Título do trabalho: Improving the Delayed Stochastic Simulation Algorithm

Nome do aluno: Camillo de Lellis Falcão da Silva

Nome do(s) orientador(es): Itamar Leite Oliveira

Área de Pesquisa do Mestrado: Inteligência Computacional

Email de contato do aluno: camillofalcao@ice.ufjf.br

Email de contato do(s) orientador(es): itamar.leite@ufjf.edu.br

Ano de ingresso no programa: 2012

Abstract:

In non-delayed stochastic models, while transcription is instantaneous, it takes a considerable time for an RNAp (RNA polymerase) to create an RNA (ribonucleic acid) molecule. In delayed stochastic models, transcription and translation can be represented by a single delayed reaction. Recently, there have been several algorithms for extending the SSA (stochastic simulation algorithm) for incorporate delays (DSSA - delay SSA). Although there is a performance gain of DSSA comparing to SSA for gene transcription and translation, stochastic simulation of biochemical networks has a performance that prevents its use to simulate gene regulatory network on a genome-scale. Therefore, there is a great need to improve the DSSA performance and this paper presents two proposals for that: a new dependency graph and a delayed product list.

Palavras-chave: stochastic simulation, gene regulatory networks, delayed stochastic simulation algorithm.

Improving the Delayed Stochastic Simulation Algorithm

Camillo L. Falcão da Silva, Itamar L. Oliveira

Departamento de Ciência da Computação – Universidade Federal de Juiz de Fora (UFJF)

camillofalcao@ice.ufjf.br, itamar.leite@ufjf.edu.br

Abstract. *In non-delayed stochastic models, while transcription is instantaneous, it takes a considerable time for an RNAp (RNA polymerase) to create an RNA (ribonucleic acid) molecule. In delayed stochastic models, transcription and translation can be represented by a single delayed reaction. Recently, there have been several algorithms for extending the SSA (stochastic simulation algorithm) for incorporate delays (DSSA - delay SSA). Although there is a performance gain of DSSA comparing to SSA for gene transcription and translation, stochastic simulation of biochemical networks has a performance that prevents its use to simulate gene regulatory network on a genome-scale. Therefore, there is a great need to improve the DSSA performance and this paper presents two proposals for that: a new dependency graph and a delayed product list.*

1. The Theoretical Foundation

Biochemical network simulations using the ordinary differential equations (ODE) approach do not work well when concentration of some species is extremely low. Gene expression and gene regulatory network (GRN) are typical examples where ODE system does not work properly due to stochasticity and the discrete nature of these kinds of networks and where, notoriously, the involved species concentration are low. Therefore, the stochastic simulation algorithm (SSA) is the most suitable method for simulating GRNs, since this method captures the stochasticity and discreteness of the GRN. However, the SSA can be computationally expensive for large systems. In order to accelerate the SSA, several approaches have been proposed. The most notable improvements include the use of dependency graphs (DG). Employing this graph significantly reduces the number of propensity calculations required by the SSA for most biochemical models. Another way to improve the SSA performance is to reduce the reactions number of the biochemical system [Gibson and Bruck 2000], whether we are intending simulate a GRN.

In non-delayed stochastic models, while transcription is instantaneous, it takes a considerable time for an RNAp **(**RNA polymerase**)** to create an RNA **(**ribonucleic acid**)** molecule. In delayed stochastic models, transcription and translation can be represented by a single delayed reaction [Ribeiro 2010]. Recently, there have been several algorithms for extending the SSA for incorporate delays (DSSA). One DSSA algorithm version [Barrio et. al. 2006] is presented in Algorithm 1.

2. The Problem Characterization

Although there is a performance gain of DSSA comparing to SSA algorithms for gene transcription and translation, stochastic simulation of biochemical networks has a performance that prevents its use to simulate gene regulatory network on a genome-scale.

Algorithm 1. Delay Stochastic Simulation Algorithm

Data : stoichiometry, reaction rates, initial state, simulation time, delay
Result : state dynamics
begin
[1] while t < T do
[2] generate $U1$ and $U2$ as $U(0, 1)$ random variables
[3] $a_0(X(t)) = \sum_{j=1}^{m} a_j(X(t))$
[4] $\theta = \frac{1}{a_0(X(t))} \ln(1/U_1)$
[5] select j such that
[6] $\sum_{k=1}^{j-1} a_k(X(t)) < U_2 a_0(X(t)) \leq \sum_{k=1}^{j} a_k(X(t))$
[7] If delayed reactions are scheduled within $(t, t + \theta]$ then let k be
[8] the delayed reaction scheduled next at time $t + \tau$
[9] $X(t + \tau) = X(t) + v_k$
[10] $t = t + \tau$
[11] else
[12] if j is not a delayed reaction then
[13] $X(t + \theta) = X(t) + v_j$
[14] else
[15] record time, $t + \theta + \tau$, for delayed reaction j
[16] $t = t + \theta$
end

The dependency graph for SSA proposed by Gibson and Bruck (2000) is useful in DSSA, but it is not optimized for delayed reactions, thus it is recalculated more propensities than are needed.

In DSSA, the delayed reactions are scheduled in a product pendency list represented by a data structure called minimum heap (for handling the statements in line 7 and 8 in algorithm 1). Although the heap has a good performance, we propose in this work a new method for improving the performance of the delayed product list (step 7 and 8 in algorithm 1) so that product recovery from the list can be faster than when using minimum heap. Thus, there is a great need to improve the performance of DSSA and this paper presents two proposals for that: a new dependency graph and a delayed product list.

3. The Contribution Characterization

3.1. Delay Dependency Graph - DDG

For the first contribution, a new dependency graph called delay dependency graph (DDG) is proposed. It replaces the dependency graph (DG) for delayed reactions.

The reaction occurring is the event that triggers the propensities updating in SSA using DG. By using DDG, such event is the change in molecules number of the species involved in the selected reaction at the current time. In multi-delayed reactions system, with incorporated DG, propensities updating should be performed several times – at the time that the reaction occurs and at all other times when a delayed reaction product should be changed. Let A be a chemical species whose molecules number has been modified by DSSA using an incorporated DDG. Thus, the DSSA will calculate only the propensities for those reactions whose the species A is a reactant.

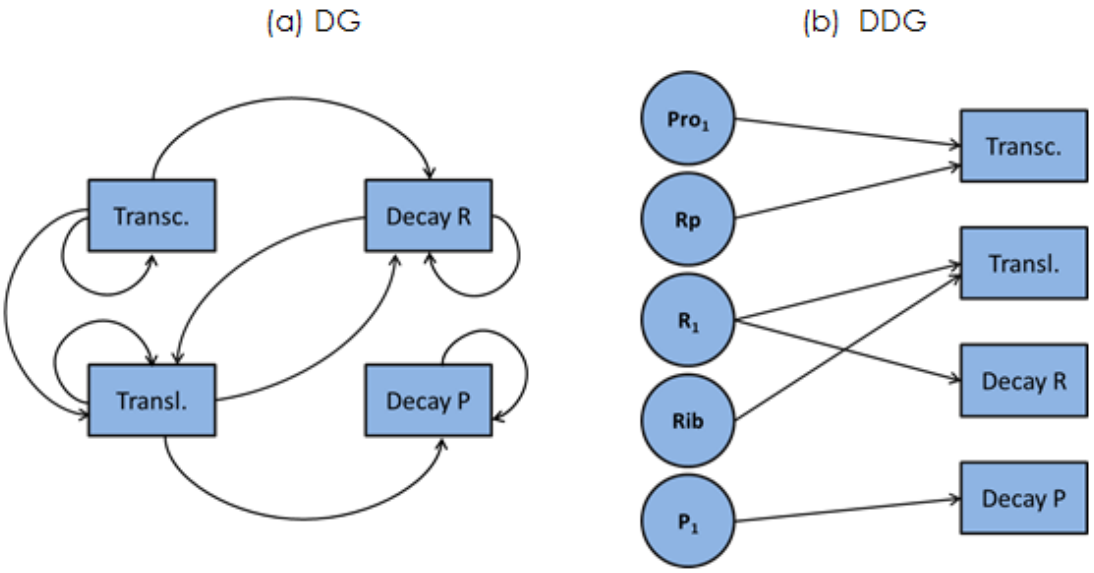

Figure 1. Examples of: dependency graphs – DG (a) and delay dependency graph – DDG (b). Both graphs are for the Toggle Switch model

3.2. Delayed Products List

The second contribution is a new data structure for controlling the delayed-products, called delayed products list (DPL). This structure can be a circular list, where each node represents a time interval. The total number of nodes of DPL is the sum of one and the ratio of maximum delay by the fixed time interval. This time interval of DPL is defined by the user in each problem. During the simulation time, the algorithm controls the DPL index by adding to the control variable the new time step. Then the program controls the time like a traditional analogical clock. In most cases the time does not force the DPL to move for the next node due to the fact that the time interval is usually large. Considering the delays distribution that we have founded in practice, it is likely that few conflicts in each DPL node occur. The advantage of using DPL instead the traditional minimum heap, is that in the DPL do not exists the overhead of extracting its top node. The Figure 2 presents the delayed product list proposed in this work.

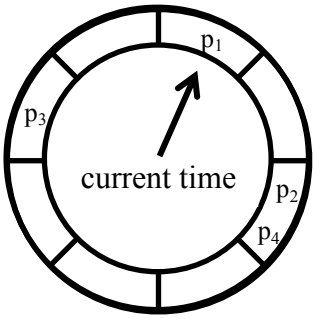

Figure 2. Delayed product list

4. The Current State of Work

4.1. Delay Dependency Graph

We presented in the X-Meeting 2012 Congress the poster: A New Dependency Graph for Stochastic Simulation Algorithm With Delay. The Toggle Switch model is a gene regulatory network that we used for testing both methods: DG-DSSA and DDG-DSSA. This model is shown in figure 3 and comprises N genes. Each one is transcribed and translated by using the first two delayed reactions, see Figure 3.

$$Pro_i + RP \xrightarrow{k_t} Pro_i(d_2) + Rp(d_2) + R_i(d_1)$$

$$R_i + Rib \xrightarrow{k_{tr}} R_i(d_3) + Rib(d_4) + P_i(d_5)$$

$$R_i \xrightarrow{d_{RBS}} \emptyset$$

$$P_i \xrightarrow{k_d} \emptyset$$

Figure 3. The Toggle Switch model

4.1.1. Results

We carried out 30 samples of execution time (in seconds) for the Toggle Switch model (Figure 3) using both algorithms. The difference between the run times of each algorithm was measured. Taking into account a model with 40 reactions and a simulation time interval of 100,000 seconds, the mean time that DDG-DSSA concludes a simulation for the proposed model is 43 seconds. On the other hand, the DG-DSSA concludes a simulation at 73 seconds. The sample deviation mean was calculated, since its value is only an estimate of the true population mean deviation, we also calculate the interval of the deviation population mean by using the Z distribution with a 99% of confidence level. This interval (-30.0674, -29.666) shows us that the DDG-DSSA has a better time performance than DG-DSSA.

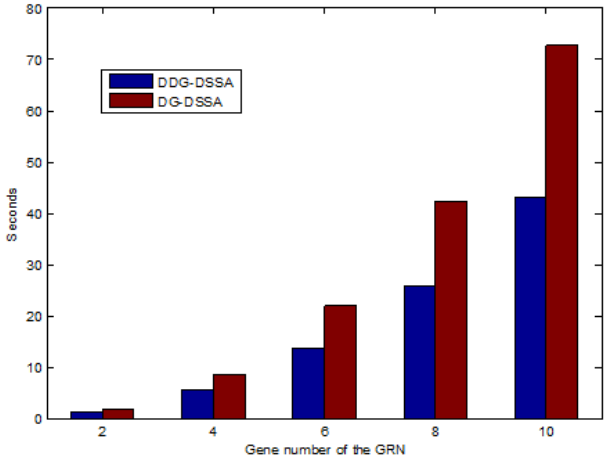

Figure 4. Performance comparison between DDG-DSSA and DG-DSSA

4.1.2. Conclusion

Through a GRN model composed of many delayed reactions, we carried out several comparisons of a DSSA with either DG or DDG. In all simulations the DDG proved more efficient than DG. In all simulation tests that we have done, the DDG-DSSA is faster than DG-DSSA about 40%. These results were obtained taking account a model composed of many reactions with delay. In models with few delayed reactions the results are very close between both methods. The performance time ratio between both methods increases non-linearly as the genes number of GRN increases (see Figure 4).

4.2. Delayed Product List - DPL

The DPL algorithm is under construction. We intend to submit DCPL, together with the DDG, to a journal in the next year.

5. Related Work

Gibson and Bruck (2000) proposed a dependency graph for propensity recalculation optimization. Roussel and Zhu (2006) and Barrio et. al. (2006) are pioneers in delayed stochastic simulations of gene regulatory networks. We not found papers introducing a minimum heap to control the scheduled of delayed reactions, but we think that this is a natural choice.

References

Gibson, M. A. and Bruck, J. (2000) "Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels", In: J. Phys. Chem.

Ribeiro, A. (2006) "Stochastic and delayed stochastic models of gene expression and regulation", In: Mathematical Biosciences.

Roussel, M. R. and Zhu, R. (2006) "Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression", In: Physical Biology.

Barrio et. al. (2006) "Oscillatory Regulation of Hes1: Discrete Stochastic Delay Modeling and Simulation", In: Computational Biology.