The Range of State Complexities of Languages Resulting from the Cut Operation

Michal Hospodár

Mathematical Institute, Slovak Academy of Sciences, Košice, Slovakia

Joint work with Markus Holzer

(Institut für Informatik, Universität Giessen, Germany)

LATA 2019, St. Petersburg, Russia, March 27, 2019

Outline

1. The Cut Operation – Definition and Examples
2. Range of State Complexities of Languages Resulting from Cut (Magic Number Problem for Cut)
3. Unary Case
4. Binary (General) Case
5. Summary

The Cut Operation

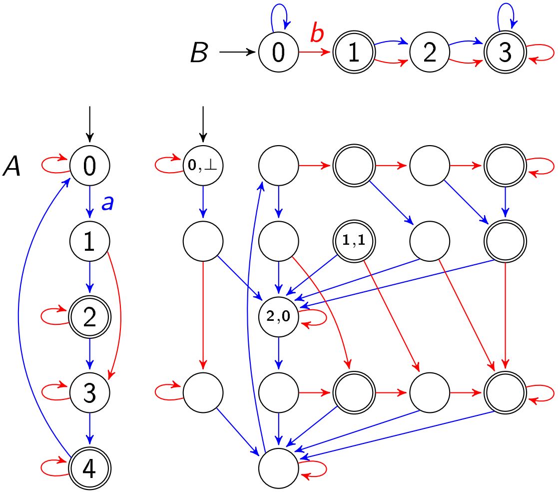
- Concatenation: $KL = \{uv \mid u \in K \text{ and } v \in L\}$
- Cut operation: machine implementation on Unix processors

$$K \,!\, L = \{uv \mid u \in K, v \in L, uv' \notin K \text{ for every nonempty prefix } v' \text{ of } v\}$$

Example (Cut vs. Concatenation)

$K = \{\mathrm{LA}, \mathrm{LATA}\}$
$L = \{\mathrm{TABLE}, \mathrm{CHAIR}\}$
$KL = \{\mathrm{LA}{\cdot}\mathrm{TABLE}, \mathrm{LATA}{\cdot}\mathrm{TABLE}, \mathrm{LA}{\cdot}\mathrm{CHAIR}, \mathrm{LATA}{\cdot}\mathrm{CHAIR}\}$
$K \,!\, L = \{\mathrm{LATA}{\cdot}\mathrm{TABLE}, \mathrm{LA}{\cdot}\mathrm{CHAIR}, \mathrm{LATA}{\cdot}\mathrm{CHAIR}\}$

An Example of the Cut Automaton



DFAs A and B and the cut automaton $A \,!\, B$; notice that the state $(1,1)$ is unreachable.

Definition

- State complexity of a regular language L:
  number $\mathrm{sc}(L) = \min\{n \mid L \text{ is accepted by an } n\text{-state } \mathrm{DFA}\}$
- State complexity of a binary operation $\circ$:
  function $(m, n) \mapsto \max\{\mathrm{sc}(K \circ L) \mid \mathrm{sc}(K) \leq m \text{ and } \mathrm{sc}(L) \leq n\}$
- Range of state complexities resulting from the operation $\circ$:
  $(m, n) \mapsto$ set $\{\mathrm{sc}(K \circ L) \mid \mathrm{sc}(K) = m \text{ and } \mathrm{sc}(L) = n\}$

  A number representing a "hole" in this set
  is called a magic number for the operation $\circ$

The state complexity, or the range of state complexities,
may depend on the size of alphabet over which K and L are defined.

Motivation and History

Cut operation was examined by

- Berglund et al. (2013) – definition, regularity preserving
- Drewes et al. (2017) – state complexity

Magic number problem was investigated by

- Iwama et al. (2000): formulation of the magic number problem for determinization of binary NFAs
- Geffert (2007): there exist magic numbers for determinization of unary NFAs
- Jirásková (2011): no magic numbers for determinization of ternary NFAs
- Holzer et al. (2012): determinization on subregular classes
- Čevorová (2013): Kleene star on unary DFAs

This talk – the magic number problem for cut – complete solution

The Range of Complexities for Cut: Unary Case

Known result: Drewes, Holzer, Jakobi, van der Merwe (2017)

The state complexity of the cut operation on unary languages:

$$f_1(m,n) = \begin{cases} 1, & \text{if } m = 1; \\ m, & \text{if } m \geq 2 \text{ and } n = 1; \\ 2m-1, & \text{if } m,n \geq 2 \text{ and } m \geq n; \\ m+n-2, & \text{if } m,n \geq 2 \text{ and } m < n. \end{cases}$$

Our results in the unary case

Let K, L be unary languages with $\mathrm{sc}(K) = m$ and $\mathrm{sc}(L) = n$.

Condition	Range of attainable complexities for $K\,!\,L$
$m \geq 1, n = 1$	$[1, m]$
$m, n \geq 2$, K infinite	$[1, 2m-1]$
$m, n \geq 2$, K finite	$[n, m+n-2]$

- What about the interval $[2m, n-1]$?

Lemma

There do not exist minimal unary m- and n-state DFAs A and B such that the minimal DFA for $L(A) ! L(B)$ has α states if $2m \leq \alpha \leq n - 1$.

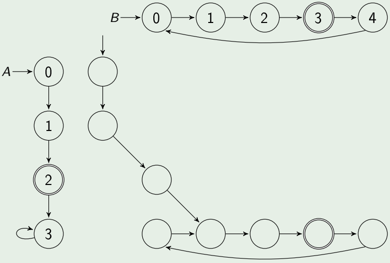
Proof

If $L(A)$ is finite

- DFA A has a final state before its sink state
- in the last row, there is a copy of B

$\Rightarrow$ $\geq n$ reachable and distinguishable states

Example

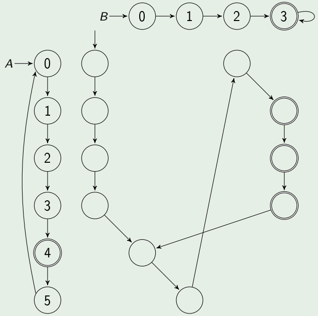


The Values from $2m$ up to $n-1$ Are Magic

Lemma

There do not exist minimal unary m- and n-state DFAs A and B such that the minimal DFA for $L(A)\,!\,L(B)$ has α states if $2m \le \alpha \le n-1$.

Example



Proof (cont.)

If $L(A)$ is infinite

- at most $m-1$ states are in the tail of $A\,!\,B$
- DFA A has only one loop
  $\Rightarrow$ at most m states are in the loop of $A\,!\,B$

$\Rightarrow$ $\le 2m-1$ reachable states

Theorem (Unary Case)

For every $m, n, \alpha \geq 1$ *such that*

1. $\alpha = 1$ *if* $m = 1$,
2. $1 \leq \alpha \leq m$ *if* $m \geq 2$ *and* $n = 1$, *or*
3. $1 \leq \alpha \leq 2m - 1$ *or* $n \leq \alpha \leq m + n - 2$ *if* $m, n \geq 2$,

there exist minimal unary m-state and n-state DFAs *A and B such that the minimal* DFA *for* $L(A) \,!\, L(B)$ *has* α *states.*

In the case of $m, n \geq 2$ *and* $2m \leq \alpha \leq n - 1$, *there do not exist minimal unary m-state and n-state* DFAs *A and B such that the minimal* DFA *for* $L(A) \,!\, L(B)$ *has* α *states.* □

Known result: Drewes, Holzer, Jakobi, van der Merwe (2017)

The state complexity of the cut operation on regular languages:

$$f(m,n) = \begin{cases} m, & \text{if } n = 1; \\ (m-1)n + m, & \text{if } n \geq 2; \end{cases}$$

with binary witnesses.

Our results: No magic numbers for cut in the general case

Let K, L be languages with $\mathrm{sc}(K) = m$ and $\mathrm{sc}(L) = n$.

Condition	Range of attainable complexities for $K\,!\,L$
$n = 1$	$[1, m]$
$n \geq 2$	$[1, (m-1)n + m]$

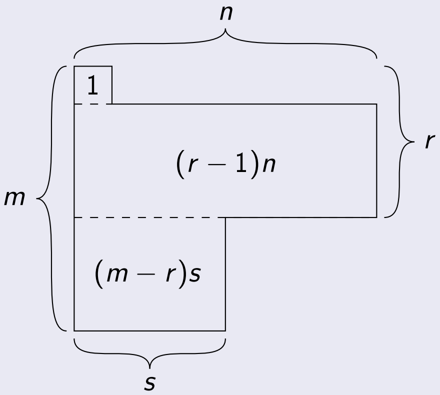
- the most interesting case is if $\alpha \in [m + n - 1, (m-1)n + 1]$

The Case $\alpha \in [m + n - 1, (m - 1)n + 1]$

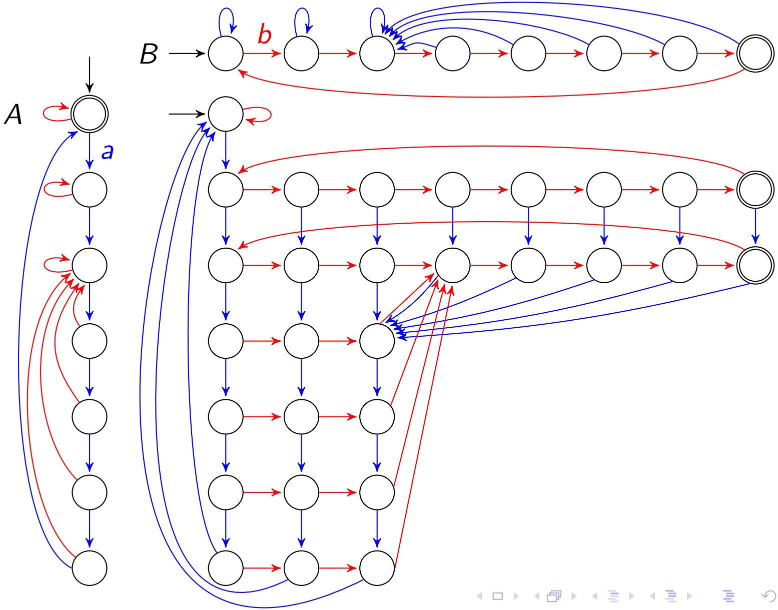
The complexities from $m + n - 1$ up to $(m - 1)n + 1$

- can be written as $\alpha = 1 + (r - 1)n + (m - r)s$ for some r, s
- or cannot

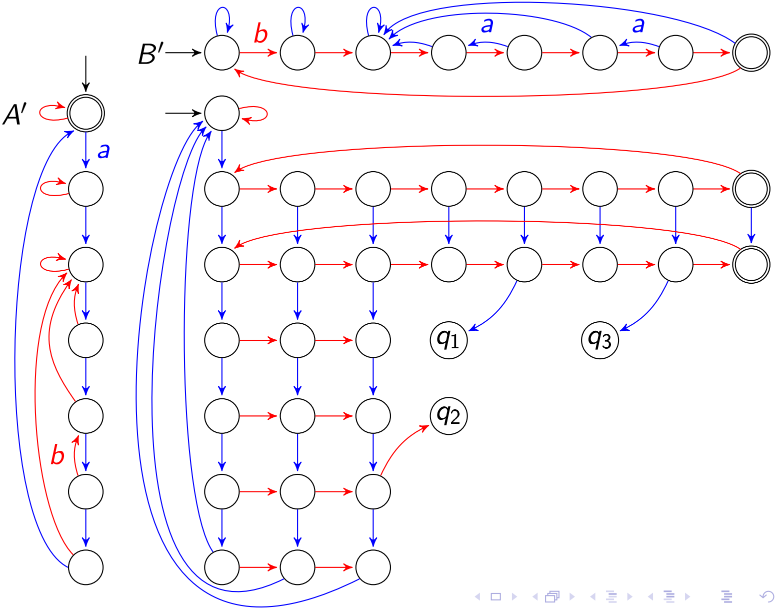
Schematic drawing of a "skeleton" for $\alpha = 1 + (r - 1)n + (m - r)s$



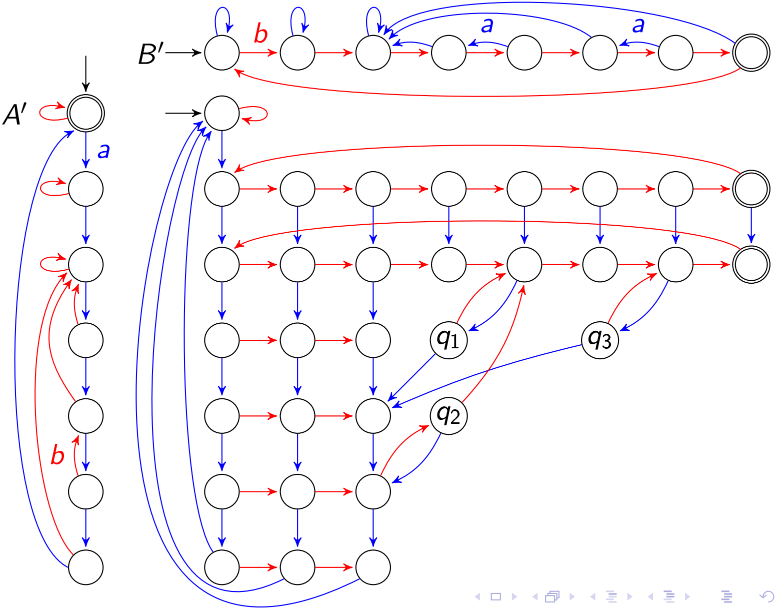
Example: $m = 7$, $n = 8$, and $\alpha = 1 + 2 * 8 + 4 * 3 = 29$



Example: $m = 7$, $n = 8$, and $\alpha = 1 + 2 * 8 + 4 * 3 \mathbf{+3} = 32$



Example: $m = 7$, $n = 8$, and $\alpha = 1 + 2 * 8 + 4 * 3 + \mathbf{3} = 32$



Magic Number Problem for Cut in the General Case

Recall that the state complexity of cut is

$$f(m,n) = \begin{cases} m, & \text{if } n = 1; \\ (m-1)n + m, & \text{if } n \geq 2. \end{cases}$$

Theorem (General Case)

For each α such that $1 \leq \alpha \leq f(m,n)$, there exist minimal binary *m-state and n-state DFAs A and B such that the minimal DFA for $L(A)\,!\,L(B)$ has α states.* □

- binary case $\Rightarrow$ every larger alphabet is solved (dummy letters)

Summary: Magic Number Problem for Cut

Unary case

Condition	Range of complexities for cut
$m = 1$	$\{1\}$
$n = 1$	$[1, m]$
$m, n \geq 2$	$[1, 2m-1] \cup [n, m+n-2]$

- if numbers from $2m$ up to $n-1$ exist, they are not attainable (are magic)
- for every number, we know whether it is or is not attainable

Binary case

Condition	Range of complexities
$n = 1$	$[1, m]$
$n \geq 2$	$[1, (m-1)n + m]$

- all numbers are attainable (not magic)

- dummy letters $\Rightarrow$ complete solution for every alphabet size
- we do not know any other operation where the magic number problem is completely solved and magic numbers exist

Thank You For Your Attention

Ďakujem za pozornosť

Danke	Arigato	Köszönöm
Dziękuję	Namaste	Grazie
Merci	Tack	Spasibo
Gràcies	Děkuji	Dank u wel
Kiitos	Gracias	Teşekkür
Xièxiè	Shokran	Obrigado

See You in Košice, Slovakia

DCFS 2019	July 17-19	deadline for submissions: April 1
CIAA 2019	July 22-25	submissions are already closed