Information and Computation 241 (2015) 142–176

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Entropy of regular timed languages [☆,☆☆]

Eugene Asarin [a,*], Nicolas Basset [b,1], Aldric Degorre [a]

[a] LIAFA, University Paris Diderot and CNRS, France
[b] Department of Computer Science, University of Oxford, UK

ARTICLE INFO

Article history:
Received 17 February 2014
Available online 5 March 2015

Keywords:
Timed automata
Timed languages
Entropy

ABSTRACT

To study the size of regular timed languages, we generalize a classical approach introduced by Chomsky and Miller for discrete automata: count words having n symbols, and compute the exponential growth rate of their number (entropy). For timed automata, we replace cardinality by volume and define (volumetric) entropy similarly. It represents the average quantity of information per event in a timed word of the language. We exhibit a criterion for telling apart "thick" timed automata with non-vanishing entropy, for which typical runs are non-Zeno and discretizable, from "thin" automata for which all runs behave in a Zeno-like way, implying a quick volume collapse. We associate to every timed automaton a positive integral operator; the entropy equals the logarithm of its spectral radius. This operator has a spectral gap, thus allowing for fast converging numerical procedures to approximate entropy. In a special case, entropy is even characterized symbolically.

1. Introduction

1.1. Our main problem: size of timed languages

Since early 90s, timed automata and timed languages are extensively used for modeling and verification of real-time systems, and thoroughly explored from a theoretical standpoint. However, two important, and closely linked, aspects have never been addressed before our first related papers: quantitative analysis of the size of these languages and of the information content of timed words. In this paper, we formalize and solve these problems for deterministic timed automata.

Recall that a timed word describes a behavior of a system, taking into account delays between events. For example, $2a3.11b$ means that an event a happened 2 time units after the system start, and b happened 3.11 time units later. A timed language, which is just a set of timed words, may represent all such potential behaviors. Our aim is to measure the size of such a language. For a fixed number n of events, we can consider the language as a subset of $\Sigma^n \times \mathbb{R}^n$ (that is of finitely many copies of the space $\mathbb{R}^n$). A natural measure in this case is just Euclidean volume V_n of this subset. When the number of events is not fixed, we can still consider for each n all the timed words with n events belonging to the language and their

☆ The support of Agence Nationale de la Recherche under the project EQINOCS (ANR-11-BS02-004) is gratefully acknowledged.
☆☆ This article is partially based on conference papers [1–5].
* Corresponding author.
E-mail address: asarin@liafa.univ-paris-diderot.fr (E. Asarin).
URLs: http://www.liafa.univ-paris-diderot.fr/~asarin (E. Asarin), http://www.liafa.univ-paris-diderot.fr/~nbasset (N. Basset), http://www.liafa.univ-paris-diderot.fr/~adegorre (A. Degorre).
1 The author performed most of this work at LIGM, University Paris-Est Marne-la-Vallée and CNRS, France.

http://dx.doi.org/10.1016/j.ic.2015.03.003
0890-5401/

volume V_n. It turns out that in most cases V_n asymptotically behaves as $2^{n\mathcal{H}}$ for some constant $\mathcal{H}$ that we call entropy of the language.

We believe that exploration of entropy of timed languages is theoretically and practically appealing for several reasons.

- Symbolic dynamics approach (including the entropy-based analysis) to finite automata has made its proofs in theory of languages, theory of codes etc. Porting it to an important new class of infinite-state automata: timed ones, is a natural research problem.
- The information-theoretic meaning of $\mathcal{H}$ can be stated as follows: for a small ε, if the delays are measured with a finite precision ε, then using the words of the language L with entropy $b\mathcal{H}$ one can transmit $\mathcal{H}+\log(1/\varepsilon)$ bits of information per event.
  In [1] we formalize this idea in terms of Kolmogorov complexity. In [6] we relate $\mathcal{H}$ to the capacity estimation of a time-based information transmission channel.
- In model-checking of timed systems, it is often interesting to know the size of the set of all behaviors violating a property or of a subset of those presented as a counter-example by a verification tool. In the same context of verification, when one overapproximates a timed language L_1 by a simpler timed language L_2 (using, for example, some abstractions as in [7]), it is important to assess the quality of the approximation. Comparison of entropies of L_1 and L_2 provides such an assessment.
- Entropy analysis provides new insights into traditional topics of the theory of timed automata: Zeno behaviors, pumping lemmata, discretization etc. These insights, developed below, were also recently used by some authors in [8,9].
- Last but not least, the main technical tool in entropy analysis, a positive integral operator associated to a timed automaton, seems to be an important and useful characteristic of the automaton. We have already successfully applied it to several problems: computing generating functions of timed languages [10], finding a natural probability distribution on a given timed automaton [11], and, more surprisingly, counting sets of permutations defined by regular expressions and randomly generating their elements [12].

In this paper, we explore and solve the following problem: given a timed language accepted by a deterministic timed automaton, find the volume V_n of the set of accepted words of a given length n and the entropy $\mathcal{H}$ of the whole language.

1.2. Classical works: entropy of regular languages

Our problems and techniques are inspired by works concerning the entropy of finite-alphabet languages (cf. [13,14]). There the cardinality of the set L_n of all elements of length n of a prefix-closed regular language also behaves as $2^{n\mathcal{H}}$ for some entropy $\mathcal{H}$. The characterization of the entropy is based on the Perron–Frobenius theory for positive matrices.

Let us sketch how it works. Given a finite deterministic automaton with state set Q, let $L_n(q)$ be the language of n-letter words recognized from state q. Consider the $|Q|$-dimensional vector $\mathbf{x}_n$ whose coordinates are the cardinalities of $L_n(q)$, $q \in Q$. It is easy to see that:

$$\mathbf{x}_n = A^n \mathbf{x}_0, \tag{1}$$

where A is the adjacency matrix of the automaton. Under some additional hypotheses (strong connectedness, aperiodicity), by Perron–Frobenius theorem for positive matrices, all the components of $\mathbf{x}_n$ grow as ρ^n where $\rho = \rho(A)$ is the spectral radius of the matrix A (which coincides with its maximum eigenvalue). Hence, the entropy can be computed as follows (all the logarithms in this paper are base 2):

$$\mathcal{H} = \log \rho(A). \tag{2}$$

1.3. On techniques used

In this paper, we extend the technique sketched above to timed automata. For a deterministic timed automaton $\mathcal{A}$, we define a timed language $L_n(q, \mathbf{x})$ of all the timed words (with n events) accepted from the state $(q, \mathbf{x})$. We denote by $v_n(q, \mathbf{x})$ the volume of this language (now it is a function on $Q \times \mathbb{R}^{\mathrm{d}}$, with d the number of clocks), and we generalize Eq. (1) as follows

$$v_n = \Psi^n v_0, \tag{3}$$

but now the role of the adjacency matrix is played by some positive matrix integral operator Ψ associated to the timed automaton.

The analog of Perron–Frobenius theory for such operators is much more involved than for matrices, and we have to use some advanced functional analysis from [15,16] to explore properties of Ψ useful for study of the entropy; the most important of them being a spectral gap.