Timed Modelling of Gene Networks with Arbitrarily Precise Expression Discretization

S. Van Goethem, J.-M. Jacquet

Faculty of Computer Science, University of Namur
Namur, Belgium

L. Brim, D. Šafránek

Faculty of Informatics, Masaryk University
Brno, Czech Republic

Abstract
In this paper, a novel approach to discrete modeling of gene regulatory networks is presented. The approach is based on timed automata and is new in: (i) reflecting discrete abstraction of gene expression with arbitrary granularity, (ii) combining boolean logic with approximation of Hill kinetics. This is achieved by introducing delays that change dynamically with respect to current activity levels of regulating genes. The approach is implemented in UPPAAL and evaluated on benchmark models and on a biological case study.

1 Introduction
The central topic in systems biology is analysis of dynamics imposed by complex interaction networks. The elementary layer of these networks is formed by genes and their mutual interactions through transcriptional regulation [15]. This phenomenon is captured by gene regulatory networks (GRNs). These networks control gene expression dynamics running on a relatively slow time-scale while determining functional modes of the cell.

There are several conceptually different approaches for modelling of GRN dynamics. The most common approach is that of ordinary differential equations (ODE) describing deterministic (population average) behaviour in continuous manner. Since even a simple interaction among two genes introduces a necessary nonlinear term into the ODE of the affected gene, analytical solution of ODE models of GRNs is impossible, thus leaving simulation as the only practical method. Moreover, continuous models require quantitative knowledge in terms of kinetic coefficients, which are unknown and very difficult to measure in vitro. As a consequence, various abstraction approaches have been developed to make GRN models more convenient for analysis under data uncertainty [21,18,20,10,8].

1 The work has been partially supported by the Czech Grant Agency grant no. GAP202/11/0312.

This paper is electronically published in
Electronic Notes in Theoretical Computer Science
URL: www.elsevier.nl/locate/entcs
All of these models are purely qualitative (and discrete) provided that the aspect of time (resp. velocity of the dynamics) is entirely abstracted out, thus leading to strong approximations. However, although rates of individual processes in transcription are not known, they can be estimated synthetically to achieve behaviour observed in vitro. To this end, it appeared important to make a step back in the abstraction by extending discrete models with the quantitative aspect of time.

The most direct approach that appeared recently is that of Siebert et al. [17] further extended by Batt et al. [3]. It relies on timed automata implemented, in the former case, by employing the UPPAAL tool [4], and in the latter case, by employing IF tool suite [7]. The approach employs multi-value discrete model which is extended by time-delays representing deadlines for exiting individual discrete configurations of genes. Every discretized gene expression level in the system is statically assigned a minimal and maximal delay determining the allowed time range of exit deadline. Such a static treatment of delays is sufficient for boolean models with minimal number of levels considered in discrete abstraction of gene expression. However, to reflect the dynamics appropriately when the model precision is allowed to be set to arbitrary number of discrete levels, time-delays defined for a regulated gene should be dynamically sensitive to actual configuration (expression values) of regulating genes (regulators configuration).

In [16], a general framework for abstracting continuous systems by timed automata is introduced. Under that framework, an ODE model of dynamical system is rigorously transformed into a timed automaton representing over-approximation of continuous behaviour. However, supported non-linear ODE models must fit in the class of multi-affine systems. Since GRNs are modeled by means of higher non-linear systems based on Hill kinetics, the approach cannot be directly employed.

In a different context, another piece of work [6] closely related to our treatment presents a timed automata based approach to model signalling pathways with UPPAAL. The approach is (under)approximative in terms of maximal and minimal delay being degraded to an exact delay. A significant contribution of the underlying model is that it makes delays sensitive to the actual configuration of signal states. A technical problem of the timed model employed is spurious instability coming from non-determinism in parallel processes controlling regulations. In particular, regulation processes generate discrete events driving the regulated genes to increase or decrease expression levels provided that production competes with degradation. Each event is repetitively generated with a particular non-zero rate. For example, a positively regulated gene has a high rate of production, but also a small non-zero rate of degradation. Non-deterministic generation of increase and decrease events causes irregular gene expression profiles with oscillations over two adjacent expression levels. A detailed analysis of this problem is provided in [12] including extensions of the formalism that attempt to overcome the problem.

In this paper we present a new approach to timed modeling of GRNs that avoids problems mentioned above. In contrast to [16], our contribution lies on the computational side. In particular, we investigate construction of the timed model for Hill kinetics by using the ideas formalized in the piece-wise affine approximation. However, this paper is a preliminary step – we do not present a formal abstraction
technique, but provide an experimental evaluation of our treatment supported by a prototype tool chain. From the technical viewpoint, features of [3] and [6] are unified under our approach. In particular, we support arbitrarily precise discretization of gene expression while treating delays dynamically. Conceptually, our model is based on incorporating rate control into the original discrete model [21] according to piece-wise affine model [18]. Instability problems encountered in [6] are eliminated. Additionally, our approach allows mixing of boolean operations over regulatory interactions with algebraic summation of their effects.

Many approaches to formalizing dynamics of gene networks have been identified in the domain of hybrid systems. The main idea is based on the observation that regulation function of a gene has a sigmoidal shape and therefore can be approximated by step-functions or ramp-functions [11,13]. Overall continuous dynamics of gene expression is abstracted into a finite number of local modes through which the (hybrid) system traverses according to the approximated sigmoidal regulatory control [10,2]. Mathematically, these approaches are based on piece-wise affine or piece-wise multi-affine models. Naturally, analysis methods developed for hybrid systems can be employed [9]. However, efficient quantitative analysis of hybrid systems is computationally demanding. Lesser precise models fitting the domain of timed automata can therefore present a suitable solution meeting the current biological knowledge.

2 Background

2.1 Gene Regulatory Networks

2.1.1 Topology of gene regulatory networks

Let $G = \{g_1, ..., g_n\}$ be a set of genes, and let $P = \{p_1, ..., p_n\}$ be a set of proteins. We consider that a protein $p_i$ is a product expressed from $g_i$ by transcription. Expression of $g_i$ can depend on concentration of some proteins in $P$ which are transcription factors (TFs) regulating expression of $g_i$. TFs that enhance expression of $g_i$ are called positive regulators (activators). TFs that inhibit expression of $g_i$ are called negative regulators (inhibitors).

We formally represent a GRN by a directed graph $(G, E)$ with edges displaying regulations, $E \subseteq G \times G$. Each edge is labelled with a sign $+$, resp. $-$, indicating the regulation type (activation, resp. inhibition). Considering an edge $\langle g_i, g_j \rangle$, $g_i$ represents the regulator and $g_j$ the regulated gene. An edge $\langle g_i, g_j \rangle \in E$ is denoted $e_{ij}$, its label is denoted $\sigma_{ij}$, $\sigma_{ij} \in \{+, -\}$. An example of a GRN is depicted in Fig. 1(ab). We use the notation $P^+_j = \{p_i \in P \mid e_{ij} \in E \land \sigma_{ij} = '+'\}$ and $P^-_j = \{p_i \in P \mid e_{ij} \in E \land \sigma_{ij} = '-'\}$ for the set of all activators (resp. inhibitors) of $g_j$.

2.1.2 Discrete semantics of gene regulatory networks

For discrete (untimed) semantics of GRNs, we consider the boolean model of René Thomas [21]. In this model, the concentration of protein $p_i$ is discretely abstracted by the activity level $l_i$ with $l_i \in [0, r_i] \subset \mathbb{N}$ where $r_i$ is the maximal activity level of $p_i$. Each edge $e_{ij} \in E$ is associated a threshold $t_{ij}$ ($0 < t_{ij} \leq r_i$) indicating minimal $l_i$ such that $e_{ij}$ is active ($e_{ij}$ is active iff $l_i \geq t_{ij}$). For $r_i > 1$ it is assumed
The model based on piecewise-affine differential equations (PADE) provides a discrete abstraction with exact relation to continuous ODE models. As a base for our timed model, we use Snoussi’s model [18]. Here the concentration of a protein $p_j$ is considered continuous and denoted by $[p_j]$, $[p_j] \in [0, r_i] \subseteq \mathbb{R}$, where $r_i \in \mathbb{R}^+$ is maximal concentration. Regulation thresholds are also interpreted in $\mathbb{R}$. Addi-
tionally, for each regulation \( e_{ij} \), a rate \( \beta_{ij} \) is specified, representing contribution of regulator \( p_i \) to positive change (production) of \([p_j]\). For each protein \( p_j \), rates \( \alpha_j \) and \( \kappa_j \) are defined where \( \alpha_j \) is dilution/degradation rate of \( p_j \), and \( \kappa_j \) is basal production rate. The dynamics of a protein \( p_j \) is defined by the following differential equation:

\[
\frac{d[p_j]}{dt} = \kappa_j + \sum_{i \in \mathcal{P}^+} \beta_{ij} \cdot \theta([p_i] > t_{ij}) + \sum_{i \in \mathcal{P}^-} \beta_{ij} \cdot \theta([p_i] < t_{ij}) - \alpha_j [p_j]
\]  

where \( \theta(\cdot) \) returns 1, if the condition \( \cdot \) is true, or 0, otherwise.

In comparison to boolean model, PADE model does not abstract from time and protein concentration, the abstraction approximates continuous Hill kinetics by regulatory logic while time and its effect on concentration dynamics is preserved in the form of rates. For a given initial state, there is a unique PADE trajectory representing time behaviour of all proteins. In contrast, from the corresponding initial state in the boolean model, there are several paths providing untimed abstractions which may represent unrealistic behaviours.

### 2.1.4 Timed models

To discard unrealistic paths in the boolean model, time constraints reflecting transition rates have been introduced and implemented using timed automata [17,3]. For activity level \( l \) \((0 \leq l \leq r_i)\) of \( p_i \), two delays \( \delta_{l,i}^{\downarrow} \) and \( \delta_{l,i}^{\uparrow} \) are specified. \( \delta_{l,i}^{\downarrow} \) (resp. \( \delta_{l,i}^{\uparrow} \)) is the delay necessary for \( p_i \) to switch from \( l \) to \( l - 1 \) (resp. \( l + 1 \)). E.g., when considering the state \( \langle 1,0 \rangle \) in Fig. 1c, the decision between switching to \( \langle 2,0 \rangle \) or \( \langle 1,1 \rangle \) depends on the rate of \( p_A \) and \( p_B \) production which can be different. In this case, \( \delta_{A,1}^{\uparrow} \) is compared with \( \delta_{B,0}^{\downarrow} \) and transition with the shortest delay is chosen.

Nevertheless, the approach mentioned above still relies on strong abstractions. First, each pair of delays \( (\delta_{l,i}^{\downarrow}, \delta_{l,i}^{\uparrow}) \) is shared for all regulatory contexts of \( g_i \). Moreover, to precisely model time behaviour we also need to distinguish delays for individual regulators configurations activating a given context. E.g., for \( p_B \) in Fig. 1 and the context \( \{A\} \), we should distinguish situation \( l_A = 1 \) and \( l_A = 2 \). Both activate the regulatory context \( \{A\} \) of \( p_B \). However, the rate of \( p_B \) production will be probably faster in the latter case due to stronger influence of the activator.

Second, protein time behaviour is not exactly reflected. In particular, each level change causes the history of protein behaviour to be lost. For example, assume \( p_i \) at level \( l_i \) is associated with delays satisfying \( \delta_{l_i,i}^{\downarrow} < \delta_{l_i,i}^{\uparrow} \). Then exact production time of \( p_i \) is lost once \( \delta_{l_i,i}^{\downarrow} \) is reached and \( p_i \) level decreased. This problem is overcome in [17,3] by generalizing each delay \( \delta \) to interval \([0, \delta]\) at the price of adding many unrealistic paths including Zeno paths.

### 3 A New Timed Approach for GRNs

To overcome the problems mentioned in Section 2.1.4, we propose an approach that extends boolean model with time aspects adopted from the PADE model. In
contrast to the boolean model, our approach works with arbitrary number of arbitrarily distributed activity levels discretizing the protein concentration domain. This refinement brings the abstraction of regulatory modes closer to a continuous model, namely, the effect of an activation can be considered stronger with increasing level of the respective regulator (as modelled in Hill kinetics). Each particular configuration of regulators is assigned with an individual time delay affecting the activity of regulated gene. For a given configuration of regulators, the time delay is obtained as a value reciprocal to the corresponding rate in PADE model.

Delays are generalized to intervals of the form \([\delta^{\min}_j, \delta^{\max}_j]\) representing delay uncertainty. Moreover, for \(\delta^{\min} = 0\) delay intervals ensure conservative approximation of continuous concentration. In the continuous system, exiting the range of a discrete level from two different concentration values under contrast to the boolean model, our approach works with arbitrary number of arbitrarily distributed activity levels discretizing the protein concentration domain. This refinement brings the abstraction of regulatory modes closer to a continuous model, namely, the effect of an activation can be considered stronger with increasing level of the respective regulator (as modelled in Hill kinetics). Each particular configuration of regulators is assigned with an individual time delay affecting the activity of regulated gene. For a given configuration of regulators, the time delay is obtained as a value reciprocal to the corresponding rate in PADE model.

3.1 Formalization of Time Delays

For a GRN \((G,E)\), each gene product \(p_j\) is assigned a maximal number of activity levels \(r_j\) which is now arbitrary. Timed semantics of \(p_j\) dynamics is based on delay determined for current regulators configuration and is given as a value reciprocal to the rate \(\frac{dp_j}{dt}\) in the corresponding PADE model.

Formally, let \(\omega \in \{\text{\textquoteleft}min', \text{\textquoteleft}max'\}\}\) be a label denoting the minimal/maximal delay. For \(p_j\) at current activity level \(l_j\), delay to next change of \(p_j\) activity level, \(\delta_j^{\omega}\), is determined by the following equation (\(\text{sgn}(\delta_j^{\omega})\) indicates the direction of the expected change):

\[
\frac{1}{\delta_j^{\omega}} = \frac{1}{\delta_{k,j}^{(i)}} + \sum_{i \in P_j^+} \frac{l_i}{\delta_{k,j}^{(ii)}} + \sum_{i \in P_j^-} \frac{l_i}{\delta_{k,j}^{(ii)}} + \sum_{c \in C_j} \frac{\theta(\Theta_{\text{cop}_k})}{\delta_{j,\text{cop}_k}^{\omega}} + \sum_{k=1}^{l_j} \frac{1}{\delta_{k,\alpha_j}^{\omega}} \tag{2}
\]

where

(i) \(\delta_{k,j}^{(i)}\) is the delay implied by the (constant) basal production rate of \(p_j\), \(\delta_{k,j}^{(i)} > 0\).

(ii) \(\delta_{k,\beta_{ij}}^{(ii)}\) is the delay implied by the regulator \(p_i\) satisfying \(l_i = k\). If \(p_i \in P_j^+\) then \(\forall k \in [1,r_i] : \delta_{k,\beta_{ij}}^{(ii)} > 0\) and \(\sum_{k=1}^{r_i} \frac{1}{\delta_{k,\beta_{ij}}^{(ii)}} > 0\). If \(p_i \in P_j^-\) then \(\delta_{0,\beta_{ij}}^{(ii)} > 0\), \(\forall k \in [1,r_i] :\)
\( \delta_{k,\alpha,j} < 0. \)

(iii) effect of cooperation \( \epsilon \) is determined by delay \( \delta_{\text{cop},j}^{\epsilon} \) satisfying \( \delta_{\text{cop},j}^{\epsilon} > 0. \)

(iv) \( |\delta_{k,\alpha,j}^{\epsilon}| \) is the delay implied by the dilution/degradation of \( p_j \) at level \( l_j = k, \forall k \in [1, r_i] : \delta_{k,\alpha,j}^{\epsilon} < 0. \)

For any neglected delay in a particular configuration, the value is considered diverging (\(+\infty \) or \(-\infty \)). It is also supposed that every delay satisfies \( \delta^{\min} \leq \delta^{\max} \).

### 3.2 Implementation in Extended Timed Automata Framework

The timed model is represented compositionally in terms of UPPAAL processes (communicating timed automata) provided that there is an individual process reflecting the behaviour of each protein. For protein \( p_j \), the respective process is associated with clocks \( c_{\text{down}}^{j} \) and \( c_{\text{up}}^{j} \), controlling level decrease and increase, respectively. Clocks are employed to stopwatch the delay to the next change of \( p_j \) activity level. We employ two strategies in modeling clocks – stopwatch and non-stopwatch, both described in more detail in Section 3.2.1. Time delay is specified by the interval \( [\delta_{\text{min}}^{j}, \delta_{\text{max}}^{j}] \). Current values of \( \delta_{\text{min}}^{j}, \delta_{\text{max}}^{j} \) are atomically updated whenever some protein changes its activity level. There are two strategies for updating \( \delta_{\text{min}}^{j}, \delta_{\text{max}}^{j} \) – selecting from a precomputed array or computing delays on-the-fly. We have implemented both strategies in our UPPAAL timed automata framework.

A general process scheme is displayed in Fig. 2(left). The scheme is displayed for 3-level protein. It can be trivially generalized to \( n \) levels. A concrete instance is shown in Fig. 2(right). The process is assigned an output broadcast channel to which information about each level change of \( p_j \) is sent. For each regulator in \( P_j^+ \cup P_j^- \), the process is assigned an input channel for receiving messages notifying (potential) changes of regulators configuration. Self-regulations are managed by self-communication through local process variables.

For each activity level \( k, k \in [0, r_i] \), there is a level location \( L_k \) representing \( p_j \) at level \( l_j = k \), and an update location \( U_k \), where regulators configuration changes are processed. Level changes are represented by red (increasing level) and blue (decreasing level) transitions in Fig. 2. Under constant regulators configuration, level updates are achieved directly by transitions among level locations. When the regulatory configuration is changed, level update is obtained indirectly by switching to the respective update location (where delays are updated) and then to a particular level location. Red and blue transitions inform other processes through the output channel. After any level change, time delays \( \delta_{\text{max}}^{j}, \delta_{\text{min}}^{j} \) are updated to reflect the newly achieved regulators configuration. Therefore each such a transition is accompanied by procedure \( rA() \) that performs all necessary updates and clock resets (depending on the clock strategy). Considering update locations committed ensures delays to be updated atomically.

A level location is switched to the respective update location if a message is received on some input channel (black transitions). Effect of these transitions is immediate update of \( \delta_{\text{min}}^{j} \) and \( \delta_{\text{max}}^{j} \) according to Formula 2. In update state \( U_k \) it is decided whether the changed regulators configuration will lead to a level change or
S. Van Goethem et al.

Fig. 2. (left) Automata template for 3-level protein processes. (right) Instance for a protein $p_Z$ regulated by $p_X, p_Y$; self-transitions are ignored for the sake of lucidity. Level changing locations are denoted blue (decrease level) and red (increase level). Regulators configuration information is received on broadcast channels $X_{br}, Y_{br}$. Grey and orange transitions manage the received information and implement update of the process delay variables.

not. In the negative case, a regret transition (orange) leads back to $L_k$ causing clock manipulation managed by procedure $r()$ (depending on the clock strategy).

Update locations have self-transitions (black) that ensure reupdating of delays whenever the regulators configuration changes. Self-transitions on level locations (green) are employed only in the case of stopwatch clock strategy. Their meaning is explained in Section 3.2.1. Finally, dotted transitions treat “blind” level updates ensuring the system never exceeds the range $[0, r_j]$.

3.2.1 Time Constraints

Next we describe the mechanism of time constraints reasoning about clocks $c_{up}^j, c_{down}^j$ wrt actual delays $\delta_{\text{min}}^j, \delta_{\text{max}}^j$. Under stopwatch strategy, for each $p_j$ both clocks $c_{up}^j, c_{down}^j$ are considered as stopwatches satisfying that just one is active in any point in time provided that $c_{up}^j$ (resp. $c_{down}^j$) keeps the total time $p_j$ spent increasing (resp. decreasing) since the initial state. This strategy precisely reflects timed behaviour at the extent of limiting the analysis tasks to simulation. Under this treatment, clock manipulation procedures $r(), rA()$ ensure activation of the clock appropriate for current values of delays while freezing the second clock. In non-stopwatch strategy, the respective clocks are reset by procedures $r(), rA()$ invoked with any level update and thus measure only time from the last level change. Within this strategy timed behaviour is strongly approximated but entirely fits the basic timed automata framework which allows formal analysis, e.g., model checking. In our UPPAAL framework we have developed a model supporting the application of both strategies. Figure 3 shows the difference between the two strategies.

In any system configuration, $\delta_{\text{min}}^j, \delta_{\text{max}}^j$ satisfy either $\delta_{\text{max}}^j \geq \delta_{\text{min}}^j > 0$ (1), $\delta_{\text{min}}^j \leq \delta_{\text{max}}^j < 0$ (2), $\delta_{\text{min}}^j < 0 < \delta_{\text{max}}^j$ (3), or $\delta_{\text{min}}^j \rightarrow -\infty, \delta_{\text{max}}^j \rightarrow \infty$ (4). To detect these situations, we supply the $p_j$ process with local boolean variables $up_j, down_j$ such that $up_j$ is true iff $\infty > \delta_{\text{max}}^j > 0$ and $down_j$ is true iff $-\infty < \delta_{\text{min}}^j < 0$. Additionally,
Fig. 3. Evolution of the difference $c_{up} - c_{down}$ in the stopwatch strategy (left) compared with the corresponding approximation in the non-stopwatch strategy (right). Red and blue segments of X-axis correspond to times when just a single clock is active. Concurrent activity is denoted by black segments.

**Table 1**

Settings of state invariants and transition time constraints (left). Settings of delay variables for individual situations of delay values (right).

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Stopwatch</th>
<th>Non-stopwatch</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_k$ invar.</td>
<td>$\left( c_{upj} - c_{downj} \leq \Delta_{upj} \right) \land \left( c_{downj} - c_{upj} \leq \Delta_{downj} \right)$</td>
<td>$\left( c_{upj} \leq \Delta_{upj} \right) \land \left( c_{downj} \leq \Delta_{downj} \right)$</td>
</tr>
<tr>
<td>Red trans.</td>
<td>$c_{upj} - c_{downj} \geq \Delta_{upj}$</td>
<td>$c_{upj} \geq \Delta_{upj}$</td>
</tr>
<tr>
<td>Blue trans.</td>
<td>$c_{downj} - c_{upj} \geq \Delta_{downj}$</td>
<td>$c_{downj} \geq \Delta_{downj}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\delta_{downj}$</th>
<th>$\delta_{upj}$</th>
<th>$\delta_{minj}$</th>
<th>$\delta_{maxj}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>ff</td>
<td>tt</td>
<td>$\delta_{minj}$</td>
<td>$\delta_{maxj}$</td>
</tr>
<tr>
<td>(2)</td>
<td>tt</td>
<td>ff</td>
<td>$\delta_{minj}$</td>
<td>$\delta_{maxj}$</td>
</tr>
<tr>
<td>(3)</td>
<td>tt</td>
<td>tt</td>
<td>$\delta_{minj}$</td>
<td>$\delta_{maxj}$</td>
</tr>
<tr>
<td>(4)</td>
<td>ff</td>
<td>ff</td>
<td>$\delta_{minj}$</td>
<td>$\delta_{maxj}$</td>
</tr>
</tbody>
</table>

Constraints for situations (1), (2), (4) directly implement expected behaviour. Situation (3) is more intricate. It happens just when delay uncertainty allows both degradation and production. In non-stopwatch strategy, the behaviour is naturally modelled by non-determinism. In stopwatch case, concurrent activity of both clocks is solved by non-deterministically selecting just one of the clocks $c_{upj}, c_{downj}$ to be active. This is achieved by the effect of green self-transitions placed on level locations and guarded by condition $upj \land \downarrow_{downj}$.

Since a level change in any process may affect the current regulators configuration, a particular clock manipulation procedure is invoked in both clock strategies immediately after the respective update of delays.

### 3.3 Example

In Fig. 4, a simple gene interaction model is presented. Table in the middle gives values for degradation delays of $p_B$ and production delays guarded by the activation $e_{AB}$. Minimal and maximal delays are assumed to coincide. Actual delay $\delta_B$ is computed from the equation (according to Formula 2):

$$\frac{1}{\delta_B} = \frac{1}{\delta_{1,xy}} \Theta(l_X \geq 1) + \frac{1}{\delta_{2,xy}} \Theta(l_X \geq 2) - \frac{1}{\delta_{1,y}} \Theta(l_Y \geq 1) - \frac{1}{\delta_{2,y}} \Theta(l_Y \geq 2)$$
Table on the right shows $\delta_Y$ evaluated for each system configuration.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\alpha_{AB}$</th>
<th>$\beta_{LAB}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>1</td>
<td>-15</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>-6</td>
<td>4</td>
</tr>
</tbody>
</table>

$|r_A| = |r_B| = 3$

Fig. 4. (left) Delays for the example model. (right) Delay $\delta_B$ computed for particular configurations of $p_B$ levels (columns) and $p_A$ levels (rows).

### 4 Evaluation and Case Study

To automatize the process of translating models into UPPAAL language, we have developed a tool *model-builder* [12]. The model-builder takes as input an XML file describing the GRN together with the settings of the timed model. Output of model-builder is a UPPAAL timed automaton representing the GRN timed model. In addition to this, the tool can parse the traces returned by UPPAAL simulation/model-checking in order to visualize the respective trajectories.

In order to evaluate expressiveness of our timed model, we have conducted simulations of a two-component regulatory motif [14] producing a nontrivial oscillatory behaviour. The motif, depicted in Fig. 5, has the advantage of being very sensitive to perturbations in rate coefficients, therefore, it has been chosen as a challenging model to test our methodology.

**Fig. 5.** Two-component oscillatory motif and its continuous simulations.

In the continuous framework, the following ODEs describe this model:

$$\frac{d[p_X]}{dt} = \frac{0.6 \cdot [p_X]^{10}_{\text{self-activation}}}{4.6^{10} + [p_X]^{10}_{\text{inhibition by Y}}} + \frac{8}{1 + \frac{[p_Y]^{10}_{\text{degradation}}}{5^{10}}} - 0.1 \cdot [p_X]_{\text{degradation}}$$

$$\frac{d[p_Y]}{dt} = \frac{2 \cdot [p_X]^{4}_{\text{activation by X}}}{5^{4} + [p_X]^{4}_{\text{degradation}}} - 0.1 \cdot [p_Y]_{\text{degradation}}$$

Table 2 (first row) shows the original parameters. Respective simulated trajectories are given in Fig. 5 (presented in two variants – with and without the $X$ self-activation). In order to prepare the original continuous model for discrete approximation by a timed automaton, we first normalize (second row) the concentration domain by dividing the values by the maximal concentration value reached $(26 M \cdot s^{-1})$. Values of degradation parameters remain unchanged, because they are independent of molar volume. Since we implement our timed model with on-the-fly delay computation in UPPAAL where time is approximated in integer domain,
we upscale the time domain by a certain integer factor which will reduce the imprecision caused by integer rounding of rational rates when evaluating the equation (2). For our particular example we choose the factor 40 which implies that a single time unit (1s) in the original model is rescaled to 40s thus making the dynamics 40 times slower. This implies that rate parameters are divided by 40 (third row in Table 2).

After uniformly discretizing the normalized concentration domains of both proteins to 11 discrete levels, we proceed to transforming rate constants to delays. We consider the rate parameters of the normalized model and try to identify for each particular discrete level of $p_X, p_Y$ the delays for the next level increase/decrease.

Table 3(left) summarizes delays for our model. Note that delays are always specified to mimic the normalized rates. Upscaling of time is done implicitly by model-builder and is not included in the timed model specification.

After running model-builder and UPPAAL we have obtained simulations depicted in Fig. 6. The results qualitatively reflect simulations achieved with the continuous model as can be seen by comparing with Fig. 5. We have also considered a variant discretizing the concentration domain by 3 levels. The timed model is given by delays specified in Table 3(right). Simulation results are depicted in Fig. 7. When comparing the 3-level model with the 11-level model, it can be seen that increasing granularity of discretization naturally leads to a more precise reproduction of the original continuous model behaviour. However, when comparing the two tables in Table 3, the 3-level model requires significantly less parameters (delays) to be specified.

<table>
<thead>
<tr>
<th>k</th>
<th>$d_{pX}$</th>
<th>$d_{pY}$</th>
<th>$d_{pX}$ &amp; $d_{pY}$</th>
<th>$d_{pY}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>300</td>
<td>500</td>
<td>350</td>
</tr>
<tr>
<td>2</td>
<td>75</td>
<td>200</td>
<td>250</td>
<td>150</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>200</td>
<td>150</td>
<td>125</td>
</tr>
<tr>
<td>4</td>
<td>=</td>
<td>200</td>
<td>125</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>=</td>
<td>200</td>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>=</td>
<td>200</td>
<td>75</td>
<td>38</td>
</tr>
<tr>
<td>7</td>
<td>=</td>
<td>200</td>
<td>38</td>
<td>25</td>
</tr>
<tr>
<td>8</td>
<td>=</td>
<td>200</td>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>=</td>
<td>200</td>
<td>19</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>=</td>
<td>200</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 3
Parameters for the oscillatory pattern with 11 levels (left) and 3 levels (right).
Please note that a comprehensive set of further case studies as well as performance benchmarks results is available in [12]. As regards the performance of UPPAAL reachability analysis for both model variants, the analysis of the 3-level model took 0.14 s with 2491 states visited. For the 11-level model, the computation time was 0.73 s with 6059 states visited. Results were performed on a computer based on Intel T7200 (dual core) 2GHz CPU with 2GB RAM.

To show a biological case study, we have investigated a model of a central GRN module governing the G1/S cell cycle transition in mammalian cells [19]. The model considers a two-gene network describing interaction of the tumor suppressor protein pRB and the central transcription factor E2F1 (see Fig. 8 (left)). The model demonstrates the feature of bistability, i.e., the occurrence of two stable states characterized by different concentration of E2F1. The first stable state represents the low level of the E2F1 protein concentration. In this state, the cell stays in G1 phase. When increased enough, the concentration of E2F1 grows higher, to the level of the second stable state, which causes the cell approaches to S phase.

We reformulated the model in the timed automata framework and we employed UPPAAL model checking algorithm to identify bistability. By abstracting the original continuous model [19], we obtained PADE model depicted in Fig. 8. Domains of both proteins were naturally discretized into 3 levels by the thresholds. The considered rate coefficients as shown in [19] impose bistability of E2F1 (converging either to concentration $E2F1 = 8$ or $E2F1 = 0$ with the bistable switch at $E2F1 = 1.5$). This bistable switch is sensitive to the degradation coefficient $\alpha_{pRB}$.

We built the timed model for the considered GRN by directly transforming the rates to delays as defined in Section 3. We set the respective delays to re-
We presented a new approach to incorporate time into discrete models of regulatory networks. Employing our prototype tool chain, we transformed the timed model into UPPAAL language. In consequence, we have successfully evaluated the model on simulation and model checking experiments.
Evaluation showed that the more activity levels are distinguished, the more precise are the results. However, increasing resolution of discrete levels arises in technical difficulty of delineating proper values for time delays. The results were achieved on automata without stopwatches, the entailed approximation did not qualitatively affect the expected (approximate) results.

For future work we plan to develop a method automatizing setting of delay values wrt rates in PADE model. This can be achieved easily for exponential decays in the degradation term. However, for delays of simple and even cooperated regulations the task is non-trivial and remains to be solved. An interesting question is to employ the inverse problems method [1] to determine constraints on delays compatible wrt time-series measurements on micro-arrays. On the theoretical side, we plan to study rigorously the relation of timed automata to the approximated ODE model. The abstraction has been only informally defined in this paper.

References


   URL http://essay.utwente.nl/59184/


URL http://anna.fi.muni.cz/~xsafran1/Work/cs2bio/

URL http://dl.acm.org/citation.cfm?id=2032305.2032336


