SABRE: A Tool for Stochastic Analysis of Biochemical Reaction Networks

Frederic Didier
School of Computer and Communication Sciences, EPFL, Switzerland

Thomas A. Henzinger
IST Austria
Institute of Science and Technology Austria

Maria Mateescu
School of Computer and Communication Sciences, EPFL, Switzerland

Verena Wolf
Department of Computer Science, Saarland University, Saarbrücken, Germany

***Abstract*—The importance of stochasticity within biological systems has been shown repeatedly during the last years and has raised the need for efficient stochastic tools. We present SABRE, a tool for stochastic analysis of biochemical reaction networks. SABRE implements fast adaptive uniformization (FAU), a direct numerical approximation algorithm for computing transient solutions of biochemical reaction networks. In addition to the stochastic analysis, SABRE may also conduct deterministic analysis.**

I. Introduction

Markov chains are an omnipresent modeling approach in the applied sciences. Often, they describe *population processes*, that is, they operate on a multidimensional discrete state space, where each dimension of a state represents the number of individuals of a certain type. Depending on the application area, "individuals" may be customers in a queuing network, molecules in a chemically reacting volume, servers in a computer network, actual individuals in a population, etc. Here, we are particularly interested in dynamical models of biochemical reaction networks, such as signaling pathways, gene expression networks, and metabolic networks.

Within the setting of stochastic analysis, biochemical reaction networks are modeled as discrete-state continuous-time Markov processes (CTMCs) according to Gillespie's theory of stochastic chemical kinetics [3]. The transient evolution of a CTMC is given by a system of linear ordinary differential equations, known as the *chemical master equation* (CME). The solution of the CME is then used to derive certain measures of interest such as expected populations and event probabilities.

In the case of discrete-time Markov chains (DTMCs), the stochastic analysis gives the probability distribution over all states after k discrete steps. For instance, in the case that the populations are numbers of molecules of different type, it is the probability distribution after k chemical reactions.

SABRE differs from other tools such as PRSIM [5], for instance, in that it focuses on the analysis of large and well-structured Markov population models [4], i.e., models that describe the interactions of different objects of certain types. Typically, the transitions rates in a Markov population model are state-dependent. SABRE can even be used for the numerical approximation of infinite-state models with possibly unbounded transition rates.

This research has been partially funded by the Swiss National Science Foundation under grant 205321-111840 and by the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University.

II. Tool

SABRE is a tool for the transient analysis, that is, SABRE computes the behaviour of the input model until a given time horizon t. SABRE may execute either a stochastic analysis or a deterministic analysis of the input system; in the stochastic case the state of the system at time t is given by a probability distribution over the discrete states of the system, while in the deterministic case the analysis is done over a continuous state space, and its result is a single state of this continuous space. The result of the deterministic analysis, also known as mean field analysis, is an approximation of the expectation of the stochastic analysis. Each of the two analysis (stochastic and deterministic) may be applied for each of the two semantics (CTMC and DTMC).

For stochastic analysis, SABRE implements four algorithms: fast standard uniformization, fast adaptive uniformization[6], explicit Runge-Kutta fourth order method where the dynamics are given by a differential equation, and Gillespie simulation[**?**]. The different configurations in which SABRE may operate are depicted in Figure 1. The focus of the tool is on the fast adaptive uniformization (FAU) method, while the remaining methods are given for completeness and comparison (in some situation, these other methods may be more useful that FAU, e.g. simulation based method offer a better way to compromise when the state space is beyond the limits of FAU).

Fast adaptive uniformization is a variant of the uniformization method[6] which is, an efficient method to compute probability distributions if the number of states of the Markov process is manageable. Fast adaptive uniformization[2] improves the original uniformization method at the cost of a small approximation error. The main ideas for this improvement are the on-the-fly construction of the state space and the restriction imposed on the state space to contain only states with significant probabilities, e.g. states that have a probability larger than 10^{-15}, thus obtaining an important reduction of the state space.

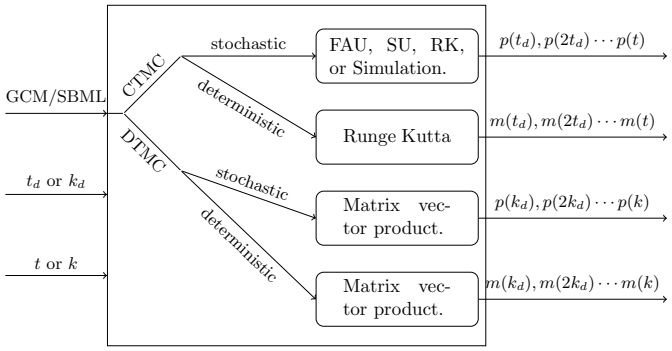

Fig. 1. Software architecture. Depending on the chosen semantics, the analysis mode and, eventually, the type of stochastic solution, SABRE computes the desired results. The vector $p(t)$ is the probability distribution after time t, while the vector $p(k)$ is the probability distribution after k steps. For the deterministic analysis, the values $m(t)$ and $m(k)$ correspond to the mean field of the corresponding CTMC, respectively DTMC.

TABLE I
CASE STUDIES SUMMARY

Analysis	Model	Time	Error	States
Stochastic	Exclusive switch	$94s$	$9e-8$	3047
Stochastic	Moran's model	$49s$	0	1001

III. INTERFACE

SABRE accepts two input formalism. The input models can be given in SBML format or as guarded-command models (GCM). SBML is a standardized format for representing models of biological processes, such as metabolism or cell signaling and is the input to SABRE's core program, while GCMs are a textual description of processes and are given in the style of Dijkstra's guarded-command language[4].

Both formalisms may express CTMC or DTMC models. The difference between the two interpretations comes from the semantic given to the rate function of each command. In the case of CTMCs, for a given reaction j, the rate function assigns to each state a positive real value that represents the rate of the outgoing transition j. For DTMCs, the rate functions are normalized such that, they define, for each state s, a probability distribution over its successors.

SABRE is available on-line, as a web tool, at http://mtc.epfl.ch/~mateescu/sabre.

IV. CASE STUDIES

We present the results of two case studies for stochastic analysis of CTMCs and DTMCs. For more and larger experiments on stochastic analysis of CTMCs we refer the reader to the paper giving the fast adaptive uniformization algorithm[2]. All our experiments are performed on a 3GHz Intel Linux PC, with 6 GB of RAM. We give the results of our experiments in Table I.

A. Genetic exclusive switch

The exclusive genetic switch we analyse involves two species of proteins that may bound to the same promoter site. We denote the unbounded proteins by N_1 and N_2 and the bounded ones by r_1 and r_2[1]. When it is bounded to the promotor site, a protein represses the production of the other protein. And so, for example, production of N_1 happens at rate g_1 only if no molecule of type 2 is bounded to the promoter site ($r_2 = 0$). This is illustrated in the GCM code by the following command:

```
r2=0 |- g1 -> N1:=N1+1;
```

We set the initial state to 25 molecules N_1 and 0 of all the other species, and run the analysis for a period of time of 10000 and with the same rates as in [1].

B. Moran's population model

As a simple example of how SABRE operates on DTMC models we choose Moran's genetic population model, which can be seen as a set of biochemical reactions, more specifically as one reversible reaction.

For a population of N individuals, with two alleles, A_1 and A_2, we are interested to find the probability of fixation of A_1, that is, the probability for A_1 individuals to be equal to N after a certain time. We have two reactions: $A_2 \rightarrow A_1$ and $A_1 \rightarrow A_2$. For x_{A_1} individuals with A_1 allele and x_{A_2} individuals with A_2 allele, the probability of the first reaction is $\frac{1-s}{2} + s \cdot \frac{x_{A_1}}{N}$, where s is a small constant. As for the second reaction, its probability is $\frac{1-s}{2} + s \cdot \frac{x_{A_2}}{N}$.

We choose $N = 1000$ and $s = 0.002$, the initial state of $x_{A_1} = 1$ and we perform a transient analysis until time $k = 10^6$, at this time, the probability of fixation is 0.00049. In this case the error we obtain is 0 because no cutting is performed, the state space is kept at its complete size of 1001.

V. CONCLUSION

We have introduced SABRE, a tool for stochastic analysis of biochemical reaction networks and of population models in general. SABRE currently is an accessible web tool, allowing us to deliver our algorithms and optimizations in a fast and portable way. However, an offline version release is planned for the future. For completeness and comparison, SABRE also performs deterministic analysis of the input system.

ACKNOWLEDGMENT

We thank Marius Mateescu for valuable advices on the web interface and Nick Barton for an introduction to population genetics.

REFERENCES

[1] B. Barzel and O. Biham. Calculation of switching times in the genetic toggle switch and other bistable systems. *Phys. Rev. E*, 78(4):041919, Oct 2008.

[2] F. Didier, T. A. Henzinger, M. Mateescu, and V. Wolf. Fast adaptive uniformization of the chemical master equation. In *Proceedings of the 1st workshop on High Performance Computational Systems Biology (HIBI'09)*. IEEE Computer Societys Conference Publishing Services, 2009.

[3] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.

[4] T. Henzinger, B. Jobstmann, and V. Wolf. Formalisms for specifying markovian population models. In Springer, editor, *LIX Colloquium Reachability Problems'09*, pages 3–23, 2009.

[5] M. Z. Kwiatkowska, G. Norman, and D. Parker. Prism 2.0: A tool for probabilistic model checking. In *QEST*, pages 322–323, 2004.

[6] A. van Moorsel and W. Sanders. Adaptive uniformization. *ORSA Communications in Statistics: Stochastic Models*, 10(3):619–648, 1994.