

Subtree and Tree Pattern Pushdown Automata for Trees In Prefix Notation

JAN JANOUŠEK AND BOŘIVOJ MELICHAR

{JANOUSEJ, MELICHAR}@FEL.CVUT.CZ

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ELECTRICAL ENGINEERING
CZECH TECHNICAL UNIVERSITY
CZECH REPUBLIC

MOTIVATION

STRINGOLOGY

String suffix and factor automata.

PROPERTIES:

- 1 ACCEPT ALL OCCURENCES OF AN INPUT SUFFIX AND AN INPUT FACTOR, RESPECTIVELY, IN A TEXT OF SIZE n .
- 2 SEARCH PHASE FOR ALL OCCURENCES OF AN INPUT SUFFIX OR AN INPUT FACTOR OF SIZE m IN TIME $\mathcal{O}(m)$, AND NOT DEPENDING ON n .
- 3 ALTHOUGH THE NUMBER OF FACTORS IN THE TEXT CAN BE QUADRATIC IN n , THE TOTAL SIZE OF THE DETERMINISTIC FACTOR AUTOMATON IS LINEAR IN n .

MOTIVATION

ARBOLOGY

Subtree and tree pattern pushdown automata – ANALOGOUS TO STRING SUFFIX AND FACTOR AUTOMATA.

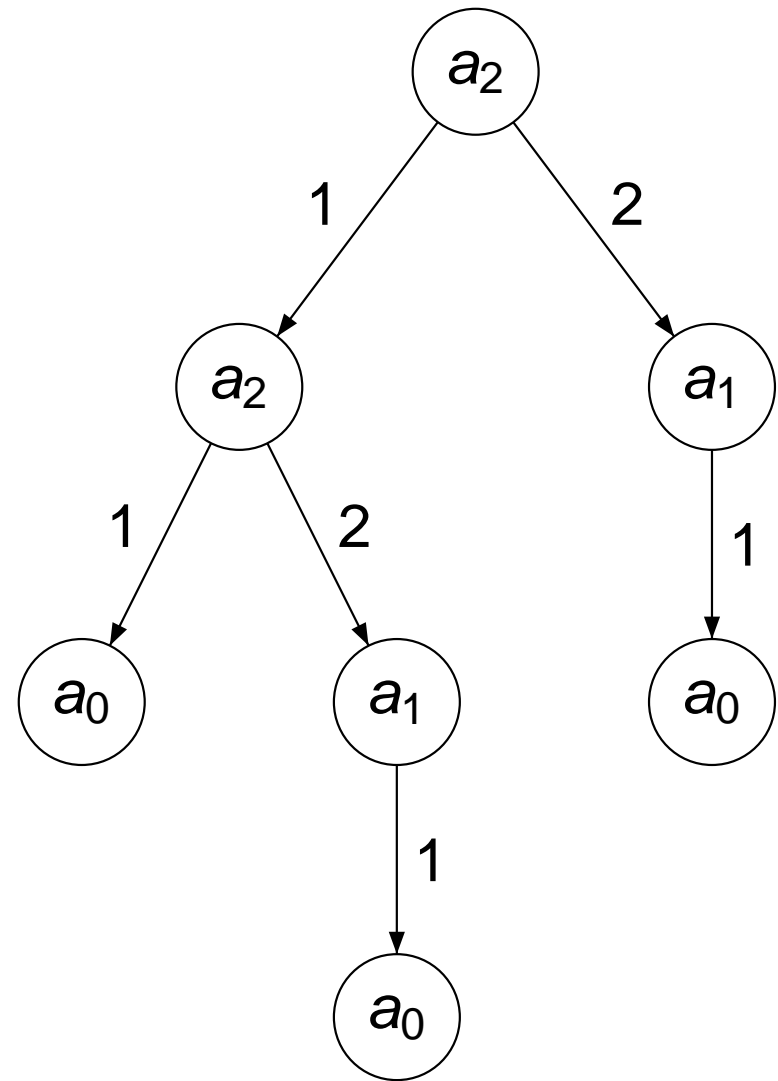
PROPERTIES:

- 1 ACCEPT ALL OCCURENCES OF AN INPUT SUBTREE AND OF SUBTREES MATCHING AN INPUT TREE PATTERN, RESPECTIVELY, IN A TREE OF SIZE n .
- 2 SEARCH PHASE FOR ALL OCCURENCES OF AN INPUT SUBTREE OR AN INPUT TREE PATTERN OF SIZE m IN TIME $\mathcal{O}(m)$, AND NOT DEPENDING ON n .
- 3 ALTHOUGH THE NUMBER OF TREE PATTERNS MATCHING THE TREE CAN BE EXPONENTIAL IN n , THE TOTAL SIZE OF THE DETERMINISTIC TREE PATTERN PUSHDOWN AUTOMATON IS LINEAR IN n .

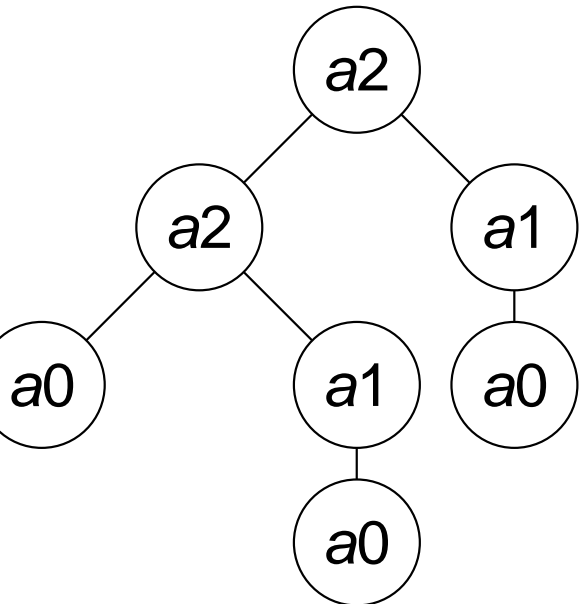
SUBTREE PDA

EXAMPLE 1

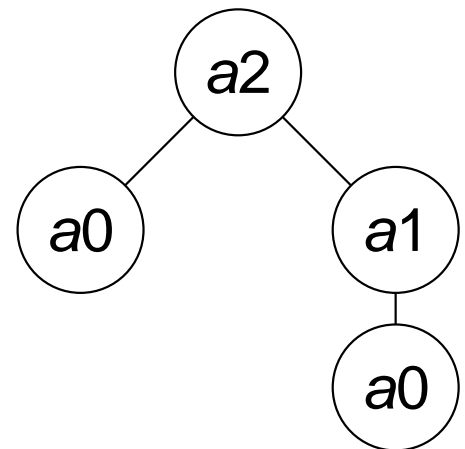
- RANKED ALPHABET
 $\mathcal{A} = \{a_2, a_1, a_0\}$
- TREE t_1
PREFIX NOTATION IS
 $pref(t_1) =$
 $a_2 a_2 a_0 a_1 a_0 a_1 a_0$
- SUBTREES OF t_1 IN
PREFIX NOTATION ARE:
 - ① $a_2 a_2 a_0 a_1 a_0 a_1 a_0$
 - ② $a_2 a_0 a_1 a_0$
 - ③ $a_1 a_0$
 - ④ a_0



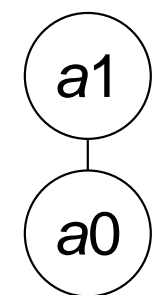
ALL SUBTREES OF TREE t_1 AND THEIR PREFIX NOTATION



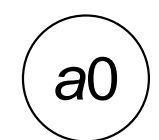
$a2\ a2\ a0\ a1\ a0\ a1\ a0$



$a2\ a0\ a1\ a0$



$a1\ a0$



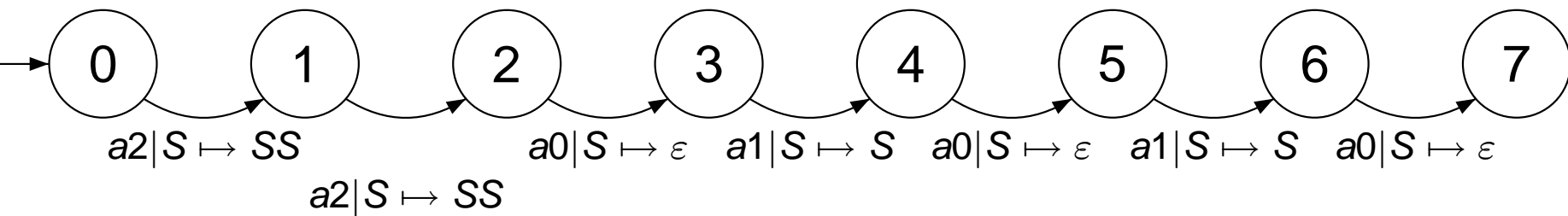
$a0$

THEOREM 1

GIVEN A TREE t AND ITS PREFIX NOTATION $pref(t)$, ALL SUBTREES OF t IN PREFIX NOTATION ARE SUBSTRINGS OF $pref(t)$.

EXAMPLE 1, CONTD.

TRANSITION DIAGRAM OF DETERMINISTIC PDA $M_p(t_1)$ ACCEPTING $pref(t_1) = a_2 a_2 a_0 a_1 a_0 a_1 a_0$ BY EMPTY PUSHDOWN STORE



INITIAL CONTENTS OF PUSHDOWN STORE IS S .

TRACE OF DETERMINISTIC PDA $M_p(t_1)$ FOR INPUT STRING $pref(t_1) = a2 a2 a0 a1 a0 a1 a0$

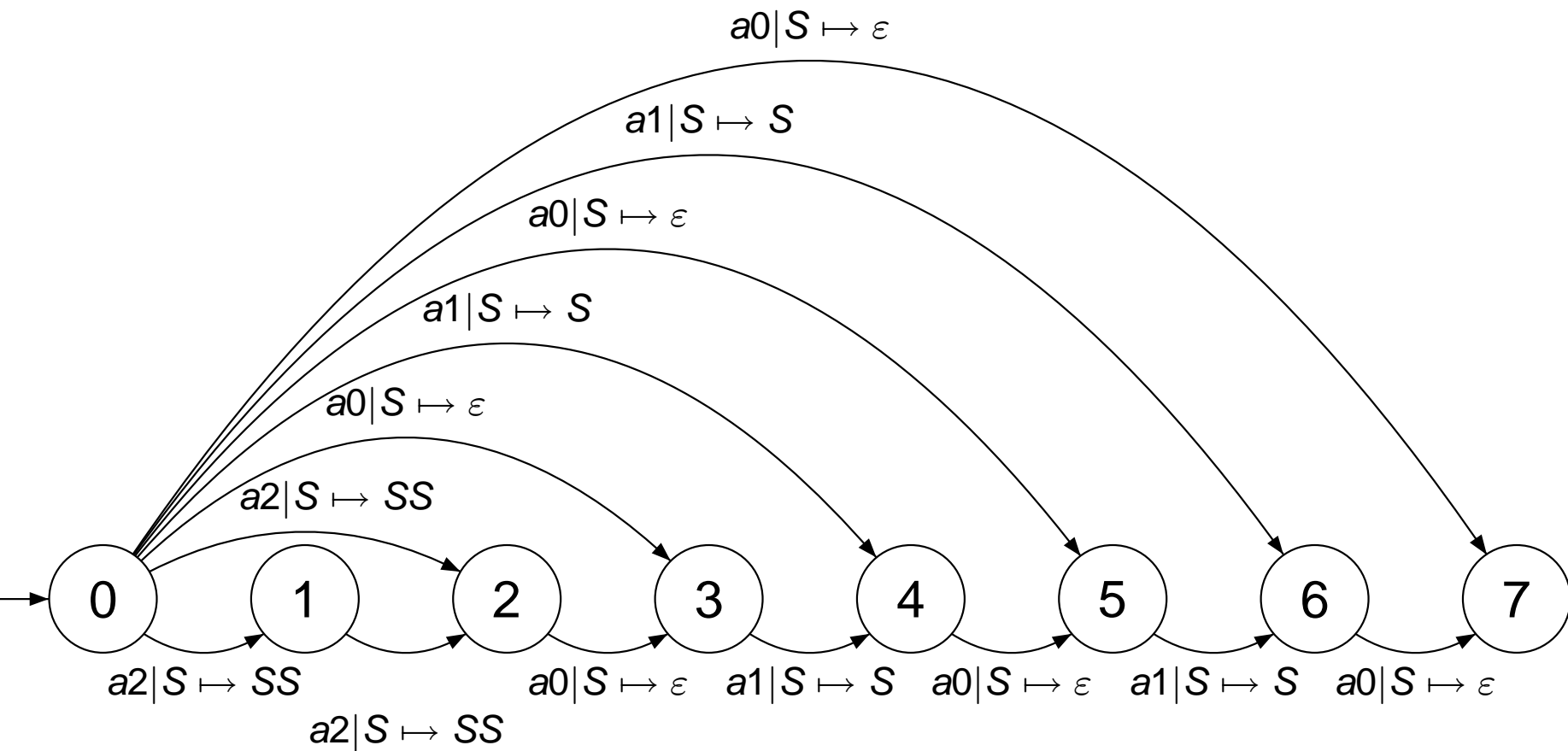
STATE	PUSHDOWN STORE	INPUT
0	S	a2 a2 a0 a1 a0 a1 a0
1	S S	a2 a0 a1 a0 a1 a0
2	S S S	a0 a1 a0 a1 a0
3	S S	a1 a0 a1 a0
4	S S	a0 a1 a0
5	S	a1 a0
6	S	a0
7	ϵ	ϵ

ACCEPT

ACCEPT BY EMPTY PUSHDOWN STORE.

NONDETERMINISTIC SUBTREE PDA $M_{nps}(t_1)$ FOR TREE t_1 IN PREFIX NOTATION

$pref(t_1) = a_2 a_2 a_0 a_1 a_0 a_1 a_0$



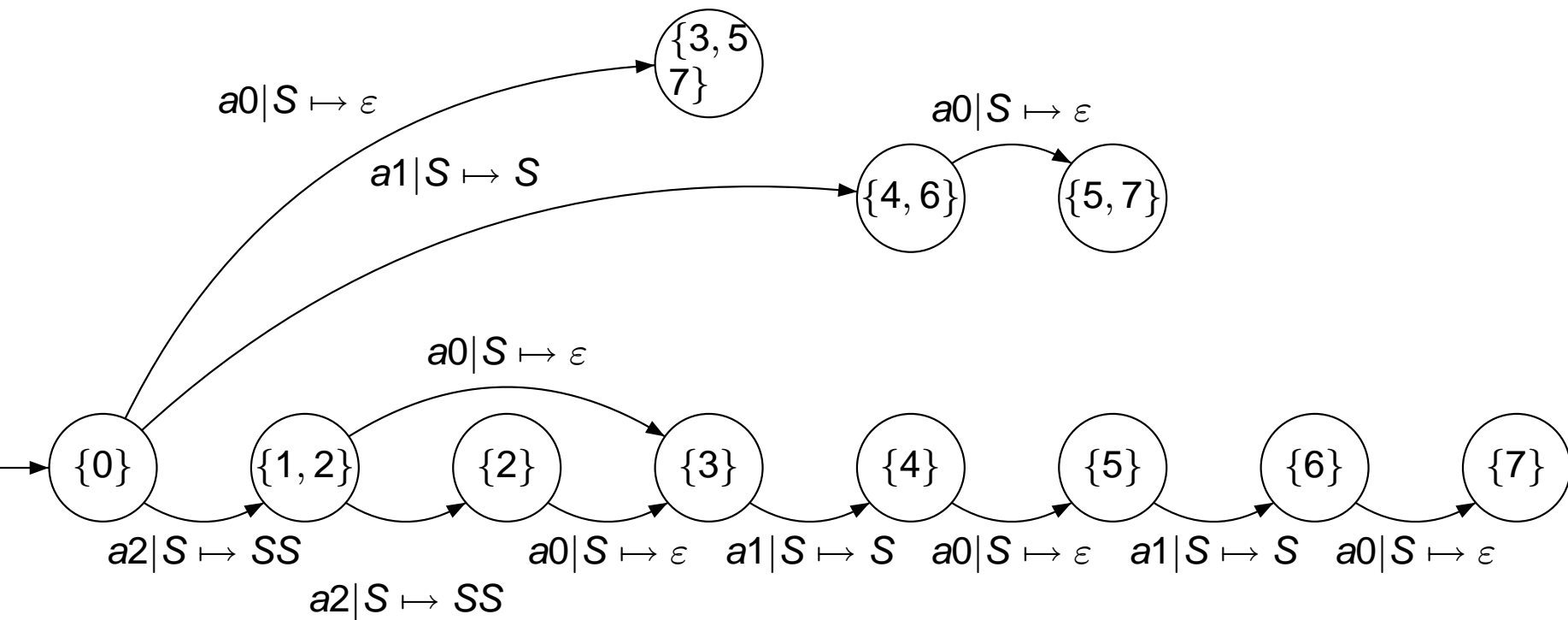
TRANSFORMATION TO DETERMINISTIC PDA

INPUT-DRIVEN PDA – PUSHDOWN STORE OPERATIONS ARE DETERMINED BY THE INPUT SYMBOL.

ANY NONDETERMINISTIC INPUT-DRIVEN PDA CAN BE DETERMINISED SIMILARLY AS IN THE CASE OF FINITE AUTOMATA – THE STATES OF THE DETERMINISTIC PDA CORRESPOND TO SUBSETS OF STATES OF THE NONDETERMINISTIC PDA (D-SUBSETS).

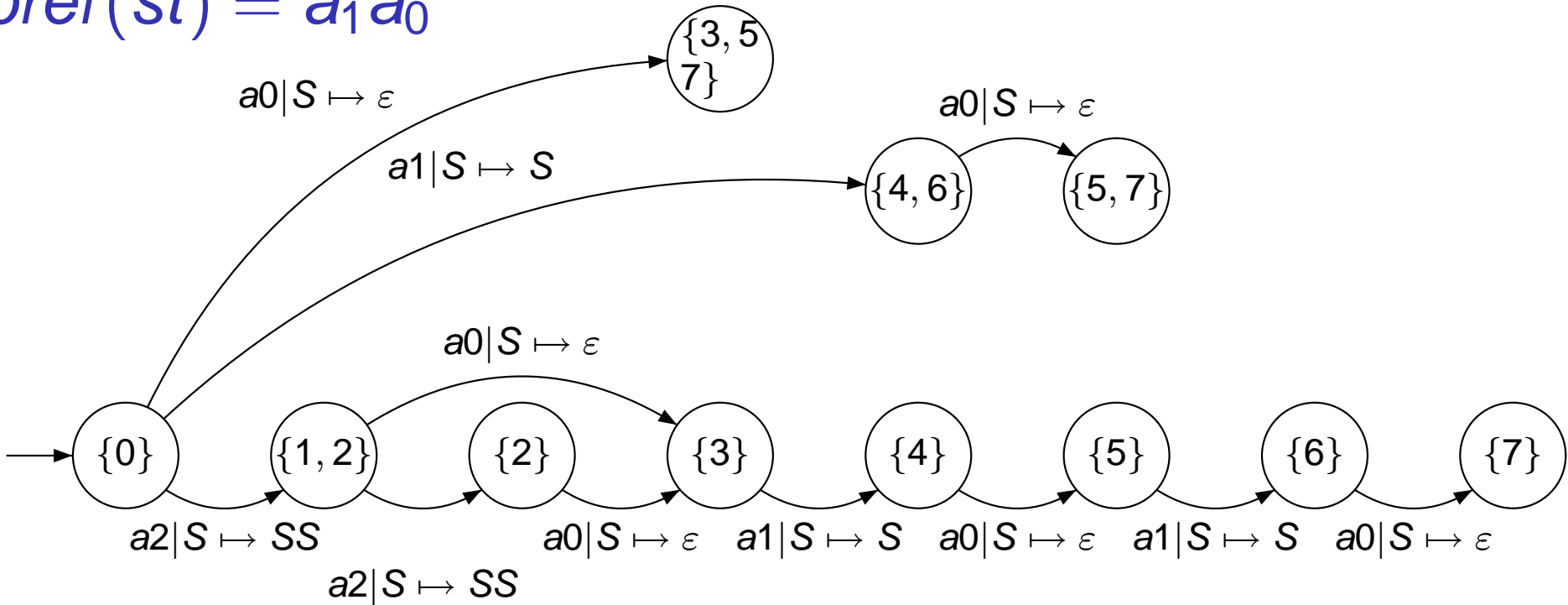
MOREOVER, NONDETERMINISTIC ACYCLIC INPUT-DRIVEN PDA – THE CONTENTS OF THE PUSHDOWN STORE CAN BE PRECOMPUTED, AND ONLY TRANSITIONS AND STATES WITH POSSIBLE PUSHDOWN OPERATIONS ARE SELECTED.

DETERMINISTIC SUBTREE PDA $M_{dps}(t_1)$ FOR TREE IN PREFIX NOTATION $pref(t_1) = a2 a2 a0 a1 a0 a1 a0$



TRACE OF DETERMINISTIC SUBTREE PDA $M_{dps}(t_1)$ FOR AN INPUT SUBTREE st IN PREFIX NOTATION

$$pref(st) = a_1 a_0$$



STATE	PDS	INPUT
{0}	S	a1 a0
{4, 6}	S	a0
{5, 7}	ϵ	ϵ
ACCEPT		

INPUT SUBTREE

a1

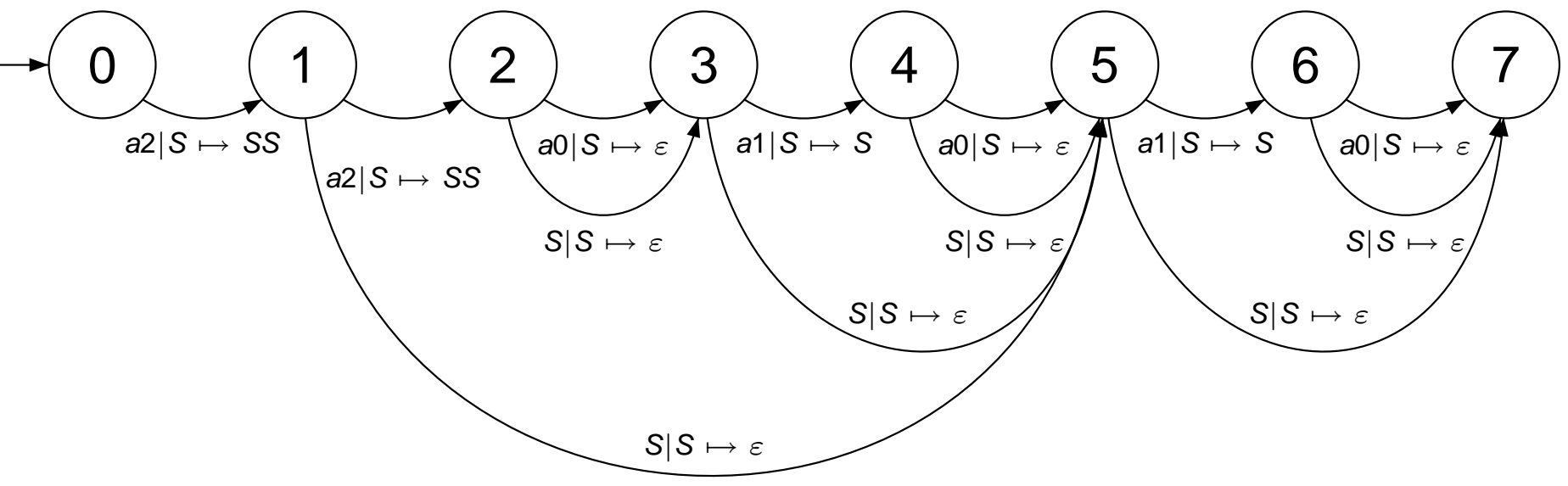
|

a0

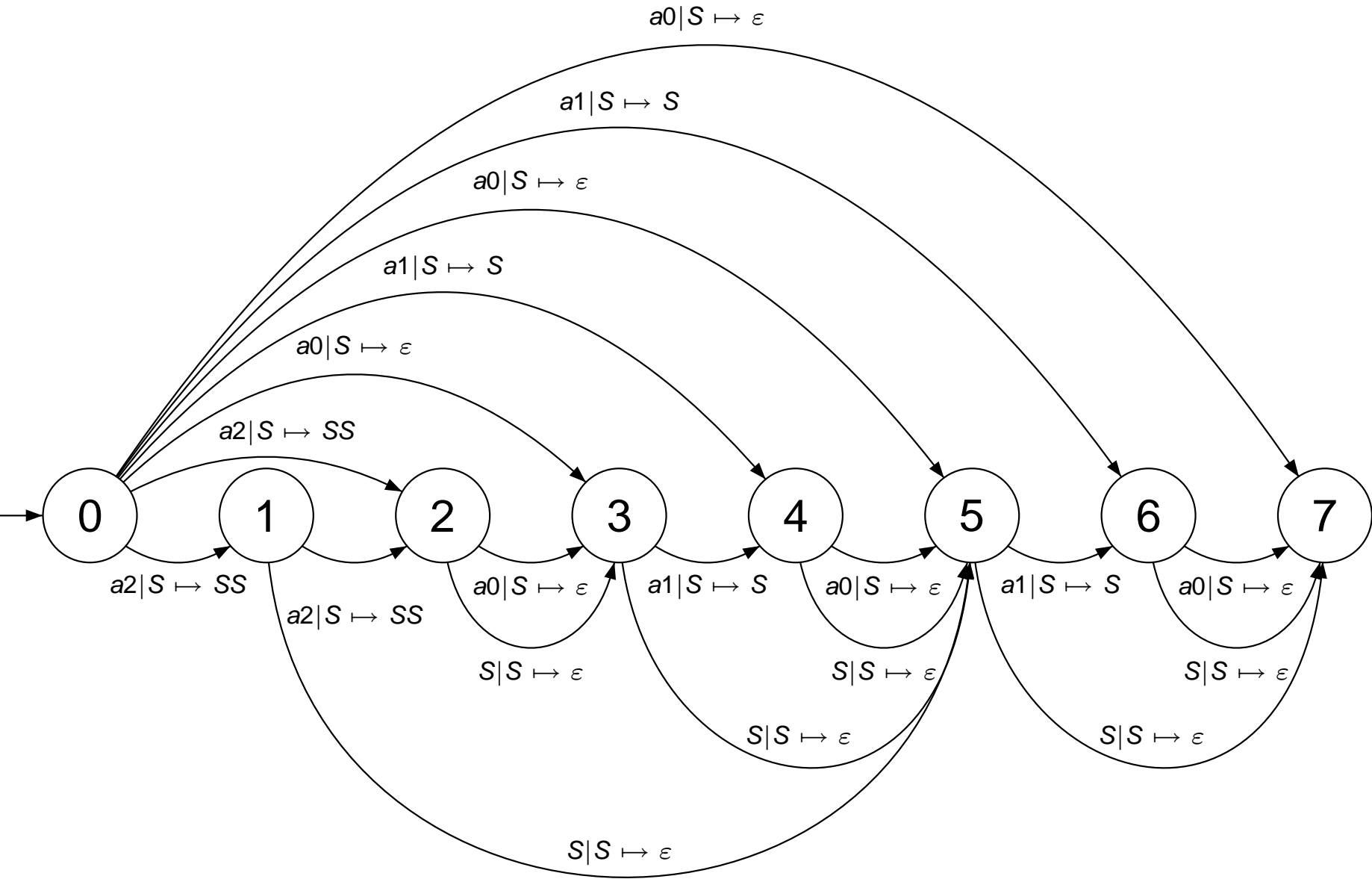
12

TREE PATTERN PDA

DETERMINISTIC TREETOP PDA $M_{pt}(t_1)$ FOR TREE t_1 IN PREFIX NOTATION $pref(t_1) = a2 a2 a0 a1 a0 a1 a0$

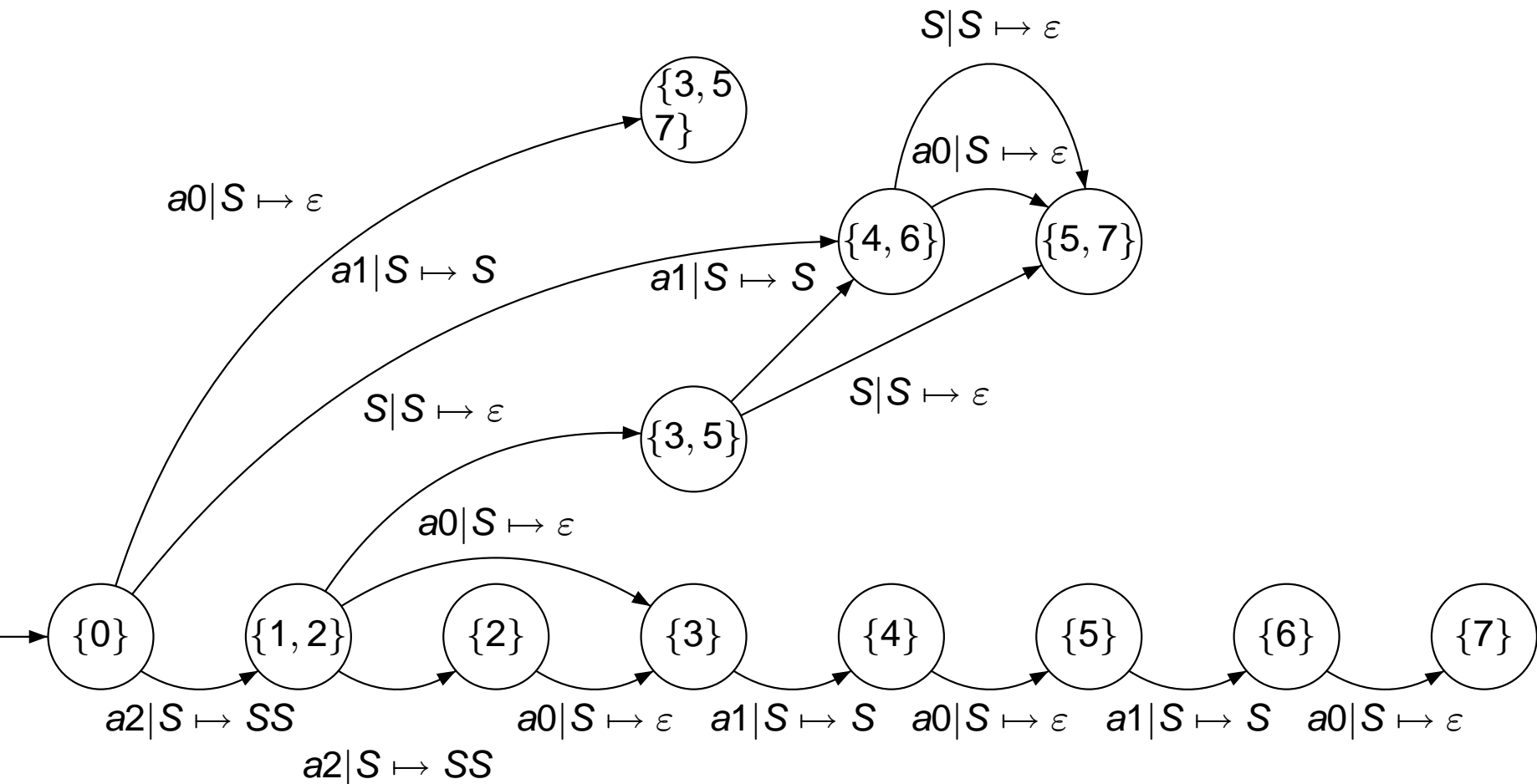


NONDETERMINISTIC TREE PATTERN PDA $M_{npg}(t_1)$ FOR $pref(t_1) = a2 a2 a0 a1 a0 a1 a0$



DETERMINISTIC TREE PATTERN PDA $M_{dpg}(t_1)$ FOR TREE t_1 IN PREFIX NOTATION

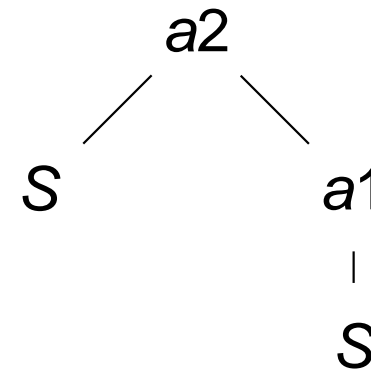
$pref(t_1) = a_2 a_2 a_0 a_1 a_0 a_1 a_0$



TRACE OF DETERMINISTIC PDA M_{dpg} FOR PREFIX NOTATION OF TREE PATTERN $a_2 S a_1 S$

STATE	PDS	INPUT
$\{0\}$	S	$a_2 S a_1 S$
$\{1, 2\}$	SS	$S a_1 S$
$\{3, 5\}$	S	$a_1 S$
$\{4, 6\}$	S	S
$\{5, 7\}$	ϵ	ϵ
ACCEPT		

INPUT TREE TEMPLATE



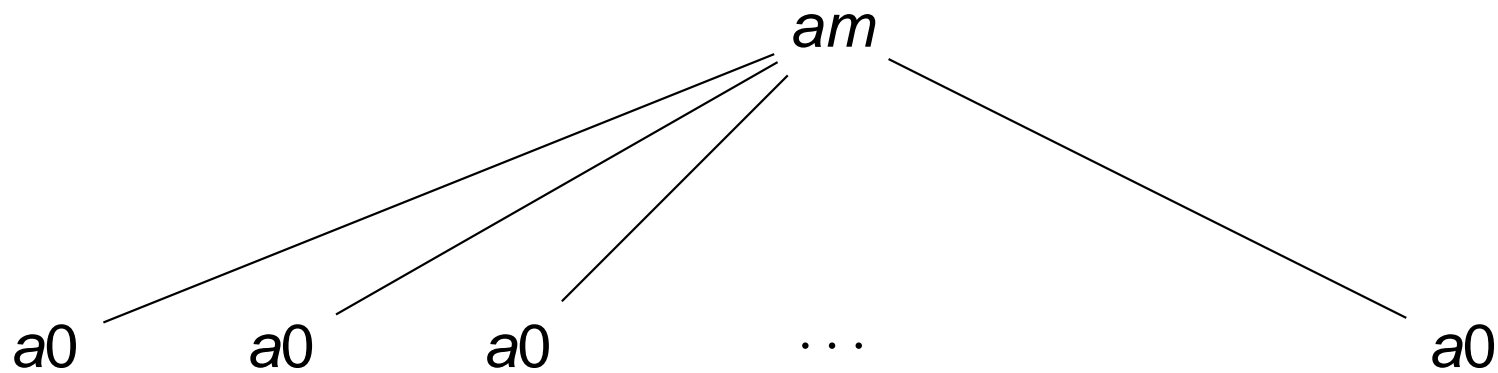
THEOREM 3

GIVEN A TREE t WITH n NODES AND ITS PREFIX NOTATION $pref(t)$, THE NUMBERS OF STATES AND TRANSITIONS OF THE DETERMINISTIC TREE PATTERN PDA $M_{dpg}(t)$ ARE $\mathcal{O}(n)$.

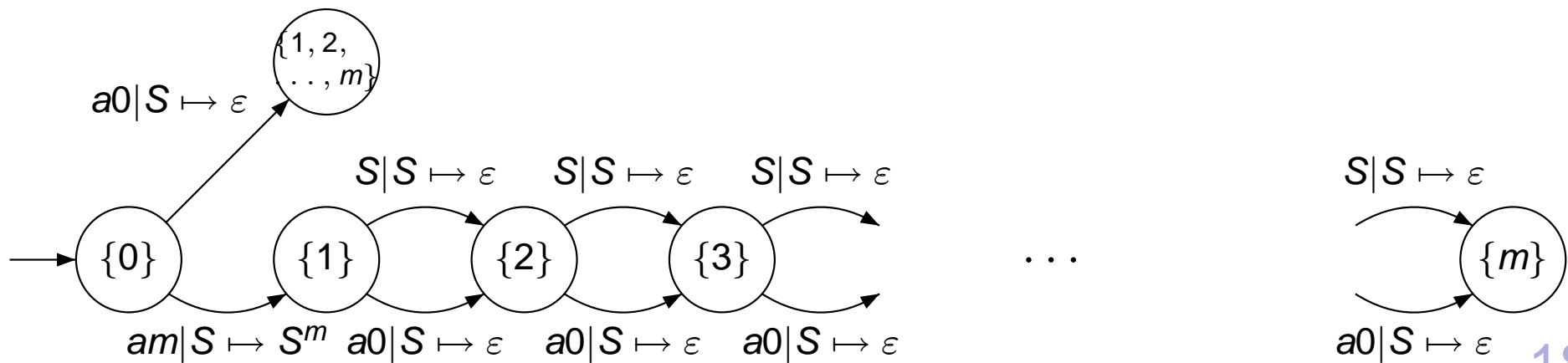
(ALTHOUGH THE NUMBER OF DISTINCT TREE TEMPLATES MATCHING THE TREE IS LESS OR EQUAL 2^{n-1})

EXAMPLE 2

TREE t_2 , $pref(t_2) = am a0^m$



DETERMINISTIC TREE PATTERN PDA FOR $pref(t_2)$:

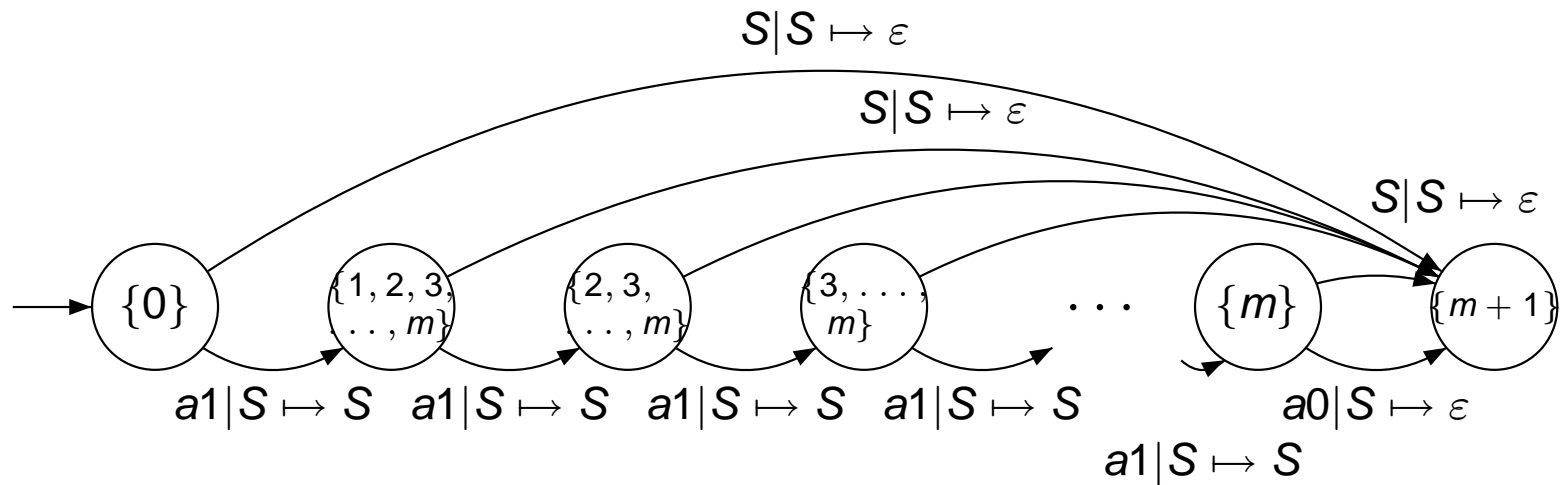


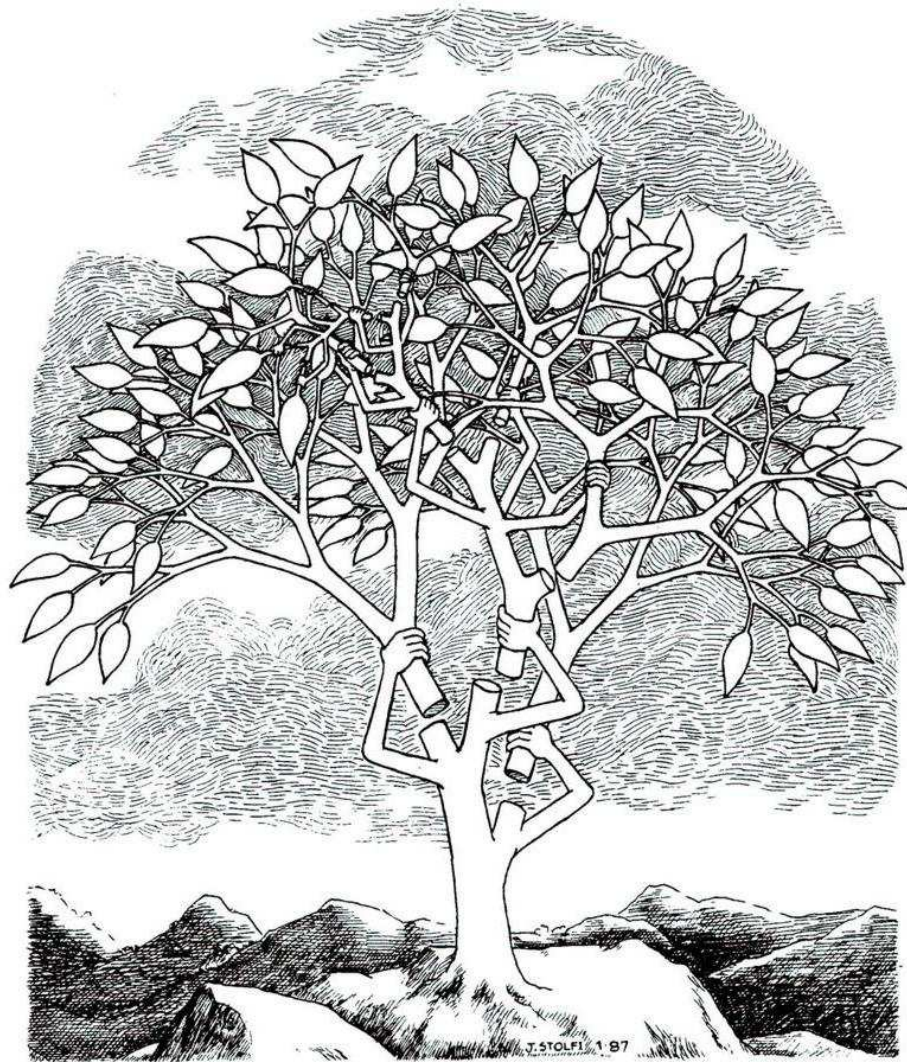
EXAMPLE 3

TREE t_3 , $pref(t_3) = a1^m a0$

$a1$
|
 $a1$
|
 $a1$
|
⋮
|
 $a1$
|
 $a0$

DETERMINISTIC TREE PATTERN PDA FOR $pref(t_3)$:





WEB PAGES <http://www.arbology.org>
<http://www.arbology.com>
COMING SOON ...

TREE LANGUAGES, TREE AUTOMATA AND DETERMINISTIC PUSHDOWN AUTOMATA

REGULAR TREE LANGUAGES ARE ACCEPTED BY **FINITE TREE AUTOMATA**.

DETERMINISTIC PUSHDOWN AUTOMATA ACCEPT A PROPER SUPERCLASS OF THE REGULAR TREE LANGUAGES IN PREFIX OR POSTFIX NOTATION.

THIS IS PROVED IN:

JANOŮŠEK, J., MELICHAR, B.: *On Regular Tree Languages and Deterministic Pushdown Automata*. SUBMITTED. 2008.

THIS PAPER CONTAINS ALSO ALGORITHM OF TRANSFORMATION OF ANY FINITE TREE AUTOMATON TO AN EQUIVALENT DETERMINISTIC PUSHDOWN AUTOMATON.