Universal Witnesses for State Complexity of Basic Operations Combined with Reversal*

Janusz Brzozowski\textsuperscript{1} and David Liu\textsuperscript{2}

\textsuperscript{1} David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, Canada N2L 3G1
\texttt{brzozo@uwaterloo.ca}

\textsuperscript{2} Department of Computer Science, University of Toronto
Toronto, ON, Canada M5S 3G4
\texttt{liudavid@cs.toronto.edu}

\textbf{Abstract.} We study the state complexity of boolean operations, concatenation, and star, with one or two of the argument languages reversed. We derive tight upper bounds for the symmetric differences and differences of such languages. We prove that the previously discovered bounds for union, intersection, concatenation and star of such languages can all be met by the recently introduced universal witness and its variants.

\textbf{Keywords:} basic operation, boolean operation, regular language, reversal, state complexity, universal witness.

\section{Introduction}

For background on state complexity see \cite{2, 3, 11}. The \textit{state complexity of a regular language} is the number of states in the minimal deterministic finite automaton (DFA) recognizing the language. The \textit{state complexity of an operation} on regular languages is the maximal state complexity of the result of the operation as a function of the state complexities of the arguments. We refer to state complexity simply as \textit{complexity}.

The state complexity of basic operations combined with reversal was studied in 2008 by Liu, Martin-Vide, A. Salomaa, and Yu \cite{9}. For regular languages $K, L \subseteq \Sigma^*$ with state complexities $m$ and $n$, the basic operations considered in \cite{9} were union ($K \cup L$), intersection ($K \cap L$), product (concatenation or concatenation) ($KL$), and star ($L^*$), combined with reversal ($LR$). It was shown that $(2^m-1)(2^n-1)+1$ is a tight upper bound for $(K \cup L)^R = K^R \cup L^R$ and $(K \cap L)^R = K^R \cap L^R$. Gao and Yu \cite{8} found the tight upper bound $m2^n-(m-1)$ for $K \cup L^R$ and $K \cap L^R$. It was also proved in \cite{9} that $3 \cdot 2^{m+n-2} - (2^n - 1)$ is an upper bound for $(KL)^R = L^RK^R$, but the question of tightness was left open. Cui, Gao, Kari and Yu \cite{5} answered this question positively, and also showed that $3 \cdot 2^{m+n-2}$ is an upper bound for

\textsuperscript{*} This work was supported by the Natural Sciences and Engineering Research Council of Canada under grant No. OGP0000871, and was done while the second author was at the University of Waterloo.
Universal Witnesses for Operations with Reversal 73

Fig. 1. DFA $U_n(a, b, c)$ of a complex language $U_n(a, b, c)$

$K^{RL}$. In another paper [6], they proved that $(m - 1)2^n + 2^{n-1} - (m - 1)$ is a tight upper bound for $K^L$. Gao, K. Salomaa, and Yu [7] demonstrated that $2^n$ is a tight upper bound for $(L^*)^R = (L^R)^*$. Thus eight basic operations with reversal added have been considered so far.

There are two usual steps in finding the state complexity of an operation. First, establish an upper bound; then, find a stream $(L_n | n \geq k)$ of languages, where $k$ is some small positive integer, to act as witnesses to show that the bound is tight. In the literature witnesses for binary operations have usually been two distinct streams.

The family of DFAs $U_n(a, b, c) = (Q, \Sigma, \delta, q_0, F)$ for $n \geq 3$ illustrated in Fig. 1 and the corresponding language stream $(U_n(a, b, c) | n \geq 3)$ were proposed in [3] as the “universal witness.”

Throughout this paper we use the notation $w: D \rightarrow t$ to indicate that the map $\delta(\cdot, w)$ corresponding to word $w \in \Sigma^*$ causes the transformation $t$ on the set of states of the DFA $D$, omitting the $D$ if it is clear from the context. The inputs to $U_n(a, b, c)$ perform the following transformations on $Q = \{0, \ldots, n-1\}$: the cycle of all $n$ states, $a : (0, \ldots, n-1)$; the transposition of 0 and 1 (leaving other states unchanged), $b : (0, 1)$; and the singular transformation sending state $n-1$ to state 0 (and not affecting any other states), $c : (n-1 \rightarrow 0)$. It is well known that the inputs of $U_n(a, b, c)$ perform all $n^n$ possible transformations of states, and also that the state complexity of the reverse of $U_n(a, b, c)$ is $2^n$; the latter result follows by a theorem from [10].

In [3] Brzozowski defined two languages $K$ and $L$ over $\Sigma$ to be permutationally equivalent if one can be obtained from the other by permuting the letters of the alphabet. The permutationally equivalent language of $U_n(a, b, c)$ obtained by interchanging $a$ and $b$ is denoted by $U_n(b, a, c)$. The restriction of the language to alphabet $\{a, b\}$ is denoted by $U_n(a, b)$. These definitions and notation are extended to DFAs in the natural way.

It was proved in [3] that the bound $mn$ for union, intersection, difference ($K \setminus L$) and symmetric difference ($K \oplus L$) is met by $(U_m(a, b, c) | m \geq 3)$ and $(U_n(b, a, c) | n \geq 3)$. The bound $2^{n-1} + 2^{n-2}$ for star is met by $U_n(a, b)$, and the bound $(m - 1)2^n + 2^{n-1}$ for product, by $U_m(a, b, c)U_n(a, b, c)$. This justifies the use of $U_n(a, b, c)$ as a “universal witness” for the basic operations.

For some operations we require extensions of the universal witness stream. A dialect of $U_n(a, b, c)$ is the language of any DFA with three inputs $a$, $b$, and $c$, where $a$ is a cycle as above, $b$ is the transposition of any two states $(p, q)$, and $c$ is a singular transformation $(r \rightarrow s)$ sending any state $r$ to any state $s \neq r$. The