and has $m - k + 1$ states. □

In what follows, we assume that the state complexity of the second language is at least two. We inspect three worst-case examples from the literature, and modify them by making some states in the first automaton final. To simplify the proofs, we use the property of all these witnesses that the letter a performs the permutation $q_i \cdot a = q_{(i+1) \bmod m}$ in A and a permutation in B. If these two conditions are satisfied, then we get the following observation.

Lemma 4.2. *Let $A = (\{q_0, q_1, \ldots, q_{m-1}\}, \Sigma, \cdot, q_0, \{q_i \mid m - k \leq i \leq m - 1\})$ and $B = (Q_B, \Sigma, \cdot, 0, F_B)$, where $Q_B = \{0, 1, \ldots, n - 1\}$, be* DFA*s. Assume that there is a symbol a in Σ such that $q_i \cdot a = q_{(i+1) \bmod m}$ and the symbol a performs a permutation on Q_B. Let N be an* NFA *for $L(A)L(B)$ from Construction 3.1. Then in the subset automaton of N, we have*

1. *For each subset S of Q_B with $0 \in S$, the set $\{q_{m-k}\} \cup S$ is reachable from a set $\{q_{m-k-1}\} \cup S'$, where $S' \subseteq Q_B$ and $|S'| = |S| - 1$;*
2. *For each subset S of Q_B and each $i = 1, 2, \ldots, m - k - 1$, the set $\{q_i\} \cup S$ is reachable from a set $\{q_0\} \cup S'$, where $S' \subseteq Q_B$ and $|S'| = |S|$;*
3. *Moreover, if $0 \cdot a = 0$, then for each subset S of Q_B with $0 \in S$ and for each $i = 0, 1, \ldots, m - 1$, the set $\{q_i\} \cup S$ is reachable from a set $\{q_{m-k-1}\} \cup S'$, where $S' \subseteq Q_B$ and $|S'| = |S| - 1$.*

Proof. Since a is a permutation on Q_B, we can use $q \cdot a^{-1}$ to denote the state p with $p \cdot a = q$. Next, we can extend a^{-1} to subsets of Q_B and to a^{-i} for every positive integer i.

1. Let $S' = (S \setminus \{0\}) \cdot a^{-1}$. Then $|S'| = |S| - 1$ and the set $\{q_{m-k}\} \cup S$ is reached from $\{q_{m-k-1}\} \cup S'$ by a.
2. Let $S' = S \cdot a^{-i}$ where $i = 1, 2, \ldots, m - k - 1$. Then $|S'| = |S|$ and the set $\{q_i\} \cup S$ is reached from $\{q_0\} \cup S'$ by a^i.
3. Let $S' = (S \setminus \{0\}) \cdot a^{-(k+1+i)}$ where $i = 0, 1, \ldots, m - 1$. Then $|S'| = |S| - 1$ and $\{q_i\} \cup S$ is reached from $\{q_{m-k-1}\} \cup S'$ by a^{k+1+i} since $0 \cdot a = 0$. □

Ternary witness languages meeting the upper bound $m2^n - 2^{n-1}$ for concatenation are described in Theorem 2.1 of [14]. We modify these languages by making k states final in the first DFA. Then we prove that the state complexity of the resulting concatenation meets the upper bound $m2^n - k2^{n-1}$.

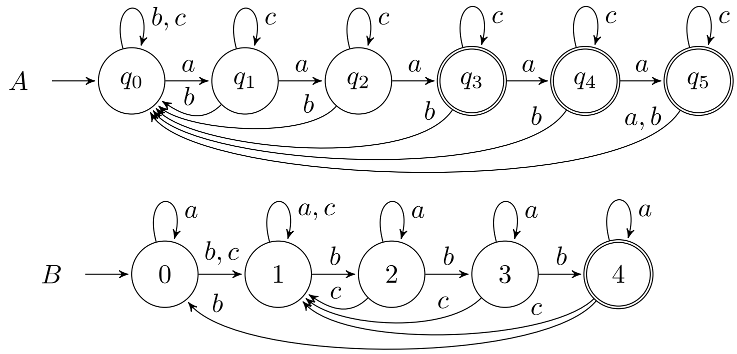

FIGURE 1. Ternary witnesses for concatenation meeting the upper bound $m2^n - k2^{n-1}$; $m = 6$, $k = 3$, and $n = 5$.

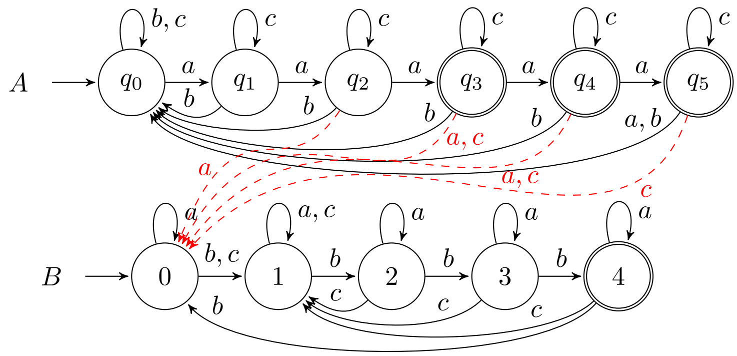

FIGURE 2. An NFA N for $L(A)L(B)$, where DFAs A and B are shown in Figure 1.

Lemma 4.3 (Ternary Witness Automata with $|F_A| = k$ and $|F_B| = 1$). *Let $m, n \geq 2$ and $1 \leq k \leq m - 1$. There exist a ternary m-state* DFA *A with k final states and a ternary n-state* DFA *B such that* $\mathrm{sc}(L(A)L(B)) = m2^n - k2^{n-1}$.

Proof. Define an m-state DFA $A = (\{q_0, q_1, \dots, q_{m-1}\}, \{a, b, c\}, \cdot, q_0, F_A)$, where $F_A = \{q_i \mid m - k \leq i \leq m - 1\}$ and for each i in $\{0, 1, \dots, m - 1\}$, we have

$$q_i \cdot a = q_{(i+1) \bmod m}, \qquad q_i \cdot b = q_0, \qquad \text{and} \qquad q_i \cdot c = q_i.$$

Define an n-state DFA $B = (Q_B, \{a, b, c\}, \cdot, 0, \{n-1\})$, where $Q_B = \{0, 1, \dots, n-1\}$ and for each j in Q_B, we have

$$j \cdot a = j, \qquad j \cdot b = (j + 1) \bmod n, \qquad \text{and} \qquad j \cdot c = 1.$$

The DFAs A and B, where $m = 6, k = 3$, and $n = 5$, are shown in Figure 1.

Construct an NFA N for $L(A)L(B)$ from DFAs A and B by adding transitions $(q_{i-1}, a, 0)$ and $(q_i, c, 0)$ for each i with $m - k \leq i \leq m - 1$; the initial state of N is q_0, and the set of final states is $\{n - 1\}$. Figure 2 shows the NFA N resulting from DFAs A and B from Figure 1. Let $\mathcal{R}$ be the following family of $m2^n - k2^{n-1}$ subsets of states of the NFA N:

$$\mathcal{R} = \big\{\{q_i\} \cup S \mid 0 \le i \le m-k-1 \text{ and } S \subseteq Q_B\big\} \cup$$
$$\big\{\{q_i\} \cup S \mid m-k \le i \le m-1,\ S \subseteq Q_B \text{ and } 0 \in S\big\}.$$

To prove the lemma, we only need to show that each subset in $\mathcal{R}$ is reachable in the subset automaton of N, and that all these subsets are pairwise distinguishable.

We first prove reachability. The proof is by induction on $|\{q_i\} \cup S|$. The basis, $|\{q_i\} \cup S| = 1$, holds true since $\{q_0\}$ is the initial subset of the subset automaton, and it goes to the subset $\{q_i\}$ by a^i if $1 \le i \le m-k-1$. Let $1 \le t \le n$, and assume that each subset in $\mathcal{R}$ of size t is reachable. Notice that the symbol a performs the permutation $q_i \cdot a = q_{(i+1) \bmod m}$ on states of A and a permutation on states of B and moreover $0 \cdot a = 0$. By Lemma 4.2 case 3, each set $\{q_i\} \cup S$ of size $t+1$, where $m-k \le i \le m-1$ and $S \subseteq Q_B$ with $0 \in S$, can be reached from a set of size t. Next, by Lemma 4.2 case 2, each set $\{q_i\} \cup S$ of size $t+1$ where $1 \le i \le m-k-1$ is reached from a set $\{q_0\} \cup S'$ of size $t+1$. Hence it is enough to show the reachability of sets $\{q_0\} \cup S$ for every subset S of Q_B such that $|\{q_0\} \cup S| = t+1$. We have

$$\{q_{m-1}\} \cup (S \ominus \min S) \cdot a^{-1} \xrightarrow{a} \{q_0\} \cup (S \ominus \min S) \xrightarrow{b^{\min S}} \{q_0\} \cup S,$$

where $0 \in S \ominus \min S$ and the set $\{q_{m-1}\} \cup (S \ominus \min S)$ can be reached from a set of size t by Lemma 4.2 case 3. This proves reachability.

To prove distinguishability, let $\{q_i\} \cup S$ and $\{q_j\} \cup T$ be two distinct subsets in $\mathcal{R}$. Notice that the state $n-1$ is uniquely distinguishable in NFA N since it is a unique final state. Next, the state $n-1$ is reached from each state of Q_B in the subgraph of unique in-transitions $(t, b, t+1)$ where $0 \le t \le n-2$. It follows that each state in Q_B is uniquely distinguishable. By Proposition 2.1, if $S \ne T$, then $\{q_i\} \cup S$ and $\{q_j\} \cup T$ are distinguishable. Now let $S = T$. Then $i \ne j$, and without loss of generality, $0 \le i < j \le m-1$. There are three cases:

1. Let $i < m-k \le j$, that is, q_i is non-final and q_j is final in A. Then $0 \notin (\{q_i\} \cup S) \cdot c$, but $0 \in (\{q_j\} \cup S) \cdot c$, so after reading c, the resulting sets differ in state 0 and are distinguishable as shown above.
2. Let $m-k \le i < j$, that is, both q_i and q_j are final in A. Then we read a^{m-j} and get the sets $\{q_0\} \cup S$ and $\{q_{m-j+i}\} \cup S$ which are considered in case 1.
3. Let $i < j < m-k$, that is, both q_i and q_j are non-final in A. Then we read a^{m-k-j} and get the sets $\{q_{m-k-j+i}\} \cup S$ and $\{q_{m-k}\} \cup \{0\} \cup S$ which either differ in state 0 or are considered in case 1.

This proves distinguishability and concludes our proof. $\square$

Yu *et al.* [14] left the binary case open. Later, a paper by Maslov [10] was found, in which the author describes binary witnesses meeting the upper bound $m2^n - 2^{n-1}$ assuming that $n \ge 3$. Let us show that his witnesses, modified to have k final states in A as shown in Figure 3 for $m = 6, k = 3$, and $n = 5$, meet the upper bound $m2^n - k2^{n-1}$ whenever $n \ge 3$.

Lemma 4.4 (Binary Witnesses with $|F_A| = k$ and $|F_B| = 1$; $n \ge 3$). *Let $m \ge 2$, $n \ge 3$, and $1 \le k \le m-1$. There exist a binary m-state* DFA *A with k final states and a binary n-state* DFA *B such that* $\mathrm{sc}(L(A)L(B)) = m2^n - k2^{n-1}$.

Proof. Define an m-state DFA $A = (\{q_0, q_1, \ldots, q_{m-1}\}, \{a, b\}, \cdot, q_0, F_A)$, where $F_A = \{q_i \mid m-k \le i \le m-1\}$ and for each i in $\{0, 1, \ldots, m-1\}$, we have $q_i \cdot a = q_{(i+1) \bmod m}$ and $q_i \cdot b = q_i$.

Define an n-state DFA $B = (\{0, 1, \ldots, n-1\}, \{a, b\}, \cdot, 0, \{n-1\})$, where for each state j of B, we have $j \cdot a = j$ if $j \le n-3$, $(n-2) \cdot a = n-1$, $(n-1) \cdot a = n-2$, and $j \cdot b = j+1$ if $j \le n-2$, $(n-1) \cdot b = n-1$.

The DFAs A and B, where $m = 6, k = 3$, and $n = 5$, are shown in Figure 3.

Construct an NFA N for $L(A)L(B)$ from DFAs A and B by adding transitions $(q_{i-1}, a, 0)$ and $(q_i, b, 0)$ for each i with $m-k \le i \le m-1$; the initial state of N is q_0, and the set of final states is $\{n-1\}$. Let $\mathcal{R}$ be the same family of $m2^n - k2^{n-1}$ subsets as in the previous proof. We need to show that all sets in $\mathcal{R}$ are reachable

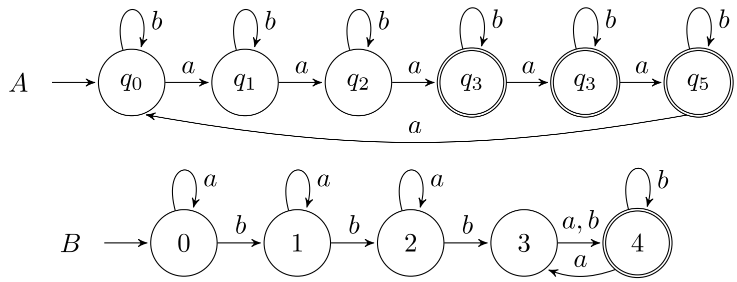

FIGURE 3. Binary witnesses for concatenation meeting the upper bound $m2^n - k2^{n-1}$ assuming that $n \geq 3$; $m = 6$, $k = 3$, $n = 5$.

and pairwise distinguishable in the subset automaton of N. The proof of reachability is exactly the same as in the proof of Lemma 4.3.

To prove distinguishability, let $\{q_i\} \cup S$ and $\{q_j\} \cup T$ be two distinct subsets in $\mathcal{R}$. Notice that the state $n-1$ is uniquely distinguishable since it is a unique final state in N. Next, the state $n-1$ is uniquely reachable from each state in $\{0, 1, \ldots, n-1\}$ through the following unique in-transitions $0 \xrightarrow{b} 1 \xrightarrow{b} \cdots \xrightarrow{b} n-2 \xrightarrow{a} n-1$. It follows that each state in $\{0, 1, \ldots, n-1\}$ is uniquely distinguishable. By Proposition 2.1, if $S \neq T$, then $\{q_i\} \cup S$ and $\{q_j\} \cup T$ are distinguishable. Now let $S = T$. Then $i \neq j$, and without loss of generality, $0 \leq i < j \leq m-1$. There are four cases:

1. Let $i < m-k \leq j$, so $0 \in S$. Then we read b and get $\{q_i\} \cup (S \cdot b)$ and $\{q_j\} \cup \{0\} \cup (S \cdot b)$, which differ in state 0 since $0 \notin S \cdot b$.
2. If $m-k \leq i < j$, then we read a^{m-j} and get $\{q_{m-j+i}\} \cup (S \cdot a^{m-j})$ and $\{q_0\} \cup (S \cdot a^{m-j})$, which are considered in case 1.
3. If $i < j < m-k$ and $0 \in S$, then we read a^{m-k-j} and get $\{q_{m-k-j+i}\} \cup (S \cdot a^{m-k-j})$ and $\{q_{m-k}\} \cup (S \cdot a^{m-k-j})$, which are considered in case 1.
4. If $i < j < m-k$ and $0 \notin S$, then we read a^{m-k-j} and get $\{q_{m-k-j+i}\} \cup (S \cdot a^{m-k-j})$ and $\{q_{m-k}\} \cup \{0\} \cup (S \cdot a^{m-k-j})$, which differ in state 0.

This concludes our proof. $\square$

While the ternary witnesses from Lemma 4.3 require $m \geq 2$ and $n \geq 2$, the binary witnesses from Lemma 4.4 do not work if $n = 2$. In Theorem 1 from [5], binary witnesses for $m \geq 1$ and $n \geq 2$ are described. However, the proof of Theorem 1 from [5] does not work. For example, it is claimed that the set $\{q_{m-k-1}, j_2 - 1, \ldots, j_s - 1\}$ goes to $\{q_{m-k+1}, 0, j_2, \ldots, j_s\}$ by aab^{n-1}; *cf.* line -4 on page 515. In fact it goes to $\{q_{m-k+1}, 0\}$. Such an error occurs several times in the proof, namely, on line -2 on page 515, and on lines 2 and 8 on page 516. The authors overlooked that ab^{n-1} does not perform an identity on $\{0, 1, \ldots, n-1\}$, but moves this set to $\{0\}$. Here we provide a correct proof.

Lemma 4.5 (**[5], Binary Witness Automata with $|F_A| = k$ and $|F_B| = 1$).** *Let $m \geq 1$ and $n \geq 2$. Let $k = 1$ if $m = 1$, and $1 \leq k \leq m-1$ otherwise. There exist a binary m-state* DFA *A with k final states and a binary* DFA *B such that* $\mathrm{sc}(L(A)L(B)) = m2^n - k2^{n-1}$.

Proof. Define an m-state DFA $A = (\{q_0, q_1, \ldots, q_{m-1}\}, \{a, b\}, \cdot, q_0, F_A)$, where $F_A = \{q_i \mid m-k \leq i \leq m-1\}$ and for each i in $\{0, 1, \ldots, m-1\}$, we have $q_i \cdot a = q_{(i+1) \bmod m}$ and $q_i \cdot b = q_i$.

Define an n-state DFA $B = (\{0, 1, \ldots, n-1\}, \{a, b\}, \cdot, 0, \{n-1\})$, where for each state j of B, we have $j \cdot a = (j+1) \bmod n, 0 \cdot b = 0,$ and $j \cdot b = (j+1) \bmod n$ if $j \geq 1$. The DFAs A and B, where $m = 6, k = 3$, and $n = 5$, are shown in Figure 4.

First let $m = 1$, so $L(A) = \{a, b\}^*$. Construct an NFA N for $L(A)L(B)$ from the DFA B by adding the transition $(0, a, 0)$. In the subset automaton of N, the singleton set $\{0\}$ is the initial subset, and each

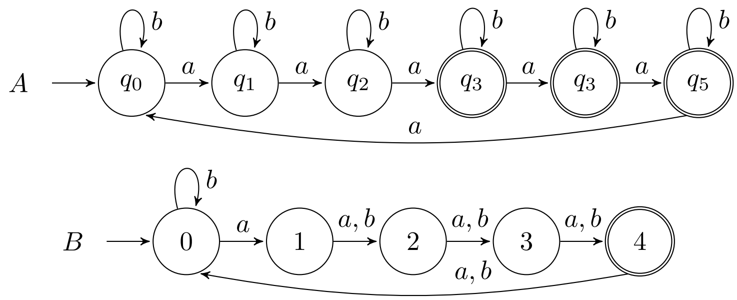

FIGURE 4. Binary witnesses meeting the bound $m2^n - k2^{n-1}$; $m = 6$, $k = 3$, and $n = 5$ [5].

subset S of size $t + 1$ such that $0 \in S$ is reached from the subset $(S \setminus \{0\}) \ominus \min(S \setminus \{0\})$ of size t by the string $ab^{\min(S\setminus\{0\})-1}$. Since the state $n - 1$ is uniquely distinguishable and uniquely reachable from every other state in $\{0, 1, \ldots, n - 1\}$, all the states of the subset automaton of N are pairwise distinguishable by Proposition 2.1. Hence $\mathrm{sc}(L(A)L(B)) = 2^{n-1}$.

Now let $m \geq 2$. Construct an NFA N for $L(A)L(B)$ from DFAs A and B as in the Construction 3.1. Let $\mathcal{R}$ be the same family of $m2^n - k2^{n-1}$ subsets as in the proof of Lemma 4.3. Let us show that each subset $\{q_i\} \cup S$ in $\mathcal{R}$ is reachable in the subset automaton of N. The proof is by induction on $|\{q_i\} \cup S|$.
The basis, with $|\{q_i\} \cup S| \leq 2$, holds true, since we have

$$\begin{aligned}
&\{q_0\} \xrightarrow{a^i} \{q_i\} \quad (1 \leq i \leq m - k - 1),\\
&\{q_{m-k-1}\} \xrightarrow{a} \{q_{m-k}, 0\} \xrightarrow{(ab^n)^i} \{q_{m-k+i}, 0\} \quad (1 \leq i \leq k - 1),\\
&\{q_{m-1}, 0\} \xrightarrow{a} \{q_0, 1\} \xrightarrow{b^{j-1}} \{q_0, j\} \quad (2 \leq j \leq n - 1),\\
&\{q_0, n - 1\} \xrightarrow{b} \{q_0, 0\},\\
&\{q_0, j \ominus i\} \xrightarrow{a^i} \{q_i, j\} \quad (1 \leq i \leq m - k - 1).
\end{aligned}$$

Let $1 \leq t \leq n$, and assume that each set in $\mathcal{R}$ of size t is reachable. By Lemma 4.2 case 1, every set $\{q_{m-k}\} \cup S$ in $\mathcal{R}$ of size $t+1$ is reachable from a set in $\mathcal{R}$ of size t. Now let $\{q_i\} \cup S$ be a set in $\mathcal{R}$ of size $t+1$ with $i \neq m-k$. Consider four cases:

(*i*) Let $m - k + 1 \leq i \leq m - 1$, so $0 \in S$. Take $S' = S \setminus \{0\}$. Then

$$\{q_{i-1}\} \cup (S' \ominus \min S') \xrightarrow{a} \{q_i\} \cup \{0\} \cup (S' \ominus (\min S' - 1)) \xrightarrow{b^{\min S' - 1}} \{q_i\} \cup S;$$

notice that $0 \in S' \ominus \min S'$. This proves this case by induction on i.

(*ii*) Let $i = 0$ and $0 \notin S$. Then $\{q_{m-1}\} \cup (S \ominus \min S) \xrightarrow{ab^{\min S - 1}} \{q_0\} \cup S$, where the former set is considered in case (i).

(*iii*) Let $i = 0$ and $0 \in S$. Take $S' = S \setminus \{0\}$. Then $\{q_{m-1}\} \cup (S' \ominus \min S') \cup \{n - 1\} \xrightarrow{a} \{q_0\} \cup \{0\} \cup (S' \ominus (\min S' - 1)) \xrightarrow{b^{\min S' - 1}} \{q_0\} \cup S$, where the first set is considered in case (i).

(*iv*) Let $1 \leq i \leq m - k - 1$. Then $\{q_i\} \cup S$ is reachable by Lemma 4.2 case 2.

To prove distinguishability, let $\{q_i\} \cup S$ and $\{q_j\} \cup T$ be two distinct subsets in $\mathcal{R}$. Notice that the state $n - 1$ is uniquely distinguishable since it is a unique final state of N. Next, the state $n - 1$ is uniquely reachable from states in Q_B since for every $j = 0, 1, \ldots, n - 2$ the transition $(j, a, j + 1)$ is a unique in-transition. By Proposition 2.1, if $S \neq T$, then the sets $\{q_i\} \cup S$ and $\{q_j\} \cup T$ are distinguishable.

Table 1. The state complexity of concatenation if the witness languages from [5, 10, 14] have the second half of their states final; in rows we have m, in columns n.

	Upper bound			Maslov [10]			YZS [14]			JJS [5]		
	2	4	6	2	4	6	2	4	6	2	4	6
2	6	24	96	5	4	18	6	14	27	6	22	84
4	12	48	192	10	5	35	12	28	54	12	42	156
6	18	72	288	15	6	52	18	42	81	18	63	225

Now let $S = T$, so $i < j$. If $S = \emptyset$, we read a^{m-k-j} and get $\{q_{m-k-j+i}\}$ and $\{q_{m-k}, 0\}$. If $S \neq \emptyset$, we first read $a^{m-1-j}b^n$ to get $\{q_{m-1-j+i}, 0\}$ and $\{q_{m-1}, 0\}$. Now we read a. There are two sub-cases:

1. If $m - j + i \geq m - k$, then we get $\{q_{m-j+i}, 0, 1\}$ and $\{q_0, 1\}$ which are distinguishable.
2. If $m - j + i < m - k$, then we get $\{q_{m-j+i}, 1\}$ and $\{q_0, 1\}$. Then we read $a^{(m-k-1)-(m-j+i)}b^n$ and get $\{q_{m-k-1}, 0\}$ and $\{q_{j-i-k-1}, 0\}$. Finally we read a and get $\{q_{m-k}, 0, 1\}$ and $\{q_{j-i-k}, 1\}$, which are distinguishable. ☐

Our next goal is to describe, for all m, n, k, ℓ with $n \geq 2$, two DFAs of m and n states, and k and ℓ final states, respectively, meeting the upper bound $m2^n - k2^{n-1}$ on the complexity of the concatenation of their languages. We try to modify the witness automata in all cases, by making the second half of their states final. The upper bound in such a case is $3m \cdot 2^{n-2}$.

Table 1 shows that none of the three witnesses presented in [5, 10, 14] meets this bound. Even making two states final in DFA B, results in a complexity of concatenation less that $m2^n - 2 \cdot 2^{n-1}$ in all three cases. Therefore we present new pairs of witness languages. To cover all possible values of m, n, k, ℓ, we modified the witness from Theorem 1 of [5] by defining transitions on a new symbol c. Notice that making some states final in DFA B does not play any role in the proof of reachability. We use the new symbol c only in the proof of distinguishability.

Theorem 4.6 **(Ternary Witness Languages with $|F_A| = k$ and $|F_B| = \ell$).** *Let $m \geq 1$ and $n \geq 2$. Let $k = 1$ if $m = 1$ and $1 \leq k \leq m - 1$ otherwise. Let $1 \leq \ell \leq n - 1$. There exist a ternary* DFA *A with m states and k final states and a ternary* DFA *B with n states and ℓ final states such that* $\mathrm{sc}(L(A)L(B)) = m2^n - k2^{n-1}$.

Proof. Define an m-state DFA $A = (\{q_0, q_1, \ldots, q_{m-1}\}, \{a, b, c\}, \cdot, q_0, F_A)$, where $F_A = \{q_i \mid m - k \leq i \leq m - 1\}$ and for each i in $\{0, 1, \ldots, m - 1\}$, we have $q_i \cdot a = q_{(i+1) \bmod m}, q_i \cdot b = q_i,$ and $q_i \cdot c = q_i$,

Define an n-state DFA $B = (\{0, 1, \ldots, n - 1\}, \{a, b, c\}, \cdot, 0, \{n - 1\})$, where $F_B = \{j \mid n - \ell \leq j \leq n - 1\}$ and for each state j of B,

$$
\begin{aligned}
&j \cdot a = (j + 1) \bmod n, \\
&0 \cdot b = 0, \quad j \cdot b = (j + 1) \bmod n \text{ if } j \geq 1, \\
&j \cdot c = 0 \text{ if } j \leq n - 2, \quad (n - 1) \cdot c = n - 1.
\end{aligned}
$$

The DFAs A and B, where $m = 6, k = 3, n = 5$, and $\ell = 2$ are shown in Figure 5. Notice that the transitions on a and b are the same as in Theorem 1 of [5].

Construct an NFA for $L(A)L(B)$ from DFAs A and B as described in Construction 3.1. Since the transitions on a and b are the same as in DFAs in the proof of Lemma 4.5, the proof of reachability is the same; notice that making some states final in DFA B does not play any role in the proof of reachability.

We only need to prove distinguishability. To this aim, let $\{q_i\} \cup S$ and $\{q_j\} \cup T$ be two distinct reachable subsets. Notice that the state $n - 1$ is uniquely distinguishable by the string c since we have $\ell \leq n - 1$, so state 0 is not final. Next, the state $n - 1$ is uniquely reachable from all states in Q_B through unique in-transitions on a.

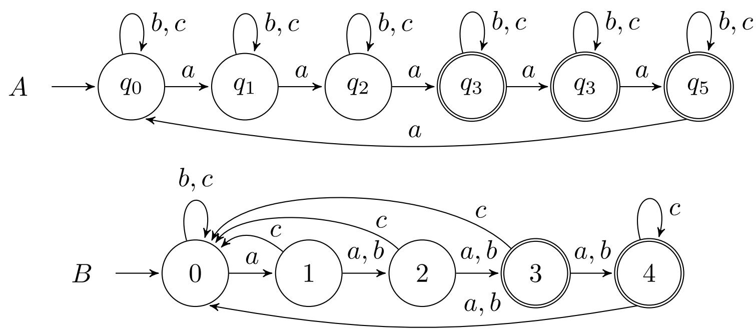

FIGURE 5. Ternary witnesses meeting the bound $m2^n - k2^{n-1}$; $m = 6$, $k = 3$, $n = 5$, and $\ell = 2$.

It follows that $\{q_i\} \cup S$ and $\{q_j\} \cup T$ are distinguishable if $S \neq T$. Now let $S = T$. In this case we continue exactly the same way as in the proof of Lemma 4.5. $\square$

We have proven that for every number of states in A and B, except for one state in B, and every number of final states in A and B, except for none or all, there exist ternary automata meeting the upper bound $m2^n - k2^{n-1}$ for concatenation of their languages. We might ask whether there are binary languages with more final states in B meeting this bound. We provide a positive answer in the next theorem. However, notice that we require $k \leq m - 2$ here, that is, the first DFA must have at least two non-final states.

Theorem 4.7 (**Binary Witness Automata with $|F_A| \leq m - 2$).** *Let $m \geq 3$, $n \geq 4$, $1 \leq k \leq m-2$, and $1 \leq \ell \leq n-1$. There exist a binary* DFA *A with m states and k final states and a binary* DFA *B with n states and ℓ final states such that* $\mathrm{sc}(L(A)L(B)) = m2^n - k2^{n-1}$.

Proof. Define an m-state DFA $A = (Q_A, \{a, b\}, \cdot, q_0, F_A)$, where we have $Q_A = \{q_0, q_1, \dots, q_{m-1}\}$, $F_A = \{q_i \mid m-k \leq i \leq m-1\}$, and for each i in $\{0, 1, \dots, m-1\}$,

$$q_i \cdot a = q_{(i+1) \bmod m}, \quad q_0 \cdot b = q_0 \quad \text{and} \quad q_i \cdot b = q_{i-1} \text{ if } 1 \leq i \leq m-1.$$

Define an n-state DFA $B = (Q_B, \{a, b\}, \cdot, 0, \{n-1\})$, where $Q_B = \{0, 1, \dots, n-1\}$, $F_B = \{n-\ell, n-\ell+1, \dots, n-1\}$ if $\ell \leq n-2$, and $F_B = Q_B \setminus \{1\}$ if $\ell = n-1$. For each state j of B, we have

$$\begin{aligned} &0 \cdot a = 0, && j \cdot a = j+1 \text{ if } 1 \leq j \leq n-2, && \text{and} && (n-1) \cdot a = 1, \\ &0 \cdot b = 1, && 1 \cdot b = 2, && \text{and} && j \cdot b = j \text{ if } 2 \leq j \leq n-1. \end{aligned}$$

The DFAs A and B, where $m = 6, k = 3, n = 5$, and $\ell = 2$ are shown in Figure 6; notice that the DFA B is the same as in [2]. Since $k \leq m-2$, the states q_0 are q_1 are never final. By definition of B, state 1 is not final either.

Construct an NFA N for $L(A)L(B)$ as described in Construction 3.1. We prove that the subset automaton of N has $m2^n - k2^{n-1}$ reachable and pairwise distinguishable states. The proof of reachability is by induction on $|\{q_i\} \cup S|$.

The base, with $|\{q_i\} \cup S| = 1$, holds true since $\{q_0\} \xrightarrow{a^i} \{q_i\}$ for $1 \leq i \leq m-k-1$. By Lemma 4.2, we need only to prove that every set $\{q_0\} \cup S$ of size $t+1$ and $0 \notin S$ is reachable. Let $S' = ((S \ominus (\min S - 1)) \setminus \{1\}) \cup \{0\}$. Then $|S'| = |S|$ and $0 \in S'$. By Lemma 4.2 case 3, the set $\{q_0\} \cup S'$ is reachable from a set of size t. Next we have

$$\{q_0\} \cup S' \xrightarrow{b(ab)^{\min S - 1}} \{q_0\} \cup S;$$

notice that $1 \in S \ominus (\min S - 1)$ and $q_0 \xrightarrow{ab} q_0$, because $q_1 \notin F_A$ since $k \leq m-2$. This proves reachability.

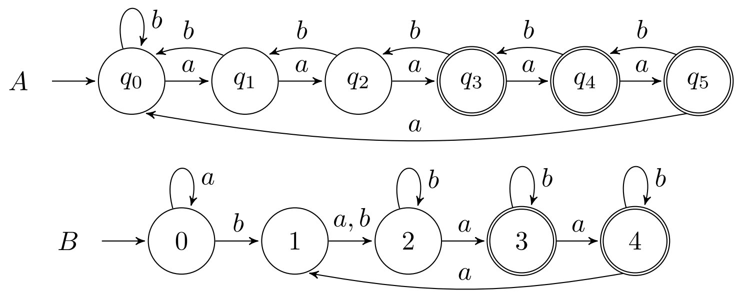

FIGURE 6. Binary witnesses meeting the bound $m2^n - k2^{n-1}$ in the case $k \le 2$ (modified from [2]); $m = 6$, $k = 3$, $n = 5$, $\ell = 2$.

To prove distinguishability, we use the string $w = \prod_{i=0}^{n-4} a^{n-3-i} b^m a^{i+2}$. We have

$$\{2\} \cdot w = \{2\}; \quad (Q_B \setminus \{2\}) \cdot w = \{1\}; \quad Q_A \cdot w \subseteq Q_A \cup \{0,1\}.$$

We now use these properties to prove distinguishability. Let $\{q_i\} \cup S$ and $\{q_j\} \cup T$ where $i, j \in \{0, 1, \ldots, m-1\}$ and $S, T \subseteq \{0, 1, \ldots, n-1\}$ be two reachable states of the subset automaton of N. We consider several cases:

1. If $2 \in S$ and $2 \notin T$, then

$$\{q_i\} \cup S \xrightarrow{wab^m} \{q_0\} \cup \{2, 3\} \xrightarrow{(ba)^{n-2}} \{q_1\} \cup \{1, 3\}, \text{ and}$$
$$\{q_j\} \cup T \xrightarrow{wab^m} \{q_0\} \cup \{2\} \xrightarrow{(ba)^{n-2}} \{q_1\} \cup \{1\}.$$

   If $3 \in F_B$, we have distinguished the sets. If not, we read $a^{n-\ell-3}$ and distinguish the sets since all states j with $j < n - \ell$ are non-final.
2. If $1 \le s \le n-1$ and $s \ne 2$, $s \in S$ and $s \notin T$, we read a^{n+1-s} to get the case 1.
3. If $0 \in S$ and $0 \notin T$, we read b to get the case 2.
4. If $S = T$ and $1 \le i < m - k \le j$, we read ba and get $\{q_i\} \cup S \cdot ba$ and $\{q_j\} \cup \{0\} \cup S \cdot ba$. Since $0 \notin S \cdot ba$, we get the case 3. If $0 = i < m - k \le j$, we read ba and get $\{q_1\} \cup S \cdot ba$ and $\{q_j\} \cup \{0\} \cup S \cdot ba$. We again get the case 3.
5. If $m - k \le i < j$, we read the string a^{m-j} and get $\{q_{m-j+i}\} \cup S \cdot a^{m-j}$ and $\{q_0\} \cup S \cdot a^{m-j}$, which is considered in the case 4.
6. If $i < j < m - k$, we read the string a^{m-k-j} and get $\{q_{m-k-j+i}\} \cup S \cdot a^{m-k-j}$ and $\{q_{m-k}\} \cup \{0\} \cup S \cdot a^{m-k-j}$. If $0 \notin S \cdot a^{m-k-j}$, we get the case 3. If $0 \in S \cdot a^{m-k-j}$, we get the case 4. □

5. BINARY CONCATENATION; $|F_A| = m - 1$ AND $2 \le |F_B| \le n - 1$

Now we turn our attention to the concatenation of binary languages represented by m-state DFA with $m-1$ final states and n-state DFA with more than one final state. In the general case, the upper bound is $(m+1)2^{n-1}$. The next theorem provides a lower bound that is smaller just by one. Our computations show that no pair of binary languages meets the bound $(m+1)2^{n-1}$ in the case of $m, n \le 4$.

Theorem 5.1 (Binary Concatenation with $|F_A| = m - 1$; Lower Bound). *Let $m, n \ge 3$ and $2 \le \ell \le n - 1$. There exist a binary* DFA *A with m states and $m - 1$ final states and a binary* DFA *B with n states and ℓ final states such that* $\mathrm{sc}(L(A)L(B)) \ge (m+1)2^{n-1} - 1$.

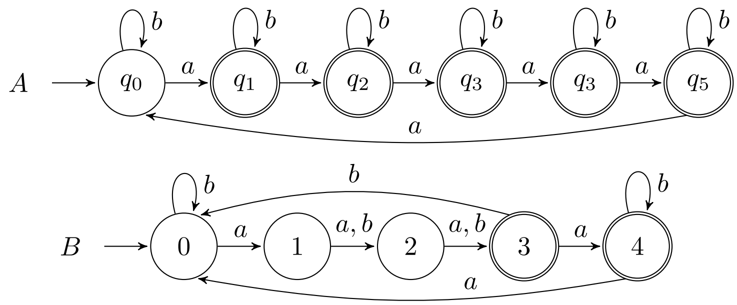

FIGURE 7. The binary DFAs meeting the bound $(m+1)2^{n-1}-1$; $m=6$, $k=5$, $n=5$, and $\ell=2$.

Proof. Define an m-state DFA $A = (\{q_0, q_1, \ldots, q_{m-1}\}, \{a, b\}, \cdot, q_0, F_A)$, where $F_A = \{q_1, q_2, \ldots, q_{m-1}\}$, and for each i in $\{0, 1, \ldots, m-1\}$,

$$q_i \cdot a = q_{(i+1) \bmod m} \qquad \text{and} \qquad q_i \cdot b = q_i.$$

Define an n-state DFA $B = (\{0, 1, \ldots, n-1\}, \{a, b\}, \cdot, 0, F_B)$, where we have $F_B = \{j \mid n-\ell \le j \le n-1\}$, and for each j in $\{0, 1, \ldots, n-1\}$,

$$\begin{aligned} &j \cdot a = (j+1) \bmod n, \\ &0 \cdot b = 0, \quad j \cdot b = (j+1) \bmod (n-1) \text{ if } 1 \le j \le n-2, \quad (n-1) \cdot b = (n-1). \end{aligned}$$

The DFAs A and B for $m=6, k=5, n=5$, and $\ell=2$ are shown in Figure 7.

Construct an NFA N for $L(A)L(B)$ from DFAs A and B as described in Construction 3.1. We prove that the subset automaton of N has $(m+1)2^{n-1}-1$ reachable and pairwise distinguishable states. To this aim, consider the following family of $(m+1)2^{n-1}-1$ subsets:

$$\begin{aligned} \mathcal{R} = &\{\{q_0\} \cup X \mid X \subseteq \{0, 1, \ldots, n-1\} \text{ and } X \ne \{n-1\}\} \ \cup \\ &\{\{q_i\} \cup X \mid 1 \le i \le n-1, X \subseteq \{0, 1, \ldots, n-1\}, \text{ and } 0 \in X\}. \end{aligned}$$

First we prove that each set in $\mathcal{R}$ is reachable. The proof of reachability is by induction on $|q_i \cup X|$.

The basis, with $|S| \le 2$, holds true since $\{q_0\}$ is the initial subset, $\{q_0\} \xrightarrow{a} \{q_1, 0\}$, and $\{q_{(i-1) \bmod m}, 0\} \xrightarrow{ab} \{q_i, 0\}$ for $i = 0, 1, \ldots, m-1$.

Let $2 \le t \le n$ and assume that each subset in $\mathcal{R}$ of size t is reachable. By Lemma 4.2, we only need to prove that every set $\{q_0\} \cup S$ of size $t+1$ and $0 \notin S$ is reachable. To show that $\{q_0\} \cup S$ is reachable, recall that $\{q_{m-1}\} \cup (S \ominus \min S)$ is reachable by Lemma 4.2 case 3 since $0 \in S \ominus \min S$. Next we have

$$\{q_{m-1}\} \cup (S \ominus \min S) \xrightarrow{a} \{q_0\} \cup (S \ominus (\min S - 1)) \xrightarrow{b^{\min S - 1}} \{q_0\} \cup S.$$

This proves reachability.

To prove distinguishability, let $\{q_i\} \cup S$ and $\{q_j\} \cup T$ be two distinct sets in $\mathcal{R}$. In a similar way as in the proof of Lemma 4.5, we can show that for every state t, $0 \le t \le n-1$, the string $a^{n-1-t}b(ab^{n-2})^{n-3}$ is accepted by NFA N only from t. It follows that the sets $\{q_i\} \cup S$ and $\{q_j\} \cup T$ are distinguishable if $S \ne T$.

If $S = T$, then we may assume that $S \ne \emptyset$ because $\{q_0\}$ is the only reachable set of size one. Thus we need to distinguish the sets $\{q_i\} \cup S$ and $\{q_j\} \cup S$, where $S \ne \emptyset$ and $i \ne j$. We first read $(ab^{n-2})^{n-2}$ to get $\{q_x, 0\}$

and $\{q_y, 0\}$ with $x \neq y$; notice that both symbols a and b perform a permutation on the states of DFA A. We may assume that $x < y$. Consider two cases:

1. If $y = m-1$, we use a to get $\{q_{x+1}, 0, 1\}$ and $\{q_0, 1\}$, which differ in state 0.
2. If $y < m-1$, then we use $(ab)^{m-1-y}$ to get $\{q_{x+m-1-y}, 0\}$ and $\{q_{m-1}, 0\}$ which are considered in case 1.

This completes our proof. □

6. Concatenation on alternating finite automata

In this section, we consider the concatenation operation on alternating finite automata (AFAs) [3]. Our aim is to describe languages K and L accepted by an m-state and n-state AFA, respectively, such that the minimal AFA for the language KL requires $2^m + n + 1$ states. This solves an open problem stated by Fellah, Jürgensen, and Yu in [3], where the upper bound is proven to be the same. First, let us give some basic definitions and notations. For details, we refer the reader to [1, 3, 7–9, 12].

An *alternating finite automaton* (AFA) is a quintuple $A = (Q, \Sigma, \delta, s, F)$, where Q is a finite non-empty set of states, $Q = \{q_1, \ldots, q_n\}$, Σ is an input alphabet, δ is the transition function that maps $Q \times \Sigma$ into the set $\mathcal{B}_n$ of boolean functions over the n variables $q_1, \ldots, q_n$, $s \in Q$ is the initial state, and $F \subseteq Q$ is the set of final states. For an example, consider AFA $A_1 = (\{q_1, q_2\}, \{a, b\}, \delta, q_1, \{q_2\})$, where transition function δ is given in Table 2.

Table 2. The transition function of the alternating finite automaton A_1.

δ	a	b
q_1	$q_1 \vee q_2$	q_1
q_2	q_2	$\overline{q_1} \wedge q_2$

The transition function δ is extended to the domain $\mathcal{B}_n \times \Sigma^*$ as follows: For all g in $\mathcal{B}_n$, a in Σ, and w in Σ^*, $\delta(g, \varepsilon) = g$; if $g = g(q_1, \ldots, q_n)$, then $\delta(g, a) = g(\delta(q_1, a), \ldots, \delta(q_n, a))$; $\delta(g, wa) = \delta(\delta(g, w), a)$.

Next, let $f = (f_1, \ldots, f_n)$ be the Boolean vector with $f_i = 1$ iff $q_i \in F$. The language accepted by the AFA A is the set $L(A) = \{w \in \Sigma^* \mid \delta(s, w)(f) = 1\}$.

In our example we have $\delta(q_1, ab) = \delta(\delta(q_1, a), b) = \delta(q_1 \vee q_2, b) = q_1 \vee (\overline{q_1} \wedge q_2) = q_1 \vee q_2$. To determine whether $ab \in L(A_1)$, we evaluate $\delta(q_1, ab)$ at the vector $f = (0, 1)$. We obtain 1, hence $ab \in L(A_1)$.

Recall that the state complexity of a regular language L, $\mathrm{sc}(L)$, is the smallest number of states in any DFA accepting L. Similarly, the alternating state complexity of L, $\mathrm{asc}(L)$, is the smallest number of states in any AFA for L. It follows from Theorem 4.1, Corollary 4.2 of [3] and Lemma 1, Lemma 2 of [7] that a language L is accepted by an n-state AFA if and only if L^R is accepted by a DFA with 2^n states and 2^{n-1} final states. As this is a crucial observation for this section, we restate these results and provide proof ideas.

Lemma 6.1 ([3, 7])**.** *Let L be a language accepted by an n-state* AFA. *Then the reversal L^R is accepted by a* DFA *of 2^n states, of which 2^{n-1} are final.*

Proof Idea. Let $A = (\{q_1, q_2, \ldots, q_n\}, \Sigma, \delta, q_1, F)$ be an n-state AFA for L. Construct a 2^n-state NFA $A' = (\{0,1\}^n, \Sigma, \delta', S, \{f\})$, where

- for every $u = (u_1 \ldots, u_n) \in \{0,1\}^n$ and every a in Σ, $\delta'(u, a) = \{u' \in \{0,1\}^n \mid \delta(q_i, a)(u') = u_i \text{ for } i = 1, \ldots, n\}$;
- $S = \{(b_1, \ldots, b_n) \in \{0,1\}^n \mid b_1 = 1\}$;
- $f = (f_1, \ldots, f_n) \in \{0,1\}^n$ with $f_i = 1$ iff $q_i \in F$.

Then $L(A) = L(A')$, NFA A' has 2^{n-1} initial states and $(A')^R$ is deterministic. It follows that L^R is accepted by a DFA with 2^n states, of which 2^{n-1} are final. □

Corollary 6.2. *For every regular language L, $\mathrm{asc}(L) \geq \lceil \log(\mathrm{sc}(L^R)) \rceil$.*

Lemma 6.3 (*cf.* [7], Lem. 2)**.** *Let L^R be accepted by a* DFA *A of 2^n states, of which 2^{n-1} are final. Then L is accepted by an n-state* AFA*.*

Proof Idea. Consider 2^n-state NFA A^R for L which has 2^{n-1} initial states and exactly one final state. Let the state set Q of A^R be $\{0,1,\dots,2^n-1\}$ with initial states $\{2^{n-1},\dots,2^n-1\}$ and final state k. Let δ be the transition function of A^R. Moreover, for every $a\in\Sigma$ and for every $i\in Q$, there is exactly one state j such that j goes to i on a in A^R. For a state $i\in Q$, let $\operatorname{bin}(i)=(b_1,\dots,b_n)$ be the binary n-tuple such that $b_1b_2\cdots b_n$ is the binary notation of i on n digits with leading zeros if necessary.

Define an n-state AFA $A'=(Q',\Sigma,\delta',q_1,F')$, where $Q'=\{q_1,\dots,q_n\}$, $F'=\{q_\ell\mid \operatorname{bin}(k)_\ell=1\}$, and for each i in Q and a in Σ, $(\delta'(q_1,a),\dots,\delta'(q_n,a))(\operatorname{bin}(i))=\operatorname{bin}(j)$ where $i\in\delta(j,a)$. Then $L(A')=L(A^R)$. □

By Corollary 6.2, we have $\operatorname{asc}(L)\ge\lceil\log(\operatorname{sc}(L^R))\rceil$. The upper bound for concatenation on AFAs is 2^m+n+1, as proven by Fellah *et al.* Theorem 9.3 of [3]. They conjectured that this bound is tight.

In [7], the lower bound 2^m+n was proven, however, the witnesses from Theorem 1 of [5] with half of states final in both automata were used. As we mentioned above, *cf.* Table 1, these witness languages do not meet the upper bound for concatenation on DFAs. Hence the proof in Theorem 5 of [7] is not correct, so the problem is still open. Our next aim is to prove the tightness of the upper bound 2^m+n+1 for concatenation on AFAs. We might use the ternary witness from Theorem 4.6, but, as we show below, for $\operatorname{asc}(K)\ge 2,\ \operatorname{asc}(L)\ge 2$, it is sufficient to use the binary witness languages described in the proof of Theorem 4.7 to get languages that meet the upper bound 2^m+n+1 for concatenation on AFAs. The following lemma not only proves the claim in Theorem 5 of [7], but also solves the open problem mentioned above.

Lemma 6.4 (**Concatenation on AFAs: Lower Bound**)**.** *Let $m,n\ge 2$. There exist binary languages K and L accepted by an m-state and n-state* AFA*, respectively, such that* $\operatorname{asc}(KL)=2^m+n+1$.

Proof. Let L^R be the binary regular language accepted by the minimal DFA A from the proof of Theorem 4.7, with 2^n states and 2^{n-1} final states. Let K^R be the binary regular language accepted by the minimal DFA B from the proof of Theorem 4.7, with 2^m states and 2^{m-1} final states. Then, by Lemma 6.3, the languages K and L are accepted by an m-state and n-state AFA, respectively. Using Theorem 4.7, we get

$$\operatorname{sc}((KL)^R)=\operatorname{sc}(L^RK^R)=2^n\cdot 2^{2^m}-2^{n-1}\cdot 2^{2^m-1}=2^{n-1}\cdot 2^{2^m}(1+1/2).$$

By Corollary 6.2, we have $\operatorname{asc}(KL)\ge\lceil\log(2^{n-1}\cdot 2^{2^m}(1+1/2))\rceil=2^m+n$.

Our next aim is to show that $\operatorname{asc}(KL)\ge 2^m+n+1$. Suppose for a contradiction that KL is accepted by an AFA of 2^m+n states. Then $(KL)^R$ is accepted by a 2^{2^m+n}-state DFA with 2^{2^m+n-1} final states. It follows that the minimal DFA for $(KL)^R$ has at most 2^{2^m+n-1} final states. However, the minimal DFA for $(KL)^R$ has $2^n2^{2^m}-2^{n-1}2^{2^m-1}$ states, of which $2^{n-1}2^{2^{m-1}}+2^{n-1}2^{2^{m-1}-1}$ are non-final; notice that $\{q_i\}\cup S$ is non-final iff $i\le 2^{n-1}-1$ and $S\subseteq\{0,1,\dots,2^{m-1}-1\}$ or $2^{n-1}\le i\le 2^n-1$ and $S\subseteq\{0,1,\dots,2^{m-1}-1\}$ with $0\in S$. Thus the number of final states in the minimal DFA for $(KL)^R$ is

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

and since $m\ge 2$, we get

$$\begin{aligned}
&2^{n-1}(2^{2^m}+2^{2^m-1})-2^{n-1}(2^{2^{m-1}}+2^{2^{m-1}-1})=\\
&2^{2^m}2^{n-1}\left(1+\frac{1}{2}-\frac{1}{2^{2^{m-1}}}-\frac{1}{2^{2^{m-1}+1}}\right)>\\
&2^{2^m+n-1}\left(1+\frac{1}{2}-\frac{1}{4}-\frac{1}{4}\right)=2^{2^m+n-1}.
\end{aligned}$$

Hence, the minimal DFA for $(KL)^R$ has more than 2^{2^m+n-1} final states, a contradiction. It follows that $\mathrm{asc}(KL) \geq 2^m + n + 1$, which proves the theorem. $\square$

We continue with examining the complexity of concatenation of unary AFA languages. Since the reverse of every unary language is the same language, we get that a unary language is accepted by an n-state AFA if and only if it is accepted by a 2^n-state DFA with 2^{n-1} final states. So in order to have languages K and L accepted by an m-state and n-state AFA, respectively, we only need to find unary languages K and L represented by a 2^m-state and 2^n-state DFA with half states final, respectively. The next lemma shows that the upper bound for binary AFAs cannot be met in the unary case. Then we provide a lower bound.

Lemma 6.5 (**Concatenation of Unary AFAs; Upper Bound**). *Let $m, n \geq 1$. Let K and L be unary languages accepted by an m-state and n-state* AFA*, respectively. Then* $\mathrm{asc}(KL) \leq m + n + 1$.

Proof. By Lemma 6.1, the unary language K is accepted by a 2^m-state DFA with 2^{m-1} final states and L is accepted by a 2^n-state DFA with 2^{n-1} final states. It follows that KL is accepted by a DFA with $2^m \cdot 2^n$ states, as is proven in Theorem 5.5 of [14]. By adding some final or non-final unreachable states, we can construct an equivalent DFA with 2^{m+n+1} states and 2^{m+n} final states. By Lemma 6.3, the language KL is accepted by an $(m + n + 1)$-state AFA. $\square$

Lemma 6.6 (**Concatenation of Unary AFAs; Lower Bound**). *Let $m, n \geq 1$. There exist unary languages accepted by an m-state and n-state* AFA*, respectively, such that* $\mathrm{asc}(KL) \geq m + n - 1$.

Proof. We have $\gcd(2^{m-1}, 2^{n-1} + 1) = 1$.

Consider DFA $A = (\{0, 1, \ldots, 2^m - 1\}, \{a\}, \cdot, 0, \{i \mid 2^{m-1} - 1 \leq i \leq 2^m - 2\})$, where $i \cdot a = i + 1$ if $0 \leq i < 2^{m-1} - 1$, and $i \cdot a = 0$ otherwise. Thus A has 2^m states and 2^{m-1} final states, so $L(A)$ is accepted by an m-state AFA; notice that only 2^{m-1} states are reachable in A.

Next, consider DFA $B = (\{0, 1, \ldots, 2^n - 1\}, \{a\}, \cdot, 0, \{j \mid 2^{n-1} \leq j \leq 2^n - 1\})$, where $j \cdot a = j + 1$ if $0 \leq j < 2^{n-1}$, and $j \cdot a = 0$ otherwise. Similarly as above, $L(B)$ is accepted by an n-state AFA, and this time only $2^{n-1} + 1$ states are reachable in B. As shown in Theorem 5.4 of [14] $\mathrm{sc}(L(A)L(B)) = 2^{m-1} \cdot (2^{n-1} + 1) = 2^{m+n-2} + 2^{m-1}$. By Lemma 6.3, we have $\mathrm{asc}(L(A)L(B)) \geq m + n - 1$. $\square$

As a corollary of Lemmas 6.4, 6.5, and 6.6, we state the following theorem.

Theorem 6.7 (**Concatenation on AFAs**). *Let $m, n \geq 2$. Let K and L be languages over an alphabet Σ accepted by an m-state and n-state* AFA*, respectively. Then* $\mathrm{asc}(KL) \leq 2^m + n + 1$*, and this bound is tight if $|\Sigma| \geq 2$. If $|\Sigma| = 1$, then* $\mathrm{asc}(KL) \leq m + n + 1$*. There exist unary m-state and n-state* AFA *languages meeting the bound $m + n - 1$.*

7. Conclusions

We studied the state complexity of the concatenation of languages represented by deterministic and alternating finite automata. First, we described ternary languages meeting the upper bound $m2^n - k2^{n-1}$ for all possible values of m, n, k, ℓ, where m and k is the number of states and the number of final states in the minimal DFA for the first language, and n and ℓ is the number of states and the number of final states in the minimal DFA for the second language. Then, in the case of $k \leq m - 2$, that is, if the first automaton has at least two non-final states, we described appropriate binary languages. Finally, we considered the case of $k = m - 1$ and $\ell \geq 2$ over a binary alphabet, and obtained a lower bound that is smaller than the corresponding upper bound just by one. We strongly conjecture that our lower bound is tight in this case.

We used our binary witnesses for the concatenation on DFAs to define binary languages K and L accepted by an m-state and n-state AFA, respectively, such that the minimal AFA for KL requires $2^m + n + 1$ states. This proves that the upper bound $2^m + n + 1$ from [3] is tight, and solves the open problem stated in Theorem 9.3 of [3]. We also proved that this upper bound cannot be met by unary AFA languages, where we get upper bound $m + n + 1$ and lower bound $m + n - 1$.

References

[1] J. Brzozowski and E. Leiss, On equations for regular languages, finite automata, and sequential networks. *Theoret. Comput. Sci.* **10** (1980) 19–35.

[2] E. Dorčák, *Concatenation of regular languages and state complexity*, ŠVK thesis. P. J. Šafárik University, Košice (2015)

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

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

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

[6] G. Jirásková, State complexity of some operations on binary regular languages. *Theoret. Comput. Sci.* **330** (2005) 287–298.

[7] G. Jirásková, Descriptional complexity of operations on alternating and boolean automata. In *CSR 2012*, edited by E. Hirsch *et al.* Vol. 7353 of *Lect. Notes Comput Sci.* Springer (2012) 196–204.

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

[9] E. Leiss, On generalized language equations. *Theoret. Comput. Sci.* **14** (1981) 63–77.

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

[11] M. Rabin and D. Scott, Finite automata and their decision problems. *IBM Res. Develop.* **3** (1959) 114–129.

[12] M. Sipser, Introduction to the theory of computation. Cengage Learning, Boston (2012)

[13] S. Yu, Regular languages, Chapter 2. In vol. I of *Handbook of Formal Languages*, edited by G. Rozenberg and A. Salomaa. Springer, Heidelberg (1997) 41–110.

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