Transition Complexity of Language Operations

Michael Domaratzki *
Jodrey School of Computer Science
Acadia University
Wolfville, NS B4P 2R6 Canada
mike.domaratzki@acadiau.ca

Kai Salomaa†
School of Computing
Queen's University
Kingston, ON K7L 3N6 Canada
ksalomaa@cs.queensu.ca

Abstract

The number of transitions required by a nondeterministic finite automaton (NFA) to accept a regular language is a natural measure of the size of a that language. There has been a significant amount of work on considering the trade-off between the number of transitions and other descriptional complexity measures for regular languages. In this paper, we consider the effect of language operations on the number of transitions required to accept a regular language. This work extends previous work on descriptional complexity of regular language operations, in particular, under the measures of deterministic state complexity, nondeterministic state complexity and regular expression size.

1 Introduction

The examination of the descriptive complexity of operations on the regular languages has a long history, with a large emphasis being on the deterministic state complexity of an operation: what is the increase in the sizes of deterministic finite automata (DFA, see Section 2 for definitions) when a given operation is applied? For a survey of these results, see Yu [22, 23]. However, recent work has also focused on other measures of the descriptional complexity of operations on regular languages, most notably the nondeterministic state complexity [5, 9, 10], but also including regular expressions size [6], radius and nondeterministic radius [3, 5].

Measuring the size of a nondeterministic finite automaton (NFA) by the number of transitions has received a significant amount of attention in the literature [7, 13, 14, 15, 18, 21]. The rationale for this research is that the number of transitions in an NFA M is more likely to be the dominant term in an expression of the total storage required for M. Much of the research using the number of transitions as a descriptional complexity measure has been concerned with trade-offs in descriptional complexity: for instance, bounds on the number of transitions in an NFA accepting the same language as a regular expression of some given size.

In this paper, we examine the change in the number of transitions when applying operations which preserve regularity, including boolean operations, catenation, Kleene

*Research supported in part by a grant from NSERC.
†Research supported in part by a grant from NSERC.

closure and reversal. We also investigate morphic operations, where the change in the number of transitions is a more interesting descriptional complexity problem than in the case of state complexity. For some of these operations, we can more precisely quantify the change in the number of transitions by considering refinements on the number of transitions, including the number of transitions leaving the initial state of an NFA and the number of transitions entering all final states of an NFA.

2 Preliminary Definitions

For additional background in formal languages and automata theory, please see Rozenberg and Salomaa [20] or Hopcroft and Ullman [11]. Let Σ be a finite set of symbols, called *letters*. Then Σ^* is the set of all finite sequences of letters from Σ, which are called *words*. The empty word ϵ is the empty sequence of letters. The *length* of a word $w = w_1 w_2 \cdots w_n \in \Sigma^*$, where $w_i \in \Sigma$, is n, and is denoted $|w|$. Given a word $w \in \Sigma^*$ and $a \in \Sigma$, $|w|_a$ is the number of occurrences of a in w. A *language* L is any subset of Σ^*. By $\overline{L}$, we mean $\Sigma^* - L$, the complement of L.

The reversal of a word $w = x_1 x_2 \cdots x_n$ ($x_i \in \Sigma$), denoted w^R, is defined by $w^R = x_n \cdots x_2 x_1$. By extension, $L^R = \{x^R \ : \ x \in L\}$. Let Σ, Δ be alphabets and $h : \Sigma \to \Delta^*$ be a function. Then h can be extended to a morphism $h : \Sigma^* \to \Delta^*$ via the condition that $h(uv) = h(u)h(v)$ for all $u, v \in \Sigma^*$. If $L \subseteq \Sigma^*$, then $h(L) = \{h(x) \ : \ x \in L\}$ and if $L' \subseteq \Delta^*$, $h^{-1}(L') = \{x \ : \ h(x) \in L'\}$.

An NFA is denoted as $M = (Q, \Sigma, \delta, q_0, F)$ where Q is the finite set of states, Σ is the input alphabet, $q_0 \in Q$ is the distinguished start state and $F \subseteq Q$ is the set of final states. Further, $\delta \subseteq Q \times \Sigma \times Q$. Given a word $w = w_1 w_2 \cdots w_n \in \Sigma^*$, where $w_i \in \Sigma$, we say that w is accepted by M if there exist $q_1, q_2, \ldots, q_n \in Q$ such that $(q_{i-1}, w_i, q_i) \in \delta$ for all $1 \leq i \leq n$ and $q_n \in F$. The language $L(M) \subseteq \Sigma^*$ accepted by M is the set of all words which are accepted by M. An NFA is deterministic (a DFA) if, for all pairs $(q, a) \in Q \times \Sigma$, there exists at most one $q' \in Q$ such that $(q, a, q') \in \delta$. A DFA in which there exists exactly one $q' \in Q$ for each $(q, a) \in Q \times \Sigma$ such that $(q, a, q') \in \delta$ is called complete; a DFA which is not complete is called incomplete.

Recall that a state q is useful if q is reachable from the start state and there exists a final state which is reachable from q. In this paper, we only consider NFAs where all states are useful. Unless otherwise mentioned, by an NFA we mean an automaton without ϵ-transitions. The state complexity of a regular language is the minimal number of states in any DFA accepting L. Similarly, the nondeterministic state complexity of a language L is the minimal number of states in any NFA accepting L. The state complexity (respectively, nondeterministic state complexity) of a regular language L is denoted as $\mathrm{sc}(L)$ (resp., $\mathrm{nsc}(L)$).

We now define the transition complexity of a regular language:

Definition 2.1 *Let L be a regular language. The (nondeterministic) transition complexity of L, $tc(L)$, is the smallest number of transitions of any NFA that recognizes L.*

We also require the following notation to discuss additional properties of transition complexity of regular languages. For any regular language $L \subseteq \Sigma^*$, let

$$\mathcal{M}(L) = \{M = (Q, \Sigma, \delta, q_0, F) \ : \ M \text{ is an NFA, } L(M) = L \text{ and } |\delta| = \mathrm{tc}(L)\}.$$

Thus, $\mathcal{M}(L)$ is the set of all NFAs accepting L which have the minimal number of transitions. We call any $M \in \mathcal{M}(L)$ a *transition-minimal* NFA for L.

3 Transition Complexity of Boolean Operations

We first examine the transition complexity of the operations of union, intersection and complementation. To precisely discuss the transition complexity of union, we require some additional notation. Let L be a regular language and let $s(L) = \min_{M \in \mathcal{M}(L)}\{|\delta \cap (\{q_0\} \times \Sigma \times Q)|\}$. That is, $s(L)$ is the minimal number of transitions leaving the start state of M for any transition-minimal NFA M accepting L.

Theorem 3.1 *Let L_1, L_2 be regular languages with $tc(L_i) = n_i$ and $s(L_i) = s_i$. Then $tc(L_1 \cup L_2) \leq n_1 + n_2 + s_1 + s_2$.*

Proof. Let $M_i = (Q_i, \Sigma, \delta_i, q_i, F_i)$ be a transition-minimal NFAs with s_i transitions leaving q_i for $i = 1, 2$. Let $M = (Q_1 \cup Q_2 \cup \{q_0\}, \Sigma, \delta, q_0, F)$ be the NFA defined by

$$\delta = \delta_1 \cup \delta_2 \cup \{(q_0, a, q) \ : \ (q \in Q_1 \text{ and } (q_1, a, q) \in \delta_1) \text{ or } (q \in Q_2 \text{ and } (q_2, a, q) \in \delta_2)\}$$

and F defined by $F = F_1 \cup F_2$ if $\epsilon \notin L_1 \cup L_2$ and $F = F_1 \cup F_2 \cup \{q_0\}$ otherwise. Thus, q_0 has the union of the transitions leaving q_1 and q_2. Therefore, M can either simulate M_1 or M_2 and $L(M) = L_1 \cup L_2$. Further, M has $n_1 + n_2 + s_1 + s_2$ transitions. ■

Corollary 3.2 *Let $L_1, L_2 \subseteq \Sigma^*$ be regular languages with $tc(L_i) = n_i$. Then $tc(L_1 \cup L_2) \leq 2(n_1 + n_2)$.*

We now consider whether the bound in Theorem 3.1 is tight:

Lemma 3.3 *For all $n_1, n_2 \geq 1$, there exist regular languages L_1, L_2 such that $tc(L_i) = n_i$, and $tc(L_1 \cup L_2) \geq n_1 + n_2 + s(L_1) + s(L_2)$.*

Proof. Let $n_1, n_2 \geq 1$. Let $L_1 = (a^{n_1})^*$ and $L_2 = (b^{n_2})^*$ where a, b are distinct letters. Then we can verify that $\text{tc}(L_i) = n_i$: each requires n_i states by a counting argument (see, e.g., Holzer and Kutrib [10, Lemma 3]), and must also contain a cycle, so each must also have n_i transitions. Note that $s(L_i) = 1$.

Now, consider any NFA M accepting $L_1 \cup L_2$. It is known that M requires $n_1 + n_2 + 1$ states [10, Thm. 5]. Further, $n_1 + n_2$ of these states must be reachable from the start state, which requires $n_1 + n_2$ transitions. However, since $L_1 \cup L_2$ is infinite, we additionally require that there is one further transition to create a cycle. Assume without loss of generality that this transition is labelled a. Then we note that if these were the only transitions in M, every sufficiently long word in $L_1 \cup L_2$ would contain an occurrence of a, since it must pass through the cycle. However, this is a contradiction. Thus, we require at least one more transition, for a total of $n_1 + n_2 + 2$ transitions. This establishes the result. ■

The case of intersection is interesting, as it requires us to consider the labels of the transitions:

Theorem 3.4 *Let $L_1, L_2 \subseteq \Sigma^*$ be regular languages with $tc(L_1) = n$ and $tc(L_2) = m$. Further, let $\Sigma = \{a_1, \dots, a_k\}$, $n = n_1 + \cdots + n_k$ and $m = m_1 + \cdots + m_k$ where n_i (resp., m_i) is the number of transitions labelled by a_i in some fixed transition-minimal NFA accepting L_1 (resp., L_2). Then $tc(L_1 \cap L_2) \le \sum_{i=1}^{k} m_i n_i$.*

Proof. Let $M_i = (Q_i, \Sigma, \delta_i, q_i, F_i)$ be any transition-minimal NFA accepting L_i, for $i = 1, 2$. Then consider the standard cross-product construction for an NFA accepting $L_1 \cap L_2$: let $M = (Q_1 \times Q_2, \Sigma, \delta, (q_1, q_2), F_1 \times F_2)$, where

$$\delta = \{((q_i, q_j), a, (q_k, q_\ell)) \ : \ a \in \Sigma, (q_i, a, q_k) \in \delta_1, (q_j, a, q_\ell) \in \delta_2\}.$$

From this, we note that a transition with label a_i is present in δ for every ordered pair of transitions from δ_1 and δ_2 labelled by a_i. Thus, we get exactly $n_i m_i$ such transitions for each a_i. This gives the result. ■

Thus, the minimal transition complexity of intersection requires us to examine the decompositions of n and m as in Theorem 3.4 which minimize $\sum_{i=1}^{k} n_i m_i$. We now prove that Theorem 3.4 is tight for decompositions in which the number of transitions is balanced:

Lemma 3.5 *Let $n, m \ge 1$ and $\Sigma = \{a, b\}$. There exist languages $L_1, L_2 \subseteq \Sigma^*$ with $tc(L_1) \le 2n$ and $tc(L_2) \le 2m$ such that $tc(L_1 \cap L_2) \ge 2nm$.*

Proof. Let $L_1, L_2 \subseteq \Sigma^*$ be given by

$$L_1 = \{w \in \Sigma^* \ : \ |w|_a \equiv 0 \pmod{n}\}, \quad L_2 = \{w \in \Sigma^* \ : \ |w|_b \equiv 0 \pmod{m}\}.$$

Note that L_1 can be accepted by an NFA with n states, n transitions labelled a and n transitions labelled b. Similarly, L_2 can be accepted by an NFA with m states, m transitions labelled a and m transitions labelled b.

Let $M = (Q, \Sigma, \delta, q_0, F)$ be any NFA for $L_1 \cap L_2$. Let $B(i, j) \subseteq Q$ be the set of states that can be reached by some word w where $|w|_a \equiv i \pmod{n}$ and $|w|_b \equiv j \pmod{m}$. As all states are useful, the sets $B(i, j)$ and $B(i', j')$ must be disjoint for all pairs $1 \le i, i' \le n$, $1 \le j, j' \le m$, $(i, j) \ne (i', j')$.

Now, consider an arbitrary set $B(i, j)$. There must be some state $q_i \in B(i, j)$ such that (q_i, a, q_i') where $q_i' \in B(i + 1, j)$ (where addition is performed modulo m). Similarly, there must be a transition labelled b from some state in $B(i, j)$ to some state in $B(i, j + 1)$ (again, $j+1$ is interpreted modulo n). Thus, in total, each $B(i, j)$ must have two transitions leaving it, and we get $2nm$ transitions in M. Note that nm transitions are labelled a and nm are labelled b. ■

We note also that trivial intersections also achieve the upper bound in Theorem 3.4: if $L_1 = \{a^n\} \subseteq \{a, b\}^*$ and $L_2 = \{b^m\} \subseteq \{a, b\}^*$, then $L_1 \cap L_2 = \emptyset$, and $\mathrm{tc}(\emptyset) = 0 = n \cdot 0 + 0 \cdot m$. Proving that Theorem 3.4 is tight for more complex decompositions of n and m is a topic for further research. For complementation, we get the following bounds:

Theorem 3.6 *Let $L \subseteq \Sigma^*$ be a regular language over a k-letter alphabet Σ with $tc(L) = n$. Then $tc(\overline{L}) \le k 2^{n+1}$. Further, for all $n \ge 1$, there exists a regular language $L_n \subseteq \{a, b\}^*$ with $tc(L_n) \le n$ and $tc(\overline{L_n}) \ge 2^{n/2-5} - 1$.*

Proof. If $M = (Q, \Sigma, \delta, q_0, F)$ is an NFA accepting L with n transitions, then $|Q| \leq n+1$. By the subset construction, there exists a DFA $M' = (2^Q, \Sigma, \delta', q_0', F')$ with at most 2^{n+1} states accepting L. By exchanging final and non-final states in M', we get a DFA accepting $\overline{L}$. This DFA again has at most 2^{n+1} states and $k2^{n+1}$ transitions.

For the lower bound, consider the language $L_k = \{a, b\}^* a\{a, b\}^k a\{a, b\}^*$ for $k \geq 1$. Holzer and Kutrib [10] have shown that $\text{nsc}(L_k) = k$ while $\text{nsc}(\overline{L_k}) = 2^{k-2}$. Now, consider any NFA M accepting $\overline{L_k}$, and assume that it has fewer than $2^{k-2}-1$ transitions. Since we can assume that M is initially connected, this means that M has strictly fewer than 2^{k-2} states, a contradiction. Thus, $\text{tc}(\overline{L_k}) \geq 2^{k-2}-1$. Now, we can observe that $\text{tc}(L_k) \leq 2k+6$. This gives the result. ∎

It is interesting to note that Jirásková [16] gives a language $L_n \subseteq \{a, b\}^*$ (for all $n \geq 1$) which demonstrates a tight lower bound on the blow-up for nondeterministic state complexity of complementation, i.e., $\text{nsc}(L_n) = n$ and $\text{nsc}(\overline{L_n}) = 2^n$. However, the NFA recognizing L_n has $4n - 4$ transitions, which does not imply as strong a result as the witness languages of Holzer and Kutrib.

For the unary case, we have a tight result due to a powerful result of Mera and Pighizzini [19]. In this case, we can immediately conclude that for all $n \geq 1$ there exists a unary regular language L_n such that $\text{tc}(L_n) \leq n$, while $\text{tc}(\overline{L_n}) \geq g(n)$, where $g(n)$ is Landau's function, which satisfies $g(n) \in e^{\Theta(\sqrt{n \lg n})}$. For unary alphabets, this is a tight bound on the effect of complementation on transition complexity.

4 Transition Complexity of Catenation Operations

To discuss the transition complexity of catenation, we require further notation. Let L be a regular language and let $f(L) = \min_{M \in \mathcal{M}(L)}\{|\delta \cap (Q \times \Sigma \times F)|\}$. That is, $f(L)$ is the minimal number of transitions entering the final states of M for any transition-minimal NFA M accepting L.

Theorem 4.1 *Let L_1, L_2 be regular languages with $tc(L_i) = n_i$ for $i = 1, 2$. Then $tc(L_1L_2) \leq n_1 + n_2 + f(L_1)$.*

Proof. First, assume that $f(L_1) > 0$. Let $M_i = (Q_i, \Sigma, \delta_i, q_i, F_i)$ be a transition-minimal NFA for L_i $(i = 1, 2)$ with the additional condition that M_1 has $f(L_1) > 0$ transitions entering F_1. Let $M = (Q_1 \cup Q_2, \Sigma, \delta, q_1, F)$ be the NFA defined by $\delta = \delta_1 \cup \delta_2 \cup \{(q, a, q_2) : q \in Q_1, (q, a, q_f) \in \delta_1, q_f \in F_1\}$. Further, $F = F_2$ if $q_2 \notin F_2$ and $F = F_2 \cup F_1$ if $q_2 \in F_2$. From this, we can verify that $L(M) = L_1L_2$ and M has $n_1 + n_2 + f(L_1)$ transitions.

If $f(L_1) = 0$, then $L_1 = \{\epsilon\}$ or $L_1 = \emptyset$. In the first case, $L_1L_2 = L_2$, and $\text{tc}(L_1L_2) \leq n_2$. In the second case, $L_1L_2 = \emptyset$, and $\text{tc}(L_1L_2) = 0$. Thus, the inequality holds in both cases. ∎

We can make Theorem 4.1 more precise with some additional notation. For any NFA M, let $P = \{q \in Q : \exists q' \in F \text{ such that } (q, a, q') \in \delta\}$. That is, P is the set of prefinal states in M. Then for any transition-minimal NFA M over an alphabet $\Sigma = \{a_1, a_2, \ldots, a_k\}$ with $f(L)$ transitions entering the final states, let us write $f(L)$ as

$$f(L) = \sum_{q \in P} \sum_{i=1}^{k} n_{q,i} \tag{1}$$

where $n_{q,i} = |\{q' \in F \ : \ (q, a_i, q') \in \delta\}|$. We note that it is possible that $n_{q,i}$ is greater than one. We now employ the decomposition of $f(L)$ to prove another upper-bound on the transition complexity of catenation:

Theorem 4.2 *Let $L_1, L_2 \subseteq \Sigma^*$ be regular languages over Σ ($|\Sigma| = k$) with $tc(L_i) = n_i$ for $i = 1, 2$. If L_1 is accepted by some transition-minimal NFA M where $f(L_1)$ is as in (1), then $tc(L_1 L_2) \leq n_1 + n_2 + \sum_{q \in P} \sum_{i=1}^{k} \min\{n_{q,i}, 1\}$.*

Proof. We consider only the case where $f(L_1) > 0$ (the case where $f(L_1) = 0$ remains the same as in the proof of Theorem 4.1). Let $M_i = (Q_i, \Sigma, \delta_i, q_i, F_i)$ be a transition-minimal NFA for L_i ($i = 1, 2$) with the additional condition that M_1 has $f(L_1)$ transitions entering F_1. Let $M = (Q_1 \cup Q_2, \Sigma, \delta, q_1, F)$ be the NFA defined by $\delta = \delta_1 \cup \delta_2 \cup \{(q, a, q_2) \ : \ q \in Q_1, (q, a, q_f) \in \delta_1, q_f \in F_1\}$. Further, $F = F_2$ if $q_2 \notin F_2$ and $F = F_2 \cup F_1$ if $q_2 \in F_2$.

Let $q_p \in P$ and $1 \leq i \leq k$. If $n_{q,i} > 1$, then δ_1 (in M, the original NFA accepting L) has transitions (q_p, a_i, q') and (q_p, a_i, q'') for some $q', q'' \in F_1$. However, in our original estimation, we count each of these transitions as contributing one transition to the component $\{(q, a, q_2) \ : \ (q, a, q_f) \in \delta, q_f \in F\}$ of δ (in the NFA M accepting $L_1 L_2$). However, only one transition is actually created: it is labelled a and goes from q_p to q_2. This leads us to the refined upper bound. ■

Corollary 4.3 *Let $L_1, L_2 \subseteq \Sigma^*$ be regular languages with $tc(L_i) = n_i$. Then $tc(L_1 L_2) \leq 2n_1 + n_2$.*

We can now consider lower bounds on the transition complexity of catenation:

Lemma 4.4 *For all $n_1, n_2 \geq 1$, there exist regular languages L_1, L_2 such that $tc(L_i) = n_i$, and $tc(L_1 L_2) \geq n_1 + n_2 + f(L_1)$.*

Proof. Let $L_1 = (a^{n_1})^*$ and $L_2 = (b^{n_2})^*$. Then note that $\mathrm{tc}(L_i) = n_i$ and $f(L_1) = 1$. Now, consider $L_1 L_2 = (a^{n_1})^* (b^{n_2})^*$. Clearly, we require $n_1 + n_2$ transitions for the distinct loops accepting $(a^{n_1})^*$ and $(b^{n_2})^*$. Further, since the state sets for these loops must be disjoint (otherwise a word of the form $a^+ b^+ a^+$ could be accepted) one additional transition must connect these two loops, for a total of $n_1 + n_2 + 1$ transitions. ■

The situation for the transition complexity of Kleene closure is slightly more complex, since the measures $f(L)$ and $s(L)$ are not necessarily minimized by the same NFA. Let fs be defined by

$$fs(L) = \min_{M \in \mathcal{M}(L)} \{|\delta \cap ((\{q_0\} \times \Sigma \times Q) \cup (Q \times \Sigma \times F))|\}.$$

Note that $fs(L)$ is the minimum number of transitions leaving the initial state and entering the final states for any one transition-minimal NFA.

Lemma 4.5 *Let Σ be an alphabet with $|\Sigma| = k$, and $L \subseteq \Sigma^*$ be a regular language with $tc(L) = n$.*

(a) If $\epsilon \notin L$, then $tc(L^*) \leq n + k + fs(L)$.

(b) If $\epsilon \in L$, then $tc(L^*) \le n + f(L)$.

Proof. We assume that $fs(L) > 0$, as $fs(L) = 0$ implies $L = \emptyset$ or $L = \{\epsilon\}$, which are easily handled.

Consider case (a) first. Let $M = (Q, \Sigma, \delta, q_0, F)$ be a transition-minimal NFA for L with a total of $fs(L)$ transitions entering the final states and leaving the initial states. Let $M^* = (Q \cup \{q_0'\}, \Sigma, \delta', q_0', F')$ be the NFA defined by $F' = F \cup \{q_0'\}$ and $\delta' = \delta \cup \{(q_0', a, q) : (q_0, a, q) \in \delta\} \cup \{(q, a, q_0') : (q, a, q_f) \in \delta, q_f \in F\} \cup \{(q_0', a, q_0') : (q_0, a, q_f) \in \delta, q_f \in F\}$. Then $L(M^*) = L^*$.

For case (b), if $\epsilon \in L$ then the state q_0' in the construction of M^* is not necessary. Thus, in this case let $M^* = (Q, \Sigma, \delta', q_0, F)$ be defined by $\delta' = \delta \cup \{(q, a, q_0) : (q, a, q_f) \in \delta, q_f \in F\}$. In this case, $L(M^*) = L^*$. ■

Lemma 4.6 *For all $n \ge 0$, there exists a regular language $L_n \subseteq \{a, b\}^*$ with $\epsilon \notin L_n$ (resp., with $\epsilon \in L_n$) such that $tc(L_n) \le n$ and $tc(L_n^*) = n + fs(L_n)$ (resp., $tc(L_n^*) = n + f(L_n)$).*

Proof. First, let's consider the case when $\epsilon \notin L_n$. Let $k \ge 2$ and choose $L_k = a^{k-1}b(a^k b)^*$. Then $\mathrm{tc}(L_k) = k + 1$ and $fs(L_k) = 2$. Let M be a transition-minimal NFA for L_k^*. By Lemma 4.5, M has at most $k + 3$ transitions.

Consider an accepting computation of M on the word $w = a^{k-1}ba^k ba^k b$ and let q_i, $i = 0, \ldots, k$ be the state that M has reached after reading the prefix $a^{k-1}ba^i$ of w. If $q_i = q_j$ for some $0 \le i < j \le k$, then M will accept a word that has a subword a^x with $x > k$. Since the latter is impossible, all states q_i, $i = 0, \ldots, k$ are distinct.

Since M is transition-minimal, the following symbol b must take the state q_k to q_0. It is easy to verify that otherwise M either accepts illegal words or has more than $k + 3$ transitions. Thus we have seen that M has a cycle C of length $k+1$ having k a-transitions and one b-transition.

Since M accepts the empty word, the initial state p_0 of M is a final state. From this it follows that p_0 cannot be part of the cycle C and there must be an a-transition from p_0 to the state q_2 in C. Otherwise, M could not accept $a^{k-1}b$ using at most 2 transitions in addition to the transitions of C. The NFA with transition (p_0, a, q_2) and the cycle C cannot accept the word $(a^{k-1}b)^2 \in L_k^*$ and consequently M needs at least one more transition.

Now, consider the case when $\epsilon \in L_n$. Let $k \ge 1$. Choose $L_k = \epsilon + a^k b(b^2)^*$. Now $\mathrm{tc}(L_k) = k + 2$ and $f(L_k) = 1$. It can be verified that $\mathrm{tc}(L_k^*) = k + 3$. ■

We can also deal with the transition complexity of the positive Kleene closure of a language:

Corollary 4.7 *Let $L \subseteq \Sigma^*$ be a regular language with $tc(L) = n$. Then $tc(L^+) \le n + f(L)$.*

Proof. We can verify that the construction of case (b) of Lemma 4.5 also gives a construction for L^+. ■

We note that Lemma 4.5 and Corollary 4.7 can be further refined using the decompositions of $f(L)$ given by (1).

5 Transition Complexity of Morphic Operations

Generally, the issue of state complexity of morphism and inverse morphism have not been examined. This is expected since the constructions for morphism and inverse morphism appear routine. However, when counting the number of transitions, we find the situation is somewhat more interesting. We begin with the transition complexity of morphisms:

Theorem 5.1 *Let $h : \Delta^* \to \Sigma^*$ be a morphism and $L \subseteq \Delta^*$ be a regular language. Let $tc(L) = n$ and $\ell = \max\{|h(a)| \ : \ a \in \Sigma\}$. Then $tc(h(L)) \leq \ell n$. Furthermore, for all values of $\ell \geq 1$ and $n \geq 1$, this bound is reachable.*

Proof. Let $M = (Q, \Sigma, \delta, q_0, F)$ be an NFA accepting L with n transitions. Then consider the NFA M' obtained by adding, for each transition $(q, a, q') \in \delta$, a chain of new states from q to q' connected by $|h(a)|$ transitions, which causes M' to move from q to q' on input $h(a)$. Thus, every transition in M is replaced by at most ℓ transitions in M'. This gives the upper bound.

For the lower bound, for any $n \geq 1$ and $\ell \geq 1$, let $\Delta = \Sigma = \{a\}$, and $h(a) = a^\ell$. Further, let $L = \{a^n\}$. Then $h(L) = \{a^{\ell n}\}$. Clearly, L requires n transitions, while $h(L)$ requires ℓn, which demonstrates that the upper bound is tight. ■

For inverse morphism, we find an interesting observation: even though the standard construction for demonstrating that the regular languages are closed under inverse morphism does not increase the number of states in an NFA, it can increase the number of transitions.

Theorem 5.2 *Let Σ, Δ be alphabets with $|\Delta| = k$. If $L \subseteq \Sigma^*$ be a regular language with $tc(L) = n$, and $h : \Delta^* \to \Sigma^*$ is a morphism, then $tc(h^{-1}(L)) \leq k(n+1)^2$.*

Proof. Let $M = (Q, \Sigma, \delta, q_0, F)$ be an NFA for L with n transitions. As $\mathrm{tc}(L) = n$, we have $\mathrm{nsc}(L) \leq n + 1$, since we can assume that M is initially connected. Let $M' = (Q, \Delta, \delta', q_0, F)$ be the NFA defined by $(q_1, a, q_2) \in \delta'$ if and only if there is a path from q_1 to q_2 labelled by $h(a)$ in M. This standard construction gives $L(M') = h^{-1}(L)$. Now, consider that M' has $n + 1$ states, and thus has at most $k(n + 1)^2$ transitions. This gives the result. ■

Theorem 5.2 applies to any morphism, including those that may map letters to the empty string. Recall that we say that a morphism $h : \Delta^* \to \Sigma^*$ is ϵ-free if $h(a) \neq \epsilon$ for all $a \in \Delta$. We now show a relationship between the transition complexity of morphisms which are not ϵ-free and the complexity of what we call their ϵ-free restrictions.

Let $h : \Delta^* \to \Sigma^*$ be an arbitrary morphism. Let $\Delta_+, \Delta_\epsilon \subseteq \Delta$ be defined by $\Delta_+ = \{b \in \Delta \ : \ h(b) \neq \epsilon\}$. and $\Delta_\epsilon = \Delta - \Delta_+$. Further, let $h_+ = h|_{\Delta_+}$, i.e., the restriction of h to letters of Δ_+.

Lemma 5.3 *Let $L \subseteq \Sigma^*$ be a regular language, and $h : \Delta^* \to \Sigma^*$ be a morphism. Then $tc(h^{-1}(L)) \leq tc(h_+^{-1}(L)) \cdot (|\Delta_\epsilon| + 1) + |\Delta_\epsilon|$.*

Proof. Let $M_+ = (Q, \Sigma, \delta_+, q_0, F)$ be an NFA for $h_+^{-1}(L)$. If $M = (Q, \Sigma, \delta, q_0, F)$ is defined by $\delta = \delta_+ \cup \{(q, b, q) \ : \ q \in Q, b \in \Delta_\epsilon\}$, then note that $L(M) = h^{-1}(L)$. This holds since M can now read any number of letters $b \in \Delta_\epsilon$, which are mapped to ϵ by h, at any time during the computation. Now, the number of transitions in M is at most $|\delta_+| + |Q||\Delta_\epsilon|$. As $|Q| \leq |\delta_+| + 1$, the result follows. ■

6 Transition Complexity of Reversal

The transition complexity of reversal relies on the standard construction: reverse the transitions and exchange the roles of final and initial states. When counting the number of transitions, we find again that we need to employ the measure of the number of transitions entering the final states. However, the transition complexity itself is dependent on the number of final states: For all regular languages, let $F(L) = \min_{M \in \mathcal{M}(L)}\{|F| \ : \ M = (Q, \Sigma, \delta, q_0, F)\}$.

Theorem 6.1 *Let L be a regular language. If $F(L) = 1$ then $tc(L^R) = tc(L)$, and otherwise $tc(L^R) \leq tc(L) + f(L)$.*

Proof. Let L be a regular language accepted by a transition-minimal NFA $M = (Q, \Sigma, \delta, q_0, F)$. Assume first that $|F| = 1$. Let $F = \{q_f\}$. Let $M' = (Q, \Sigma, \delta', q_f, \{q_0\})$ be the NFA defined by $\delta' = \{(q_2, a, q_1) \ : \ (q_1, a, q_2) \in \delta\}$. Thus, M' has all transitions reversed, and accepts exactly L^R. Note that since reversal is an involution, this implies that $\mathrm{tc}(L) = \mathrm{tc}(L^R)$.

Consider now the case where $F(L) > 1$. Then we may assume that $M = (Q, \Sigma, \delta, q_0, F)$ is a transition-minimal NFA with $f(L)$ transitions entering the final states in F. Let $M' = (Q \cup \{q_0'\}, \Sigma, \delta', q_0', F')$ be the NFA defined by

$$\delta' = \{(q_2, a, q_1) \ : \ (q_1, a, q_2) \in \delta\} \cup \{(q_0', a, q) \ : \ (q, a, q_f) \in \delta, q_f \in F\}$$

and $F' = \{q_0\}$ if $q_0 \notin F$ and $F' = \{q_0, q_0'\}$ if $q_0 \in F$. Again, M' has all transitions reversed. However, we require an additional state q_0' which allows us to simulate all the final states in F. ■

We now consider lower bounds on the results in Theorem 6.1. The case of $F(L) = 1$ is handled by Theorem 6.1. For the case of $F(L) > 1$, we have the following result:

Theorem 6.2 *For all $n \geq 1$ there exists a regular language L_n such that $F(L_n) > 1$, $tc(L_n) \leq n$ and $tc(L_n^R) \geq n + f(L_n)$.*

Proof. Let $k \geq 1$ and define $L_k = (a^k)^*((b^2)^+ \cup (c^2)^+)$. We claim that

(a) $F(L_k) = 2$,

(b) $f(L_k) = 2$,

(c) $\mathrm{tc}(L_k) \leq k + 6$,

(d) $\mathrm{tc}(L_k^R) \geq k + 8$.

First, note that the minimal incomplete DFA for L_k has $k + 6$ transitions, so (c) follows.

Let M be an arbitrary transition-minimal NFA for L_k. It is easy to verify that M must have a cycle C_1 of k a-transitions, a cycle C_2 of two b-transitions and a cycle C_3 of two c-transitions. (If the length of C_2 or C_3 would be a proper multiple of two, condition (c) could not hold.) Also clearly the cycles C_1, C_2 and C_3 cannot have any states in common.

Since all the cycles must be connected, we need at least two transitions to connect the cycles and have "used up" all the available $k + 6$ transitions, which means that there cannot be any states not belonging to one of the cycles C_1, C_2 or C_3. Thus both C_2 and C_3 must have a final state and this gives (a) and (b).

Finally, we argue that (d) holds. Let M' be an arbitrary NFA for L_k^R. Again it is easy to verify that M' must have a cycle C of a-transitions, where the length of C is k (or a multiple of k). Thus it is sufficient to show that M' must have at least 8 transitions not belonging to the cycle C.

Clearly M' must have a cycle C_b of b-transitions that has a length that is a multiple of 2 and the states of C_b are disjoint from the states of C. If the length of C_b is greater than two, we have four b-transitions.

The other case is that M' has states q and p such that there are transitions

$$(q, b, p) \text{ and } (p, b, q). \tag{2}$$

If one of q or p is the initial state of M', the automaton M' will necessarily accept words in b^+c^+ which is impossible. Hence q and p are not the initial state.

Since M' must accept the word b^2a^k, either there must be two b-transitions in addition to (2), or there must be a b-transition from the initial state to q or p, and an a-transition from the other of these states to the cycle C.

Thus there must be either four b-transitions or three b-transitions and an a-transition between cycles C_b and C. Using a completely analogous argument we see that M' must have either four c-transitions or three c-transitions and an a-transition that connects a c-cycle to the states of C. In all cases the total number of transitions is at least $k + 8$. ■

However, we can tighten Theorem 6.1 in general, by appealing again to decomposition of $f(L)$ as in (1):

Theorem 6.3 *Let L be a regular language with $F(L) > 1$, accepted by some transition-minimal NFA M where $f(L)$ is as in (1). Then* $tc(L^R) \leq tc(L) + \sum_{q \in P} \sum_{i=1}^{k} \min\{1, n_{q,i}\}$.

Proof. Let L be a regular language accepted by a transition-minimal NFA $M = (Q, \Sigma, \delta, q_0, F)$. with $f(L)$ transitions entering the final states in F.

Consider the construction $M' = (Q \cup \{q_0'\}, \Sigma, \delta', q_0', F')$ defined by

$$\delta' = \{(q_2, a, q_1) \ : \ (q_1, a, q_2) \in \delta\} \cup \{(q_0', a, q) \ : \ (q, a, q_f) \in \delta, q_f \in F\}$$

and $F = \{q_0\}$ if $q_0 \notin F$ and $F = \{q_0, q_0'\}$ if $q_0 \in F$.

Let $q_p \in P$ and $1 \leq i \leq k$. If $n_{q,i} > 1$, then δ (in M, the original NFA accepting L) has transitions (q_p, a_i, q') and (q_p, a_i, q'') for some $q', q'' \in F$. However, in our original estimation, we count each of these transitions as contributing one transition to the component $\{(q_0', a, q) \ : \ (q, a, q_f) \in \delta, q_f \in F\}$ of δ' (in the NFA M' accepting L^R). Again however, only one transition is actually created: it goes from q_0' to q_p. This leads us to the refined upper bound. ■

The results of Theorem 6.3 can, in general, fail to lead to any advantage over Theorem 6.1, if $n_{q,i} = 1$ for all $q \in P$ and $1 \leq i \leq k$ (this is the case for languages L_n used in the proof of Theorem 6.2).

7 Open Problems

We note some open problems concerning the transition complexity of language operations raised here. First, we note that in Section 5, we have given an upper bound of $k(n+1)^2$

transitions to recognize the inverse morphic image of a language requiring n transitions. However, we do not have a matching lower bound given by a morphism $h : \Delta^* \to \Sigma^*$ and a language L where $|\Delta|$ and $\max_{a \in \Delta} |h(a)|$ do not depend on $\mathrm{tc}(L)$ (a lower bound of kn, where $k = |\Delta|$ is easily obtained by considering $L = \{a^n\}$ and the morphism $h : \{b_1, \ldots, b_k\}^* \to \{a\}^*$ which maps $h(b_i) = a$ for all $1 \leq i \leq k$).

We note some additional operations whose descriptional complexities have been considered in the literature, but which we have not examined here. Jirásková and Okhotin [17] have considered the cyclic shift of a language: $\mathrm{cycle}(L) = \{xy \ : \ x, y \in \Sigma^*, yx \in L\}$. The state complexity of cycle is $2^{\Theta(n^2)}$ and the nondeterministic state complexity is $2n^2 + 1$. If $\mathrm{tc}(L) = m$, $s(L) = s$ and $f(L) = f$, the NFA construction of Jirásková and Okhotin gives an upper bound of $2m^2 + mfs + m$ on the transition complexity of $\mathrm{cycle}(L)$.

We also note that the state complexity of the operation $\frac{1}{2}(L) = \{x \ : \ xy \in L, |x| = |y|\}$ has been considered [2]. The upper bound on the deterministic state complexity is $ne^{O(\sqrt{n \lg n})}$ (the nondeterministic state complexity of $\frac{1}{2}(\cdot)$ has not been formally studied, but the standard NFA construction gives a upper bound of n^3 states). Further, we have not considered the case of shuffle [1] or shuffle on trajectories [4].

Further, the interaction between the different measures $f(L), s(L)$ and $\mathrm{tc}(L)$ provide additional research questions concerning the operations we have studied. For instance, is it true that for all n_1, n_2, f_1 with $f_1 \leq n_1$, there exists a finite alphabet Σ (whose size does not depend on n_1, n_2 or f_1) and regular languages $L_1, L_2 \subseteq \Sigma^*$ with $\mathrm{tc}(L_i) = n_i$ and $f(L_1) = f_1$ such that $\mathrm{tc}(L_1 L_2) = n_1 + n_2 + f_1$?

8 Conclusions

In this paper, we have examined the transition complexity of several basic operations on regular languages. Additional measures for giving upper-bounds on the transition complexity of operations are used, including the number of final states, $F(L)$, the number of transitions leaving the initial state, $s(L)$, and the number of transitions entering the final states, $f(L)$, to refine the upper bounds on the effect of operations on transition complexity. Further, we can often improve the estimates on the worst case behaviour of the operations by examining a more precise decomposition of $f(L)$.

The lack of a general-purpose tool for proving lower bounds on the number of transitions required to accept a regular language, like the Myhill-Nerode theorem for DFAs and fooling-set methods for NFAs (see, e.g., Hromkovič [12] or Glaister and Shallit [8]) make obtaining results about the minimal number of transitions challenging. Additional work is necessary to develop suitable, general-purpose tools for proving sharp lower bounds on the number of transitions required to accept a regular language.

References

[1] C. Câmpeanu, K. Salomaa, and S. Yu. Tight lower bound for the state complexity of shuffle of regular languages. *J. Autom. Lang. Comb.*, 7(3):303–310, 2002.

[2] M. Domaratzki. State complexity of proportional removals. *J. Autom. Lang. Comb.*, 7:455–468, 2002.

[3] M. Domaratzki, K. Ellul, J. Shallit, and M.-W. Wang. Non-uniqueness and radius of cyclic unary NFAs. *Intl. J. Found. Comp. Sci.*, 16:883–896, 2005.

[4] M. Domaratzki and K. Salomaa. State complexity of shuffle on trajectories. *J. Autom. Lang. Comb.*, 9:217–232, 2004.

[5] K. Ellul. Descriptional complexity measures of regular languages. M.Math thesis, University of Waterloo, 2002.

[6] K. Ellul, B. Krawetz, J. Shallit, and M.-W. Wang. Regular expressions: New results and open problems. *J. Autom. Lang. Comb.*, 9:233–256, 2004.

[7] V. Geffert. Translation of binary regular expressions into nondeterministic ϵ-free automata with $O(n \log n)$ transitions. *J. Comput. Sys. Sci.*, 66:451–472, 2003.

[8] I. Glaister and J. Shallit. A lower bound technique for the size of nondeterministic finite automata. *Inf. Proc. Letters*, 59:75–77, 1996.

[9] M. Holzer and M. Kutrib. Unary language operations and their nondeterministic state complexity. In *DLT 2002: Developments in Language Theory*, pages 241–254, 2002.

[10] M. Holzer and M. Kutrib. Nondeterministic descriptional complexity of regular languages. *Intl. J. Found. Comp. Sci.*, 14:1087–1102, 2003.

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

[12] J. Hromkovič. *Communication Complexity and Parallel Computing.* Springer, 1997.

[13] J. Hromkovič. Descriptional complexity of finite automata: Concepts and open problems. *J. Autom. Lang. Comb.*, 7:519–531, 2002.

[14] J. Hromkovič and G. Schnitger. NFAs with and without ϵ-transitions. In *Proceedings of ICALP 2005*, volume 3580 of *LNCS*, pages 385–396. Springer, 2005.

[15] J. Hromkovič, S. Seibert, and T. Wilke. Translating regular expressions into small ϵ-free nondeterministic finite automata. *J. Comput. System Sci.*, 62:565–588, 2001.

[16] G. Jirásková. State complexity of some operations on regular languages. *Theor. Comp. Sci.*, 330(2):287–298, 2005.

[17] G. Jirásková and A. Okhotin. State complexity of cyclic shift. In C. Mereghetti, B. Palano, G. Pighizzini, and D. Wotschke, editors, *Proceedings of DCFS 2005*, pages 182–193, 2005.

[18] Y. Lifshits. A lower bound on the size of ϵ-free NFA corresponding to a regular expression. *Inf. Proc. Lett.*, 85:293–299, 2003.

[19] F. Mera and G. Pighizzini. Complementing unary nondeterministic automata. *Theor. Comp. Sci.*, 330(2):349–360, 2005.

[20] G. Rozenberg and A. Salomaa, Editors. *Handbook of Formal Languages.* Springer, 1997.

[21] G. Schnitger. Regular expressions and NFAs without ϵ-transitions. In *Proc. of STACS 2006*, volume 3884 of *LNCS*, pages 432–443. Springer, 2006.

[22] S. Yu. State complexity of the regular languages. *J. Autom. Lang. Comb.*, 6:221–234, 2001.

[23] S. Yu. State complexity of finite and infinite regular languages. *Bull. EATCS*, 76:142–152, 2002.