ARTICLE IN PRESS

Nano Communication Networks ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Nano Communication Networks

journal homepage: www.elsevier.com/locate/nanocomnet

Computing equivalences on model abstractions representing multiscale processes

Krishnendu Ghosh

Department of Computer and Information Technology, Miami University, 301C Mosler, 1601 University Blvd, Hamilton, OH 45011, United States

ARTICLE INFO

Article history:
Received 12 March 2015
Received in revised form 3 May 2015
Accepted 21 May 2015
Available online xxxx

Keywords:
Formal modeling
Systems biology
Model checking
Algorithms
Probabilistic modeling

ABSTRACT

Querying by temporal logic as a reasoning mechanism on a system representing multiscale processes is important in understanding the details of a large and complex system, in particular in the models of biochemical pathways. A novel formalism representing a system of multiscale biochemical pathways is described. The definitions of multiscale model in discrete domains are represented in the form of a labeled transition system. A polynomial time algorithm is constructed for identification of systems representing multiscale pathway. A probabilistic variant of the multiscale formalism is stated.

Published by Elsevier B.V.

1. Introduction

Model checking in systems biology [1] has received significant interest in recent years. The emphasis in formal modeling in biology is representation of biological processes by finite state machines (FSM) and apply temporal logics to verify interesting biological relationships. A key computational problem in model checking is state explosion [2] and often, abstractions become computational infeasible for large problems. The challenge in modeling biological systems is imprecise and incomplete information. Additionally, the abstractions of biological models become large in terms of state space to account for interconnected biological processes executing at different orders of time scales. Biological processes executing at different orders of time scales in a system are *multiscale* processes. The system representing multiscale processes is a *multiscale system*. Communication between molecular processes occurs at different time scales than in a cellular processes [3]. The need to create an integrated system for studying biological entities executing at multiple time scales, namely molecular, cellular and organic levels is essential for a detailed understanding of biology of all the processes. In this paper, we describe modeling of biological multiscale processes in a system. We construct an algorithm to identify partial ordering of identical processes in two multiscale systems. The processes are represented as *labels* on the transitions of a labeled transition system.

We motivate the construction of our formalism for a biological system representing multiscale processes.

Example 1. In the example, the processes are biochemical pathways. Consider four biochemical pathways, $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ with chemicals V, W, X, Y and Z:

$$\mathcal{A} : X \xrightarrow{\epsilon} Y.$$
$$\mathcal{B} : Y \xrightarrow{M} X + Z.$$
$$\mathcal{C} : Y \xrightarrow{N} X + V.$$
$$\mathcal{D} : Y \xrightarrow{P} X + W.$$

The notation, $C_s \xrightarrow{\alpha} C_p$ denotes a set of substrates, C_s in the presence of a catalyst (chemical), α produces a set of products, C_p. Also, $\alpha \in \{M, N, P, \epsilon\}$ where M, N, P denote catalysts and ϵ represents absence of a catalyst. Fig. 1(a)

E-mail address: ghoshk@miamioh.edu.

http://dx.doi.org/10.1016/j.nancom.2015.05.002
1878-7789/Published by Elsevier B.V.

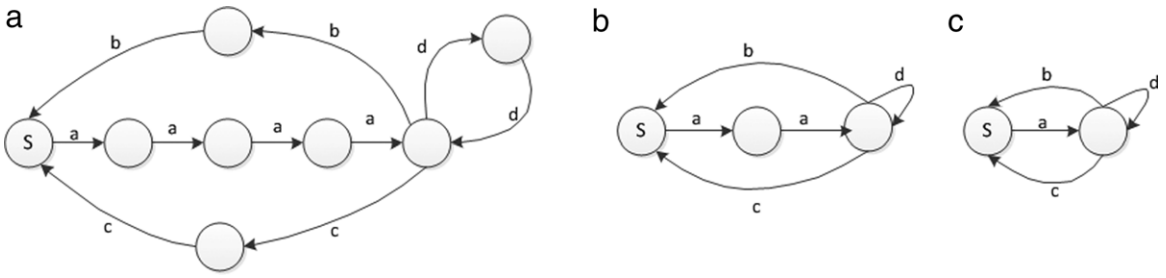

Fig. 1. Finite state machine representing identical partial ordering of pathways $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ represented by edge labels a, b, c and d, respectively. (a) System of pathways with transitions representing 0.25 mole of the substrate consumption. (b) System of pathways with transitions representing 0.5 mole of the substrate consumption. (c) System of pathways without identical successive pathways.

shows a finite state machine (FSM) representing pathways $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$. The initial state, S contains one mole of the chemicals X, M, N and P. The transition, with label a represents pathway $\mathcal{A}$ being executed consuming 0.25 moles of chemical, X. Similarly, transitions with labels b, c and d represent execution of pathways, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$, respectively. Each transition represents consumption of 0.25 mole of the substrate of each of the reaction. Pathways, $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ execute nondeterministically after completion of pathway $\mathcal{A}$ executing four times successively. Each state in the FSM stores the concentration of the chemicals formed or consumed in the form of labels on the states. Each time step is in a pathway represented by the number of moles consumed in the pathway. For simplicity, the state labels are not shown in Fig. 1. The quantities of the chemicals present in the system are stored in the states and represented by labels on the states. In Fig. 1(b), a FSM represents the system of four pathways similar to Fig. 1(a). The transitions in Fig. 1(b) represent consumption of 0.5 moles of substrates and hence, the number of transitions and states in the FSM are less. For example, the two successive transitions labeled, a in Fig. 1(a) can be collapsed to one transition in Fig. 1(b). Hence, the state between successive transitions labeled with a in Fig. 1(a) is not in Fig. 1(b). The FSMs of Fig. 1(a) and (b) represent identical partial ordering of the pathways but the state space is different. A path in a FSM is of the form $s_1, e_1, s_2, e_2 \ldots$ where $s_1, s_2, \ldots$ and $e_1, e_2, \ldots$ represent states and edges, respectively. Fig. 1(c) represents a FSM without successive pathways in any path. Therefore, for every path in the FSMs of Fig. 1(a) and (b) representing successive pathways as a single pathway there exists a path in Fig. 1(c) and vice versa. Formally, we construct a preorder relation to capture the partial ordering of the pathways represented by the edges on the FSMs in Fig. 1(a)–(c). We consider only the edge labels on the FSM to evaluate the identifiability of partial ordering of pathways. The example shows the FSM in Fig. 1(a) is detailed and has a larger state space than the FSM in Fig. 1(b). Temporal logics queries posed on FSM in Fig. 1(b) will be efficient because of smaller state space. The temporal logics queries represented by formula, φ when posed on the FSMs in the example would be identical if **X** (next state) operator is not in φ.

The example is a drastic simplification of reality because there could be more than four pathways and the exact number of moles of substrates before one or multiple different pathways initiate is not known. The concentration of the substrates is represented in the form of intervals (discretized) such that the model is tractable [4]. The consumption of moles represented by the transitions in the FSMs in Fig. 1(a) and (b) are implicitly modeling time taken by a pathway to consume the substrate of a pathway before the execution of the next pathway that uses the products of the first pathway as its substrate. We design a novel formalism that is natural and succinct for modeling multiscale processes in a system. The incomplete and imprecise information of initial concentrations of the species could lead to different multiscale processes with different state labels (though the partial ordering of the edge labels may be the same). We formulate the *problem* to identify the partial ordering on edge labels of two FSMs representing the multiscale scale systems. Identification of the processes is based on the edge labels of the FSMs only. Incomplete information of biological data is addressed by nondeterminism of the transitions from a state in the FSM. The contributions of this work: (i) Design of a nondeterministic system model representing multiscale pathways, (ii), a polynomial time algorithm that is able to identify the partial ordering (equivalences) of pathways of two multiscale systems and (iii) a probabilistic formulation of the multiscale system is stated.

A preliminary version [5] of this work was published that did not include details and the probabilistic version of the multiscale formalism.

2. Background and prior work

We review the literature on stuttering on systems, algorithms for computing equivalences on Kripke structures, asynchronous modeling and temporal logics in biological systems. These are different theories but form the basis of our work. Stuttering on systems had been mentioned by Lamport [6] for modeling concurrent programs and reasoning by temporal logics.

Definition 1 (*Kripke Structure*)**.** Given a set of propositions, AP, a *Kripke structure*, $\mathcal{K} = \langle S, S_0, E, L\rangle$ consists of

(1) S is the set of states.
(2) $S_0 \subseteq S$ is the initial set of states.
(3) $E \subseteq S \times S$ is the transition relation.
(4) $L : S \to 2^{AP}$ where L is the labeling function that labels each state with a subset from the set, AP.

Please cite this article in press as: K. Ghosh, Computing equivalences on model abstractions representing multiscale processes, Nano Communication Networks (2015), http://dx.doi.org/10.1016/j.nancom.2015.05.002