Theoretical Computer Science 646 (2016) 86–108

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Efficient algorithms for membership in boolean hierarchies of regular languages [☆]

Christian Glaßer [a,*], Heinz Schmitz [b], Victor Selivanov [c,1]

[a] Universität Würzburg, Theoretische Informatik, Am Hubland, 97074 Würzburg, Germany
[b] Hochschule Trier, Fachbereich Informatik, Schneidershof, 54293 Trier, Germany
[c] A.P. Ershov Institute, Siberian Branch of the Russian Academy of Sciences, Novosibirsk State University, Russia

ARTICLE INFO

Article history:
Received 29 October 2014
Received in revised form 19 January 2016
Accepted 11 July 2016
Available online 26 July 2016
Communicated by J. Karhumaki

Keywords:
Automata and formal languages
Computational complexity
Dot-depth hierarchy
Boolean hierarchy
Decidability
Efficient algorithms

ABSTRACT

This paper provides *efficient* algorithms that decide membership for classes of several Boolean hierarchies for which efficiency (or even decidability) were previously not known. We develop new forbidden-chain characterizations for the single levels of these hierarchies and obtain the following results:

- The classes of the Boolean hierarchy over level Σ_1 of the dot-depth hierarchy are decidable in NL (previously only the decidability was known). The same remains true if predicates mod d for fixed d are allowed.
- If modular predicates for arbitrary d are allowed, then the classes of the Boolean hierarchy over level Σ_1 are decidable.
- For the restricted case of a two-letter alphabet, the classes of the Boolean hierarchy over level Σ_2 of the Straubing–Thérien hierarchy are decidable in NL. This is the first decidability result for this hierarchy.
- The membership problems for all mentioned Boolean-hierarchy classes are logspace many-one hard for NL.
- The membership problems for quasi-aperiodic languages and for d-quasi-aperiodic languages are logspace many-one complete for PSPACE.

1. Introduction

The study of decidability and complexity questions for classes of regular languages is a central research topic in automata theory. Its importance stems from the fact that finite automata are fundamental to many branches of computer science, e.g., databases, operating systems, verification, hardware and software design.

There are many examples for decidable classes of regular languages (e.g., locally testable languages), while the decidability of other classes is still a challenging open question (e.g., generalized star-height). Moreover, among the decidable classes

[☆] A preliminary version of this paper appeared as [14] at the conference STACS 2008.
* Corresponding author.
E-mail addresses: glasser@informatik.uni-wuerzburg.de (C. Glaßer), h.schmitz@hochschule-trier.de (H. Schmitz), vseliv@iis.nsk.su (V. Selivanov).
[1] This work was done during a stay of the third author at the University of Würzburg, supported by DFG Mercator program (GL 325/5-1) and by RFBR grant 07-01-00543a.

http://dx.doi.org/10.1016/j.tcs.2016.07.017

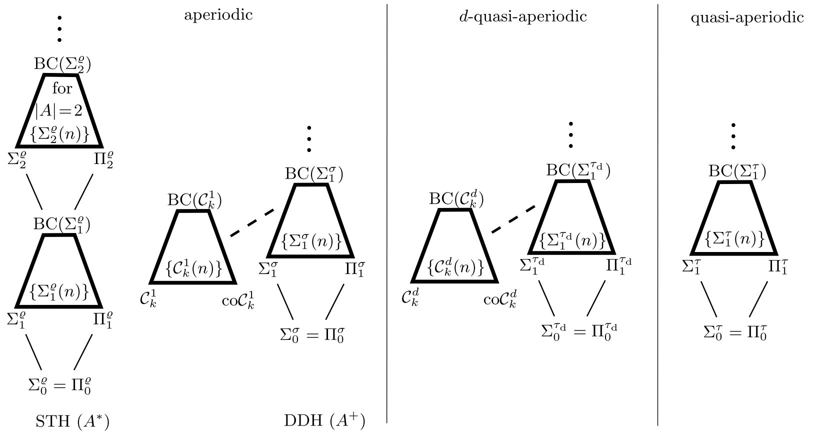

Fig. 1. Boolean hierarchies considered in this paper. Dashed lines indicate the inclusions $\mathcal{C}_0^1(n) \subseteq \cdots \subseteq \mathcal{C}_k^1(n) \subseteq \cdots \subseteq \Sigma_1^\sigma(n)$ and $\mathcal{C}_0^d(n) \subseteq \cdots \subseteq \mathcal{C}_k^d(n) \subseteq \cdots \subseteq \Sigma_1^{\tau_d}(n)$.

there is a broad range of complexity results. For some of them, e.g., the class of piecewise testable languages, efficient algorithms are known that work in nondeterministic logarithmic space (NL) and hence in polynomial time. For other classes, a membership test needs more resources, e.g., deciding the membership in the class of star-free languages is PSPACE-complete.

The purpose of this paper is to provide *efficient* algorithms that decide membership for classes of several Boolean hierarchies for which efficiency (or even decidability) were not previously known. Many of the known efficient decidability results for classes of regular languages are based on so-called forbidden-pattern characterizations. Here a language belongs to a class of regular languages if and only if its deterministic finite automaton does *not* have a certain subgraph (the forbidden pattern) in its transition graph. Usually, such a condition can be checked efficiently, e.g., in nondeterministic logarithmic space [38,7,12,13].

However, for the Boolean hierarchies considered in this paper, the design of efficient algorithms is more involved, since here no forbidden-pattern characterizations are known. More precisely, wherever decidability is known, it is obtained from a characterization of the corresponding class in terms of *forbidden* alternating chains of word extensions. Though the latter also is a forbidden property, the known characterizations are not efficiently checkable in general. (Exceptions are the special 'local' cases $\Sigma_1^\varrho(n)$ and $\mathcal{C}_k^1(n)$ where decidability in NL is known [30,29].) To overcome these difficulties, we first develop alternative forbidden-chain characterizations (they essentially ask only for certain reachability conditions in transition graphs). From our new characterizations we obtain efficient algorithms for membership tests in NL. For two of the considered Boolean hierarchies, these are the first decidable characterizations at all, i.e., for the classes $\Sigma_2^\varrho(n)$ for the alphabet $A = \{a, b\}$, and for the classes $\Sigma_1^\tau(n)$).

Definitions. We sketch the definitions of the Boolean hierarchies considered in this paper; formal definitions can be found in section 2. Σ_1^ϱ denotes the class of languages definable by first-order Σ_1-sentences over the signature $\varrho = \{\le, Q_a, \dots\}$ where for every letter $a \in A$, $Q_a(i)$ is true if and only if the letter a appears at the i-th position in the word. Σ_1^ϱ equals level 1/2 of the Straubing–Thérien hierarchy (STH for short) [40,45,41,21]. Σ_2^ϱ is the class of languages definable by similar first-order Σ_2-sentences; this class equals level 3/2 of the Straubing–Thérien hierarchy. Let σ be the signature obtained from ϱ by adding constants for the minimum and maximum positions in words and adding functions that compute the successor and the predecessor of positions. Σ_1^σ denotes the class of languages definable by first-order Σ_1-sentences of the signature σ; this class equals level 1/2 of the dot-depth hierarchy (DDH for short) [8,46]. Let τ_d be the signature obtained from σ by adding the unary predicates $P_d^0, \dots, P_d^{d-1}$ where $P_d^j(i)$ is true if and only if $i \equiv j \pmod d$. Let τ be the union of all τ_d. $\Sigma_1^{\tau_d}$ (resp., Σ_1^τ) is the class of languages definable by first-order Σ_1-sentences of the signature τ_d (resp., τ). $\mathcal{C}_k^d$ is the generalization of Σ_1^ϱ where neighborhoods of $k+1$ consecutive letters and distances modulo d are expressible (Definition 2.5). For a class $\mathcal{D}$ (in our case one of the classes Σ_1^ϱ, Σ_1^σ, $\mathcal{C}_k^d$, $\Sigma_1^{\tau_d}$, Σ_1^τ, and Σ_2^ϱ for $|A| = 2$), the *Boolean hierarchy over* $\mathcal{D}$ is the family of classes

$$\mathcal{D}(n) \stackrel{df}{=} \{L \mid L = L_1 - (L_2 - (\ldots - L_n)) \text{ where } L_1, \ldots, L_n \in \mathcal{D} \text{ and } L_1 \supseteq L_2 \supseteq \cdots \supseteq L_n\}.$$

The Boolean hierarchies considered in this paper are illustrated in Fig. 1.

Our contribution. The paper contributes to the understanding of Boolean hierarchies of regular languages in two ways:

1. For the classes $\Sigma_1^\sigma(n)$, $\Sigma_1^{\tau_d}(n)$, and $\Sigma_2^\varrho(n)$ for the alphabet $A = \{a, b\}$ we prove new characterizations in terms of forbidden alternating chains. In case of $\Sigma_2^\varrho(n)$ for the alphabet $A = \{a, b\}$, this is the first characterization of this class.