

Kleene Closure and State Complexity

Galina Jirásková^{1,*} and Matúš Palmovský^{2,†}

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

² Institute of Computer Science, Faculty of Science, P.J. Šafárik University
Jesenná 5, 040 01 Košice, Slovakia
matp93@gmail.com

Abstract: We prove that the automaton presented by Maslov [Soviet Math. Doklady 11, 1373–1375 (1970)] meets the upper bound $3/4 \cdot 2^n$ on the state complexity of Kleene closure. This fixes a small error in this paper that claimed the upper bound $3/4 \cdot 2^n - 1$. Our main result shows that the upper bounds $2^{n-1} + 2^{n-1-k}$ on the state complexity of Kleene closure of a language accepted by an n -state DFA with k final states are tight for every k in the binary case. We also present some results of our calculations. We consider not only the worst case, but we study all possible values that can be obtained as the state complexity of Kleene closure of a regular language accepted by a minimal n -state DFA. Using the lists of pairwise non-isomorphic binary automata of 2,3,4, and 5 states, we compute the frequencies of the resulting complexities for Kleene closure, and show that every value in the range from 1 to $3/4 \cdot 2^n$ occurs at least ones. In the case of $n = 6, 7, 8$, we change the strategy, and consider binary automata, in which the first symbol is a circular shift of the states, and the second symbol is generated randomly. We show that all values from 1 to $3/4 \cdot 2^n$ are attainable, that is, for every m with $1 \leq m \leq 3/4 \cdot 2^n$, there exists an n -state binary DFA A such that the state complexity of $L(A)^*$ is exactly m .

1 Introduction

Kleene closure is a basic operation on formal languages which is defined as

$$L^* = \{w \mid w = v_1 v_2 \cdots v_k, k \geq 0, v_i \in L \text{ for all } i\}.$$

It is known that if L is recognized by an n -state deterministic finite automaton (DFA), then the language L^* is recognized by a DFA of at most $3/4 \cdot 2^n$ states [8, 13]. The first worst-case example meeting this upper bound was presented already by Maslov in 1970 [8]. However, he did a small error and did not give any proof in his paper.

Later, Yu, Zhuang, and Salomaa [13] proved that the size of the minimal DFA for Kleene closure depends on the number of final states of a given DFA, and that the upper bound is $2^{n-1} + 2^{n-1-k}$, where k is the number of final and non-initial states.

In this paper we give a proof of Maslov's result and we fix an error in his paper [8] by proving that Maslov's automaton meets the upper bound $3/4 \cdot 2^n$. Then we show that the upper bounds $2^{n-1} + 2^{n-1-k}$ are tight for every n and k with $1 \leq k \leq n-1$. This is the main result of our paper. The witness automata are defined over a binary alphabet. The size of the alphabet is optimal since the state complexity of Kleene closure over a unary alphabet is only $(n-1)^2 + 1$.

In the second part of our paper we consider not only the worst case, but rather study all possible values that can be obtained as the number of states of the minimal DFA recognizing the Kleene closure of a regular language represented by a minimal n -state DFA. The problem is known as "the magic number problem" in the literature, and so called "magic numbers" are exactly the "holes" in the hierarchy that cannot be obtained in such a way.

The problem was first stated for NFA to DFA conversion by Iwama, Kambayashi, and Takaki in [5]. It is known that in the ternary case, no magic numbers exist, that is, each value from n to 2^n may be obtained as the size of the minimal DFA equivalent to a given minimal n -state NFA [7]. On the other hand, it is known that in the unary case, magic numbers exist [3], but we do not know which values are magic. The binary case is still open.

For Kleene closure, the possible resulting values are in the range from 1 to $3/4 \cdot 2^n$, for an alphabet of at least two symbols, and in the range from 1 to $(n-1)^2 + 1$ for a unary alphabet, and it is known that for a growing alphabet of size 2^n , no magic numbers exist [6].

Here we study the binary case. Using the lists of pairwise non-isomorphic automata of 2,3,4, and 5 states, we compute the frequencies of the resulting complexities for Kleene closure, and show that every value in the range from 1 to $3/4 \cdot 2^n$ occurs at least ones. We display our results in graphs, and compute the average complexity.

In the case of $n = 6, 7, 8$, we change the strategy, and consider binary automata, in which the first symbol is a circular shift of the states, and the second symbol is generated randomly. We consider an arbitrary number of final states. We show that all values from 1 to $3/4 \cdot 2^n$ are attainable, and we show that for every m with $1 \leq m \leq 3/4 \cdot 2^n$, there exists an n -state binary DFA A such that the state complexity of $L(A)^*$ is exactly m .

*Research supported by grants VEGA 2/0183/11, APVV-0035-10.

†Research supported by grants VEGA 2/0183/11, APVV-0035-10.

Thus our calculations show, that in the binary case, up to $n = 8$, no magic numbers exists. Moreover, for every n , the numbers $1, n$, and $2^{n-1} + 2^{n-1-k}$ with $1 \leq k \leq n-1$ are attainable by the complexity of Kleene closure. The situation is completely different in the case of a unary alphabet, where two holes of length n exist for every n [2].

2 Preliminaries

Let Σ be a finite alphabet and Σ^* the set of all strings over Σ . The empty string is denoted by ε . The length of a string w is $|w|$. A *language* is any subset of Σ^* . We denote the size of a set A by $|A|$, and its power-set by 2^A .

A *deterministic finite state automaton* is a quintuple $A = (Q, \Sigma, \delta, s, F)$, where Q is a finite set of states; Σ is a finite set of input symbols; δ is the transition function that takes as arguments a state and an input symbol and returns a state; s is an element of Q called the initial state; F is the set of final states (or accepting states), $F \subseteq Q$. The language accepted or recognized by the DFA A is defined as the set $L(A) = \{w \in \Sigma^* \mid \delta(s, w) \in F\}$.

A *nondeterministic finite automaton* is a quintuple $A = (Q, \Sigma, \delta, s, F)$, where Q, Σ, s , and F are the same as for a DFA, and δ is the transition function that takes a state in Q and an input symbol in Σ as arguments and returns a subset of Q . The language accepted or recognized by the NFA A is defined as the set $L(A) = \{w \in \Sigma^* \mid \delta(s, w) \cap F \neq \emptyset\}$.

Two automata are *equivalent* if they recognize the same language.

A DFA A is *minimal* if every equivalent DFA has at least as many states as A . It is known that every regular language has a unique, up to isomorphism, minimal DFA, and that a DFA $A = (Q, \Sigma, \delta, s, F)$ is minimal if and only if

- (i) all its states are reachable, that is, for every state q in Q , there exists a string w in Σ^* such that $\delta(s, w) = q$; and
- (ii) no two distinct states are equivalent; two states p and q are equivalent if for every string w in Σ^* , $\delta(p, w) \in F$ if and only if $\delta(q, w) \in F$.

The *state complexity* of a regular language L , denoted by $sc(L)$, is number of states in the minimal DFA accepting the language L .

Every NFA can be converted to an equivalent DFA by the subset construction [10] as follows. Let $A = (Q, \Sigma, \delta, s, F)$ be an NFA. Construct the DFA $A' = (2^Q, \Sigma, \delta', \{s\}, F')$, where $F' = \{R \subseteq Q \mid R \cap F \neq \emptyset\}$, and $\delta'(R, a) = \bigcup_{r \in R} \delta(r, a)$ for each R in 2^Q and each a in Σ . The DFA A' is called the *subset automaton* of the NFA A . The subset automaton need not be minimal since some of its states may be unreachable or equivalent.

To prove that states of a DFA are not equivalent, we will use the following observation.

Proposition 1. *Let N be an NFA. Let for every state q of the NFA N , there exists a string w_q such that w_q is accepted*

by N only from the state q . Then the subset automaton corresponding to the NFA N does not have equivalent states.

Proof. Let S, T be subsets of states of N , where $S \neq T$. Without loss of generality, there exists a state q such that $q \in S$ and $q \notin T$. Then the string w_q is accepted from S but w_q is not from T . Hence S and T are not equivalent. \square

For languages K and L the *concatenation* $K \cdot L$ is defined as $K \cdot L = \{uv \mid u \in K, v \in L\}$. The language L^k with $k \geq 0$ is defined inductively by $L^0 = \{\varepsilon\}$, $L^1 = L$, $L^{i+1} = L^i \cdot L$.

Definition 1. *The Kleene closure of a language L is the language L^* defined as*

$$L^* = \bigcup_{i \geq 0} L^i.$$

3 NFA for Kleene Closure

In this section we describe the construction of a nondeterministic automaton recognizing the Kleene closure of a given language represented by DFA.

Let $A = (Q, \Sigma, \delta, s, F)$ be the minimal DFA accepting a language L . Construct an NFA A^* for the language L^* from DFA A as follows:

- For each state q in Q and each symbol a in Σ such that $\delta(q, a) \in F$, add the transition on a from q to s .
- If $s \notin F$, then add a new start state q_0 to Q and make this state accepting. For each symbol a in Σ add the transition on a

from q_0 to $\delta(s, a)$ if $\delta(s, a) \notin F$, and

from q_0 to $\delta(s, a)$ and from q_0 to s if $\delta(s, a) \in F$.

We illustrate this construction in the following example.

Example 1. Consider the DFA A shown in Fig. 1. In Fig. 2, we add the following transitions: the transition from s to the state s on the letter b because A has the transition from s to the *final state* 2; the transition from 1 to s on a because A has the transition from 1 to the *final state* 2; the transition from 2 to s on b because A has the transition from 2 to the *final state* 2.

Since s is non-final, we add the new initial state q_0 , make this state final, and we add transitions from q_0 as follows. Since there is a transition from the old initial state s to state 1 on the letter a in A , and 1 is non-final, we add the new transition from the state q_0 to 1 on a , and since there is transition from s to the state 2, which is final in A , we add the new transition from state q_0 to 2 and transition from q_0 to s on the letter b .

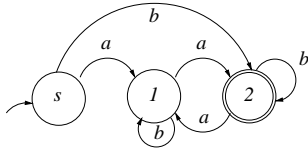


Figure 1: A 3-state DFA A.

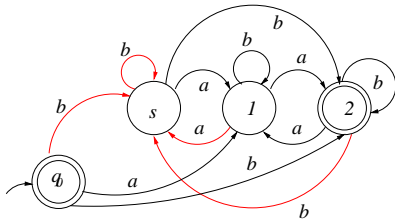


Figure 2: The NFA for the language $L(A)^*$, where A is shown in Fig. 1.

4 State Complexity of Kleene Closure

The state complexity of Kleene closure is defined as the minimal number of states that are sufficient and necessary in the worst case for a DFA to accept the Kleene closure of a regular language represented by an n -state DFA. The following upper bound is from [13]. For, the sake of completeness we give a simplified proof here.

Lemma 1 (Upper Bound [13]). *Let $A = (Q, \Sigma, \delta, s, F)$ be an n -state DFA such that $|F \setminus \{s\}| = k$. Then the minimal DFA for the language $L(A)^*$ has at most $2^{n-1} + 2^{n-1-k}$ states.*

Proof. Construct the NFA N for the language $L(A)^*$ as described above. Consider the subset automaton of the NFA N . Let S be a reachable subset of automaton. Notice that if a final state of N is in S , than the state s is also in S . It follows that only the following subsets can be reachable in the subset automaton:

1. $\{q_0\}$;
2. $S \subseteq Q$ with $s \in S$,
3. $S \subseteq Q \setminus (F \cup \{s\})$ and $S \neq \emptyset$.

This gives at most $1 + 2^{n-1} + 2^{n-1-k} - 1$ reachable sets, which gives the desired upper bound. \square

Notice that the number $2^{n-1} + 2^{n-1-k}$ is maximal if $k = 1$. For $k = 1$, we have $2^{n-1} + 2^{n-1-k} = 2^{n-1} + 2^{n-2} = 3/4 \cdot 2^n$. Thus we get the following upper bound.

Corollary 1. *Let L a language accepted by an n -state DFA. Then the minimal DFA for the language L^* has at most $3/4 \cdot 2^n$ states.*

Yu, Zhuang, and Salomaa [13] presented the witness language accepted by DFA shown in Fig. 3, and they proved that it meets the upper bound $3/4 \cdot 2^n$.

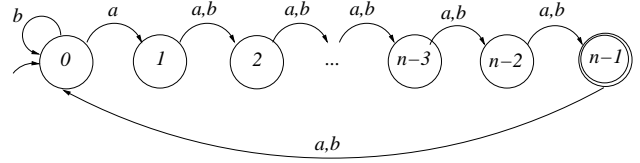


Figure 3: YYS'94 automaton meeting the bound $3/4 \cdot 2^n$ for Kleene closure.

The first witness language was presented already by Maslov [8] in 1970. Maslov claimed, without any proof, the upper bound for Kleene closure is $3/4 \cdot 2^n - 1$ and that the DFA from Fig. 4 meets this bound. However, Maslov's automaton, in fact, meets the bound $3/4 \cdot 2^n$. Here we fix this error and provide a proof.

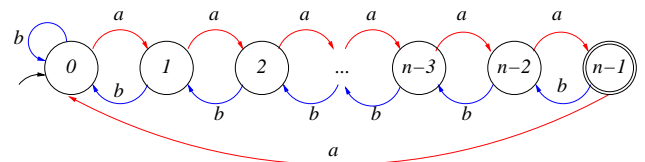


Figure 4: The Maslov's DFA A.

First, construct an NFA N for the language $L(A)^*$ by adding the transition on a from $n - 2$ to 0 , by adding a new initial and final state q_0 , and by adding the transition on a from q_0 to 1 and the transition on b from q_0 to 0 . The NFA N is shown in Fig. 5.

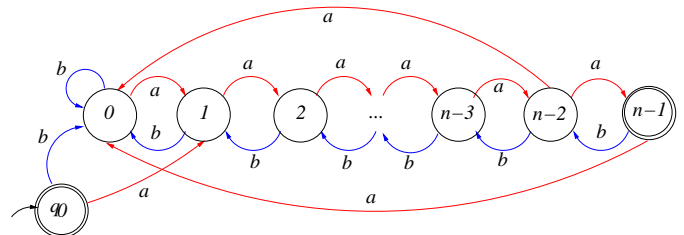


Figure 5: An NFA N for $L(A)^*$.

The next two lemmata show that the subset automaton of the NFA N has $3/4 \cdot 2^n$ reachable and pairwise distinguishable states.

Lemma 2. *The subset automaton of the NFA N shown in Fig. 5 has $3/4 \cdot 2^n$ reachable states.*

Proof. By induction on $|S|$, we prove that every subset S of $\{0, 1, \dots, n - 1\}$, such that $n - 1 \in S$ implies $0 \in S$, is reachable. The base is $|S| = 1$. The set $\{q_0\}$ is reachable since it is the initial state of the subset automaton. The set $\{i\}$, where $0 \leq i \leq n - 2$, is reached from $\{q_0\}$ by the string ba^i since we have $\{q_0\} \xrightarrow{b} \{0\} \xrightarrow{a^i} \{i\}$.

Assume that every set S with $|S| = k$, where $1 \leq k \leq n - 1$, is reachable. Let $S = \{i_1, i_2, i_3, \dots, i_k, i_{k+1}\}$, where $0 \leq i_1 < i_2 < \dots < i_k < i_{k+1} \leq n - 1$, be set of size $k + 1$. Consider three cases:

(i) $i_1 = 0$ and $i_{k+1} = n - 1$.

Take $S' = \{i_2 - 1, i_3 - 1, \dots, i_k - 1, n - 2\}$. Then $|S'| = k$ and therefore S' is reachable by the induction hypothesis. Since $S' \xrightarrow{a} \{0, i_2, i_3, \dots, i_k, n - 1\} = S$, the set S is reachable.

(ii) $i_1 = 0$ and $i_{k+1} < n - 1$.

Take $S' = \{0, i_2 + x, i_3 + x, \dots, i_k + x, n - 1\}$, where $x = n - 1 - i_{k+1}$. Then $|S'| = k + 1$ and S' contains states 0 and $n - 1$. Therefore, the set S' is reachable as shown in case (i). Since $S' \xrightarrow{b^x} \{0, i_2, i_3, \dots, i_k, i_{k+1}\} = S$, the set S is reachable.

(iii) $i_1 > 0$ and $i_{k+1} < n - 1$.

Take $S' = \{0, i_2 - i_1, i_3 - i_1, \dots, i_k - i_1, i_{k+1} - i_1\}$. Then $|S'| = k + 1$ and S' contains state 0. Therefore the set S' is reachable as shown in cases (i) and (ii). Since we have $|S'| \xrightarrow{a^{i_1}} \{i_1, i_2, i_3, \dots, i_k, i_{k+1}\} = S$, the set S is reachable.

We have shown that the subset automaton has $3/4 \cdot 2^n$ reachable states. \square

Lemma 3. *All the reachable states of the subset automaton corresponding to the NFA N shown in Fig. 5 are pairwise distinguishable.*

Proof. Notice that the string a^{n-1-i} is accepted by the NFA N only from the state i . By Proposition 1, no two distinct subsets of $\{0, 1, \dots, n - 1\}$ are equivalent.

Next, we need to show that $\{q_0\}$ and some final subset S are distinguishable. If S is a final subset, then $n - 1 \in S$. Consider the string a^n . The set $\{q_0\}$ goes on a^n to $\{0, 1\}$, which is non-final set since $n \geq 3$. However, the state $n - 1$ goes on a^n to $n - 1$ in the NFA. It follows that a^n is accepted by the subset automaton from S . This the string a^n distinguishes $\{q_0\}$ and S .

Hence all reachable states of the subset automaton of N are pairwise distinguishable. \square

As a corollary of the two lemmata above, we get the following result.

Theorem 1. *Let L be the language accepted by the Maslov's automaton shown in Fig. 4. Then the minimal DFA for the language L^* has $3/4 \cdot 2^n$ states.*

Proof. Let N be the NFA for the language L^* shown in Fig. 5. By Lemma 2, the subset automaton of N has $3/4 \cdot 2^n$ reachable states. By Lemma 3, these states are distinguishable. It follows that the minimal DFA for L^* has $3/4 \cdot 2^n$ states, which meets the upper bound given by Corollary 1. \square

Notice that the upper bound given by Lemma 2 depends on the number of final states in a given DFA. Now, in the main result of our paper, we present automata with k final states that meet the upper bound $2^{n-1} + 2^{n-1-k}$.

To this aim, consider an n -state DFA $A = (Q, \Sigma, \delta, s, F)$, where

- $Q = \{0, 1, \dots, n - 1\}$;
- $\Sigma = \{a, b\}$;
- $s = 0$;
- $F = \{n - k, n - k + 1, n - k + 2, \dots, n - 1\}$;
- $\delta(i, a) = (i + 1) \bmod n$,
 $\delta(0, b) = 0$,
 $\delta(i, b) = i + 1$ if $1 \leq i \leq n - 3$,
 $\delta(n - 2, b) = 0$,
 $\delta(n - 1, b) = n - 1$.

The DFA A with $k = 3$ is shown in Fig. 6. Notice that this automaton is obtained by a modification of YZS'94 automaton in Fig. 3 presented in [13]. These two automata differ only in transitions on b in the states $n - 2$ and $n - 1$.

Construct an NFA N for the language $L(A)^*$ as described in Section 3. For $k = 3$, the NFA N shown in Fig. 7. Consider the subset automaton of N , and let as show that this subset automaton has $2^{n-1} + 2^{n-1-k}$ reachable and pairwise distinguishable states.

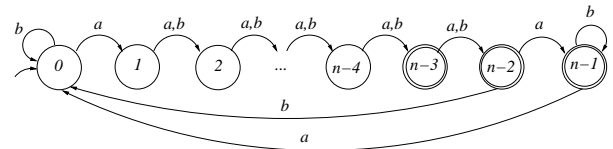


Figure 6: Modified YZS'94 automaton; $k = 3$.

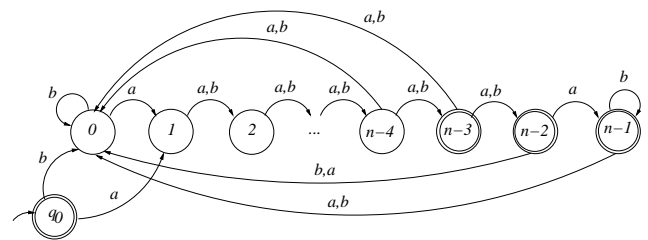


Figure 7: The NFA N for modified YZS'94 automaton; $k = 3$.

Lemma 4. *The subset automaton of the NFA N has $2^{n-1} + 2^{n-1-k}$ reachable state.*

Proof. Notice that if a reachable set contains a final state of N , then it must contain also the state 0.

The set $\{q_0\}$ is reachable since it is the initial state of subset automaton. The set $\{0\}$ is reached from $\{q_0\}$ by

b , and since we have $\{0\} \xrightarrow{a^i} \{i\}$ if $1 \leq i \leq n - k - 1$, all subsets S with $|S| = 1$ are reachable.

Next we have

$$\begin{aligned} \{n - k - 1\} &\xrightarrow{a} \{0, n - k\} \xrightarrow{b} \dots \xrightarrow{b} \{0, n - 2\}, \\ \{0, n - 2\} &\xrightarrow{a} \{0, 1, n - 1\} \xrightarrow{b^{n-3}} \{0, n - 2, n - 1\} \xrightarrow{b} \{0, n - 1\}, \\ \{0, n - 1\} &\xrightarrow{a} \{0, 1\} \xrightarrow{b^{i-1}} \{0, i\} \text{ if } 1 \leq i \leq n - k - 1. \end{aligned}$$

Finally, if $1 \leq i < j \leq n - k - 1$ then $\{i, j\}$ is reached from $\{0, j - i\}$ by a^i . Thus all S with $|S| = 2$ are reachable.

Assume that every set S with $|S| = t$, where $2 \leq t \leq n - 1$, is reachable. Let $S = \{i_1, i_2, \dots, i_t, i_{t+1}\}$, where $0 \leq i_1 < i_2 < \dots < i_t < i_{t+1} \leq n - 1$, be set of size $t + 1$.

Consider three cases:

(i) $i_1 = 0$ and $i_{t+1} = n - 1$.

Let $S' = \{0, i_3 - i_2, i_4 - i_2, \dots, i_t - i_2, n - 2\}$. Then S' is of size t , thus it is reachable by the induction hypothesis. Since we have

$$S' \xrightarrow{a} \{0, 1, i_3 - i_2 + 1, \dots, i_t - i_2 + 1, n - 1\} \xrightarrow{b^{i_2-1}} S,$$

the set S is reachable.

(ii) $i_1 = 0$ and $i_{t+1} < n - 1$.

Let $S' = \{0, i_3 - i_2, \dots, i_{t+1} - i_2, n - 1\}$. Then S' is of size $t + 1$ and contains 0 and $n - 1$, thus S' is reachable by (i). Since we have

$$S' \xrightarrow{a} \{0, 1, i_3 - i_2 + 1, \dots, i_t - i_2 + 1, i_{t+1} - i_2 + 1\} \xrightarrow{b^{i_2-1}} S,$$

the set S is reachable.

(iii) $i_1 > 0$ and $i_{t+1} < n - k$.

Take $S' = \{0, i_2 - i_1, i_3 - i_1, \dots, i_t - i_1, i_{t+1} - i_1\}$. Then $|S'| = t + 1$ and S' contains state 0. Therefore the set S' is reachable as shown in cases (i) and (ii).

Since we have $S' \xrightarrow{a^{i_1}} \{i_1, i_2, i_3, \dots, i_t, i_{t+1}\} = S$, the set S is reachable.

This proves the reachability of $2^{n-1} + 2^{n-1-k}$ states. \square

Lemma 5. *All reachable states of the subset automaton of the NFA N are pairwise distinguishable.*

Proof. Notice that in the NFA N , the string $a^{n-1-i}b^n$ is accepted only from the state i . By Proposition 1 this proves the distinguishability of subsets of $\{0, 1, \dots, n - 1\}$. Now we need to show that $\{q_0\}$ is not equivalent to any final subset S of $\{0, 1, \dots, n - 1\}$. If S is final, then there is a final state $i \geq n - k$ such that $i \in S$. Then $a^{n-1-i}b^n$ is accepted by the subset automaton from S and rejected from $\{q_0\}$. This concludes the proof. \square

Now, we can state our main result.

Theorem 2. *Let $n \geq 3$ and $1 \leq k \leq n - 1$. There exists an n -state DFA A with k final states such that the minimal DFA for the language $L(A)^*$ has $2^{n-1} + 2^{n-1-k}$ states. \square*

We conclude this section with two observations showing that the numbers 1 and n can be attained by the complexity of Kleene closure.

Proposition 2. *For every n , there exists a binary language L accepted by a minimal n -state DFA such that the language L^* has state complexity 1.*

Proof. Let $L = \{a, b\} \cup \{w \mid |w| \geq n - 1\}$. The minimal DFA for L has n states. Since $a \in L, b \in L$, we have $L^* = \{a, b\}$, and therefore the state complexity of L^* is 1. \square

Proposition 3. *For every n , there exists a binary language L accepted by a minimal n -state DFA such that the language L^* has state complexity n .*

Proof. Let $L = ((a + b)^n)^*$. The minimal DFA for L has n states. Next, we have $L = L^*$ and therefore the state complexity of L^* is n . \square

5 Two to Five-State Automata: Frequency of Possible Complexities for Kleene Closure

In this section, we consider not only the worst case, but rather study all possible values that can be obtained as the number of states of the minimal DFA recognizing the Kleene closure of a regular language represented by a minimal n -state DFA.

For Kleene closure, the possible resulting values are in the range from 1 to $3/4 \cdot 2^n$, and it is known that for a growing alphabet of size 2^n , no gaps in the hierarchy of possible complexities exist [6].

Here we study the binary case. Using the lists of pairwise non-isomorphic minimal deterministic finite automata of 2, 3, 4, and 5 states, we computed the frequencies of the resulting complexities for Kleene closure, and showed that every value in the range from 1 to $3/4 \cdot 2^n$ occurs at least ones.

5.1 Results for Two to Five-State Automata with Average Value of Complexity for Kleene Closure

Our results for $n = 2, 3, 4, 5$ concerning the frequency of the resulting complexities, including the average complexity, are displayed in the four graphs shown in Figures 8-11 on the next page.

Notice that for $n = 4, 5$ the complexity one has the highest frequency. On the other hand, there are only a four DFA's whose Kleene closure has complexity 2. Starting with complexity 5, the frequency has a decreasing tendency. The average values approximately n , which corresponds to the fact that the high complexities occur very rarely.

Although, in the worst case, the Kleene closure is a hard operation with an exponential complexity, its average complexity is only n , which allows the operation to be effectively used in practical applications.

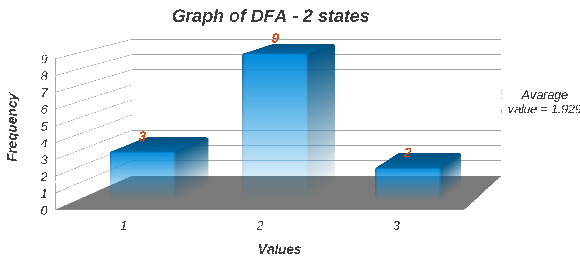


Figure 8: The frequency of complexities of Kleene closure for 2-state DFAs; average = 1.929

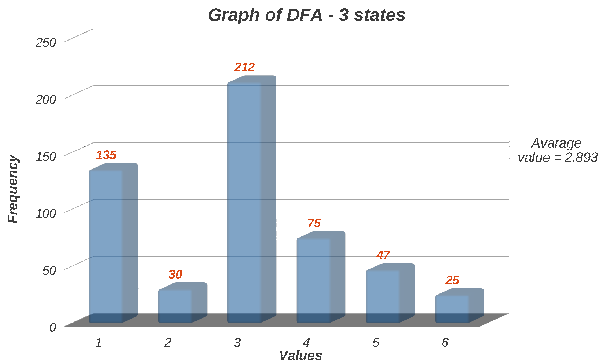


Figure 9: The frequency of complexities of Kleene closure for 3-state DFAs; average = 2.893

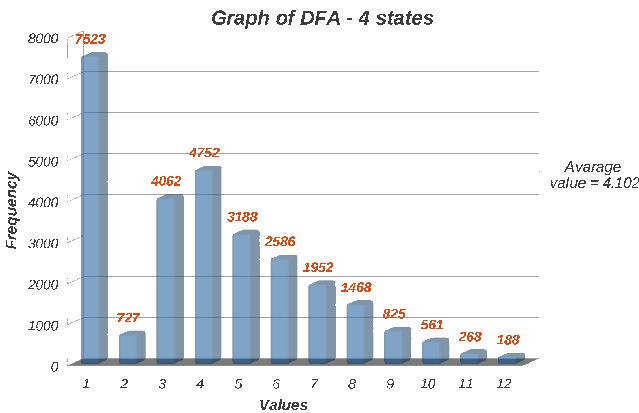


Figure 10: The frequency of complexities of Kleene closure for 4-state DFAs; average = 4.102

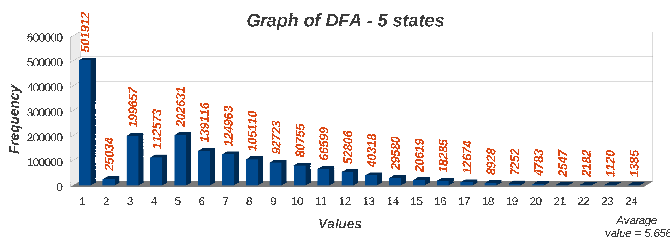


Figure 11: The frequency of complexities of Kleene closure for 5-state DFAs; average = 5.656

6 Randomly Generated Automata and Complexities of Kleene Closure

This subsection is different from the previous one. For $n \geq 6$ we do not have input text files. We change the strategy, and consider binary automata, in which the first symbol is a circular shift of the states, and the second symbol is generated randomly. We consider an arbitrary number of final states. We run our application on such randomly generated automaton. We consider an arbitrary number of final states. We show that all values from 1 to $3/4 \cdot 2^n$ are attainable, and for every m with $1 \leq m \leq 3/4 \cdot 2^n$, we provide an n -state binary DFA A such that the state complexity of $L(A)^*$ is exactly m . The lists of these automata for $n = 6, 7, 8$ follow.

Thus our computations show, that in the binary case, up to $n = 8$, no holes in the state complexity of Kleene closure exist. Moreover, for every n , the numbers 1, n , and $2^{n-1} + 2^{n-1-k}$ with $1 \leq k \leq n - 1$ are attainable by the complexity of Kleene closure.

7 Conclusions

We studied the complexity of languages that results from the Kleene closure operation on regular languages. First, we proved that the n -state automata presented by Maslov in his 1970 paper meets the upper bound $3/4 \cdot 2^n$ on the state complexity of Kleene closure. We fixed a small error in the Maslov's paper [8], which claimed the upper bound $3/4 \cdot 2^n - 1$.

Then, in the main result of our paper, we provided the n -state binary automata with k final states, that meet the upper bound $2^{n-1} + 2^{n-1-k}$ on the state complexity of Kleene closure.

In the second part of the paper, we considered all possible values of the complexity of Kleene closure in the binary case. Using our application and the lists of pairwise non-isomorphic minimal automata of 2,3,4, and 5 states, we computed the frequency of the resulting complexities of Kleene closure and the average complexity of Kleene closure. We showed that each possible complexity occurs at least once.

For $n = 6, 7, 8$, we considered automata, in which the first symbol is a circular shift of the states, the second symbol is generated randomly, and the number of final states is arbitrary. For every possible value m in the range from 1 to $3/4 \cdot 2^n$, we found an n -state DFA accepted a language such that the minimal DFA for the Kleene closure of this language has exactly m states.

Thus for $n \leq 8$, every value in the range from 1 to $3/4 \cdot 2^n$ is attainable by the complexity of the Kleene closure in the binary case. Whether this is true for larger values of n remains open. Also getting the whole range of complexities from 1 to $3/4 \cdot 2^n$ for any fixed alphabet, or at least for an alphabet that grows at most linearly with n , is of great interest to us.

References

- [1] Brzozowski, J., Leiss, E.: On equations for regular languages, finite automata, and sequential networks, *Theoret. Comput. Sci.* 10, 19–35 (1980)
- [2] Čevorová, K.: Kleene star on unary regular languages. DCFS 2013, to appear.
- [3] Geffert, V.: (Non)determinism and the size of one-way finite automata. In: Mereghetti, C., Palano, B., Pighizzini, G., Wotschke D. (eds.) 7th International Workshop on Descriptive Complexity of Formal Systems, pp. 23–37. University of Milano, Italy (2005)
- [4] Hopcroft, J.: An $n \log n$ algorithm for minimizing states in A finite automaton STAN-CS-71-190. *Computer Science* (1971)
- [5] Iwama, K., Kambayashi, Y., Takaki, K.: Tight bounds on the number of states of DFAs that are equivalent to n -state NFAs. *Theoret. Comput. Sci.* 237, 485–494 (2000). Preliminary version in: Bozapalidis, S. (ed.) 3rd International Conference on Developments in Language Theory. Aristotle University of Thessaloniki (1997)
- [6] Jirásková, G.: On the state complexity of complements, stars, and reversals of regular languages. In: Ito M., Toyama M. (eds.) DLT 2008. LNCS, vol. 5257, pp. 431–442. Springer (2008)
- [7] Jirásková, G.: Magic numbers and ternary alphabet. *Int. J. Found. Comput. Sci.* 22(2): 331–344 (2011)
- [8] Maslov, A.N.: Estimates of the number of states of finite automata. *Soviet Math. Doklady* 11, 1373–1375 (1970)
- [9] Sipser, M.: *Introduction to the theory of computation*. PWS Publishing Company, Boston (1997)
- [10] Rabin, M., Scott, D.: Finite automata and their decision problems. *IBM Res. Develop.* 3, 114–129 (1959)
- [11] Šebej J.: Reversal of regular language and state complexity. Master's thesis. P.J. Šafárik University in Košice, Slovakia (2012)
- [12] Yu, S.: Chapter 2: Regular languages. In: Rozenberg, G., Salomaa, A. (eds.) *Handbook of Formal Languages - Vol. I*, pp. 41–110. Springer, Heidelberg (1997)
- [13] Yu, S., Zhuang, Q., Salomaa, K.: The state complexity of some basic operations on regular languages. *Theoret. Comput. Sci.* 125, 315–328 (1994)