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)

DLT Satellite Workshop
September 6, 2018, Kyoto, Japan

Basic Notions

Cut Operation

- Concatenation: $KL = \{uv \mid u \in K \text{ and } v \in L\}$
- Cut operation: machine implementation on Unix processors
  $K\,!\,L = \{\, uv \mid$
  $u \in K,\ v \in L, \text{ and } uv' \notin K \text{ for every nonempty prefix } v' \text{ of } v\,\}$

State Complexity

- State complexity of a regular language L:
  value $\mathrm{sc}(L) = \min\{n \mid L \text{ is accepted by a DFA with } n \text{ states}\}$
- State complexity of a binary operation $\circ$:
  function $(m,n) \mapsto \max\{\mathrm{sc}(K \circ L) \mid \mathrm{sc}(K) \le m \text{ and } \mathrm{sc}(L) \le n\}$
- 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 $\circ$

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): determinization of binary NFAs
- Van Zijl (2005): determinization of unary XNFAs
- Geffert (2007): determinization of unary NFAs
- Holzer et al. (2012): determinization on subregular classes
- Čevorová (2013): Kleene star on unary DFAs
- Šebej (2013): reversal on DFAs over a growing alphabet
- ...

This talk – complexity of languages resulting from the cut operation (the magic number problem for cut) – unary and binary alphabets

Range of Complexities for the Cut Operation (1)

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

The state complexity of the cut operation on unary languages:

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

Result of this paper (Theorem 6)

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 $[2m, n-1]$?

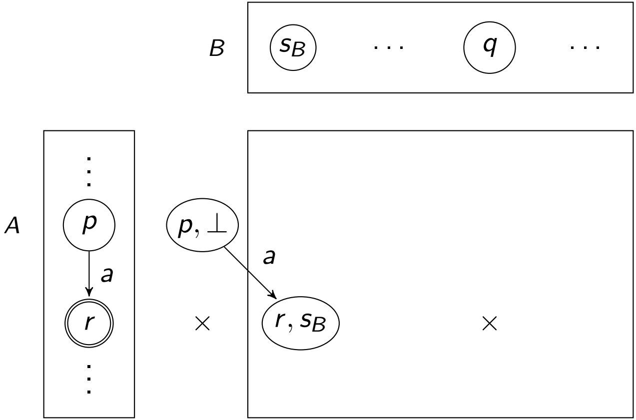


Figure: The DFAs A and B and the cut automaton $A\,!\,B$.
We have $r = \delta_A(p, a)$.
The state r is the first final state in a computation.

The Construction of the Cut Automaton

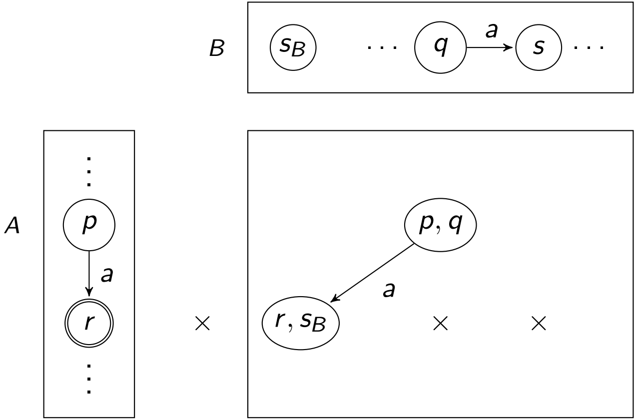


Figure: The DFAs A and B and the cut automaton $A\,!\,B$.
We have $r = \delta_A(p,a)$, $s = \delta_B(q,a)$, and the state r is final, hence no state except for (r, s_B) is reachable in the row r.

The Construction of the Cut Automaton

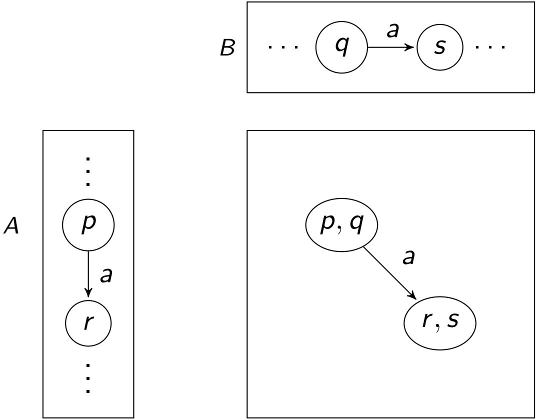


Figure: The DFAs A and B and the cut automaton $A ! B$.
We have $r = \delta_A(p, a)$, $s = \delta_B(q, a)$, and the state r is non-final.

The Construction of the Cut Automaton

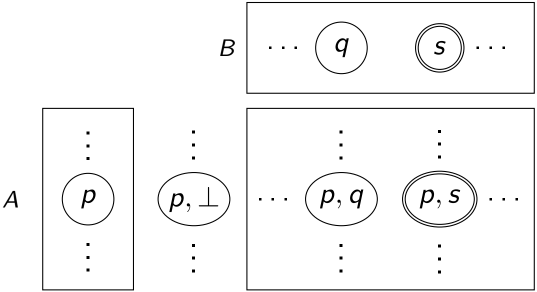
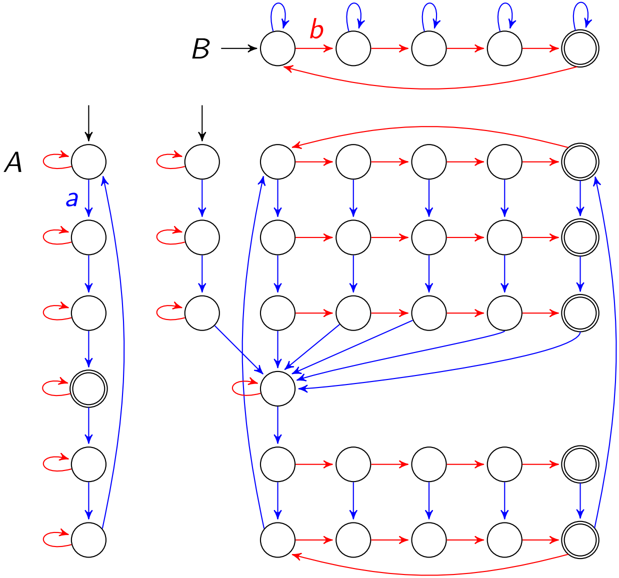


Figure: The DFAs A and B and the cut automaton $A\,!\,B$.
Since the state s is final in B, each state (p, s) is final in $A\,!\,B$.

An Example of the Cut Automaton



Lemma 5.

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 $2m \leq \alpha \leq n-1$

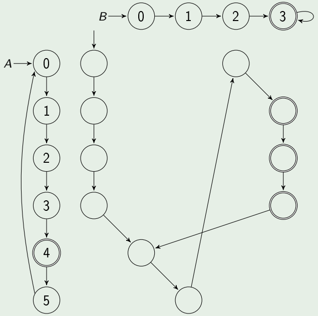
Proof.

If $L(A)$ is infinite

- at most $m-1$ states are out of the product part
- DFA A has only one loop $\Rightarrow$ at most m states are in the product part

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

Example

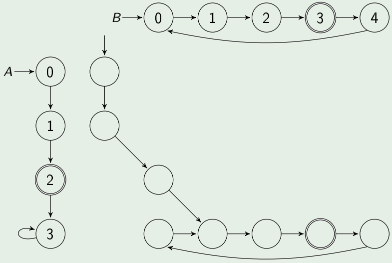


The Values from $2m$ up to $n-1$ Are Not Attainable

Lemma 5.

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 finite

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

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

Range of Complexities for the Cut Operation (2)

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

The state complexity of the cut operation on regular languages:

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

Result of this paper (Theorem 11)

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

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

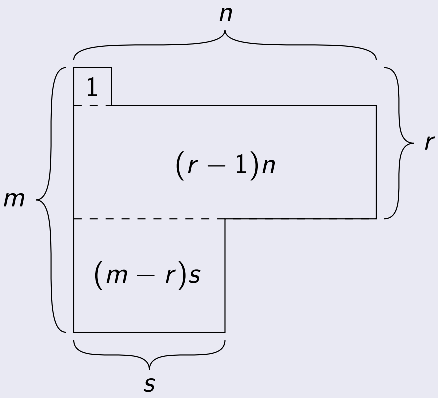
We divide the range to smaller ranges which are proven separately:
$[1, 2m-1]$, $[2m, m+n-2]$, $[m+n-1, (m-1)n+1]$,
$[(m-1)n+2, (m-1)n+m]$

Range $[m + n - 1, (m - 1)n + 1]$

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

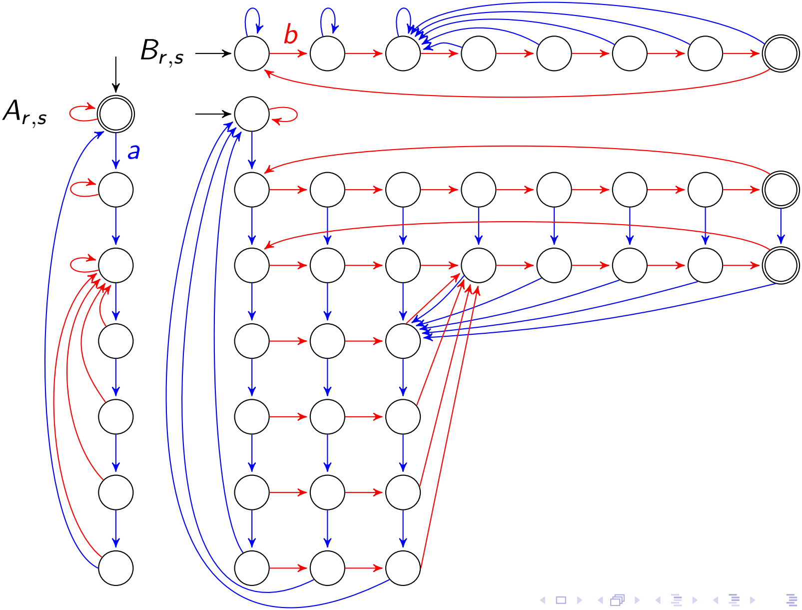
- either can be written as $1 + (r - 1)n + (m - r)s$ for some r, s
- or cannot

Schematic drawing of a "skeleton"



- We have $2 \leq r \leq m$ and $1 \leq s \leq n$
- Let $\alpha = 1 + (r - 1)n + (m - r)s$
  - If $r := r + 1$, then $\alpha := \alpha + n - s$
  - If $s := s + 1$, then $\alpha := \alpha + m - r$

Example of a Skeleton for $m = 7$, $n = 8$, $r = 3$, and $s = 3$



What For Values Between Skeletons?

Let $\alpha = 1 + (r-1)n + (m-r)s$

- If $r := r+1$, then $\alpha := \alpha + n - s$
- If $s := s+1$, then $\alpha := \alpha + m - r$

We can add $n - s$ or $m - r$ to a current number of reachable states.
To attain the values in between, we use a new variable t.

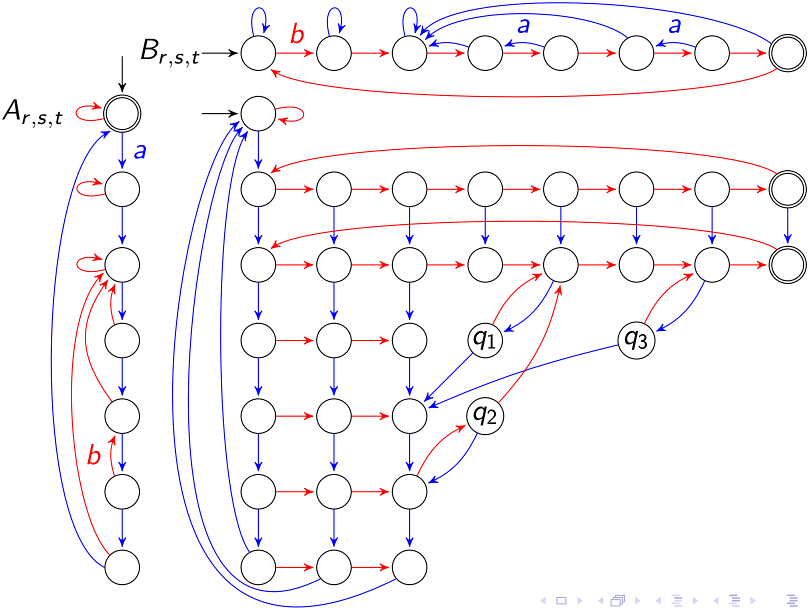
Lemma 9.

Let $2 \leq r \leq m$, $1 \leq s \leq n$, and $1 \leq t \leq \min\{n-s, m-r\} - 1$. Let $\alpha = 1 + (r-1)n + (m-r)s + t$. Then there exist a minimal binary m-state DFA $A_{r,s,t}$ and a minimal binary n-state DFA $B_{r,s,t}$ such that the minimal DFA for the language $L(A_{r,s,t})\,!\,L(B_{r,s,t})$ has exactly α states.

Proof.

We add t new reachable states ("teeth") to the skeleton $A_{r,s}\,!\,B_{r,s}$ to get α reachable states. Then, we prove distinguishability. □

Example of Teeth for $m = 7$, $n = 8$, $r = 3$, $s = 3$, and $t = 3$



Reachability of states in skeleton:

- $\rightarrow (0,0) \xrightarrow{a^i b^j} (i,j)$
- $(i,j) \xrightarrow{a} (i',j')$ where $j' \leq s-1$ if $i' \geq r$
- $(i,j) \xrightarrow{b} (i'',j'')$ where $i'' \leq r-1$ if $j'' \geq s$

Reachability of teeth:

- $(r-1, s+i) \xrightarrow{a} (r, s+i-1) = q_i$ for odd i
- $(r+i, s-1) \xrightarrow{b} (r+i-1, s) = q_i$ for even i

Distinguishability:

- $(0,0) \xrightarrow{b^*}$ *reject* but $(i,j) \xrightarrow{b^{n-1-j}}$ *accept* if $(i,j) \neq (0,0)$
- $(i,j) \xrightarrow{b^{n-1-j}}$ *accept* but $(i',j') \xrightarrow{b^{n-1-j}}$ *reject* if $j \neq j'$
- $(i,j) \xrightarrow{a^{m-i}} (0,0)$ but $(i',j') \xrightarrow{a^{m-i}} (i'',j'')$ if $i \neq i'$

Summary

Unary case

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

- if numbers from $2m$ up to $n - 1$ exist, they are not attainable (are magic)
- for every number from 1 up to $f_1(m, n)$, we know whether it is or is not attainable
  $\Rightarrow$ the problem is completely solved for unary languages

General case

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

- all numbers from 1 up to $f(m, n)$ are attainable (not magic)
  $\Rightarrow$ the problem is completely solved for binary languages
- we can duplicate letters
  $\Rightarrow$ the problem is completely solved for every alphabet size
- we do not know other operation with such result

Thank You For Your Attention

Questions?